A

Adaptive Learning Systems

Transforming Pedagogical Approaches

Inyoung Na and Volker Hegelheimer
Department of English, Iowa State University, Ames, IA, USA

Abstract

Adaptive learning systems, which tailor content, pacing, and feedback to individual students' unique needs, offer personalized learning experiences. This customization is enabled by sophisticated learning analytics that continuously collect and analyse students' performance and interactions with learning materials to optimize the learning process. As a result, students benefit from self-paced and self-directed learning opportunities, fostering greater engagement and autonomy. This article examines the architecture and functioning of adaptive learning systems with a particular focus on leveraging artificial intelligence (AI) and machine learning. It explores the practical applications of these systems in language classrooms and beyond, addressing both the opportunities and challenges they present. The article concludes with suggestions for future directions in the field.

Keywords

Adaptive learning systems · Computer-assisted language learning · Intelligent systems · Personalized learning · Artificial intelligence · Large language models · ChatGPT

Introduction

One of the critical challenges in language classrooms is addressing learner variability, as learners come with diverse knowledge levels, learning goals, and styles, making their learning needs and preferences unique (Slavuj et al., 2017). The prevalent 'one-size-fits-all' approach, in which a single instructional method or material is applied to all, can become less effective, often irrelevant, and unengaging for learners. Promoting *personalized learning* is thus essential, as research evidence has found considerable learning benefits for groups of learners who were taught in a personalized manner according to their needs (VanLehn, 2011). However, it is not feasible for an individual human instructor to tailor a curriculum to many individual learners.

Adaptivity in instructional systems has thus emerged as a practical solution, catering to each learner's unique needs and characteristics. An adaptive learning system can be broadly defined as a system that adjusts instruction based on identified characteristics of learners to improve the

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_1-1

efficiency and efficacy of learning (Oxman et al., 2014).

Elements of Adaptive Learning Systems

The definition of an adaptive learning system suggests the technological basis of adaptive systems that typically includes three elements: learner model, domain model, and adaptation model (Brusilovsky, 2012; Knutov et al., 2009). The **learner model** is the source of adaptation, representing the learner's characteristics to which the system adapts, such as learning styles, performance, and proficiency levels. The **domain model** is the target of adaptation, defining what is to be learned through a space consisting of a set of knowledge elements. The adaptation model connects these two models and determines how the learning experience should be tailored. Table 1 provides a detailed overview of these three models.

Methods of Adaptation

A common method of adaptation is that the system *selects* and o*rders* the *content* to be presented to the learner based on the learner's needs and characteristics identified. Additionally, problem-solving support is particularly advantageous in the context of language learning, as any in-time help the system offers to the learner. It provides scaffolding and hints, immediate feedback on correct and incorrect answers, and misconception/error identification on learner solutions (Slavuj et al., 2017). Applications that function in such a way include popular language learning apps.

Language Learning Apps

Language learning is well-suited for adaptive learning approaches due to the diverse needs and varying proficiency levels of learners from different backgrounds. With the popularity of informal language learning, there is a great demand for personalized approaches, both in traditional classrooms and self-directed environments. Due to space constraints, we have included only Duolingo as a representative example of language learning apps, though other apps like Babbel, Rosetta Stone, and Memrise also employ similar adaptive learning features. Duolingo, one of the most popular language learning apps, has over 500 million active users worldwide, offering courses in over 30 languages (Duolingo, 2024). It features a structured course with a sequence of lessons and exercises created by human experts and AI. Lessons are presented as a linear 'path' divided into smaller 'units' with specific learning objectives. Learners take lessons in the order presented on the path, but what each learner sees in a lesson is unique since Duolingo uses AI models to tailor each lesson to the learner's needs and knowledge level (Pajak & Bicknell, 2022).

Upon signing up, learners take a placement test, allowing them to start at an appropriate level and skip basic units. Lessons adapt in real time, adjusting difficulty based on performance. More correct answers lead to more challenging exercises, while incorrect answers prompt additional exercises focusing on those mistakes (Vazquez et al., 2020). Another adaptive feature is spaced repetition, where the algorithm schedules reviews based on the learner's retention needs, ensuring personalized practice (Jiang et al., 2024).

Artificial Intelligence and Machine Learning

Compared to well-defined learning domains like mathematics, however, language learning, being both skill- and knowledge-based and following a nonlinear process, is more complex to construct and implement into adaptive systems (Garrett, 2009; Godwin-Jones, 2017). Due to this nature, language instructional systems often employ machine learning (ML) and artificial intelligence (AI) techniques and methods. ML techniques use the observations of learner's actions and behaviour to develop the training set for making future predictions, providing adaptivity, as seen in Duolingo (Troussas et al., 2019). AI techniques,

Adaptive Learning Systems, Table 1 The architecture of an adaptive system

Component	Function
Learner model *(Adapt to what?)*	Keeps the learner model up-to-date by analysing and interpreting user behaviours and inferred learning outcomes (i.e. learner modelling)
Domain model *(Adapt what?)*	Establishes clear relationships between learning materials and domain items (i.e. knowledge indexing)
Adaptation model *(How to adapt)*	Implements the adaptation by combining the information from the learner and domain models

particularly natural language processing (NLP), are used to process a solution for a learner's free language output using such corpus analysis and grammatical error simulations as in intelligent tutoring systems (ITSs). ITSs have been built using these technologies to achieve the effect of one-on-one tutoring that a human teacher can provide. ITSs, using their 'intelligence', identify the abilities and weaknesses of each student to base their problem-solving support and sequencing learning items and tasks (Hasanov et al., 2019; Slavuj et al., 2017). ITSs have sometimes been reported to produce learning effects comparable to human tutoring (VanLehn, 2011). However, the level of engagement and feedback provided by these systems remains unsatisfactory (Kuo et al., 2023) due to challenges such as difficulties in evaluating productive skills like writing and speaking, as well as in providing nuanced feedback beyond simple correct–incorrect statements (Slavuj et al., 2015).

Recent advances in large language models (LLMs), like OpenAI's ChatGPT, Microsoft's Copilot, and Google's Gemini, suggest new ways to enhance adaptive learning mechanisms and students' learning experiences. For instance, while multiple-choice items were previously favoured in evaluating students' success with learning items due to their straightforward dichotomous (right or wrong) nature, evaluating learners' free responses to open-ended questions was challenging due to technological limitations (Kwon, 2023). Rather than providing instant and personalized feedback that could enhance students' learning, the adaptivity of these earlier systems was often merely a list of desired learning materials for an individual in their subsequent learning. LLMs, pretrained on vast amounts of data, are highly sophisticated and make it possible to understand and predict language patterns, generating humanlike text. These models now allow for automated and more detailed examination of students' extended responses and direct identification of misconceptions. More adaptive systems are starting to incorporate LLMs, with the GPTs' API accessible for constructing chatbots within the adaptive platform. These chatbots interact with students to produce meaning-focused dialogues (Meniado, 2023) and qualitative information that helps more nuanced diagnosis of students' weaknesses, which allows for scaffolding, personalized feedback for remediation, and the generation of customized assessment items (Kuo et al., 2023).

The new Duolingo Max, powered by GPT-4, offers enhanced personalized learning (Duolingo Team, 2023). The 'Explain My Answer' chat feature allows learners to ask more about the lesson and receive explanations and examples to clarify why an answer was right or wrong. The 'Roleplay' feature allows learners to practice real-world conversations in various scenarios, with world characters, such as ordering coffee at a cafe in Paris or asking a friend to go for a hike. After the interaction, learners receive feedback on the accuracy and complexity of their responses, as well as tips for future conversations. Currently, these features are available for English learners of Spanish, French, and Finnish, with plans to expand to more languages.

Future Considerations

Most adaptive systems focus on areas like vocabulary, grammar, and reading; however, systems targeting other skills (i.e. listening, speaking, and writing) are relatively scarce (Kaur et al., 2023; Slavuj et al., 2017). These systems typically adapt to learners' performance by measuring the

number of correct and incorrect answers to questions (Kaur et al., 2023). Essentially, they rely heavily on quantitative learning data, yet a more comprehensive understanding of student learning may require qualitative data, especially in dealing with language production (Ayedoun et al., 2020). That said, these systems should extend beyond performance metrics to include other individual learner profiles, such as learning styles and preferences, enhancing the learning experience. Typically, these characteristics have been identified through questionnaires (Truong, 2016); however, learning systems should aim to incorporate AI-based automatic detection methods to minimize biases in self-reported data (Kaur et al., 2023).

With the highly sophisticated LLMs and their associated chatbots now available, language learning platforms can leverage these models to dynamically generate personalized learning experiences that adjust to a learner's progress, preferences, and needs in real time. Among all, the ability of LLMs to address extended responses is particularly beneficial, providing feedback tailored to each student's writing rather than being generally applicable. This feedback adeptly diagnoses strengths and weaknesses concerning both global aspects (e.g., content, organization, and style) and local issues (e.g., language and grammar errors), offering specific remedies. However, despite these advantages, the feedback can often be overwhelming and may not feel as tailored to individual learners as human feedback (Na et al., 2025), indicating that further improvements are necessary.

To this end, systematic studies on student interactions with AI tools are crucial. For instance, Na et al. (2025) found that when university ESL students used ChatGPT for writing revisions, general prompts (e.g. 'Can you give me feedback on this essay?') often led to lengthy, less actionable feedback, and learners struggled to locate errors due to the lack of visual features (e.g. Grammarly's flagging). In contrast, more specific prompts (e.g. 'What do you think of the flow of the second paragraph?') resulted in more targeted, helpful feedback. However, many prompts still led to ChatGPT generating fully revised texts, even when not requested, encouraging copying and pasting behaviours from the learners.

More research is needed to optimize AI tools by enhancing features that promote engagement and effective learning behaviours, rather than simply providing easy shortcuts for students. Developers should consider integrating LLMs into their platforms to maximize learning benefits by customising the model to provide just the right amount of support and the type and amount of feedback learners wish to receive. However, successful implementation of LLMs in adaptive systems relies on the accuracy and precision of these models.

Pedagogical Considerations

Adaptive technology should be seen as a complement, not a replacement, for classroom instruction. The success of teaching with an adaptive platform depends on deliberate human planning, monitoring, and intervention, making the instructor's role crucial (Cavanagh et al., 2020). Instructors should guide students' self-directed learning by regularly checking in on their progress using learning analytics (e.g., login frequency, learning rates/speed, and progress) and providing supplementary materials and instructor-led feedback to address the student's weaknesses identified by the system.

Chatbots powered by LLMs offer numerous advantages for language teaching and learning. A simple prompt to ChatGPT can generate customized lesson plans, materials, and exercises tailored to students' varying needs and interests (Koraishi, 2023). Teachers can also encourage students to engage in real-time language practice with ChatGPT, receiving instant, personalized feedback on grammar and word choice (Amin, 2023). Furthermore, by specifying learning objectives, teachers can use ChatGPT to generate various assessment items, such as multiple-choice and open-ended questions, at different proficiency levels (Ockey, 2024). The benefits of formative assessment are enhanced with timely and frequent feedback from ChatGPT on student responses,

fostering increased autonomy and motivation (Warschauer & Grimes, 2008).

However, AI use in student work poses challenges to academic integrity. Perkins (2023) notes that these tools can blur the lines between legitimate academic assistance and misconduct, such as plagiarism, due to the difficulty in detecting misuse. Nevertheless, if students clearly state any AI use in their submissions, it may not necessarily be considered plagiarism (Perkins, 2023). Meanwhile, institutions should develop clear policies to address academic integrity concerns, and teachers should train students on ethical AI usage.

Conclusion

This article explored how adaptive language systems, particularly those enhanced by AI and LLMs like ChatGPT, hold transformative potential for personalized language education by providing customized learning experiences through adaptive content, assessments, and feedback to meet individual learners' needs. As these technologies continue to evolve, ongoing research and development will be crucial to maximize their potential for language learning while addressing ethical considerations, particularly around academic integrity. The role of educators is vital in ensuring these tools are integrated responsibly and effectively within educational settings. Collaborative efforts among researchers, educators, policymakers, and developers will be essential to navigate the challenges and fully realize the opportunities presented by these emerging tools.

Cross-References

- ▶ Generative AI in CALL: Evolution, Applications, and Future Directions
- ▶ Pedagogical Innovation in CALL
- ▶ The Impact of Artificial Intelligence (AI) on CALL Pedagogies

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Amin, M. Y. M. (2023). AI and Chat GPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies, 4*(4), 1–15. https://doi.org/10.33422/ijhep.v4i4.554

Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward personalized scaffolding and fading of motivational support in L2 learner–dialogue agent interactions: An exploratory study. *IEEE Transactions on Learning Technologies, 13*(3), 604–616. https://doi.org/10.1109/TLT.2020.2989776

Brusilovsky, P. (2012). Adaptive hypermedia for education and training. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 46–66). Cambridge University Press. https://doi.org/10.1017/CBO9781139049580.006

Cavanagh, T., Chen, B., Lahcen, R. A. M., & Paradiso, J. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. *The International Review of Research in Open and Distributed Learning, 21*(1), 172–196. https://doi.org/10.19173/irrodl.v21i1.4557

Duolingo. (2024). Home page. Retrieved June 1, 2024, from https://www.duolingo.com

Duolingo Team. (2023). Introducing Duolingo Max, a learning experience powered by GPT-4. Duolingo Blog. https://blog.duolingo.com/duolingo-max/

Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. *The Modern Language Journal, 93*, 719–740. https://doi.org/10.1111/j.1540-4781.2009.00969.x

Godwin-Jones, R. (2017). Scaling up and zooming in: Big data and personalization in language learning. *Language Learning & Technology, 21*(1), 4–15. https://dx.doi.org/10125/44592

Hasanov, A., Laine, T. H., & Chung, T. S. (2019). A survey of adaptive context-aware learning environments. *Journal of Ambient Intelligence and Smart Environments, 11*(5), 403–428. https://doi.org/10.3233/AIS-190534

Jiang, X., Hopman, E., Reuveni, B., & Kittredge, A. (2024). Duolingo path meets expectations for proficiency outcomes [white paper]. http://duolingo-papers.s3.amazonaws.com/reports/Duolingo_whitepaper_language_read_listen_write_speak_2024.pdf

Kaur, P., Kumar, H., & Kaushal, S. (2023). Technology-assisted language learning adaptive systems: A comprehensive review. *International Journal of Cognitive Computing in Engineering, 4*, 301–313. https://doi.org/10.1016/j.ijcce.2023.09.002

Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 years later: A comprehensive survey of adaptive hypermedia methods and techniques. *New Review of Hypermedia and Multimedia, 15*(1), 5–38. https://doi.org/10.1080/13614560902801608

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology, 3*(1).

Kuo, B. C., Chang, F. T., & Bai, Z. E. (2023). Leveraging LLMs for adaptive testing and learning in Taiwan Adaptive Learning Platform (TALP). *Artificial Intelligence in Education 2023 (AIED 2023), Tokyo, Japan.*

Kwon, T. (2023). *Interfaces for personalized language learning with generative language models* [Master's thesis, Columbia University]. https://doi.org/10.7916/26qy-3n10

Meniado, J. C. (2023). The impact of ChatGPT on English language teaching, learning, and assessment: A rapid review of literature. *Arab World English Journal, 14*(4). https://doi.org/10.24093/awej/vol14no4.1

Na, I., Duris, M., & Hegelheimer, V. (2025). ChatGPT in ESL writing: Assessing effectiveness and perceptions among international students. In Y. Wang, A. Alm, & G. Dizon (Eds.), *Insights into AI and language teaching and learning*. Castledown. https://doi.org/10.29140/9781763711600

Ockey, G. J. (2024). *Introducing second language assessment*. Cambridge University Press.

Oxman, S., Wong, W., & Innovations, D. (2014). *White paper: Adaptive learning systems* (pp. 6–7). Integrated Education Solutions.

Pajak, B., & Bicknell, K. (2022). At Duolingo, humans and AI work together to create a high-quality learning experience. Duolingo Blog. https://blog.duolingo.com/how-duolingo-experts-work-with-ai/

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice, 20*(2). https://doi.org/10.53761/1.20.02.07

Slavuj, V., Kovacic, B., & Jugo, I. (2015). Intelligent tutoring systems for language learning. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 814–819). https://doi.org/10.1109/MIPRO.2015.7160383

Slavuj, V., Meštrović, A., & Kovačić, B. (2017). Adaptivity in educational systems for language learning: A review. *Computer Assisted Language Learning, 30*(1–2), 64–90. https://doi.org/10.1080/09588221.2016.1242502

Troussas, C., Chrysafiadi, K., & Virvou, M. (2019). An intelligent adaptive fuzzy-based inference system for computer-assisted language learning. *Expert Systems with Applications, 127*, 85–96. https://doi.org/10.1016/j.eswa.2019.03.003

Truong, H. M. (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior, 55*, 1185–1193. https://doi.org/10.1016/j.chb.2015.02.014

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vazquez, G., Lopez, M., & Soto, V. (2020). Duolingo: An AI-powered language learning platform. In *Proceedings of the international conference on artificial intelligence in education (AIED)* (Vol. 1, pp. 289–295).

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal, 3*(1), 22–36. https://doi.org/10.1080/15544800701771580

E

Ergonomics in CALL Design and Teaching

Marie-Josée Hamel[1] and Catherine Caws[2]
[1]University of Ottawa, Ottawa, ON, Canada
[2]University of Victoria, Victoria, BC, Canada

Abstract

This encyclopaedia entry discusses on ergonomics in Computer-Assisted Language Learning (CALL). Ergonomics is an applied science concerned with the study of efficiency of individuals within their work or life environment. The article emphasises on two main areas where ergonomics can be effectively applied in CALL: designing learning environments and adapting language learning and teaching to learners' needs and behaviours.

Keywords

Ergonomics · CALL · Interaction · Human Factors · Instructional Design · Adaptive Teaching

Introduction

Ergonomics is a term derived from the Greek words 'ergon', meaning 'work', and 'nomos', meaning 'law'. It was introduced in the mid-twentieth century by a British psychologist, Hywel Murrell, to describe the study of people's efficiency in their working environments, focusing on the use of scientific principles to promote human well-being and performance.

Ergonomics offers many promising avenues in technology-mediated language education. In Computer-Assisted Language Learning (CALL), ergonomics plays an important role in interaction-based research by providing a conceptual framework (Bertin & Gravé, 2010; Raby, 2005) that looks specifically at the 'learner-task-tool' relationship. Moreover, ergonomics offers methods and tools enabling researchers and language practitioners to observe and analyse learner-computer interactions (Caws & Hamel, 2016). By investigating what learners *do*, the *outcome* of what they do and how *they think*, CALL ergonomics seeks to shed light on learner behaviours and perceptions, in order to optimise pedagogical interventions (Fischer, 2007). Results of these analyses can and should be recycled into the instructional design process and reused for teaching, learning and training purposes.

CALL Ergonomics in a Nutshell

CALL ergonomics is an interdisciplinary approach that is rooted in interaction-based research and software engineering. It "constitutes both a methodological and theoretical framework that seeks to describe interactions between users and instruments in a view to ameliorate these interactions so that learning or work can be

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_8-1

enhanced" (Caws & Hamel, 2016, p. 21). As such, CALL ergonomics uses scientific and engineering principles, viewing the learning environment as a 'workplace' (Benedyk et al., 2009, p. 238). In this unique context, the objective of CALL ergonomics is to better understand how learners adapt to their environment (Raby, 2005), namely how they perform, how they produce or perceive language or how they interact with others in another language through technologies. The goal is then to enhance specific 'work' (i.e., learning) situations.

While recognising the critical role that technology plays in today's society, CALL interaction-based research also seeks to inform on "specific models and interventions that are common in technology-mediated learning and teaching and that may need further development" (Caws & Hamel, 2016, p. 2). Empirical data is collected through an iterative process that typically includes two critical phases: the first phase consists of observing and analysing language learners interacting with or through technologies while the second phase consists of recycling these data to (re) design and optimise CALL learning ecosystems.

CALL ergonomics provides a learner-centred framework for instructional design and technology-mediated pedagogical interventions.

CALL Ergonomics and Instructional Design

Instructional design is described herein as the systematic process of creating educational and training programs. Human factors (such as physical, cognitive, social, etc.) play an important role in instructional design as they refer to the study of how humans (i.e., learners) *interact* with elements (such as technological tools) in a particular system (a specific learning environment); the aim being to optimise such interactions. By understanding and applying knowledge of these factors, instructional designers aim to develop systems and products that enhance learner well-being and overall system performance. User experience (UX) design tools, such as personas (Cooper, 1999) and scenarios (Carroll, 2000), can be used within CALL educational contexts when designing contents. Models derived from computer interface design, such as the ADDIE model, have also been applied to develop CALL systems (Colpaert, 2006).

The analysis of an activity (or activities) is a key component in CALL ergonomics and instructional design. It involves a systematic examination of how learning tasks are performed within a specific learning context to understand the processes, tools and interactions engaged. It aims to identify the steps, skills and resources required by learners to successfully complete a task, as well as the potential challenges and areas for improvement. Such analyses focus on the user experience (UX), which is central to the design of learner-centred CALL systems or technology-mediated learning environments. These analyses may explore individual and collaborative language-learning tasks to identify patterns of individual learner behaviour and second language learning during technology-mediated activities, learner personas (i.e., learner types) (Heift, 2013), or the effectiveness of technology-mediated affordances. The methods used to perform these analyses are mixed relying on data collected from interactions with and through various technologies or language learning systems (Blin et al., 2013). Ultimately, these activity analyses examine the quality of the experience, particularly the 'learner-task-tool' interactions (Caws & Hamel, 2016).

CALL Ergonomics for Adaptive Teaching

Within a CALL environment, adaptive teaching relies on an investigation of technological, pedagogical and linguistic affordances that enable, facilitate and sustain adaptive instruction. As expressed by Ellis and Goodyear, "Helping students learn, in contemporary higher education, involves an understanding of the complex web of relationships that give shape and meaning to students' activity and experience" (2010, p. 16). CALL ergonomics, being sensitive to individual differences, aligns well with adaptive learning approaches (Vandewaetere et al., 2010), taking into account learners' evolving knowledge and abilities, as they become more proficient users of technologies.

Several factors require investigation when trying to optimise learning in CALL: the diversity of educational contexts (online, hybrid, in-class), the wealth of digital technologies (social media, tutoring systems for different skills, etc.) and the successful integration of CALL tasks and activities into diverse language education contexts.

While the ultimate goal of CALL ergonomics is to design efficient and effective learning environments, researchers rely on various measures to better assess these environments: task analysis and recording (through screen captures) of learning processes (Park & Kinginger, 2010; Hamel & Séror, 2016; Schulze et al., in press). Such approaches to analysis of learning behaviours allow the diagnosis of specific learning needs. The results of these analyses are then recycled into instructional prescriptions to provide enhanced teaching interventions that better adapt to learners' needs.

Conclusion

CALL ergonomics continues to offer promising methods and outcomes to enhance learning and teaching in technology-mediated language learning, as well as to adapt to learners' evolving needs. At a time when intelligent learning systems, technologies and digital applications continue to expand into learners' personal and professional spheres, it becomes even more pressing to better understand how these new devices and their affordances may affect language learning and teaching. Ultimately, the goal is to optimise the 'learner-task-tool' interactions.

Cross-References

- ▶ Adaptive Learning Systems: Transforming Pedagogical Approaches
- ▶ Blended Learning and Teaching Models of CALL
- ▶ Mobile-Assisted Language Learning (MALL)
- ▶ Pedagogical Innovation in CALL
- ▶ Reflective Approaches to CALL
- ▶ Task-Based Language Learning in CALL Environments
- ▶ Virtual Reality-Assisted Language Learning (VRALL)

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Benedyk, R., Woodcock, A., & Harder, A. (2009). The Hexagon-Spindle Model for educational ergonomics. *Work, 32*(3), 237–248. https://doi.org/10.3233/WOR-2009-0822

Bertin, J. C., & Gravé, P. (2010). In favor of a model of didactic ergonomics. In J. C. Bertin, P. Gravé, & J.-P. Narcy-Combes (Eds.), *Second language distance learning and teaching: Theoretical perspectives and didactic ergonomics* (pp. 1–36). Information Science Reference. https://doi.org/10.4018/978-1-61520-707-7.ch001

Blin, F., Caws, C., Hamel, M.-J, Heift, T., Schulze, M., & Smith, B. (2013). Data and elicitation methods in interaction-based research. *Proceedings of WorldCALL 2013*, Glasgow, pp. 21–28.

Carroll, J. M. (2000). *Making use: Scenario-based design of human-computer interactions*. MIT Press.

Caws, C., & Hamel, M.-J. (2016). CALL ergonomics revisited. In C. Caws & M.-J. Hamel (Eds.), *Language learner computer interactions: Theory, methodology and CALL applications* (pp. 17–40). John Benjamins Publishing Company.

Colpaert, J. (2006). Toward an ontological approach in goal-oriented language courseware design and its implications for technology-independent content structuring. *Computer Assisted Language Learning, 19*(2–3), 109–127. https://doi.org/10.1080/09588220600821461

Cooper, A. (1999). The Inmates are Running the Asylum. In U. Arend, E. Eberleh, & K. Pitschke (Eds.), *Software-Ergonomie '99. Berichte des German Chapter of the ACM* (Vol. 53). Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-99786-9_1

Ellis, R., & Goodyear, P. (2010). *Students' experiences of e-learning in higher education*. Routledge.

Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *Computer Assisted Language Learning, 20*(5), 409–442.

Hamel, M.-J., & Séror, J. (2016). Video screen capture to document and scaffold the L2 writing process. In C. Caws & M.-J. Hamel (dir.), *Language-learner-computer interactions: Theory, methodology and CALL applications* (pp. 137–162). .

Heift, T. (2013). Learner Personas in CALL. *CALICO Journal, 25*(1), 1–10. https://doi.org/10.1558/cj.v25i1.1-10

Park, K., & Kinginger, C. (2010). Writing/thinking in real time: Digital video and corpus query analysis. *Language Learning & Technology, 14*(3), 31–50.

Raby, F. (2005). A user-centered ergonomic approach to CALL research. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 179–190). New York.

Schulze, M. Caws, C., Hamel, M.-J., & Heift, T. (in press). The Cambridge handbook of technology in language teaching and learning edited by Glenn Stockwell and Yijen Wang.

C

Critical Approaches to CALL

Ron Darvin
Department of Language and Literacy Education, The University of British Columbia, Vancouver, BC, Canada

Abstract

Critical approaches to CALL recognize how digital tools that enable language learning are never neutral. The design and use of these technologies are implicated in issues of power and ideology, shaping inequalities and modes of exclusion in various learning environments. CALL research that integrates this critical lens draws attention to the inequalities of learners, tools, and classrooms. It recognizes that learner access to digital technologies and other material, cultural and social resources is unequal, shaping diverse digital literacies. More than being aesthetic or functional, the sociotechnical design of platforms is ideologically loaded, indexing the intentions of their developers while shaping behaviours online. Adaptive learning systems and AI technologies that operate through big data and algorithms can reproduce and even amplify existing social biases. In the language classroom, the ways in which teachers integrate these tools into their pedagogy also have profound implications in terms of privileging and marginalizing certain learners. By drawing attention to the way power operates in the many layers of a digitally-mediated learning environment, critical approaches to CALL raise pivotal questions about issues of equity and inclusion, and advocate for technology use that can foster a more inclusive and socially just future.

Keywords

Agency · Design · Ideology · Inequality · Critical digital literacies

Introduction

Critical approaches to Computer-Assisted Language Learning (CALL) involve a critical examination of the way power and ideologies shape the design and use of digital technologies for language learning. These approaches recognize that digital platforms are not neutral; instead, they index the interests and intentions of their designers and the material contexts in which they operate. The extent to which learners are able to participate in online spaces and draw on the affordances of various tools is shaped by their identities, their access to resources, and the way they are positioned by others in these spaces. By examining the sociotechnical design of digital tools, critical CALL highlights how these tools can both enable and constrain learner agency, and draws attention to how learners negotiate the use of these tools to achieve their own intentions.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_11-1

This perspective is essential for pedagogical approaches to CALL as it encourages educators to critically engage with the technologies they integrate into their teaching, fostering a more equitable and inclusive learning environment. By addressing the way power operates in the use of these tools, their embedded biases, and the inequalities of learners, educators can develop teaching strategies that not only enable digital literacies but also challenge digital modes of exclusion and empower learners to participate agentively in online spaces.

Platform Designs and Ideologies

Critical approaches to CALL are grounded in the understanding that digital technologies are material artefacts shaped by the interests of their designers and the broader socioeconomic context in which they are developed and deployed. This perspective draws from critical theory, which interrogates the ways in which power and ideology are embedded in cultural and technological artefacts (Luke, 2014). Digital platforms are designed with specific ideological underpinnings that shape user behaviour and interaction. For example, Darvin (2023) discusses how the design of educational platforms can index powerful interests and material inequalities. Operating through the interaction of algorithms and big data, platforms like Google Search and Google Translate prioritize certain types of information and sources, reflecting the interests of those with the resources to optimize their content for search engines.

Algorithms play a crucial role in structuring social interactions and the distribution of knowledge online. Gillespie (2014) argues that algorithms are not neutral but are embedded with the values and interests of their creators. Critical CALL examines how algorithms can reproduce and amplify social biases, impacting learners' access to information and their online interactions. For instance, adaptive learning systems that rely on big data and machine learning can reinforce existing educational inequalities by tailoring content based on biased data sets. In a study of an online EFL class, Dooly and Barberà (2020) highlighted the importance of learner awareness of the hidden agents at the layer of digital interfaces. They demonstrated how the design of an audio-conferencing tool served as a "proxy pedagogical agent", shaping knowledge acquisition and maintaining traditional power dynamics between teachers and students.

Inequalities of Access and Use

One other concern of critical CALL is the unequal access to digital tools and other material, cultural, and social resources that shape contrasting digital practices. In a study of the digital practices of Canadian youth, Darvin (2018, 2023) demonstrated how material and relational differences of home environments, spatial configurations, parental involvement, and peer networks can shape diverse digital literacies and dispositions towards technology. He also observed that material features of devices, such as screen size and interface choices, impact learners' ability to access and process information. Well-resourced learners who use laptops can access a wider range of linguistic resources compared to those using mobile phones, highlighting the role of materiality in digital learning. In a quantitative study analysing variables that shape the informal digital learning of English, Zhang and Liu (2023) noted how differences in gender, ethnicity, and university contexts shaped individual learning experiences, highlighting the need for targeted support to address these inequalities and enhance digital learning opportunities for all learners. In an earlier study on digital inclusion, Warschauer and Matuchniak (2010) noted how students from affluent neighbourhoods who had access to the latest digital devices, high-speed internet, and supportive learning environments both at home and school were more likely to engage in creative and critical uses of technology, such as producing multimedia projects, programming, and using advanced software for learning purposes. In contrast, less-resourced students who relied on outdated or shared devices, had limited internet access, and minimal support from their home

environments, were typically restricted to basic tasks such as typing documents or conducting simple online searches, which did not foster the same level of digital literacy and critical engagement as their highly resourced peers.

Pedagogical Implications

The integration of digital tools in language classrooms has profound implications for pedagogy and learner equity. Darvin and Hafner (2022) emphasize the need for critical digital literacies, arguing that teachers need to be aware of how these relations of power are embedded in digital tools and to adopt pedagogical strategies that challenge these inequities. This critical approach involves not only teaching learners how to recognize digital genres, evaluate information, and compose digital texts to achieve their specific intentions, but also fostering critical awareness of how the mediation of digital technologies shapes the distribution of knowledge, the performance of identities, and the construction of social relations. Critical digital literacies encourage learners to create texts that subvert the power structures embedded in digital platforms and to engage critically with online content to foster more equitable and inclusive digital spaces. One of the significant challenges in developing critical digital literacies however is the complexity of the technical and sociopolitical dimensions of digital media. Pangrazio (2016) argues that the integration of digital media into learners' personal and affective realms complicates the ability to critically evaluate digital texts. Additionally, the rise of advanced AI tools, which operate through opaque and proprietary algorithms, raises questions of transparency and accountability (Liu et al., 2024). Enabling critical perspectives of technology in the language classroom requires a metalanguage to facilitate such instruction, and this approach includes theorizing online genres and their roles in understanding rhetorical situations and communicative purposes (Hafner et al., 2023).

Conclusion

Critical approaches to CALL underscore the importance of recognizing and addressing the power dynamics and ideologies embedded in digital technologies. By drawing attention to the inequalities in access to digital resources and the sociotechnical design of platforms, these approaches advocate for more equitable and inclusive practices in language learning. Integrating critical digital literacies into language education is essential for empowering learners to navigate a continually evolving digital landscape with greater reflexivity and agency.

Looking ahead, the field of critical CALL can benefit from further exploration of how emergent technologies like generative AI (GenAI) are reshaping language learning practices and how these tools might perpetuate or disrupt existing inequalities. It needs to explore how GenAI systems trained on various datasets may reproduce biases present in the data by generating content that reflects ideological notions of race, gender, ethnicity, and social class. The ethical dimensions of using these tools also need to be examined further, particularly as they involve issues of privacy, data security, and learner autonomy. GenAI can also shift power dynamics in the classroom, potentially privileging certain ways of knowing and interacting, impacting the roles of teachers and opportunities for students to engage critically with the learning process. To respond to this shifting landscape, critical CALL needs to develop new pedagogical frameworks that not only incorporate these tools, but also promote a critical awareness of such tools and assert the role of teachers in maintaining a technologically innovative and socially just classroom.

Cross-References

▶ Critical DGBLLT
▶ Critical Digital Literacy
▶ Democratization of AI/Equity in AI
▶ Digital Literacies

▶ Identity and DGBLLT
▶ Identity Construction

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Darvin, R. (2018). Social class and the unequal digital literacies of youth. *Language and Literacy, 20*(3), 26–28.

Darvin, R. (2023). Sociotechnical structures, materialist semiotics, and online language learning. *Language Learning & Technology, 27*(2), 28–45.

Darvin, R., & Hafner, C. (2022). Digital literacies in TESOL: Mapping out the terrain. *TESOL Quarterly, 56*(3), 865–882.

Dooly, M., & Barberà, E. (2020). Getting smart: Towards critical digital literacy pedagogies. *Social Semiotics, 33*(2), 326–349.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkwowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–193). MIT Press.

Hafner, C. A., Harrison, S., Ho, W. Y. J., & Kwan, B. S. (2023). Digital mediation in ESP genres. *English for Specific Purposes, 71*, 115–122.

Liu, G., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1–29.

Luke, A. (2014). Defining critical literacy. In J. Avila & J. Z. Pandya (Eds.), *Moving critical literacies forward: A new look at praxis across contexts* (pp. 19–31). Routledge.

Pangrazio, L. (2016). Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education, 37*(2), 163–174.

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education, 34*(1), 179–225.

Zhang, Y., & Liu, G. (2023). Examining the impacts of learner backgrounds, proficiency level, and the use of digital devices on informal digital learning of English: An explanatory mixed-method study. *Computer Assisted Language Learning*, 1–28.

T

Trends in Pedagogical Approaches to CALL

Robert Godwin-Jones
Virginia Commonwealth University, Richmond, VA, USA

Abstract

Learning a language through technology has been enabled in recent years through the growing availability of online communities and multimedia resources. Those include L2 use through participation in social media, affinity groups, and gaming. Mobile devices have enabled local language resources to be leveraged for situated language learning, such as through augmented reality. Phones are also widely used to access services such as streaming video or music services, geared towards socializing and entertainment, but which offer affordances for informal language learning. Virtual exchange has emerged as a widely used vehicle for the development of pragmatic competency. The acceptance of translanguaging in applied linguistics has led to greater use of bilingual activities, including having learners work with machine translation. The advent of generative AI holds considerable promise for language learning, for incidental language learning through chatbots as well as for explicit learning through personalized tutors. AI also offers a powerful addition to the ability of immersive technologies to supply a stimulating virtual learning experience.

Keywords

Multimedia · Mobile computing · Artificial intelligence · Second language acquisition · Virtual exchange · Multilingualism

Introduction

One of the signature trends in CALL, following developments in second language acquisition (SLA) theory, has been a gradual expansion from a focus on learner-internal, largely cognitive approaches to a broader perspective that emphasizes peer learning and contextual factors. This sociocognitive and ecological framework has led a move away from isolated form-focused learning activities to collaborative scenarios that involve real-world meaning-making. Accompanying that direction in CALL is a pivot from a deficit-oriented approach to SLA, which centred on discrete skill-focused practice, to an emphasis on using the second language (L2) for genuine communicative activities. Real-world learning integration has led to a greater interest in leveraging online resources in all instructed environments. The rise of virtual exchange has led to the greater prominence of pragmatic and interactional competence. At the same time, CALL has followed SLA theory in expanding its focus beyond verbal

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_12-1

language to include multimodality and multilingualism, as both have become prominently practiced in social media, online affinity groups, and mobile platforms. Shifting away from native speakerism as an end goal also aligns with developments in SLA and is reflected in CALL pedagogical practices in the increasing acceptance of translation and L1-L2 cross-fertilization in learning and teaching. The integration of learners' L1 into SLA includes making machine translation an early manifestation of what has become an important current trend in CALL: the rise of generative AI. Despite concerns about its built-in linguistic and cultural biases, AI, in its text-based and multimedia forms, promises to bring significant affordances to both independent language learners through chatbots and to explicit, formal learning environments through personalized adaptive tutors.

Greater Integration of Digital Resources

Many language instructors are finding that online and mobile environments offer rich opportunities for language learning that can supplement formal instruction. While opportunities through social media, affinity groups, streaming media, and gaming have been available for some time, studies in recent years have shown how effective for SLA such activities can be for many users (Sockett, 2014). The language acquisition process afforded by such activities is supported by usage-based linguistics theory which emphasizes the centrality of implicit language learning through the incidental acquisition of language structures and chunks of language as they are used, reused, and varied (Ellis, 2017). Participation in such online activities tends to be highly motivating, as L2 users are primarily targeting entertainment and personal interest. Many media services supply optional language comprehension assistance through subtitles, transcripts, or song lyrics, thus providing scaffolding for language learning.

For instructed SLA, online L2 activities can be organized through "bridging activities" (Thorne & Reinhardt, 2008, p. 558) that have students pursue L2 resources outside of class, then bring their experiences back to the classroom (physical or virtual) for individual or group discussion and reflection. That process can be helpful in developing learner autonomy, as L2 users gain metalinguistic awareness. Participation in online activities in the L2 has been greatly enabled by the wide use of mobile devices. Smartphones and tablets are often the preferred mode of access to online resources. As mobiles are always nearby, they offer access to digital resources at all times in all contexts, thus becoming intimate and trusted companions. As more connected devices become available—at home, in the car, as wearables—the Internet of Things creates an "ambient intelligence", featuring continual connectivity to online communities and digital artefacts (Godwin-Jones, 2023, p. 7).

Expanding Learning Goals and Approaches Through Technology

Mobile devices enable closer connections of L2 user learners to real-world contexts. A major consequence is the opportunity for L2 learning to align more closely with learners' lived experiences. One way to connect local language resources through technology is the use of augmented reality (AR), most often taking the form of superimposing textual or graphic information on views of the immediate vicinity through a smartphone camera. That technology has been used in creating language trails, integrating cultural sites, often in multiple languages, and in creating gaming scenarios (Pegrum, 2019). Mobile devices have been largely responsible for the boom in multimedia, both in everyday social media exchanges and in SLA. As multimedia becomes more widely integrated, digital communication moves beyond a focus on verbal language to incorporate a variety of meaning-making and symbolic systems. A semiotic view of SLA incorporates nonverbal signals such as gestures, body language, and paralanguage (Godwin-Jones, 2023). The mobility of smartphones also serves to anchor language use to the physical location of the user, emphasizing

the localized and contextual nature of human communication.

The recognition of the importance of context in language use (Levine, 2020) has led to a growing emphasis on situated language practice, and therefore on pragmatics, as that involves finding contextually appropriate ways, linguistically and culturally, to engage in communication. Pragmatic competence is gained through experience using language in real social contexts. Technology provides an ideal opportunity for setting up situations that allow learners to explore socio-pragmatic language use and to try out their L2 pragmatic competence (González-Lloret, 2019). Particularly effective for that purpose are situations which enable or simulate real-world conversations, such as virtual exchange, which has experienced considerable growth in recent years. One of the benefits of virtual exchange is that it breaks from the artificiality of language use in the classroom. Pragmatic failure in virtual exchange can illuminate dramatically for L2 learners the greater importance of socially and culturally appropriate language use over grammatical accuracy. That metalinguistic awareness is facilitated in tandem exchanges, in which participants are led to contrast sociolinguistic features of both languages. Another activity that aids in developing that awareness is translation. The growing acceptance of incorporating translation into L2 learning is a reflection of the acceptance of translanguaging, the idea that humans do not use a single language at a time, but that we draw on all the languages we know at the same time (Garzía & Li, 2014). We have seen in particular increased number of studies of the integration of machine translation into SLA (Godwin-Jones, 2023). The embrace of translanguaging is evident in other digital practices such as fanfiction (Sauro, 2017).

Pedagogical Transformation Through AI

Generative AI has the potential to be a disruptive technology similar to the arrival of the World Wide Web in the 1990s and the smartphone in 2007, bringing profound changes to all aspects of human activity. That effect was immediately evident after the release of *ChatGPT* for businesses which focus primarily on writing, such as journalism. AI has since sent ripples of change across most businesses. For education, it is likely to bring significant transformations, not just to writing but to teaching and assessment in all fields.

Through generative AI, language learning is likely to see new opportunities in both instructional settings and autonomous learning. Since the arrival of *ChatGPT* we have seen a steady stream of blog posts, media reports, conference presentations, and workshops on ways to integrate generative AI into language learning. Numerous YouTube videos have announced that "AI will kill traditional language learning", meaning classroom-based instruction. The hype is overblown, but it is undeniable that AI has been shown to be useful for foreign language learners and teachers in a variety of ways. It can serve as a conversation partner for practicing the target language in written or spoken form. AI can be instructed to use the language at a particular level of proficiency, matching that of the learner. Before initiating a conversation with *ChatGPT*, the learner can direct the AI to correct errors and provide explanations for particular language constructions. Teachers are able to use AI to generate lesson plans, practice exercises, or multimedia stories, as well as to help in assessment of student work (Poole & Polio, 2023). Along with the evidence of the benefits of AI for language learning, a number of concerns have been raised as well. Teachers are fearful that learners will turn over homework assignments to AI. There are a host of ethical concerns, including how training data has been obtained, cultural biases inherent in the nature of that data, and the power of the large AI companies who control access and cost (Atari et al., 2023). Since training data for AI is predominantly in English, output in other languages—especially low-resourced languages—is likely to be inferior to what it produces for English language users (Stokel-Walker, 2024). AI has been shown to have problematic performance in areas of language pragmatics, such as speech acts, figurative language, and inferred meanings (Gubelmann, 2024).

Despite those concerns, it seems inevitable that given its power, AI will become integrated in multiple ways into our social and educational fabric. The rapidly evolving AI landscape makes it foolhardy to predict just how AI will be used in language learning, but published studies provide some guidance. A starting point for teachers and students is to gain enough critical AI literacy to be able to function as informed consumers. That process involves gaining an understanding of how AI generates language through a statistical rather than linguistic model and the effect that has on the ability of AI to use language, as humans do, in socially and culturally appropriate ways (Godwin-Jones, 2024). Because AI has no lived experience in our world to go by, it can "hallucinate", producing statements that are false. A framework for developing critical AI literacy is outlined in Warschauer et al. (2023). Case studies of L2 learner interactions with AI are helpful in understanding the dynamics of AI-learner relationships. Jacob et al. (2023) describe how a Chinese graduate student ("Kailing") uses *ChatGPT* in her academic writing. Rather than relying on a single prompt to generate desired output, Kailing works through a series of prompts to generate initial ideas, to help in an initial draft, and to suggest revisions, always insisting on modifying output to align with her own voice and style. One of the takeaways from the study is the importance of working with students on the process of writing prompts and on the utility of the use of chain-of-thought prompting such as that used by Kailing. Ideas for the effective use of AI can be gained by studies of the use of machine translation in L2 teaching and learning. Studies have shown generally that students do not use *Google Translate* or other tools simply to translate a text but rather to look up individual words and phrases and to check the accuracy of individual sentences (Vinall & Hellmich, 2022). Those studies also include suggested activities for L2 writing, such as post-editing, comparing human and AI drafts, or analyzing the quality of AI-produced output. Some writing assignments may mandate the use of machine translation, while others are done in its absence, a strategy likely to be useful for AI as well.

Conclusion

The future of SLA in an AI world is unclear. We have seen AI integrated widely in commercial products and are beginning to see development in one of the promising directions, namely, in immersive technologies. Virtual reality has for some time held out immense promise for language learning. So far that promise has only partly been fulfilled: Role play activities in VR tend to be limited in terms of topics and user response (Godwin-Jones, 2023). Given the ability of AI to generate conversation on the fly on any given topic and to take on a variety of identities, one can imagine the powerful language learning potential of AI-based VR. Whether that vision becomes reality will depend on the possibility of seeing cheaper, less bulky VR headsets.

While advanced technology will take advantage of the power of AI to make language practice more dynamic, the artificiality of AI interactions will limit the real-world usefulness. Pragmatic and interactional competencies will not be easily developed through AI. That issue is even more pronounced for languages other than English. Language teachers who use AI will want to supplement their use with activities that engage learners in real human-to-human communication, such as virtual exchange or other online opportunities for meaningful human exchanges. Those non-AI, real-world activities also supply the opportunity to address issues of diversity and social justice, increasingly recognized as a legitimate goal of SLA (Ortega, 2017).

Cross-References

- ▶ A multimodal Language Pedagogy
- ▶ Approaches to Teaching L2 Pragmatics in CALL
- ▶ Augmented Reality
- ▶ Pedagogical Approaches in VE
- ▶ Pedagogical Approaches to Digital Game-Based Language Learning and Teaching
- ▶ Pedagogical Approaches to Digital Game-Based Language Learning and Teaching

▶ The Impact of Artificial Intelligence (AI) on CALL Pedagogies
▶ Virtual Reality-Assisted Language Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). Which humans? *PsyArXiv Preprints*. https://psyarxiv.com/5b26t/

Ellis, N. C. (2017). Cognition, corpora, and computing: Triangulating research in usage-based language learning. *Language Learning, 67*(S1), 40–65. https://doi.org/10.1111/lang.12215

Garzía, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan. https://doi.org/10.1057/9781137385765_4

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology, 27*(2), 6–27. https://hdl.handle.net/10125/73501

Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology, 28*(2), 5–30. https://doi.org/10125/73570

González-Lloret, M. (2019). Technology and L2 pragmatics learning. *Annual Review of Applied Linguistics, 39*, 113–127. https://doi.org/10.1017/s0267190519000047

Gubelmann, R. (2024). Large language models, agency, and why speech acts are beyond them (for now)–a Kantian-cum-pragmatist case. *Philosophy & Technology, 37*(1), 32.

Jacob, S., Tate, T., & Warschauer, M. (2023). Emergent AI-assisted discourse: Case study of a second language writer authoring with ChatGPT. *arXiv preprint*. https://doi.org/10.48550/arXiv.2310.10903

Levine, G. (2020). A human ecological language pedagogy. *Modern Language Journal, 104*(S1), 1–130. https://doi.org/10.1111/modl.12608

Ortega, L. (2017). New CALL-SLA research interfaces for the 21st century: Towards equitable multilingualism. *CALICO Journal, 34*(3), 283–316. https://doi.org/10.1558/cj.33855

Pegrum, M. (2019). Mobile AR trails and games for authentic language learning. In Y. Zhang & D. Cristol (Eds.), *Handbook of mobile teaching and learning* (2nd ed., pp. 1–16). Springer. https://doi.org/10.1007/978-981-13-2766-7_89

Poole, F. J., & Polio, C. (2023). From sci-fi to the classroom: Implications of AI in task-based writing. *TASK, 3*(2), 243–272.

Sauro, S. (2017). Online fan practices and CALL. *CALICO Journal, 34*(2), 131–146. https://doi.org/10.1558/cj.33077

Sockett, G. (2014). *The online informal learning of English*. Springer. https://doi.org/10.1057/9781137414885

Stokel-Walker, C. (2024, March 8). AI chatbot models 'think' in English even when using other languages, *NewScientist*. https://www.newscientist.com/article/2420973-ai-chatbot-models-think-in-english-even-when-using-other-languages/

Thorne, S. L., & Reinhardt, J. (2008). "Bridging activities," new media literacies, and advanced foreign language proficiency. *CALICO Journal, 25*(3), 558–572. https://doi.org/10.1558/cj.v25i3.558-572

Vinall, K., & Hellmich, E. (2022). Do you speak translate?: Reflections on the nature and role of translation. *L2 Journal, 14*(1), 4–25. https://doi.org/10.5070/l214156150

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for second language writers. *Journal of Second Language Writing, 62*, 101071. https://doi.org/10.1016/j.jslw.2023.101071

P

Pedagogical Innovation in Call

Hayo Reinders
King Mongkut's University of Technology Thonburi, Bangkok, Thailand

Abstract

The word innovation is used frequently in education circles, often without clarity around its intended meaning. In this entry we look at the different ways the term has been used and how innovation can be understood as a complex, localised construct. We look at ways in which innovation can be fostered and how it can be integrated into educational contexts.

Keywords

Innovation · Change · Novelty · Development · Stakeholders · Activity system

Introduction: What Is Innovation?

Despite its common usage in everyday language, it is surprisingly difficult to find agreement on what constitutes an innovation in CALL. 'Change' is perhaps the most commonly accepted attribute, but although every innovation involves change, of course not every change amounts to an innovation. Similarly, 'improvement' is what most innovations aspire to achieve, but this could operate at many different levels, from pedagogical practices to administrative processes, and is not always easy to operationalise. And in some cases, innovation is designed to avoid potential problems, so there is no directly observable advancement. 'Novelty' is also problematic, because what is new in one context is not new in another. Many innovations involve the 'development' of new ideas or solutions to existing problems, but an innovation can also involve the removal of existing practices or obstacles.

An early definition by Delano et al. (1994, p. 489) sees 'an innovation in a second language teaching program as an informed change in an underlying philosophy of language teaching/learning, brought about by direct experience, research findings, or other means, resulting in an adaptation of pedagogic practices such that instruction is better able to promote language learning as it has come to be understood'. In other words, innovation is characterised by conscious (and often proactive) processes of change aimed at augmenting understanding and practices that go beyond surface-level changes.

In addition, innovation is first and foremost situated in a particular educational context. For example, the introduction of tablets may be a decade past in most schools in affluent countries, but in many parts of the Global South they are just being introduced. It is also situated in the sense that it is rooted in specific pedagogical practices, which in some way need to be challenged. For example, a school that introduced tablets a long

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_14-1

time ago but never adapted its classroom practices has not achieved the same level of innovation as a school that has recently started using the devices to encourage learning beyond the classroom (one of the affordances of mobile technologies; Reinders & Pegrum, 2016). Innovation thus needs to be both implemented (otherwise it remains merely an idea) and integrated into genuine educational environments in ways that key stakeholders (teachers, learners, tech support, parents, etc.) can realise in practice and benefit from. In other words, 'solutions' are not solutions until they work in the context for which they were created. To that end, innovation is often more a dynamic process than a static product. This process begins with imagining what is possible and 'pushing the boundaries of existing practices and views of teaching and learning' (Siemens & Tittenberger, 2009, p. 20). The actual implementation has to take into account myriad factors that may facilitate or hinder its success. A process of trial and error is required to establish what these factors are. Structured approaches to this include the use of 'educational engineering' (Colpaert, 2016) and 'instructional design' (including models and methodologies such as ADDIE, SAMR and AGILE; see Sweller, 2021, for an overview). Many CALL projects have failed because they do not clearly distinguish between these different phases of the innovation process. We will focus on the integration of educational technology in the following section.

Fostering Pedagogical Innovation

Many attempted CALL innovations never launch or quickly disappear. Although figures for our field are, to the best of my knowledge, not available, Rogers (2003) estimates that across multiple disciplines on average 75% of innovations fail. Even if at a technical level innovations were successful in achieving their aims, their uptake remains low. There is now a considerable body of literature that has documented the cycles of innovation and the reasons why some projects are successful in the long term and others are not. Rogers (1995) proposes five attributes that determine whether an innovation will be adopted or rejected:

1. Measurability, of the relative advantages to be gained.
2. Compatibility, or how (dis)similar it is from current practice; projects that are evolutionary are usually more likely to be accepted (Reich, 2020).
3. Complexity, or how easy or challenging it is to adopt an innovation.
4. Trialability, or whether an innovation can be tried out before complete adoption.
5. Observability, or whether the innovation is visible to stakeholders. This also relates to the extent to which its benefits can be evaluated.

There are other factors that contribute to the likelihood that an innovation will be accepted by its intended users. For example, it is important that an innovation has multiple entry points, meaning that everybody can use the innovation for something, whether they are experts in the technology or completely new to it. Another factor is the extent to which the technology is taken out of the innovation, meaning that it seamlessly integrates into the educational environment and is seen to serve the interests and needs of its stakeholders. Of course, beyond adjusting the characteristics of the innovation and the way it is implemented, the organisation can ready itself for change.

This involves, amongst others, the provision of professional learning opportunities for staff, with an emphasis on enhancing initiative. Stoller (2009) talks about the need to see every innovation as an opportunity for professional development and a concomitant willingness to experience failure. Stoller also emphasises the need to nurture a sense of ownership, not only by those leading the innovation, but also by those who will be impacted by this. This final point reminds us of the need to recognise that people will be impacted differently by proposed innovations, with subjective perceptions playing a particularly powerful role. For example, an experienced teacher may see a junior colleague's attempt to introduce AI into the materials design process as a threat, even if

this is not at all so intended. For innovations to find fertile ground, an atmosphere of mutual trust and a willingness to collaborate needs to be cultivated across the organisation, with teachers being encouraged to assume forms of teacher leadership (Reinders, 2023).

Integrating CALL Innovation

Underpinning all of the above observations is that pedagogical innovation is best considered as a holistic and interrelated constellation of processes involving multiple stakeholders. This holistic view aligns quite closely with Engeström's description of complex organisations (such as schools) and their emergence and iteration as activity systems. Such systems have a number of key components, as shown in Fig. 1.

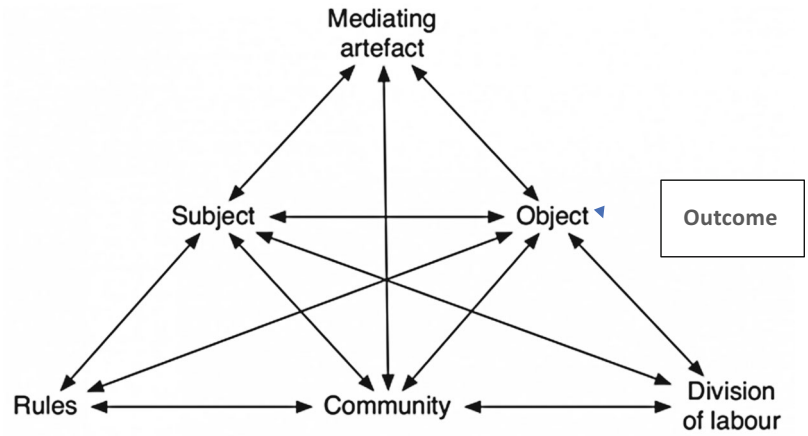
In the context of CALL innovation, in this figure 'subject' refers to the person or entity that initiates the change (e.g. a teacher, teaching team, principal, or Ministry of Education). The 'object' refers to what is to be changed and the intended result is the 'outcome'. For example, the introduction of digital storytelling in the curriculum, with the intention of enhancing learners' Willingness to Communicate. 'Community' covers the different stakeholders (e.g. learners; teachers; administrators and managers; parents; future employers; educational, legal, and regulatory bodies; teachers' associations) and 'division of labour' the roles played by different parts of the community and the activities they are expected to engage in. For example, in a development project, this concerns who is expected to be actively involved in, for example, the creation of new lesson plans using digital storytelling, whose approval needs to be obtained, who will be responsible for providing technical support, and so on. 'Rules' guide the way people are expected to work together (covering both formal educational policies and informal rules) and 'mediating artefacts' include the ways that the different elements of the organisation are supported in working together (e.g. having regular meetings; articulating shared values; encouraging open communication; creating an atmosphere of trust).

An Illustration

A lack of understanding of the system as a whole may adversely impact innovation. By way of illustration, consider one teacher's experience (vignette adapted from Reinders, 2023). Andrea (the 'subject') wants to improve language learning outcomes in her class. Specifically, after attending a conference she has become excited about the potential of using language exchange programs to encourage learners to interact more in the target language. She has learned that this can help lower learners' anxiety and enhance their L2 confidence (the 'outcome') so she decides to implement some activities around games in her 'Communication Skills' course and asks her students (the 'objects') to bring their cell phones to school.

Unfortunately, her good idea does not work out. Firstly, the WIFI at her school (one of the 'mediating artefacts' or tools, as are the cell phones) does not support 25 students all using the internet at once, something she had not anticipated. More worryingly, she receives an upset email from the IT department. Apparently, mobile phones are not supposed to be used without prior authorisation (she has broken the 'rules') and without adhering to the appropriate security and privacy protocols. The next day she is summoned to her manager's office. Apparently, all teaching innovations are coordinated by a dedicated central team (the 'Learning & Teaching Department', i.e. other members of the 'community') who do not approve of individual changes made by teachers. There is a process for new projects to be reviewed and approved, and their implementation is carefully monitored by the Quality Team ('division of labour'). Andrea is embarrassed and upset that her enthusiasm and good intentions have landed her in hot water. She vows to keep her head down and not take any more risks.

The above outcome is the worst possible, both for the teacher and the institution. To avoid this, better organisational knowledge would have helped the teacher know what processes to follow. By raising awareness and especially inducting new team members into the organisation, a great deal of confusion can be avoided. This is also



Pedagogical Innovation in Call, Fig. 1 Engeström's activity system

where collaboration between colleagues can help to anticipate challenges (and, more positively, opportunities). Adding a standing agenda item 'upcoming projects' to staff meetings can be a help, as can inviting colleagues from other departments (the wider 'community') to explain what objects and outcomes are important to them, what artefacts they use, how they divide the labour and how they understand the rules. If an organisation has a project office (or similar), its manager can be invited to explain the procedures for encouraging, inviting, approving, funding, monitoring, reporting and evaluating projects. Examples of successful initiatives can be shared.

Longer-term, CALL innovation can benefit from encouraging all stakeholders, including teachers such as Andrea, to display initiative, to raise ideas and to experiment. A rigidly hierarchical organisation where initiatives only come from the top down does not develop a culture of innovation and resilience, needed in the quickly changing landscape of education. Designating a particular course as one that can be used to rapidly generate and implement new ideas for observation can act as a type of 'sandbox' where risks can be taken safely, as can establishing a departmental project team whose purpose is to raise suggestions for improvement, and whose leader reports these 'up the chain' for feedback and (possibly) approval.

Conclusion

Pedagogical innovation in CALL clearly involves much more than simply offering an idea for a new tool or application. Beyond ideation, communication (with stakeholders), experimentation, dissemination, integration and evaluation are all critical to the long-term sustainable enhancement to the 'underlying philosophy of learning and teaching' (Delano et al., ibid.) that is sought. Activity systems theory offers one window onto the complex landscape that most educational environments comprise. Regardless of the perspective taken by stakeholders in CALL pedagogy, innovation involves a holistic process, the success of which depends on the extent to which the intended outcomes align with the community it serves.

Cross-References

- ▶ CALL Teacher Communities of Practice
- ▶ Complexities of Becoming a CALL Teacher
- ▶ Lessons Learned from Emergency Remote Teaching
- ▶ Professional Learning
- ▶ Technology Standards for Language Teachers

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Colpaert, J. (2016). Big content in an Educational Engineering approach. *Journal of Technology & Chinese Language Teaching, 7*(1), 1–14.

Delano, L., Riley, L., & Crookes, G. (1994). The meaning of innovation for ESL teachers. *System, 22*(4), 487–496.

Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.

Reinders, H. (2023). *From teacher to teacher leader*. Cambridge University Press.

Reinders, H., & Pegrum, M. (2016). Supporting language learning on the move. An evaluative framework for mobile language learning resources. In B. Tomlinson (Ed.), *Second language acquisition research and materials development for language learning* (pp. 221–233). Taylor & Francis.

Rogers, E. M. (2003). *Diffusion of innovation* (5th ed.). Free Press.

Siemens, G., & Tittenberger, P. (2009). *Handbook of emerging technologies for learning*. University of Manitoba.

Stoller, F. L. (2012). Innovation as the hallmark of effective leadership. In *Leadership in english language education* (pp. 73–84).

Sweller, J. (2021). Instructional design. In *Encyclopedia of evolutionary psychological science* (pp. 4159–4163). Springer International Publishing.

P

Project-Based Approaches to CALL

Gulbahar Beckett[1] and Michael Thomas[2]
[1]English, Iowa State University, Ames, IA, USA
[2]School of Education, Liverpool John Moore's University, Liverpool, Merseyside, UK

Abstract

With its roots in the work of the American philosopher of education John Dewey, project-based language learning (PBLL) has become increasingly prominent over the last several decades due to its interactionist and collaborative orientation, which fits well with contemporary approaches to second language acquisition. The turn towards the use of digital technologies from the start of the new millennium afforded students and teachers further opportunities to explore how PBLL can benefit from their communicative and collaborative aspects in such domains as telecollaboration. More recent research on technology-mediated project-based language learning has turned to consider how projects can engage students in addressing social justice issues and grand challenges, such as the climate crisis. While PBLL faces limitations in traditionally testing-oriented educational systems, the more widespread use of digital technologies has also challenged language educators and policymakers to consider the role of project-based approaches in defining the future role of language learning.

Keywords

Computer-assisted language learning · Digital projects · Language acquisition · Language learning · Project-based language learning

Introduction

With its roots in Dewey's experiential philosophy of learning (Dewey & Dewey, 1915), multiple definitions of project-based language learning (PBLL) have emerged over the last three decades. For Jerald and Clarke (1983), the goal of PBLL is to create extended input and output opportunities for learners to develop target language accuracy and fluency in authentic situations, typically through group work. Beckett (1999) defined PBLL as:

> a series of individual or group activities that involve language/content learning through planning, researching (empirical and/or document), analysing and synthesizing data, and reflecting on the process and product orally and/or in writing by comparing, contrasting, and justifying alternatives. (p. 4)

Most definitions agree that projects typically involve a series of related tasks that students design and implement under the guidance of their teachers by solving problems, thinking critically, and using strategies and language to learn disciplinary knowledge. According to Dooly and

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_18-1

Masats (2011), "PBLL obliges both the teacher and the students to conceptualize the *whole* learning process as one embedded in a wider context of linguistic and non-linguistic objectives, content, and output" (p. 2; italics added).

More recent definitions of PBLL recognize the growing importance of computer-assisted language learning (CALL). In this sense, technology-mediated project-based language learning aims to explore how digital technologies can be used to enhance learner creativity (Joo & Jin, 2024). This emphasis is reflected in the definition from Dooly and Masats (2020) which suggests that the use of digitally mediated PBLL is an appropriate educational approach for twenty-first century skills acquisition. Likewise, Thomas and Yamazaki (2021) defined PBLL that uses CALL technologies as a student-centred pedagogical approach that has "larger and wider, cross-disciplinary aims and objectives" (p. 4).

Key Themes in the Research Literature

Technology-Mediated Project-Based Language Learning

According to Castro Huércano (2021), students find technology-mediated project-based learning motivating. The growing prominence of studies has suggested that technology-mediated PBLL can also promote multiliteracies and multimodal learning to equip students with the knowledge and skills necessary for the rapidly changing digital world. For example, Carpenter and Matsugu (2020) illustrated how Japanese students utilized smartphones to successfully complete their drama projects. Canadian students in Dressler et al. (2020) utilized Web 2.0 tools for implementing and showcasing their German culture projects. Martin (2020) demonstrated how video-game technology-mediated projects motivated first-year medical students in Japan to learn medical knowledge and vocabulary and think critically. Saricaoglue and Geluso (2020) discussed a cross-cultural collaboration of American and Turkish students in a telecollaborative first and second language acquisition project and found that student pairs used digital tools creatively to successfully complete their projects. Two ethnographic studies conducted by Dooly and Masats (2020) in primary and middle school contexts explored classroom interactions of young EFL learners who were participating in telecollaborative projects. Analysis of naturally occurring data suggested that students completed their digital projects while practicing grammatical structures naturally, meaningfully and authentically.

Arguing that technology has the potential to transform language pedagogy, Thomas (2017) investigated a collaborative technology-mediated PBLL approach involving undergraduate English major students at a Japanese private university in one of the first research-based monographs on the subject. The participants' average English proficiency level was A2 on the Common European Framework of Reference (CEFR). To address disengagement, two learner-centred constructivist technology-mediated language learning projects composed of extended tasks were enacted as an alternative to high-stakes testing. Based on data from classroom observations, semi-structured individual interviews, information discussions, video recordings of projects, and researcher field notes, findings revealed that compared to the traditional English class activities, students found project tasks more engaging and meaningful and that technology-mediated projects using Web 2.0 technologies afforded learners the opportunity to confirm equalization in terms of agency and creativity.

On the other hand, research has also shown that technology-mediated PBLL can be challenging if students and teachers lack digital literacies (Anderson & Macleroy, 2021). Additionally, there is a paucity of research on teacher training on technology-mediated PBLL. In response, Garib (2023) conducted a phenomenological study examining Lebanese, Libyan, and Syrian EFL teachers' perceptions before and after the implementation of their projects. The participants reported that PBLL made their students' learning interactive and engaging and they thought that it could help develop learner autonomy. Garib (2023) also found that "technology in under-resourced contexts can be difficult to use due to

slow Internet connections, power outages, and the students' unfamiliarity with technology" (p. 1450), resulting in hesitations to incorporate technology.

Social Justice Project-Based Learning with Technology

To explore the potential of PBLL and the need to address social justice issues, calls for functional and critical project-based learning and teaching have also been made (Beckett, 2024; Slater & Beckett, 2019, 2024; Thomas & Yamazaki, 2021). For example, Slater and Beckett (2019) outlined a framework for aiding the functional teaching and learning of English with technology which emphasized the importance of an active approach that adopted inquiry-based learning. Beckett's (2024) framework addressed social studies content acquisition, critical thinking skills development, and associated language learning with technology. Slater and Beckett (2024) similarly offered another example of technology-mediated PBLL for content acquisition and the development of critical thinking.

While the use of digital technologies has afforded significant opportunities for teachers and learners across all sectors, their advocacy and implications have also been read against the socio-economic context of the last four decades, especially in Western societies where university education in particular has become strongly connected to a skills and employability agenda. Advancements in technology need also to be understood in the context of neoliberalism and new managerialism in education and research has explored how this context may contribute to the digital divide and socioeconomic injustices (Ortega, 2017). Additionally, students can benefit from being taught about digital literacy to build their deep knowledge of digital competency. Beckett (2024) suggests that digital literacies are required by students to learn ethics and to explore their roles and responsibilities in digital societies and proposes the use of projects in the language classroom to explore how neoliberal language policies and practices may contribute to the hegemony of the English language, displacing politically powerless indigenous languages.

Thomas and Yamazaki's (2021) edited volume was devoted to social justice topics in its exploration of technology-mediated PBLL. In it, Hernandez Avarado and Brinkwirth's (2021) telecollaborative action research project carried out by students in Mexico and the United States explored complex social issues presented in films. Students used VoiceThread, WhatsApp, Zoom, and Skype for communication and image and video creation for their social justice project covering topics such as "illegal immigration, mistreatment, objectification of women, marginalization of indigenous peoples, gang violence, human trafficking, government corruption, and other topics relevant to Mexico and the United States" (p. 215). The project work raised students' awareness of social justice issues, helped resolve misconceptions and stereotypes, and encouraged students to become more tolerant and respectful of diversity.

In the same volume, Anderson and Macleroy (2021) discussed a digital connection project developed in recognition of inadequate support for students' increasingly marginalized home languages and cultures. The project brought together young people in the UK, Algeria, Cyprus, Luxemburg, Palestine, Taiwan, and the United States and involved primary and secondary schools in mainstream and supplementary weekend schools. Digital and multilingual story ideas were created in negotiation with students who carried out drama, fantasy, science fiction, traditional tales, and documentary projects multimodally and shared in their classrooms, schools, film festivals, and on their project websites. Findings suggested that digital storytelling projects motivated students to learn new language and media skills while engaging in authentic content knowledge acquisition.

Summary and Implications for the Future

Technology has assisted and mediated PBLL for various purposes ranging from spell and grammar checks to critical thinking and raising awareness of social justice issues. The research has also shown that projects can be motivating and

promote greater equality between learners, while also serving as activities in which subject content knowledge is acquired and skills are learned.

Technology-mediated PBLL can also be used to help students address the challenges that new technologies such as artificial intelligence (AI) present. However, AI also has implications for social justice and climate change due to the complex neoliberal context of contemporary educational technology (Williamson, et al., 2023). A PBLL approach could be used to enable students to consider these issues. For example, students could design and implement projects to explore the neoliberal implications of AI or learn about the environmental impacts of the databases that support Large Language Models (LLM) necessary for AI. Teachers and students can also work on projects to learn about the ethical issues regarding the fair use of intellectual properties concerning AI (Beckett, 2024). These cross-disciplinary projects could take place in political science, language policy, and/or social studies content courses as well as aiding foreign language development.

Such approaches to technology-mediated PBLL can be undertaken across a variety of language learning levels (Thomas, 2017) and hold significant promise for situating language learning at the centre rather than the periphery of educational curricula. This is especially important at a time when language departments remain at risk of closure around the world. Cross-disciplinary projects involving digital technologies demonstrate how language learning can deal with 'big issues', 'grand challenges' and 'wicked problems', thereby bringing learners together to improve intercultural understanding, as well as develop subject-based and critical competencies.

Cross-References

- ▶ Critical Approaches to CALL
- ▶ Digital Literacies
- ▶ Immersive Literacy
- ▶ Language Teacher Education for CALL
- ▶ Pedagogical Innovation in CALL
- ▶ The Impact of AI on CALL Pedagogies
- ▶ Using Virtual Exchanges in the Second Language Classroom
- ▶ Videoconferencing in Second Language Learning
- ▶ Virtual Worlds

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anderson, J., & Macleroy, V. (2021). Stories, communities, voices: Revitalizing language learning through digital media within a project-based pedagogical framework. In M. Thomas & K. Yamazaki (Eds.), *Project-based language learning and CALL: From virtual exchange to social justice* (pp. 231–262). Equinox Publishing.

Beckett, G. H. (1999). *Project-based instruction in a Canadian secondary school's ESL classes: Goals and evaluations* (Unpublished doctoral dissertation). The University of British Columbia.

Beckett, G. H. (2024). *A paradigm shift for language education: Project-based learning and teaching*. Routledge.

Carpenter, J., & Matsugu, S. (2020). Translanguaging in project-based learning. In G. H. Beckett & T. Slater (Eds.), *Global perspectives on project-based language learning, teaching, and assessment: Key approaches, technology tools, and frameworks* (pp. 49–68). Routledge.

Castro Huércano, C. (2021). Project based learning: A quest for motivation when teaching Spanish as a Second Language at an Australian University (Unpublished doctoral dissertation). Flinders University.

Dewey, J., & Dewey, E. (1915). *Schools of tomorrow*. J. M. Dent & Sons, Ltd.

Dooly, M., & Masats, D. (2011). Closing the loop between theory and praxis: New models in EFL teaching. *ELT Journal, 65*(1), 42e51.

Dooly, M., & Masats, D. (2020). 'What do you zinc of the project': Examples of technology enhanced project-based language learning. In G. H. Beckett & T. Slater (Eds.), *Global perspectives on project-based language learning, teaching, and assessment: Key approaches, technology tools, and frameworks* (pp. 126–145). Routledge.

Dressler, R., Raedler, B., Dimitrov, K., Dressler, A., & Krause, G. (2020). Project-based learning in the advanced German class: Integrated content and language learning. In G. H. Beckett & T. Slater (Eds.), *Global perspectives on project-based language learning, teaching, and assessment: Key approaches, technology tools, and frameworks* (pp. 60–84). Routledge.

Garib, A. (2023). 'Actually, it's real work': EFL teachers' perceptions of technology assisted project-based language learning in Lebanon, Libya, and Syria. *TESOL Quarterly, 57*(4), 1434–1462.

Hernandez Avarado, M. G., & Brinkwirth, A. T. (2021). Transcultural language learning through a cinema and social justice teletandem program. In M. Thomas & K. Yamazaki (Eds.), *Project-based language learning and CALL: From virtual exchange to social justice* (pp. 201–230). Equinox Publishing.

Jerald, M., & Clarke, R. (1983). *Experiential language teaching techniques: Resources handbook number 3*. Pro Lingua Associates.

Joo, C. S., & Jin, C. L. (2024). Expanding horizons: Fostering creativity and curiosity through spherical video-based virtual reality in project-based language learning. *TESOL Quarterly, 58(2). Article e3333*.

Martin, B. (2020). Plagues infecting a medical school: A case study for authentic video games to promote motivation, engagement, and acquisition. In G. H. Beckett & T. Slater (Eds.), *Global perspectives on project-based language learning, teaching, and assessment: Key approaches, technology tools, and frameworks* (pp. 85–103). Routledge.

Ortega, L. (2017). New CALL-SLA research interfaces for the 21st century: Towards equitable multilingualism. *CALICO Journal, 34*(3), 285–316.

Saricaoglue, A., & Geluso, J. (2020). Students co-learning linguistics through PBL: A cross-cultural telecollaborative implementation. In G. H. Beckett & T. Slater (Eds.), *Global perspectives on project-based language learning, teaching, and assessment: Key approaches, technology tools, and frameworks* (pp. 104–125). Routledge.

Slater, T., & Beckett, G. H. (2019). Integrating language, content, technology, and skills development through project-based language learning: Blending frameworks for successful unit planning. *MEXTESOL Journal, 43*(1) https://www.mextesol.net/journal/index.php?page=journal&id_article=5557

Slater, T., & Beckett, G. H. (2024). Developing critical thinking and language skills through project-based learning and the knowledge framework. In S. Ali & B. M. Berg (Eds.), *The content area language and literacy open education* (pp. 117–135). University of Northern Colorado.

Thomas, M. (2017). *Project-based language learning with technology: Learner collaboration in an EFL classroom in Japan*. Routledge.

Thomas, M., & Yamazaki, K. (2021). *Project-based language learning and CAL: From virtual exchange to social justice*. Equinox Publishing.

Williamson, B., Macgilchrist, F., & Potter, J. (2023). Re-examining AI, automation and datafication in education. *Learning, Media & Technology, 48*(1), 1. https://doi.org/10.1080/17439884.2023.2167830

T

Technology Enhanced Place-Based CALL

Liudmila Klimanova[1] and
Lara Lomicka Anderson[2]
[1]University of Arizona, Tucson, AZ, USA
[2]University of South Carolina, Columbia, SC, USA

Abstract

Learning about a culture does not happen in isolation from the physical places and spaces and ecologies where culture is maintained, practiced, and/or contested. Places have a profound impact on both what and how we learn. This entry highlights place-based concepts, research and applications in language learning and discusses technology-enhanced place-based language learning (TEPBLL). Place-based teaching in CALL helps instructors and students to examine how local heritage, cultural practices, and linguistic landscapes can serve as the basis for teaching languages and intercultural concepts that constitute *L2 geoliteracy.*

Keywords

CALL · Augmented reality · Place-based learning · GIS technology · Mapping · Geoliteracy technology-enhanced place-based language learning (TEPBLL)

Introduction

Language and culture are intrinsically tied to the spaces and places where they are used, developed, and sometimes challenged. The significance of place in learning is profound, yet traditional language classrooms often present an oversimplified or abstract view of culture and language, detached from their real-world contexts (Holden & Sykes, 2011; Kramsch, 2014).

Place-based education offers a solution to this disconnect by emphasizing the importance of local environments, contexts, and places in the language learning process. This approach encourages instructors to move beyond generalized notions of monolithic cultures and instead engage students in developing a more nuanced and comprehensive understanding of the target language and culture. By exploring the specific character, history, and needs of a place via physical and digital experiences with places, students can gain a richer, more vivid perspective on the language and its speakers (Demarest, 2014; Gruenewald, 2003; Klimanova & Hellmich, 2021a). This localized approach to language education allows learners to challenge preconceived notions and stereotypes, fostering a more authentic and holistic understanding of the culture associated with the target language. By grounding language learning in real places and contexts, students can better appreciate the complexities and diversities within a culture, moving beyond simplistic generalizations and towards a more nuanced cultural competence.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_19-1

Key Concepts and Theories Underlying TEPBLL

Place-based language learning emphasizes the importance of context and environment in acquiring language skills, where "place" is understood as both a physical and cultural entity. It is seen as "physical, material, and stable (as walls, buildings, streets, billboards, and signs seem stable and concretely distinguishable 'places') and a product of institutional orders and cultural frames" (Klimanova & Hellmich, 2021a, p. 275). The concept of "physical place" is closely tied to the notion of "space," which is regarded as a nexus for cultural and linguistic practices. As Lefebvre (1992) suggests, space is a "practiced place," where language and culture are not only shaped but also experienced. In this context, *spatialization* as a model of spatial orientation in language education underscores the links between learning activities and authentic semiotic resources found in the material environment. These environments are seen as essential to embedding communicative practices within material, social, geographical, and historical contexts (Canagarajah, 2018; Malinowski et al., 2023).

The material and experiential contexts are central to *semiotic* and *ecological* perspectives on language learning, wherein environments are viewed as providing a *semiotic budget* for meaningful actions and meaningful learning experiences (van Lier, 2000, p. 252, 2003, Hauck & Satar, 2018; Satar et al., 2023). From this viewpoint, place-based language learning recognizes that language acquisition occurs through the perception, selection, and interpretation of physical signs and symbols within specific spaces (van Lier, 2004). In such environments, signs manifest in three distinct levels of signification: *iconicity, indexicality*, and *symbolicity* (van Lier, 2004; Levine, 2020). These levels of signification link language use with the physical, social, and digital worlds, forming a rich tapestry of meaning-making systems. By actively engaging learners in the process of *semiosis*—the production and interpretation of semiotic artefacts—educators can create learning environments that provide physical, social, and symbolic opportunities for meaning acquisition (Wei & Ho, 2018). Perceptive learning, therefore, becomes crucial in these contexts. Van Lier (2004) argues for an expanded definition of perception, one that includes *perceptual activity*. This view highlights how learners actively explore and engage with their environments, making connections between various perceptual inputs and language use. Learners, in this framework, are seen as active explorers, exposed to multisensory stimuli that enhance their ability to derive meaning across different linguistic experiences. Moreover, Gee (2016) describes the manipulation of semiotic resources for language teaching as "designed experiences" that are intended to elicit specific learning outcomes. These experiences often contribute to broader changes in learners' social behaviour, emotions, and attitudes. In digital games, for instance, learners are offered critical learning and aesthetic experiences that foster new insights and forms of appreciation. With the rise of digital language contexts, learning environments have become increasingly multimodal (Ho, 2019). Each mode—language, image, sound, and colour—plays a role in meaning-making and language acquisition (Kress, 2010). In computer-assisted language learning (CALL) environments, the design of learning spaces and the semiotic resources employed can significantly impact how learners access and use information (Darvin, 2023). Adopting a multimodal semiotic lens in place-based language learning research and pedagogy allows for a deeper understanding of how digital environments, with their multilingual, multimodal, and multisemiotic repertoires, enhance language and cultural learning practices.

Areas of Research in TEBPLL

Technology-enhanced place-based language learning (TEBPLL) research and practice highlight the potential of *geolocativity* to connect learners with specific spaces and places, leveraging these environments to create interactive and playful learning experiences. As learners engage with these locations, their activities and presence can transform the meaning and perception of these social spaces. The concept of *L2 geoliteracy*

reflects learners' ability to observe, notice, and interpret language use in a cultural setting within a local community. This concept involves not only understanding geographical concepts such as location, environment, spatial relationships, and cultural contexts in the target language but also interpreting how language use is influenced by the space where it occurs. *L2 geoliteracy* integrates language learning with geographic literacy, allowing learners to engage with geographical data, maps, and spatial analysis in the target language, ultimately enhancing linguistic proficiency, intercultural competence, and geographic awareness. While TEBPLL remains an emerging field in language education with limited documentation, studies are beginning to recognize the importance of teaching *L2 geoliteracy* skills and exploring various subcategories that position learning within semiotic spaces, both virtual and physical. These subcategories include GIS mapping, augmented reality/place-based learning, linguistic landscapes, and virtual reality, each offering unique approaches to immersing learners in meaningful environments and enhancing language acquisition through technology-mediated, place-based experiences.

GIS Mapping

GIS mapping is a technological tool that allows the user to capture, store, manipulate, analyse and visualize geographic data to reveal trends, patterns and relationships. Within language education, it offers significant potential for enhancing learning experiences, despite being underutilized. It allows users to add or view geographic information on specific topics, enabling deep mapping that expresses and explores spatial narratives through various media layers. This approach facilitates spatial thinking, crucial for understanding our complex world, and can be implemented in language classrooms to help students explore surroundings, create unique maps, gain geospatial thinking skills, and visualize culture in novel ways. Research on GIS mapping in language education, while limited, shows promising results. Studies have demonstrated its effectiveness in cross-disciplinary projects, such as investigating population patterns in Paris (Goldfield & Schlichting, 2007) or creating digital itineraries for study abroad programs in Germany (Apgar, 2018). These applications have enhanced students' language skills, cultural understanding, and engagement. Projects using GIS-enabled apps have also shown potential in developing cultural awareness both at home and abroad (Guerrero-Rodríguez et al., 2022). The MAPS model by Klimanova and Hellmich (2021b) further illustrates how GIS mapping can foster digital literacies and semiotic awareness. However, the field still faces challenges due to software complexity and lack of educator training, indicating a need for more focused research on the role of semiotics in GIS mapping for language learning.

Augmented Reality

Place-based education and augmented reality (AR) offer significant potential for language learners by situating tasks and communication in local spaces. AR, an interactive technology that enhances real-world interactions, overlays computer-generated digital information into a real-time environment. Place-based learning can use AR to connect students to their communities, helping them develop communication skills and critical thinking while engaging with their environment (Matthews & Holden, 2018). It contextualizes learning around real-world issues and encourages knowledge building rather than passive information consumption. Zimmerman and Land (2014) note that place-based education transforms abstract knowledge into local knowledge related to community practices. Several studies have explored place-based AR in language learning. Holden and Sykes (2011) created Mentira, a mobile AR game for Spanish learners that engaged students with local communities. Holden (2016) found that AR games cater to diverse learning styles and foster collaboration. Perry (2015) developed Explorez, an AR game for French learners that turned a campus into a virtual francophone world, creating a relevant learning environment for participants.

Linguistic Landscapes

Linguistic landscapes are "the assemblage of visible, audible, and otherwise textualized languages

of public space" (Malinowski, 2015, 2016; Malinowski et al., 2023, p. 4). Language teaching approaches utilizing linguistic landscapes can enhance students' language awareness and critical literacy skills. By examining public spaces, students can analyse how different actors influence language use and develop pragmatic competence in context-appropriate communication (Byrd Clark, 2016). Linguistic landscape research also supports the development of translingual competencies, merging semiotic resources with diverse social and material ecologies (Canagarajah, 2018).

Aladjem and Jou (2016) demonstrated the potential of using mobile devices and social media to document and interact with linguistic landscapes, finding that students assumed roles as explorers in authentic contexts while receiving virtual support from peers. Focusing on local linguistic landscapes as an activity within a telecollaborative exchange was found to provide an opportunity for learning about language diversity from an intercultural perspective (Vinagre & Llopis-Garcí, 2023). Malinowski and Dubreil (2019) suggest future research directions, including investigating how language learners acquire and appropriate discursive forms from linguistic landscapes and exploring the concept of distributed spatial repertoires for meaning-making. These approaches highlight the potential of linguistic landscapes to enhance language learning and critical awareness.

Virtual Reality

Virtual reality (VR) has emerged as a promising tool in language learning, offering immersive experiences that simulate authentic contexts beyond traditional classroom settings (Blyth, 2018; Xie et al., 2019). VR provides a sense of telepresence, allowing learners to feel as if they are sharing real spaces with others (Kramsch & Andersen, 1999). Recent studies have explored VR's potential in language education, highlighting its benefits such as promoting cultural awareness, offering personalized learning paths, and supporting learners with special needs (Berti, 2019; Lin & Lan, 2015). Research by Berti et al. (2020) and Xie et al. (2019) demonstrates VR's effectiveness in enhancing language learning experiences, particularly in cultural understanding and vocabulary acquisition. Parmaxi's (2020) review identifies 13 potential directions for future VR research in language learning, including real-life task design, cross-disciplinary approaches, and the development of fully immersive virtual experiences. The integration of VR with other technologies like 3D learning objects and mixed reality spaces offers multisensory opportunities that can engage learners in novel ways (Campbell et al., 2016; Mangina, 2017). Overall, VR's ability to unite people with spaces and places through *geolocativity* presents exciting possibilities for interactive and contextualized language learning experiences.

Conclusion

Place-based language learning offers a rich context for developing *semiotic competence*, as it allows learners to engage with semiotic resources embedded in local environments. This approach can transform learners' engagement with language, creating new layers of competence that bridge linguistic and cultural understanding. In the context of computer-assisted language learning (CALL), this perspective encourages a shift towards examining sign systems within locally construed contexts. Future research in this field could focus on developing frameworks that incorporate semiotic competence and explore semiosis in contextual spaces and places. This includes investigating how learners produce and interpret semiotic resources or artefacts within specific environments. Additionally, researchers should examine perceptive learning in L2 settings, exploring spaces for visual and auditory information that create multisensory experiences. This approach acknowledges the multimodal and visual nature of modern communication, recognizing language learning as a hyper-semiotic process where various modes interact to create meaning. By focusing on spaces and places of learning—whether in classrooms, homes, or abroad—and examining linguistic landscapes, researchers can expand our understanding of language acquisition in context. This place-based

approach to language learning research should embrace a wider diversity of theoretical, methodological, and research-design choices, paving the way for innovative ideas, frameworks, and theories that better reflect the complex, multimodal nature of language learning in real-world contexts.

Cross-References

- ▶ Augmented Reality (AR)
- ▶ Linguistic Landscapes
- ▶ Semiotic Approaches to CALL
- ▶ Virtual Reality
- ▶ Virtual Reality-Assisted Language Learning (VRALL)

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Aladjem, R., & Jou, B. (2016). The linguistic landscape as a learning space for contextual language learning. *Journal of Learning Spaces, 5*(2), 66–70.

Apgar, R. B. (2018). Emplacing Berlin: Digital itineraries and short-term study abroad in the city. *International Journal of Humanities and Arts Computing, 12*(1), 26–36. https://doi.org/10.3366/ijhac.2018.0204

Berti, M. (2019). Italian open education: Virtual reality immersions for the language classroom. In A. Comas-Quinn, A. Beaven, & B. Sawhill (Eds.), *New case studies of openness in and beyond the language classroom* (pp. 37–47). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.37.965

Berti, M., Maranzana, S., & Monzingo, J. (2020). Fostering cultural understanding with virtual reality: A look at students' stereotypes and beliefs. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 10*(1), 47–59. https://doi.org/10.4018/IJCALLT.2020010104

Blyth, C. (2018). Immersive technologies and language learning. *Foreign Language Annals, 51*, 225–232.

Byrd Clark, J. (2016). Transdisciplinary approaches to language learning and teaching in transnational times. *L2 Journal, 8*, 3–19.

Campbell, A., Santiago, K., Hoo, D., & Mangina, E. (2016). *Future mixed reality educational spaces* (pp. 1088–1093). https://doi.org/10.1109/FTC.2016.7821738

Canagarajah, S. (2018). Translingual practice as spatial repertoires: Expanding the paradigm beyond structuralist orientations. *Applied Linguistics, 39*, 31–54.

Darvin, R. (2023). Sociotechnical structures: The entanglement of multimodality and materiality in the location of information online. *Language Learning & Technology, 27*(2), 28–45.

Demarest, A. B. (2014). *Place-based curriculum design: Exceeding standards through local investigations*. Routledge.

Gee, J. P. (2016). Video games, design, and aesthetic experience. *Rivista di estetica, 63*, 149–160. https://doi.org/10.4000/estetica.1312

Goldfield, J., & Schlichting, K. (2007). Foreign language and sociology: Exploring French society and culture. In D. S. Sinton & J. Lund (Eds.), *Understanding place: GIS and mapping across the curriculum* (pp. 228–235). ESRI Press.

Gruenewald, D. A. (2003). Foundations of place: A multidisciplinary framework for place-conscious education. *American Educational Research Journal, 40*(3), 619–654.

Guerrero-Rodríguez, P., Lomicka Anderson, L., & Lord, G. (2022). SIFTR-ing through the development of cultural awareness at home and abroad. *Foreign Language Annals, 55*, 435. https://doi.org/10.1111/flan.12617

Hauck, M., & Satar, M. (2018). Learning and teaching languages in technology-mediated contexts: The relevance of social presence, copresence, participatory literacy and multimodal competence. In R. Kern & C. Develotte (Eds.), *Online multimodal communication and intercultural encounters* (pp. 133–157). Routledge.

Ho, W. Y. (2019). Self-directed language learning: A semiotic analysis of a language learning app. In *Virtual sites as learning spaces* (pp. 295–327). Springer. https://doi.org/10.1007/978-3-030-26929-6_11

Holden, D. (2016). Technology: An analysis of an original MALL game and its redesign for an ESL course. *Second Language Studies, 35*(1), 1–52.

Holden, C., & Sykes, J. (2011). Leveraging mobile games for place-based language learning. *International Journal of Game-Based Learning, 1*(2), 1–18.

Klimanova, L., & Hellmich, E. A. (2021a). Crossing transcultural liminalities with critical virtual exchange: A study of shifting border discourses. *Critical Inquiry in Language Studies, 18*(3), 273–304. https://doi.org/10.1080/15427587.2020.1867552

Klimanova, L., & Hellmich, E. A. (2021b). Putting local on the MAP: A model for engaging foreign language students with local cultures. *Foreign Language Annals, 43*(4), 158. https://doi.org/10.1111/flan.12493

Kramsch, C. (2014). Teaching foreign languages in an era of globalization: Introduction. *Modern Language Journal, 98*(1), 296–311. https://doi.org/10.1111/j.1540-4781.2014.12057.x

Kramsch, C., & Andersen, R. W. (1999). Teaching text and context through multimedia. *Language Learning & Technology, 2*(2), 31–42.

Kress, G. (2010). *Multimodality a social semiotic approach to contemporary communication*. Routledge.

Lefebvre, H. (1992). *The production of space*. Wiley-Blackwell.

Levine, G. S. (2020). A human ecological language pedagogy. *Modern Language Journal, 104*(S1, monograph), 1–130.

Lin, T. J., & Lan, Y. J. (2015). Language learning in virtual reality environments: Past, present, and future. *Educational Technology & Society, 18*(4), 486–497.

Malinowski, D. (2015). Opening spaces of learning in the linguistic landscape. *Linguistic Landscape, 1*(1), 95–113.

Malinowski, D. (2016). Localizing the transdisciplinary in practice: A teaching account of a prototype undergraduate seminar on linguistic landscape. *L2 Journal, 8*(4), 100–117.

Malinowski, D., & Dubreil, S. (2019). Linguistic landscape and language learning. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics*. Wiley.

Malinowski, D., Maxim, H. H., & Dubreil, S. (2023). Introduction: Spatializing language studies in the linguistic landscape. In *Spatializing language studies: Pedagogical approaches in the linguistic landscape* (pp. 1–17). Springer.

Mangina, E. (2017, October). 3D learning objects for augmented/virtual reality educational ecosystems. In *2017 23rd International Conference on virtual system & Multimedia (VSMM)* (pp. 1–6). IEEE.

Matthews, J., & Holden, C. (2018). The design and play of geogames as place-based education. In O. Ahlqvit & C. Schieder (Eds.), *Geogames and geoplay* (pp. 161–176). Springer.

Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*, 172. https://doi.org/10.1080/10494820.2020.1765392

Perry, B. (2015). Gamifying French language learning: A case study examining a quest-based, augmented reality mobile learning-tool. *Procedia – Social and Behavioral Sciences, 174*, 2308–2315.

Satar, M., Hauck, M., & Bilki, Z. (2023). Multimodal representation in virtual exchange: A social semiotic approach to critical digital literacy. *Language Learning & Technology, 27*(2), 72–96.

van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245–259). Oxford University Press.

van Lier, L. (2003). An ecological-semiotic perspective on language and linguistics. In C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 140–164). Continuum.

van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective* (1st ed.). Kluwer Academic.

Vinagre, M., & Llopis-Garcí, R. (2023). Multilingual landscapes in telecollaboration: A Spanish-American exchange. In *Spatializing language studies: Pedagogical approaches in the linguistic landscape* (pp. 67–86). Springer.

Wei, L., & Ho, W. Y. (2018). Language learning Sans Frontiers: A translanguaging view. *Annual Review of Applied Linguistics, 38*, 33–59. https://doi.org/10.1017/S0267190518000053

Xie, Y., Ryder, L., & Chen, Y. (2019). Using interactive virtual reality tools in an advanced Chinese language class: A case study. *TechTrends, 63*, 251–259.

Zimmerman, H. T., & Land, S. M. (2014). Facilitating place-based learning in outdoor informal environments with mobile computers. *TechTrends, 58*, 7783.

U

Unlocking Language Learning with Immersive Technology

A Dive into Extended Reality (XR)

Yu-Ju Lan
Department of Chinese as a Second Language, National Taiwan Normal University, Taipei, Taiwan

Abstract

Extended reality (XR) is an umbrella term for immersive technologies that blur the line between reality and virtuality. XR can create a dynamic, interactive, and authentic learning environment that allows language learners to engage in life-like scenarios and practice using language in context-rich settings that enhance their comprehension and retention. These unique characteristics have resulted in XR being used in language learning for more than 2 decades, and the existing literature mostly describes it as having positive effects on language learning. However, the use of XR in language learning also poses some challenges that are of concern to both educators and researchers. Therefore, this entry explored both the potential and possible challenges of utilizing XR in language education, with the aim of paving the way for further research and the wider adoption of immersive technologies in language education.

Keywords

Extended reality · Immersive technologies · Classroom ecosystem · Learner-mediated learning · Learning agent

Introduction

Extended reality (XR) is an umbrella term that covers virtual reality (VR), augmented reality (AR), mixed reality (MR), and all other immersive technologies on the virtuality continuum (Milgram et al., 1995). VR is an artificial but authentic environment that involves presenting images and sounds created with computer hardware and software to the user so as to simulate a place or situation in which they can actively participate (Lan, 2020a, b; Onyesolu, 2009). In contrast, AR provides a view of the real world augmented with digital information: either direct (e.g., using a head-mounted display) or indirect (displayed on a phone or tablet screen). AR integrates the real and virtual worlds to allow users to interact with both virtual and physical objects in real time (Azuma et al., 2001). MR is a combination of AR and VR that integrates the real and virtual worlds to create a new environment with a virtual representation that matches our natural way of seeing things. Sometimes referred to as AR 2.0 (Mubin et al., 2022), MR allows real-world objects to seamlessly coexist with digital objects, creating a real-time interactive experience (Tamura et al., 2001).

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_21-1

XR alters reality by adding any number of digital elements to the physical or real-world environment so as to blur the line between the physical and digital worlds (Chuah, 2019; Wiles, 2022). Although immersive technologies have attracted the attention of both researchers and educators for a long time, their interest appeared to increase after the owner of Facebook changed its name to "Meta" in 2021. XR can create real-life situations that allow learners to experience, explore, create, interact, and collaborate with others using their avatars that provide a sense of presence and identity. There have been numerous applications of XR, including in healthcare education (e.g., Logeswaran et al., 2021), social networking (e.g., Grant et al., 2024), art creativity (e.g., Yang et al., 2023), simulation (e.g., Rabotin et al., 2023), and other fields of education (Hwang & Chien, 2022). According to Tlili et al. (2022), the most common applications of XR in education are in science, mathematics, and engineering. However, XR could also have a huge impact within the arts and humanities on language learning (Lan, 2021). As Lan (2020a) argued, the specific features of XR—creation, immersion, and interaction—correspond very closely to the essential components of successful language learning; that is, learners can actively immerse themselves in authentic contexts and participate in social interactions.

XR has been applied to language learning for almost 2 decades, and its applications can be classified into **five categories** based on different pedagogical goals (Lan, 2020a). First, XR can extend the visual experiences of learners by prompting them to visit places they have learned about in class or cannot visit in real life (e.g., Cowie & Alizadeh, 2022; Ziyatdinova et al., 2021). Second, XR can provide entertainment. This is especially true when users engage in game-based XR, which creates an enjoyable gaming experience and often improves learner motivation and achievement (e.g., Khatoony, 2019; Lan, 2015). Third, XR can provide users with a novel platform for social networking, allowing users from different parts of the world to interact with each other via text or voice anytime and anywhere (e.g., Grant et al., 2024). They can not only communicate but also collaborate to complete tasks or solve problems (e.g., Lan et al., 2016). Fourth, authentic contexts created in XR can provide users with opportunities for hands-on practice (e.g., Yang et al., 2020), which is especially useful when the learning content is dangerous, very costly, or impossible in the real world. Fifth, users can co-create in XR (e.g., Wu et al., 2019; Yeh & Lan, 2018), which can result in foreign-language learners outperforming those who learn in traditional ways and increase their motivation.

Following the above-mentioned applications and positive effects of XR on language learning, exploring the following questions can allow both researchers and educators to gain a deeper understanding of the current state of XR research on language learning: How can XR be applied to language learning, and how does it affect language acquisition? Which language skills are improved the most when XR is applied to language learning? Which learners benefit more? What are the research methods for this kind of research? What should future research focus on? The following sections address these questions by providing a comprehensive account of the findings of systematic reviews that have examined XR-based language learning and by making recommendations for future research.

Previous Research on Using XR in Language Learning

XR has been applied to language-learning research for almost 2 decades. In addition, Meta has brought XR technology to the attention of the general public. Numerous systematic reviews on XR and language learning have been published within the past few years (e.g., Dhimolea et al., 2022; Qiu et al., 2024; Wang et al., 2020; Yang & Zhang, 2025), which are not only helpful for understanding the current situation but also for identifying future research trends. These systematic reviews and meta-analyses have produced similar results and accurately represent the current state of research on XR in language learning, as described below.

VR Prevails Over AR

As described above, XR includes VR, AR, and MR. Applications of XR in language learning have involved VR more than AR (Weng et al., 2024), with effect sizes from meta-analyses being reported (Chen et al., 2022). Another difference between the two technologies is which target skills they are applied to. A large proportion of AR applications focus on vocabulary (El Shemy et al., 2024; Shadiev & Liang, 2024; Weng et al., 2024; Yang & Zhang, 2025), while VR applications focus more on skills such as oral communication, listening, and writing (Qiu et al., 2024; Wang et al., 2020).

Different VR technologies exert different effects on language learning, and the definition of VR varies with the level of immersion or functionality offered. Two pairs of synonymous terms classify VR according to the level of immersion of VR: **immersive** versus **nonimmersive** (Robertson et al., 1993) and **high- versus low-immersive** (Parong & Mayer, 2021). **Non-immersive** and **low-immersive** VR refer to desktop or web-based VR in which users explore the virtual world through a third-person view and control their avatars using a mouse, keyboard, or game controller. In contrast, in **immersive** and **high-immersive** VR, the user wears VR goggles or a VR headset to become fully immersed in the VR environment and explore virtual worlds from a first-person perspective. Papagiannidis et al. (2008) classified VR into **game-based** and **social-based** VR based on the functionality offered. The former emphasizes the entertainment nature of VR, mainly comprising VR games that will have a set of predesigned game mechanisms, while the latter emphasizes providing a socially interactive VR environment without the need for predesigned game programs.

While high-immersive VR has recently attracted the attention of many researchers and educators, low-immersive (or nonimmersive) VR accounted for most of the literature on the use of VR for language learning (Wang et al., 2020) before the emergence of HTC headsets in 2016 (Radianti et al., 2020). However, the proportions of these two types of VR address in studies of language learning have changed slightly (Qiu et al., 2024), and recently high-immersive VR has appeared in most studies related to the use of XR in language learning (Coban et al., 2022). However, it has been found that high-immersive VR has only a small effect on language learning (Coban et al., 2022; Qiu et al., 2024), whereas low-immersive (or nonimmersive) VR exerts large effects on language learning (Qiu et al., 2024).

The variations in the effects of different types of VR on language learning are clearly due to how the activities are designed. Low-immersive VR is a more mature technology than high-immersive VR, and so the activities designed for non-immersive VR are more mature. However, this situation may change in the near future as high-immersive VR becomes more common in educational settings and more teachers become involved in using high-immersive VR for language learning.

The Lack of Effect Sizes in XR Language-Learning Research

Most systematic reviews and meta-analyses of XR in language learning do not consider effect sizes. This can be easily seen by looking at the relevant systematic reviews or meta-analyses. The volume of systematic reviews, and the number of target articles are always larger in systematic reviews than in meta-analyses. Consider two related studies performed by the present author. First, in 2014, Lan and her colleague planned to conduct a meta-analysis of the use of XR in language learning to determine its impact on the learning outcomes of foreign-language learners (Li & Lan, 2015). However, the insufficient number of articles on effect sizes resulted in the research focus shifting from a meta-analysis to a systematic review. Second, in 2019, Lan and her colleagues wanted to investigate the extent to which XR affects language-learning outcomes (Wang et al., 2020). After searching 2 databases (Web of Science and ERIC), the number of articles reporting effect sizes was not sufficient to support a meta-analysis, and so another 2 databases were searched (OpenDissertations, and Networked Digital Library of Theses and Dissertations) and finally only 13 articles were included.

Another example of the insufficient consideration of effect sizes is the number of target articles included in a systematic review or meta-analysis. Qiu et al. (2024) and Dhimolea et al. (2022) searched databases for articles published within 6- and 5-year periods, respectively. The former meta-analysis included 14 articles, while the latter systematic review included 32 articles (note also the publication years of the two articles).

The lack of studies into effect sizes directly indicates that more data are needed to confirm the exact impact of XR on language learning, in particular to address the conflicting results reported in the existing literature. For example, the meta-analysis of Qiu et al. (2024) found that XR had a negative effect on student writing performance, whereas Chen et al. (2023) found a positive effect in their systematic review. Given the conflicting findings of studies related to XR and language learning, more research is needed into how and when it is appropriate to use XR in language learning to truly benefit foreign-language learners.

Self-Reporting Is the Dominant Data-Collection Method

The phenomena described in the previous section are due to data collection mainly being performed using self-reporting methods, such as questionnaires and interviews (Chen et al., 2023; Shadiev & Liang, 2024; Weng et al., 2024). In addition, previous research has focused mainly on affective gains (Wang et al., 2020), including perceptions about using XR in language learning such as satisfaction, engagement, and learning attitudes towards immersive technologies.

Questionnaires and interviews are also used to gather participants' opinions about the benefits and impact of XR on their learning outcomes. Although some studies use objective approaches to evaluate learning outcomes, their results or conclusions are mostly based on a single perspective, such as test scores (Shadiev & Liang, 2024; Weng et al., 2024). Since language learning is a complex and dynamic process (Godwin-Jones, 2018), more-objective and multifaceted evidence should be collected to allow objective evaluations of the impact of XR on language learning, such as data on brain imaging (e.g., Legault et al., 2019; Li & Lan, 2022), eye movements (e.g., Bacca-Acosta et al., 2022), student's creative abilities (e.g., Guo & Lan, 2023; Lan et al., 2019), and the learning process and associated behaviours (e.g., Wang & Sun, 2022).

Inconsistent Findings and a Focus on Lower-Level Skills

Language skills such as speaking, reading, listening, and writing have been targeted by XR-related research, but most studies have focused on vocabulary learning (Chen et al., 2022; Dhimolea et al., 2022; Wang et al., 2020; Yang & Zhang, 2025). It is particularly interesting that almost all of the existing literature describes positive impacts of XR on vocabulary learning (El Shemy et al., 2024; Qiu et al., 2024), which is not the case for the learning of other skills. Among the above four language skills, reading has been studied the least, and the effect of XR on writing appears to be the least consistent.

Reading comprehension involves two-way information processing, from simple (i.e., low-level linguistic knowledge such as phonological knowledge and lexical knowledge) to complex (i.e., high-level linguistic knowledge such as syntactic and metacognitive knowledge and comprehension skills), and also from complex to simple (Lan et al., 2009). These two-way processes must be activated simultaneously when learners are reading so that they can properly comprehend the reading material. Creating an XR environment to satisfy the requirement for two-way information processing remains challenging. Some studies (e.g., Bulat Demir & Ilgaz Büyükbaykal, 2024) have simply moved the PDF document of an eBook into a virtual environment to enable learners to read the book in that environment. Although these studies have found positive effects, whether this design meets the above requirements requires further investigation.

Unlike reading, writing involves production and is much more challenging for language learners (Lan et al., 2019). Lan and her colleagues found positive results when foreign-language students learning Chinese as a second/foreign language used various XR tools to improve their

writing performance (Lan, 2020b; Lan et al., 2019; Lan & Tam, 2023). However, those studies focused on the prewriting stage, as did Huang et al. (2020). How XR can be used at different stages of the writing process to improve learners' writing performance requires more exploration.

It should be noted that the reported positive effects of XR on reading and writing as described above differ from the results of the meta-analyses by Dhimolea et al. (2022) and Qiu et al. (2024). Therefore, more empirical studies are needed to assess whether XR is beneficial for learning these two skills.

Lack of Diversity of Target Languages and Participants

Most previous studies of the use of XR in language learning have investigated this technology for learning English as a second/foreign language. The few studies that have investigated the impact of XR on learning other languages have included Chinese, Japanese, and Portuguese (Dhimolea et al., 2022; Shadiev & Liang, 2024; Wang et al., 2020). Most of the participants in these studies have been students at the college or university level. Although many elementary school students are involved in AR research, there are many more involved in VR. Furthermore, the vast majority of participants have been general learners, with very few studies focusing on students with special needs (Lan et al., 2018, 2024). Given that XR has been used to improve the social interactions (Ke & Im, 2013) and other skills (Kurniawati et al., 2019) of learners with special needs, there is certainly a need to explore how this technology can be used to improve language learning in that population.

The sample size is also an issue in literature on the use of XR in language learning. The samples recruited in most previous studies have been rather small (Qiu et al., 2024), which would adversely influence the accuracy of the results of meta-analyses. Thus, larger samples are needed for future studies to confirm the current findings on the effects of XR on language learning.

This section has briefly summarized the current state of research on the use of XR in language learning. Since XR is still an emerging technology, more empirical studies that have larger samples, measure effect sizes, and expand the diversity of target languages, skills, and participants will help researchers to build a more balanced and critical evidence base for the impact of XR on language learning.

Challenges and Suggestions for Future Research on XR-Based Language Learning

The previous section discussed the current state of research on XR for language, along with some suggestions for future research. There are also other challenges that need to be addressed as XR devices become more common and adopted in the classroom, as addressed below.

The Classroom Ecosystem

The classroom can be viewed as an ecosystem with many coexisting components, including students, teachers, learning materials, and the physical environment and climate. This ecosystem is a dynamic complex system in which all of its components act cooperatively, interdependently, interactively, dynamically, adaptively, and reciprocally. When all students are wearing VR goggles and immersed in virtual worlds, the associated separation from the real world and their physical surroundings can easily make them feel alienated. The accessibility issue due to the lack of a sufficient number of VR goggles should also be noted. If appropriate tasks are not assigned to contribute to the learning process, some students will waste time waiting to use the available VR goggles. Innovative instructional design becomes important in this situation to overcome alienation and ensure a harmonious classroom ecosystem.

Transfer of Learning Outcomes

The impact of XR on language learning remains unclear, but most studies have found positive results for almost all aspects of learning outcomes, including cognitive, affective, and behavioural outcomes. Many studies have used XR to address students' anxiety about public speaking (Lim et al., 2023), but there has been little research on

whether positive effects persist after XR treatment ends when learners subsequently perform public speaking in real-life settings without XR support. In addition, previous research has shown positive effects of XR in promoting learner motivation relative to learning without immersion in XR (Klimova, 2021). However, whether these positive effects on students' motivation last long enough, including after they leave school, remains to be determined. The above concerns also apply to other learning outcomes. Using XR to enhance learning is a process, but the ultimate goal should be determining how to deepen the learning effects of VR and help students to become independent learners who do not need XR as a scaffold. This is undoubtedly where considerable efforts are required from both researchers and educators.

Impact of AI on Language Learning in XR

The global artificial intelligence (AI) market was valued at around USD 200 billion in 2023, when it was projected to expand at a compound annual growth rate of 37.3% up to 2030 (Size, 2023). AI is an umbrella term for a growing set of modelling capabilities. There are three main types of AI conceptualizations (Regona et al., 2022): artificial narrow intelligence, artificial general intelligence, and artificial super intelligence. Generative AI (GAI) has grabbed global attention since the release of ChatGPT (Hwang & Chen, 2023), and it has been applied in various domains including entertainment, science, industry, healthcare, news reporting, novel writing, music production, education, transportation, social issues, business, and art.

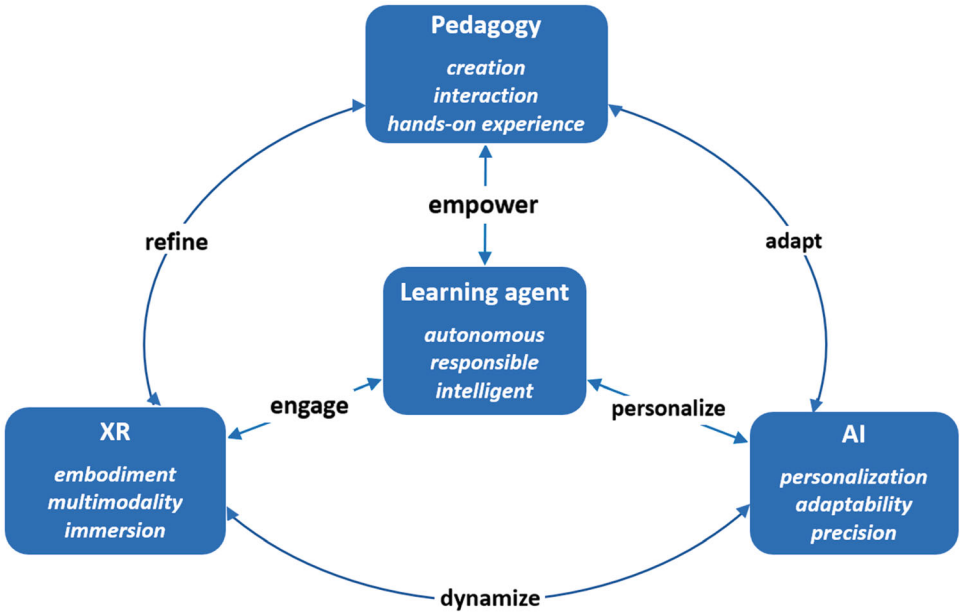
In language learning, ChatGPT can be a powerful assistant or learning partner for students to provide them with personalized learning support through human-like conversations (Lan & Chen, 2024). Combining ChatGPT with XR can not only provide appropriate real-time feedback to learners but also enhance the perception of reality during the oral communication process in an XR environment compared to using pattern-matching or dialog-flow technology. However, teachers still have concerns about students misusing ChatGPT in the classroom (e.g., Alexander et al., 2023; Lambert & Stevens, 2023). Therefore, how to appropriately combine XR and GAI to enhance students' language learning while also reducing the probability of negative effects is undoubtedly an urgent issue that needs more efforts from both researchers and educators.

Conclusion

XR holds tremendous potential for language learning by creating an authentic environment in which learners can immerse themselves in social interactions, perform exploratory and collaborative tasks, and consequently enhance their learning outcomes and motivation. Educators and researchers need to further investigate how XR—combined with other advanced technologies such as AI—can facilitate intelligent and contextualized multimodal learning to help learners to become autonomous and intelligent agents. This will enable them to collaborate effectively in future workplaces, successfully solve real-world problems they encounter, and adapt their approaches and solutions according to real situations. A focus is also needed on how the classroom ecosystem can remain balanced, beneficial, and prosocial as XR becomes more prominent in everyday learning.

To achieve these goals, learners need multifaceted and accurate information to become responsible and autonomous agents who are self-directed, self-regulated, and self-reflective. Such characteristics will make them able to define their learning goals, make plans, take action, manage their learning processes, identify problems, reflect on results, and refine their plans based on their experiences. Additionally, they need opportunities to take ownership of their learning through creative processes both individually and collaboratively.

In conclusion, language learning can be enhanced in the AI era by integrating XR affordances (embodiment, multimodality, and immersion) with innovative pedagogical approaches (e.g., creation, interaction, and hands-on experience), all supported by the AI characteristics of personalization, adaptability, and precision. Figure 1 illustrates the dynamic



Unlocking Language Learning with Immersive Technology, Fig. 1 The LeLIXR (**l**earner-**m**ediated **l**earning in **i**ntelligent XR) framework

relationship between learners, pedagogy, and technology. In the **LeLIXR** (**l**earner-**m**ediated **l**earning in **i**ntelligent **XR**) framework, learners become autonomous, responsible, and intelligent agents through the seamless integration of pedagogy and advanced technology to support learning. Simultaneously, learners' personal characteristics, learning needs, and feedback guide pedagogical methods and technologies, ensuring that they evolve to be more personalized and up-to-date. This approach will ultimately foster a learner-centred and user-friendly learning environment. Collaboration between researchers from various disciplines is essential to realize this potential.

Cross-References

- ▶ Augmented Reality
- ▶ Augmented Reality Applications and Tools
- ▶ Augmented Reality-Assisted Language Learning (ARALL)
- ▶ DGBLLT
- ▶ Extended Reality (XR)
- ▶ Immersive Literacy (Including AR/XR Literacy)
- ▶ The Impact of Artificial Intelligence (AI) on CALL Pedagogies
- ▶ Trends in Pedagogical Approaches to CALL
- ▶ Virtual Reality Assisted Language Learning
- ▶ Virtual Worlds

Acknowledgements We thank the National Science and Technology Council, Taiwan, ROC, under grant numbers MOST 110-2511-H-003-038-MY3 and MOST 111-2410-H-003-006-MY3 for financially supporting this research.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology, 23*(2), 25–43.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*(6), 34–47.

Bacca-Acosta, J., Tejada, J., Fabregat, R., Kinshuk, & Guevara, J. (2022). Scaffolding in immersive virtual reality environments for learning English: An eye tracking study. *Educational Technology Research and Development, 70*, 339–362.

Bulat Demir, S., & Ilgaz Büyükbaykal, A. C. (2024). Investigating book reading experiences in VR. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006241252659

Chen, J., Dai, J., Zhu, K., & Xu, L. (2022). Effects of extended reality on language learning: A meta-analysis. *Frontiers in Psychology, 13*, 1016519.

Chen, Y., Li, M., Huang, C., Cukurova, M., & Ma, Q. (2023). A systematic review of research on immersive technology-enhanced writing education: The current state and a research agenda. *IEEE Transactions on Learning Technologies, 17*, 919–938. https://ieeexplore.ieee.org/document/10354407

Chuah, S. H. W. (2019). Why and who will adopt extended reality technology? Literature review, synthesis, and future research agenda. *SSRN Electronic Journal (2018)*. https://doi.org/10.2139/ssrn.3300469

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review, 36*, 100452.

Cowie, N., & Alizadeh, M. (2022). The affordances and challenges of virtual reality for language teaching. *International Journal of TESOL Studies, 4*(3), 50–65. https://doi.org/10.46451/ijts.2022.03.05

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends, 66*(5), 810–824.

El Shemy, I., Jaccheri, L., Giannakos, M., & Vulchanova, M. (2024). Augmented reality-enhanced language learning for children with autism spectrum disorder: A systematic literature review. *Behaviour & Information Technology, 43*, 1–28.

Godwin-Jones, R. (2018). Chasing the butterfly effect: Informal language learning online as a complex system. *Language Learning & Technology, 22*(2), 8–27. https://doi.org/10125/44643

Grant, S., Qi, G. Y., Lan, Y.-J., & Cheng, P.-Y. (2024). Fostering academic citizenship through ubiquitous technologies in an online academic conference: A framework and its implications. *Educational Technology & Society, 27*(1), 18–34. https://doi.org/10.30191/ETS.202401_27(1).RP02

Guo, S. C., & Lan, Y.-J. (2023). Virtual world-supported contextualized multimodal EFL learning at a library. *Language Learning & Technology, 27*(2), 176–198. https://hdl.handle.net/10125/73508

Huang, H. L., Hwang, G. J., & Chang, C. Y. (2020). Learning to be a writer: A spherical video-based virtual reality approach to supporting descriptive article writing in high school Chinese courses. *British Journal of Educational Technology, 51*(4), 1386–1405.

Hwang, G. J., & Chen, N. S. (2023). Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Educational Technology & Society, 26*(2), 18.

Hwang, G. J., & Chien, S. Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence, 3*, 100082.

Ke, F., & Im, T. (2013). Virtual-reality-based social interaction training for children with high-functioning autism. *The Journal of Educational Research, 106*(6), 441–461.

Khatoony, S. (2019, December). An innovative teaching with serious games through virtual reality assisted language learning. In *In 2019 international serious games symposium (ISGS)* (pp. 100–108). IEEE.

Klimova, B. (2021). Use of virtual reality in non-native language learning and teaching. *Procedia Computer Science, 192*, 1385–1392.

Kurniawati, A., Kusumaningsih, A., & Hasan, I. (2019). Class VR: Learning class environment for special educational needs using virtual reality games. In *2019 international conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)* (pp. 1–5). IEEE.

Lambert, J., & Stevens, M. (2023). ChatGPT and generative AI technology: A mixed bag of concerns and new opportunities. *Computers in the Schools, 41*, 1–25.

Lan, Y. J. (2015). Contextual EFL learning in a 3D virtual environment. *Language Learning & Technology, 19*(2), 16–31. http://llt.msu.edu/issues/june2015/action.pdf

Lan, Y. J. (2020a). Immersion, interaction and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology, 24*(1), 1–15. http://hdl.handle.net/10125/44704

Lan, Y. J. (2020b). Immersion into virtual reality for language learning. *Psychology of Learning and Motivation: Adult and Second Language Learning, 72*, 1–26.

Lan, Y.-J. (2021). Language learning in virtual reality: Theoretical foundations and empirical practices. In Y.-J. Lan & S. Grant (Eds.), *Contextual language learning. Chinese language learning sciences*. Springer. https://doi.org/10.1007/978-981-16-3416-1_1

Lan, Y.-J., & Chen, N.-S. (2024). Teachers' agency in the era of LLM and generative AI: Designing pedagogical AI agents. *Educational Technology & Society, 27*(1), I–XVIII. https://doi.org/10.30191/ETS.202401_27(1).PP01

Lan, Y.-J., & Tam, V. T. T. (2023). The impact of 360° videos on basic Chinese writing: A preliminary exploration. *Educational Technology Research and Development, 71*, 539–562. https://doi.org/10.1007/s11423-022-10162-4

Lan, Y. J., Sung, Y. T., & Chang, K. E. (2009). Let us read together: Development and evaluation of a computer-assisted reciprocal early English reading system. *Computers & Education, 53*(4), 1188–1198.

Lan, Y. J., Kan, Y. H., Sung, Y. T., & Chang, K. E. (2016). Oral-performance language tasks for CSL beginners in Second Life. *Language Learning & Technology, 20*(3), 60–79.

Lan, Y. J., Hsiao, I. Y. T., & Shih, M. F. (2018). Effective learning design of game-based 3D virtual language learning environments for special education students. *Educational Technology & Society, 21*(3), 213–227.

Lan, Y. J., Lyu, B. N., & Chin, C. K. (2019). Does a 3D immersive experience enhance Mandarin writing by CSL students? *Language Learning & Technology, 23*(2), 125–144.

Lan, Y.-J., Shih, M.-F., & Hsiao, Y.-T. (2024). 3D immersive scaffolding game for enhancing Mandarin learning in children with ADHD. *Educational Technology & Society, 27*(2), 4–24.

Legault, J., Fang, S. Y., Lan, Y. J., & Li, P. (2019). Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain and Cognition, 134*, 90–102.

Lin, T.-J., & Lan, Y.-J. (2015). Language learning in virtual reality environments: Past, present, and future. *Educational Technology & Society, 18*(4), 486–497.

Li, P., & Lan, Y. J. (2022). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition, 25*(3), 361–378. https://doi.org/10.1017/S1366728921000353

Lim, M. H., Aryadoust, V., & Esposito, G. (2023). A meta-analysis of the effect of virtual reality on reducing public speaking anxiety. *Current Psychology, 42*(15), 12912–12928.

Logeswaran, A., Munsch, C., Chong, Y. J., Ralph, N., & McCrossnan, J. (2021). The role of extended reality technology in healthcare education: Towards a learner-centred approach. *Future Healthcare Journal, 8*(1), e79.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies* (Vol. 2351, pp. 282–292). SPIE.

Mubin, S. A., You, J. S. T., Samiraj, E. P. R., & Jaafar, S. P. (2022, November). A framework for supporting deaf and mute learning experience through extended reality. In *2022 IEEE international conference on computing (ICOCO)* (pp. 80–83). IEEE.

Onyesolu, M. O. (2009, October). Virtual reality laboratories: An ideal solution to the problems facing laboratory setup and management. In *Proceedings of the World Congress on Engineering and Computer Science* (Vol. 1).

Papagiannidis, S., Bourlakis, M., & Li, F. (2008). Making real money in virtual worlds: MMORPGS and emerging business opportunities, challenges and ethical implications in metaverses. *Technological Forecasting and Social Change, 75*(5), 610–622.

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241.

Qiu, X. B., Shan, C., Yao, J., & Fu, Q. K. (2024). The effects of virtual reality on EFL learning: A meta--analysis. *Education and Information Technologies, 29*(2), 1379–1405.

Rabotin, A., Glick, Y., Gelman, R., Ketko, I., Taran, B., Fink, N., & Furer, A. (2023). Practicing emergency medicine in the metaverse: A novel mixed reality casualty care training platform. *Surgical Innovation, 30*(5), 586–594.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Regona, M., Yigitcanlar, T., Xia, B., & Li, R. Y. M. (2022). Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *Journal of Open Innovation: Technology, Market, and Complexity, 8*(1), 45.

Robertson, G. G., Card, S. K., & Mackinlay, J. D. (1993). Three views of virtual reality: Nonimmersive virtual reality. *Computer, 26*(2), 81.

Shadiev, R., & Liang, Q. (2024). A review of research on AR-supported language learning. *Innovation in Language Learning and Teaching, 18*(1), 78–100.

Size, A. I. M. (2023). Share & trends analysis report by solution. *By technology (deep learning, machine learning), by end-use, by region, and segment forecasts, 2030*, 2022.

Tamura, H., Yamamoto, H., & Katayama, A. (2001). Mixed reality: Future dreams seen at the border between real and virtual worlds. *IEEE Computer Graphics and Applications, 21*(6), 64–70.

Tlili, A., Huang, R., Shehata, B., Liu, D., Zhao, J., Metwally, A. H. S., . . . & Burgos, D. (2022). Is Metaverse in education a blessing or a curse: A combined content and bibliometric analysis. *Smart Learning Environments*, *9*(1), 1–31.

Wang, H.-Y., & Sun, J. C.-Y. (2022). Influences of online synchronous VR co-creation on behavioral patterns and motivation in knowledge co-construction. *Educational Technology & Society, 25*(2), 31–47.

Wang, C. P., Lan, Y. J., Tseng, W. T., Lin, Y. T., & Kao, C. L. (2020). On the effects of 3D virtual worlds in language learning – A meta-analysis. *Computer Assisted Language Learning, 33*(8), 891–915. https://doi.org/10.1080/09588221.2019.1598444

Weng, Y., Schmidt, M., Huang, W., & Hao, Y. (2024). The effectiveness of immersive learning technologies in K–12 English as second language learning: A systematic review. *ReCALL, 36*, 1–20.

Wiles, J. (2022). What is a metaverse? And should you be buying in. Gartner Insights. https://www.gartner.com/en/articles/what-is-a-metaverse

Wu, Y. J. A., Lan, Y. J., Huang, S. B. P., & Lin, Y. T. R. (2019). Enhancing medical students'

communicative skills in a 3D virtual world. *Educational Technology & Society, 22*(4), 18–32.

Yang, P., & Zhang, W. (2025). Effectiveness of augmented reality on EFL learner's language gains: A meta-analysis. *Innovation in Language Learning and Teaching, 19*(1), 48–63.

Yang, F. C. O., Lo, F. Y. R., Hsieh, J. C., & Wu, W. C. V. (2020). Facilitating communicative ability of EFL learners via high-immersion virtual reality. *Journal of Educational Technology & Society, 23*(1), 30–49.

Yang, X., Cheng, P. Y., Liu, X., & Shih, S. P. (2023). The impact of immersive virtual reality on art education: A study of flow state, cognitive load, brain state, and motivation. *Education and Information Technologies, 5*, 6087–6106.

Yeh, Y. L., & Lan, Y. J. (2018). Fostering student autonomy in English learning through creations in a 3D virtual world. *Educational Technology Research and Development, 66*(3), 693–708.

Ziyatdinova, J. N., Fakhretdinova, G., Giliazova, D. R., & Pavlova, I. V. (2021). VR-technologies in foreign language learning for engineering students. In *International conference on interactive collaborative learning, September* (pp. 1027–1034). Springer International Publishing.

P

Pedagogical Approaches in VE

Tammy Jandrey Hertel[1] and Abby Dings[2]
[1]University of Lynchburg, Lynchburg, VA, USA
[2]Southwestern University, Georgetown, TX, USA

Abstract

Using virtual exchanges has become increasingly common in second-language classrooms around the world. The interactional opportunities that virtual exchanges provide can contribute to gains in learners' intercultural and linguistic competencies, among other benefits. Strategies for incorporating virtual exchanges in the curriculum are discussed, including the wide variety of technologies and formats that can be used (e.g. videoconferences, text messages, e-mail, social media, asynchronous video chats), challenges often presented by exchanges and recommendations for implementation of exchanges.

Keywords

Virtual exchange · Telecollaboration · Pedagogical approaches · Intercultural communication · Online intercultural exchange

Introduction

A virtual exchange (VE) has been defined as "a well-known pedagogical approach in foreign language (FL) education which involves engaging classes in online intercultural collaboration projects with international partners as an integrated part of their educational programmes" (O'Dowd, 2020, p. 477).

VE has become increasingly common in second-language classrooms as a means to enhance the learning experience by providing opportunities to interact and communicate with speakers of the target language. Early VE began in the 1990s and often took the form of email exchanges (cf., Cummins & Sayers, 1995; Eck et al., 1995; Warschauer, 1995) and has evolved with technological advances to include media such as videoconferences, texting and social media platforms. Other terms used for VE include "telecollaboration" and "online intercultural exchange". It has been argued that the term "virtual exchange" is preferable because it is used in other disciplines and thus can lead to more cross-disciplinary collaboration (O'Dowd, 2021a).

Benefits

The pedagogical benefits of VE in second-language education have been documented in many studies. Two large-scale reviews of VE projects (Lewis & O'Dowd, 2016; Çiftçi & Savaş, 2018) indicated that benefits of VE discussed in the literature have focused primarily on gains in three areas: intercultural competence (cf., Catalano & Muñoz Barriga, 2021; Tecedor &

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_22-1

Vasseur, 2020; Warner-Alt, 2020), learner autonomy (cf., Fuchs et al., 2012; Kong, 2022; Schwienhorst, 2000) and linguistic development (cf., Kabata & Edasawa, 2011; Priego, 2011; Sotillo, 2000). In particular, the types of interaction present in VE have been found to benefit linguistic development in terms of negotiation of meaning, peer corrective feedback and pragmatic competence development (Lewis & O'Dowd, 2016). Studies have also found that participation in VE results in an increase in learner engagement (cf., Andujar & Franco-Rodriguez, 2020; Andujar & Márbol Trapote, 2022; Oskoz & Gimeno, 2019, 2020; Vinagre & Corral-Esteban, 2018). In addition to these benefits, Henshaw and Hetrovicz (2024) point out that VE projects can provide learners with exposure to a wider variety of dialects, which can lead to more respect and appreciation for language variation. Finally, other researchers have suggested that VE can help students with the development of digital literacy, creativity and collaboration skills (Wicking, 2022).

VE Tools and Formats

VE can be either asynchronous or synchronous in nature and can include a wide variety of technological tools. Early VE projects were typically two-way email exchanges (cf., Hertel, 2003; Schenker, 2013) while more recent technologies employed include videoconferencing tools and instant messaging using mobile applications such as WhatsApp (Andujar & Franco-Rodriguez, 2020; Andujar & Márbol Trapote, 2022). Videoconferencing (using applications such as Zoom and Google Meet) has become the preferred method for VE as it has become progressively easier and more accessible (O'Dowd & O'Rourke, 2019). VE was initially conceived of as a reciprocal exchange wherein two sets of learners interacted in ways that benefited both parties, but in recent years, various companies have begun to offer service-provider VE (SPVE) with trained (and paid) language coaches.

Asynchronous Exchanges

Instructors have used a variety of tools and formats to incorporate VE in their courses. The following paragraphs present models of asynchronous VE, including email (Schenker, 2013), WhatsApp (Catalano & Muñoz Barriga, 2021), Google Blogger (Ware & Kessler, 2016) and Flipgrid (Henshaw & Hetrovicz, 2024). See Nissen (2016) and O'Dowd (2018, 2021b) for further discussion of curricular models for the incorporation of VE in language classes.

Email exchanges have been organised by instructors who had relationships with educators in countries that spoke their target language of instruction and sought to leverage those connections. Schenker (2013), for example, coordinated an email exchange between college students in the United States studying second-semester German and high school students in Germany studying English as a foreign language. Students were paired with peers in the other class, and over the span of 6 weeks, they exchanged emails, in English (due to the American students' limited German proficiency), about a variety of cultural topics, including stereotypes and popular culture. Some of the issues that became apparent in the project included student motivation, unresolved miscommunications, semester schedule misalignments and the short duration of the project. Schenker suggests ensuring that instructors monitor the emails during the project in order to intervene if needed and promote reflection on the interactions.

Catalano and Muñoz Barriga (2021), located, respectively, in the United States and Colombia, used WhatsApp to facilitate a pen pal exchange for their students, who were pre- or in-service teachers. Both cohorts of students were taking courses on intercultural communication at their respective institutions, and the WhatsApp exchanges were designed to support their course objectives, including exploration of identities, examination of stereotypes and others. The students shared texts, photos and videos to connect with each other and reflected on their exchanges via discussion boards and in-class discussion.

Like Schwenker (2013), Catalano and Muñoz Barriga (2021) also found that closer instructor oversight over the student interactions would again afford better opportunities for thoughtful interventions and would support efforts to problematise standard language ideologies and other themes related to privilege and power.

Ware and Kessler (2016) coordinated a VE project for middle school students in Spain and the United States. Individual or pairs of students in the United States were connected with individual or pairs of students in Spain to write shared private blogs using Google Blogger. The prompts that the students blogged about included an introductory letter, sharing music, slang, images that represent self and culture and diversity, and students in both cohorts responded to each other's posts. Ware and Kessler suggest that undertaking VE in the classroom entails a complex set of activities that involves substantial instructor commitment and labour.

Henshaw and Hetrovicz (2024) suggest experimenting with Flipgrid or other discussion forums that permit video postings. The model they propose involves having students in both cohorts initially posting in their native languages, thus providing input for their collaborators. Suggested video postings include simple day in the life, introduction to a favourite place or leisure activity, etc. Videos would end with a question that the other students could respond to in either their first or second language, as decided by the instructors. As the semester progresses, students could begin to post in their target language, thus moving towards output in the language. Also included in the project design are reflective activities in the form of written reports and classroom discussions.

Synchronous Exchanges

While the above examples are all asynchronous options, SPVEs tend to be synchronous and are a more recent alternative to reciprocal exchanges. The services offered by companies such as TalkAbroad, Conversifi and LinguaMeeting help alleviate many of the logistical challenges posed by reciprocal VEs. Some of the companies, such as LinguaMeeting, have their interaction options embedded within a provided course curriculum, whereas others may provide a menu of conversational topics for instructors to choose from based on their own course content and objectives.

Although there is less research on their outcomes than there is on more traditional, instructor-designed VEs, some recent studies have examined SPVE interactions. Klimanova and Vinokurova (2021) studied interactions of third-semester and advanced second-language learners of Russia engaged in sessions run by LinguaMeeting. They found that SPVE differ from reciprocal VE in the hierarchical roles that the learners and coaches take, and the interaction tends to follow the "initiation→response→evaluation" format often seen in teacher/student interactions. Providing coaches with lists of questions led to a goal-oriented interaction with little back and forth between participants.

Despite these dynamics, the results indicated that students achieved both linguistic and affective gains. Tecedor and Vasseur (2020) examined reflections written by fourth-semester students of Spanish who interacted with coaches in sessions organised by TalkAbroad, focusing specifically on evidence of the development of intercultural communicative competence. They found that SPVEs may play a role in the development of intercultural communicative competence because they provide learners with exposure to greater cultural content, particularly through the perspectives of native speakers from the target cultures. Additionally, they suggest that it is important to assign VE reflection activities in order to push learner development. Varo (2021) also found that his participants (advanced Spanish students at a US university) significantly increased their intercultural competence after participating in an SPVE. Additionally, he noted that using SPVE made the organisation and execution of the VE more manageable than traditional VE.

The choice to incorporate SPVE in the curriculum is attractive given the logistical challenges that it can reduce. Tecedor and Vasseur (2020) cite the following four advantages of SPVE: "(1) SPVEs do not require high levels of logistical

and technical involvement on the part of the instructor; (2) they eliminate the interinstitutional curricular imbalances; (3) they allow students to focus exclusively on the development of their L2 (as they are not required to speak their L1 to their interlocutor); and (4) they do not require training students on how to provide appropriate, sensitive feedback to their interlocutor" (p. 765).

However, SPVE is not without its drawbacks. The expense of SPVE, while often relatively low, is still a factor that adds to course costs. Additionally, as Klimanova and Vinokurova (2021) found, the type of interaction in SPVE is often less natural and spontaneous than interactions between peers. The scripted nature of the interactions limits the natural flow of conversation, resulting in an interview-like dynamic. While most service providers train their coaches to avoid explicit grammar teaching and translation, some coaches revert to these interactional strategies.

An alternative to the typical paid language coach model of most SPVEs and the accompanying interactional limitations can be found in the Soliya Connect Program (SCP). O'Dowd (2018) explains that the SCP facilitates small group virtual exchanges between students from the Muslim and Western worlds. These meetings, guided by trained facilitators, are designed to promote deep intercultural understanding and critical thinking through meaningful discussion and reflection on global issues. The exchanges take place synchronously over multiple weeks. While the program is not designed specifically for language learners, it has been incorporated by instructors in various countries into English language courses as a form of experiential learning. Helm (2013) found that SCP's model, which incorporates discussion of potentially divisive topics, allows participants to confront and interrogate complex cultural differences rather than avoiding them, promoting an authentic, dialogic experience that promotes deeper intercultural understanding.

When using SPVE (as with all forms of VE), it is important to prepare students ahead of time for the interactions. Strategies may involve practicing themed questions with classmates before engaging in similar interactions with their VE coach. Henshaw and Hetrovicz (2024) suggest having students listen to recorded conversation between a coach and a former student and completing listening comprehension and reflection activities in order to familiarise themselves with the format and expectations. They also suggest that students should prepare questions ahead of time and get feedback from their instructor. As with other forms of VE, reflection activities completed after the conversations support student growth in linguistic development and intercultural competence.

Challenges and Recommendations

The incorporation of VE into a course curriculum poses some challenges, of which the most obvious is logistical obstacles. This is particularly true if conducting instructor-coordinated synchronous exchange. Collaborating with individuals in different countries often involves temporal challenges such as time zone differences and unaligned class schedules, as well as potential technological challenges such as Wi-Fi and platform accessibility for all involved. Thoughtful planning and coordination well ahead of time can minimise these challenges, as can adopting hybrid asynchronous and synchronous formats. Paid synchronous exchanges typically minimise these challenges by providing defined windows of accessibility or sign-up options.

Bauler et al. (2022) suggest that both synchronous and asynchronous media should be combined in VE projects to provide students with multiple ways of expressing themselves. Stevens (2020) also recommends combining synchronous and asynchronous formats because asynchronous puts less pressure on learners, while synchronous videoconferencing is more demanding but helps learners socialise and build relationships. Guillén et al. (2020) recommend that instructors survey students to determine what media they engage with most in order to choose online platforms for VE projects.

Another challenge instructors face when incorporating VE in their curriculum is determining appropriate assessments. Henshaw and Hetrovicz (2024) stress the importance of ensuring that students are not assessed on their linguistic abilities

while completing the interaction. Instead, they suggest assessing student reflections on the exchange experiences or setting up the VE with clearly defined task-based objectives. In so doing, the instructor frames the importance of the VE as being the experience and takeaways rather than the performance during the interaction.

Henshaw and Hetrovicz (2024) suggest that instructors need to ensure that the VE experience is integrated into the course curriculum rather than being a stand-alone activity. Additionally, Catalano and Muñoz Barriga (2021) encourage instructors to pay attention to what is happening in the exchanges, both by reading and viewing interactions and by discussing and reflecting with students on a regular basis. In their experience, conflicts between learners sometimes arose, and being aware of them earlier may have allowed them to guide students towards resolution and intercultural understanding.

While the specific challenges the incorporation of VE poses vary depending on any number of factors, including those beyond the control of the participants (such as the political demonstrations in France cited by Petropoulou (2021)), thoughtful advanced planning, intentionality in the incorporation in the course curriculum and instructor involvement and attention to program progress can support a successful and valuable VE experience for all involved.

Conclusion

Well-planned incorporation of VE in second-language pedagogy can support learners in the development of their language proficiency and intercultural competence, among other benefits. VE offers students unique opportunities for engagement with native speakers of the target language through a variety of technologies and platforms with synchronous and asynchronous as well as paid and no-cost formats. The evolution of VE options over time reflects not only the technological advancements but also a recognition of their pedagogical value for second-language learners.

Cross-References

▶ Service-Provider Virtual Exchange and Its Integration in the Language Curriculum Videoconferencing in Second Language Learning
▶ Virtual Exchanges in CALL Teacher Education

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Andujar, A., & Franco-Rodriguez, J. M. (2020). WhatsApp and Jitsi to foster student engagement in an American-Spanish telecollaboration exchange. In A. Andujar (Ed.), *Recent tools for computer- and mobile-assisted foreign language learning* (pp. 60–78). IGI Global. https://doi.org/10.4018/978-1-7998-1097-1.ch003

Andujar, A., & Márbol Trapote, P. (2022). Annotating appraisal in a mobile telecollaboration project: A linguistic analysis of students' engagement. In S. Hilliker (Ed.), *Second language teaching and learning through virtual exchange* (pp. 49–72). De Gruyter Mouton. https://doi.org/10.1515/9783110727364-004

Bauler, C., Thornburg, D., Ceballos, Ó., Pineda, C., Kogan, E., & Sorri, P. (2022). Tackling problems, finding solutions: Creativity and collaboration in cross-cultural virtual exchange during a pandemic. In S. Hilliker (Ed.), *Second language teaching and learning through virtual exchange* (pp. 189–206). De Gruyter Mouton. https://doi.org/10.1515/9783110727364-010

Catalano, T., & Muñoz Barriga, A. (2021). Shaping the teaching and learning of intercultural communication through virtual mobility. *Intercultural Communication Education, 4*(1), 75–89. https://doi.org/10.29140/ice.v4n1.443

Çiftçi, E. Y., & Savaş, P. (2018). The role of telecollaboration in language and intercultural learning: A synthesis of studies published between 2010 and 2015. *ReCALL, 30*(3), 278–298. https://doi.org/10.1017/S0958344017000313

Cummins, J., & Sayers, D. (1995). *Brave new schools: Challenging cultural literacy through global learning networks*. St. Martin's Press.

Eck, A., Legenhausen, L., & Wolff, D. (1995). *Telekommunikation und Fremdsprachenunterricht: Informationen, Projekte*. AKS-Verlag.

Fuchs, C., Hauck, M., & Müller-Hartmann, A. (2012). Promoting learner autonomy through multi-literacy skills development in cross-institutional exchanges. *Language Learning & Technology, 16*(3), 82–102. http://dx.doi.org/10125/44301

Guillén, G., Sawin, T., & Avineri, N. (2020). Zooming out of the crisis: Language and human collaboration.

Foreign Language Annals, 53(2), 320–328. https://doi.org/10.1111/flan.12459

Helm, F. (2013). A dialogic model for telecollaboration. *Bellaterra Journal of Teaching & Learning Language & Literature, 6*(2), 28–48. https://doi.org/10.5565/rev/jtl3.522

Henshaw, F., & Hetrovicz, L. (2024). From the classroom to the world: Integrating virtual exchanges in the language curriculum. In F. Henshaw & K. Potowski (Eds.), *Honing our craft: World language teaching today* (pp. 108–125). Klett World Languages.

Hertel, T. J. (2003). Using an e-mail exchange to promote cultural learning. *Foreign Language Annals, 36*(3), 386–396. https://doi.org/10.1111/j.1944-9720.2003.tb02121.x

Kabata, K., & Edasawa, Y. (2011). Tandem language learning through a cross-cultural Keypal project. *Language Learning & Technology, 15*(1), 104–121. http://dx.doi.org/10125/44239

Klimanova, L., & Vinokurova, V. (2021). Service-provider virtual exchange as a viable alternative to face-to-face speaking practice. *Russian Language Journal/Русский язык, 71*(2), 121–156. https://www.jstor.org/stable/27213079

Kong, K. (2022). "Zoom" in and speak out: Virtual exchange in language learning. In S. Hilliker (Ed.), *Second language teaching and learning through virtual exchange* (pp. 97–114). De Gruyter Mouton. https://doi.org/10.1515/9783110727364-006

Lewis, T., & O'Dowd, R. (2016). Online intercultural exchange and foreign language learning: A systematic review. In R. O'Dowd & T. Lewis (Eds.), *Online intercultural exchange: Policy, pedagogy, practice* (pp. 21–68). Routledge. https://oro.open.ac.uk/47044/

O'Dowd, R. (2018). From telecollaboration to virtual exchange: State-of-the-art and the role of UNICollaboration in moving forward. *Journal of Virtual Exchange, 1*, 1–23. https://doi.org/10.14705/rpnet.2018.jve.1

O'Dowd, R. (2020). A transnational model of virtual exchange for global citizenship education. *Language Teaching, 53*(4), 477–490. https://doi.org/10.1017/S0261444819000077

O'Dowd, R. (2021a). Virtual exchange: Moving forward into the next decade. *Computer-Assisted Language Learning, 34*(3), 209–224. https://doi.org/10.1080/09588221.2021.1902201

O'Dowd, R. (2021b). What do students learn in virtual exchange? A qualitative content analysis of learning outcomes across multiple exchanges. *International Journal of Educational Research, 109*, 101804. https://doi.org/10.1016/j.ijer.2021.101804

O'Dowd, R., & O'Rourke, B. (2019). New developments in virtual exchange for foreign language education. *Language Learning & Technology, 23*(3), 1–7. https://doi.org/10125/44690.

Oskoz, A., & Gimeno, A. (2019). Engagement and attitude in telecollaboration: Topic and cultural background effects. *Language, Learning and Technology, 23*(3), 136–160. http://hdl.handle.net/10125/44700

Oskoz, A., & Gimeno-Sanz, A. (2020). Exploring L2 learners' engagement and attitude in an intercultural encounter. *Language Learning & Technology, 24*(1), 187–208.

Petropoulou, Z. (2021). Virtual classroom experiences for second language learning and cultural exchange. *International Journal of Technology in Teaching and Learning, 16*(1), 49–60. https://doi.org/10.37120/ijttl.2020.16.1.05

Priego, S. (2011). Helping each other: Scaffolding in electronic tandem language learning. *The International Journal of Technology, Knowledge and Society, 7*(2), 133–151.

Schenker, T. (2013). Intercultural competence and cultural learning through telecollaboration. *CALICO Journal, 29*(3), 449–470. https://doi.org/10.11139/cj.29.3.449-470

Schwienhorst, K. (2000). *Virtual reality and learner autonomy in second language acquisition*. Unpublished PhD dissertation, Trinity College Dublin.

Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology, 4*(1), 77–110. http://dx.doi.org/10125/25088

Tecedor, M., & Vasseur, R. (2020). Videoconferencing and the development of intercultural competence: Insights from students' self-reflections. *Foreign Language Annals, 53*(2), 761–784. https://doi.org/10.1111/flan.12495

Varo, A. (2021). Fostering intercultural competence through videoconference exchange: Using an external provider to match learners with trained native speakers and administer video calls. *Journal of Virtual Exchange, 4*, 1–13. https://doi.org/10.21827/jve.4.35981

Vinagre, M., & Corral-Esteban, A. (2018). Evaluative language for rapport building in virtual collaboration: An analysis of appraisal in computer-mediated interaction. *Language and Intercultural Communication, 18*(3), 335–350. https://doi.org/10.1080/14708477.2017.1378227

Ware, P., & Kessler, G. (2016). Telecollaboration in the secondary language classroom: Case study of adolescent interaction and pedagogical integration. *Computer Assisted Language Learning, 29*(3), 427–450. https://doi.org/10.1080/09588221.2014.961481

Warner-Alt, A. (2020). Promoting intercultural learning through synchronous video exchange: A talk abroad case study. *International Journal of Computer-Assisted Language Learning and Teaching, 10*(1), 1–14. https://doi.org/10.4018/IJCALLT.2020010101

Warschauer, M. (Ed.). (1995). *Virtual connections: Online activities & projects for networking language learners*. Second Language Teaching and Curriculum Center. Retrieved from https://files.eric.ed.gov/fulltext/ED406824.pdf

Wicking, P. (2022). Learning-oriented assessment in an international virtual exchange. In S. Hilliker (Ed.), *Second language teaching and learning through virtual exchange* (pp. 9–33). De Gruyter Mouton. https://doi.org/10.1515/9783110727364-002

A

A Multimodal Language Pedagogy

Nicolas Guichon[1] and Ciara R. Wigham[2]
[1]Interactions, Corpus, Apprentissages, Représentations (ICAR), Université Lumière Lyon 2, Lyon, France
[2]Activité, Connaissance, Transmission, éducation (ACTé), Université Clermont Auvergne, Clermont-Ferrand, France

Abstract

Multimodal pedagogy has become central to mediated language instruction (New London Group, Harv Educ Rev 66(1):60–90, 1996). As digital technologies advance, multimodal strategies are increasingly integral to language education, bridging the gap between in-class and out-of-class informal and professional practices. This entry explores multimodal pedagogy from three perspectives: reception, production, and interaction. Regarding reception, it discusses how diverse modes affect comprehension and cognitive load, with an emphasis on technologies like text-to-speech. Concerning production, it highlights the role of digital storytelling and the need for multimodal competency criteria to guide the assessment of pedagogical productions and to align them with the semiotic grammar of new and rapidly emerging digital genres. Focusing on interaction, it examines how different digital environments shape additional language communication and the need to equip learners with the necessary semio-pragmatic, and socio-cultural skills to interact online with others. The entry concludes by outlining the main facets of a multimodal language pedagogy and advocating the development of teachers' techno-semio-pedagogical skills to implement the latter.

Keywords

Multimodality · New media · Meaning-making · Techno-semio-pedagogical competence · Cognitive demand

Introduction

Since the seminal work by the New London Group (1996), multimodality has become central to mediated language pedagogy. Multimodality is defined as "the use of several semiotic modes in the design of a semiotic product or event, together with the way these modes are combined" (Kress & van Leeuwen, 2001, p. 20). For instance, in a news report, the filmed image can add information to the journalist's voiceover or complement it, with one mode (e.g., written names of protagonists on the screen) sometimes taking precedence over others (Guichon & McLornan, 2008). The seminal work of the New London Group (1996) recognized the growing impact of communication

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_27-1

technologies and proposed the theory of "multiliteracies". New media (e.g., online newspapers) had been reshaping meaning-making by mixing semiotic modes "more powerfully than was culturally the norm and even technically possible in the earlier modernity that was dominated by the book and the printed page" (Cope & Kalantzis, 2009, p. 178). This theory highlighted "the interaction of many semiotic modes in a text" (Kress, 2000, p. 158) from both the maker's and interpreter's perspectives.

Multimodality is a defining characteristic of CALL as digital technologies enable the use of different modes (text, image, sound) on the same screen (Clark & Feldon, 2005). In today's digital world, additional language (Lx) learners must navigate technology-driven scenarios, including static, dynamic, and interactive ones. They need strategies to manage cognitive demands and use multimodal approaches with cultural sensitivity to become proficient in comprehending (reception), expressing (production), and interacting (interaction) across modes. This approach can bridge in-class and out-of-class literacy practices but requires new skills from teachers, including understanding how multimodality shapes language learning and designing resources and activities. We thus define multimodal language pedagogy as the professional repertoire language teachers progressively deploy to harness the capabilities and constraints of multimodality for optimal learning experience (Guichon & Cohen, 2014).

Reception

Multimodality is valuable for language teachers as they must choose the best ways to structure and deliver learning experiences (Chang & Asselin, 2023). Implementing multimodal pedagogy requires understanding how learners engage with modes to optimize content presentation (Stockwell, 2010). For example, in a study comparing different content delivery for a listening task (audio only, audio and video, audio, video with target-language subtitles, video with subtitles in learners' dominant language), Guichon & McLornan (2008) found that the more cues available across modes while listening to a news report, the more they understood the semantic content they were asked to process. However, when modes diverged in content, such as when verbal comments differed from video images, a split-attention effect hindered comprehension. Thus, while multimodality often aids reception, it can also introduce seductive details that increase cognitive load, especially when modes are not semantically related (Schnotz, 2005).

Because most information, be it for administrative or entertainment purposes, is now habitually accessible online, language learners must learn to process information that comes in different modes and create meanings across them (Jewitt, 2009). Indeed, given the vast array of platforms and digital media, learners also need to develop their reception skills across multiple platforms and communication media (transliteracy), as well as their abilities to transfer previous knowledge across various information and technical environments and contexts (Lacelle et al. 2017). The goal of multimedia presentation is not only to provide information, but also to provide guidance for how to process the presented information—that is, for determining what to pay attention to, how to mentally organize it, and how to relate it to prior knowledge (Mayer, 2005). Accordingly, the role of the teacher is that of a cognitive guide who "provides needed guidance to support the learner's cognitive processing" (p. 12). This includes offering guidance about how to access the information (technical knowledge), what and when to access (task and strategic knowledge), and why to access the information (pedagogical knowledge).

Besides, the affordances of text-to-speech and speech-to-text technologies can now be harnessed to change the mode of texts and facilitate their reception. For instance, in a study by Guichon et al. (submitted), adult migrants are observed using these transmediative strategies to process vocal messages and the fact that an oral message is textualized both facilitates reception and reassures the learner. Transmediative strategies seem to contribute to "deepening learners' knowledge processing and memorization by strengthening the connections between audio and textual forms of language knowledge" (Zhang & Zou, 2022).

This could involve guiding learners to compare source texts (such as voice messages) and resulting texts (textualized messages) and determining the gains and gaps in understanding between the two.

Production

A growing number of personal and professional situations require producing online content and developing multimodal skills for "empowered participation in the digital world" (Council of Europe, 2022). Chang and Asselin (2023, p. 107) argue that multimodal pedagogy should offer "opportunities for students to engage in multimodal practices that are more authentic to their identities, interests, and relationships with their worlds outside of school". One way to align the language curriculum with learners' informal or professional literacy practices is through creating multimodal texts where learners use different modes to narrate stories or explore digital genres. Digital storytelling, increasingly used in Lx classrooms, involves using digital tools to combine language and imagery (Meier, 2022). While traditionally focused on personal narratives, its educational uses now include documentaries, instructional content, and socio-political commentary. Digital storytelling allows self-expression through media, engaging learners in both conventional methods like research and writing, and innovative techniques like graphics and animation. Oskoz & Elola (2016) argue that developing these skills through process-oriented or task-based methodologies helps learners create multimodal artefacts resembling sophisticated media products.

The rapid emergence of new digital genres requires language teachers, particularly those teaching language for specific purposes, to take on complex roles when teaching specialized content, often as non-content specialists in digital environments. Dressen-Hammouda and Wigham (in press) propose an evidence-based approach, including collaboration with domain professionals, to shape multimodal pedagogy that supports students' professional literacy needs. They identified key features of video-based genres, organizing them into multimodal competency criteria (MCC). The MCC were used for external assessments of student-produced videos by domain professionals and revised after focus group discussions. The revised MCC then guided students in creating video tutorials that adhered to genre conventions and allowed teachers to formally evaluate the multimodal productions in line with professional expectations.

Indeed, the assessment of multimodal productions is complex and requires educators to move beyond criteria rooted in print media and develop alternative frameworks that more effectively capture the digital composition process and its outcomes (Tan et al., 2020) as well as assess not only isolated productions but evaluate the latter with a qualitative and collaborative perspective. In an action-research project, Lacelle et al. (2017) identified native digital genres in French and conducted research with language teachers to assist students in producing texts that align with the conventions of these genres. Moving away from writing as the primary meaning-making mode in task design incited teachers to consider other modes and necessitated the inclusion of new multimodal literacy assessment criteria. Indeed, in reflective assessment tasks, learners critically review the dynamic and situated cognitive and creative processes by which they produce meanings, thus developing their semiotic awareness, defined as "critical attention to relational, multimodal aspects of meaning design" (Towndrow et al., 2013, p. 328). Reflective multimodal portfolios (Hafner & Ho, 2020), for example, acknowledge and value the qualitative and collaborative processes in assessment as well as the multimodal artefact produced.

Interaction

Meeting and interacting digitally is an increasingly frequent and ordinary communicative practice. The development of a multimodal pedagogy requires language teachers to expand their skill and knowledge base to create opportunities for learners to orchestrate semiotic resources when interacting with others online in the Lx (Guichon & Wigham, 2016). As synchronous online interaction platforms manifest their "own material properties, feel and techniques of use, affordances

and limitations" (Chun et al., 2016, p. 65), different environments exhibit distinct potentials and constraints which influence interactions and opportunities for Lx learning.

Implementing a multimodal pedagogy for interaction requires assessing the online learning environment's affordances for Lx interaction (media assessment). The design and structure of these environments impact the available semiotic modes, shaping interactions and Lx acquisition. For example, videoconferencing platforms offer gestures, facial expressions, and gaze, which can enhance communication by aiding word searches (Cohen & Wigham, 2019) and reducing corrective feedback's impact (Vidal & Wigham, 2017). Conversely, virtual worlds with avatar-based interaction can lower the affective filter, helping learners overcome anxiety and easing Lx interaction (Melchor-Couto, 2018). Text chat interactions also vary depending on whether messages are public or private (Hampel & Stickler, 2012). Additionally, software and hardware configurations affect the multimodal composition, leading to varying perceptions of the shared space (Satar & Wigham, 2023). For instance, the number of visible webcam images may differ based on device type and layout settings. Teachers must assess the cognitive demands of each mode and adjust them to align with pedagogical goals (mode assessment).

Language teachers' professional repertoire must enable them to effectively use synchronous digital learning environments in the design of tasks that encourage interaction. This involves managing electronic resources (Satar & Wigham, 2023) and equipping learners with semio-pragmatic and socio-cultural competencies. Semio-pragmatic skills include encouraging learners to frame their webcam so gestures and micro-events (e.g., smiles) are visible to support the multimodal construction of social presence for engagement (Guichon & Wigham, 2016; Satar, 2024). Socio-cultural skills involve navigating cultural dynamics, intercultural differences, and emerging in-situ intercultures during interactions (Schröder et al., 2023). A multimodal pedagogy should consider how meaning is co-constructed in interactions, the dynamic relationalities and positionalities that emerge, and how learners share their and others' identities.

Conclusion

Multimodality has fundamentally reshaped language pedagogy, moving beyond traditional literacy to include a dynamic mix of semiotic modes. Table 1 outlines aspects of multimodal pedagogy and delineates directions for equipping language teacher with techno-semio-pedagogical skills (Guichon, 2024).

A three-step process for training educators in multimodal pedagogy could be: first, for Lx educators to recognize the importance of multimodal literacy; second, to engage in reception, production, and interaction practices themselves to understand multimodal composition and

A Multimodal Language Pedagogy, Table 1 Facets of multimodal language pedagogy

Reception	Understand how learners engage with different modes and what cognitive demand it can entail Train learners to process monomodal, bimodal and multimodal messages and assess how the combination or lack thereof facilitates or complexifies reception Train learners to use text-to-speech or speech-to-text technologies to implement transmediative strategies
Production	Understand native digital genres and their semiotic grammar Incorporate tasks, such as digital storytelling, involving multimodal composition Provide guidance that fosters the development of students' semiotic awareness Incorporate multimodal literacy criteria into assessment practices
Interaction	Understand how the structure of digital learning environments shapes tools and mediates interaction Create tasks that exploit the different semiotic affordances of digital environments for interaction Equip learners with the necessary semio-pragmatic and socio-cultural skills to accomplish interaction tasks

practices, building competence through knowledge and experience; and third, to transfer these techno-semio-pedagogical skills to guide learners in developing their own multimodal literacy.

As the digital age expands information presentation and production, language educators are tasked with guiding learners through this multimodal landscape. A multimodal language pedagogy requires Lx teaching to be responsive to technological advancements and attuned to the evolving needs of the learners. It emphasizes broadening teachers' professional skills to include creating, assessing, and facilitating tasks using diverse semiotic resources for reception, production, and interaction.

Cross-References

- ▶ Multiliteracies-Based Approaches to CALL
- ▶ Multimodal Literacy
- ▶ New Literacies: A Historical Perspective
- ▶ Semiotic Approaches in CALL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Chang, L., & Asselin, M. (2023). A case study of experiences with transmediation for a student with learning disabilities. *Language and Literacy, 25*(2), 105–130.

Chun, D., Smith, B., & Kern, R. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal, 100*, 64–80.

Clark, R. E., & Feldon, D. F. (2005). Five common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 97–115). CUP.

Cohen, C., & Wigham, C. R. (2019). A comparative study of lexical word search in an audioconferencing and a videoconferencing condition. *CALL Journal, 32*(4), 448–481.

Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal, 4*(3), 164–195.

Council of Europe. (2022). *Digital citizenship education handbook*. https://rm.coe.int/prems-003222-gbr-2511-handbook-for-schools-16x24-2022-web-bat-1-/1680a67cab

Dressen-Hammouda, D., & Wigham, C. R. (in press). A multimodal assessment grid to guide the design and evaluation of instructional video tutorials. In M. Querol-Julián & I. Fortanet-Gómez (Eds.), *Multimodal literacy in English as an additional language in higher education*. Routledge.

Guichon, N. (2024). Sur les traces de Richard Kern: Acknowledging the pivotal role of technologies in language education. *The Modern Language Journal, 108*(2), 563–566.

Guichon, N., & Cohen, C. (2014). The impact of the webcam on an online L2 interaction. *Canadian Modern Language Review, 70*(3), 331–354.

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85–93.

Guichon, N., & Wigham, C. R. (2016). A semiotic perspective on webconferencing-supported language teaching. *ReCALL, 28*(1), 62–82.

Guichon, N., Bédard, V., & Fortier, V. (submitted). Digital literacy Practices Outside Language Classrooms: Insights of Adult Migrants' Language Education. CALL Journal.

Hafner, C., & Ho, W. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing, 47*, 100710. https://doi.org/10.1016/j.jslw.2020.100710

Hampel, R., & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL, 24*(2), 116–137.

Jewitt, C. (2009). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 14–27). Routledge.

Kress, G. (2000). Design and transformation: New theories of meaning. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 153–161). Routledge.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse*. Hodder Arnold.

Lacelle, N., Boutin, J.-F., & Lebrun, M. (2017). *La littératie médiatique multimodale appliquée en contexte numérique – Outils conceptuels et didactiques*. Presses de l'Université du Québec.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). CUP.

Meier, S. (2022). Digital storytelling: A didactic approach to multimodal coherence. *Frontiers in Communication, 7*. https://doi.org/10.3389/fcomm.2022.906268

Melchor-Couto, S. (2018). Virtual world anonymity and foreign language oral interaction. *ReCALL, 30*(2), 232–249.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–90.

Oskoz, A., & Elola, I. (2016). Digital stories: Bridging multimodal texts to the Spanish writing classroom. *ReCALL, 1*(3), 1–17.

Satar, M. (2024). Social presence in multimodal online learning environments. In R. Hampel & U. Stickler (Eds.), *The handbook of language learning and technology* (pp. 272–286). Bloomsbury.

Satar, M., & Wigham, C. R. (2023). *Instruction giving in online language lessons*. Routledge.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). CUP.

Schröder, U., Adami, E., & Dailey-O'Cain, J. (Eds.). (2023). *Multimodal communication in intercultural interaction*. Routledge.

Stockwell, G. (2010). Effects of multimodality in computer-mediated communication tasks. In M. Thomas & H. Reinders (Eds.), *Task-based language learning and teaching with technology* (pp. 83–104). Continuum.

Tan, L., Zammit, K., D'warte, J., & Gearside, A. (2020). Assessing multimodal literacies in practice: A critical review of its implementations in educational settings. *Language and Education, 34*(2), 97–114.

Towndrow, P. A., Nelson, M. E., & Yusuf, W. F. B. M. (2013). Squaring literacy assessment with multimodal design: An analytic case for semiotic awareness. *Journal of Literacy Research, 45*(4), 327–355.

Vidal, J., & Wigham, C. R. (2017). Fournir des rétroactions en ligne. In N. Guichon & M. Tellier (Eds.), *Enseigner l'oral en ligne: une approche multimodale* (pp. 125–150). Didier.

Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *CALL Journal, 35*(4), 696–742. https://doi.org/10.1080/09588221.2020.1744666

R

Reflective Approaches to CALL

Meei-Ling Liaw[1] and Sabrina Priego[2]
[1]National Taichung University of Education, Taichung, Taiwan
[2]Université Laval, Québec, QC, Canada

Abstract

In this entry, we first define reflective learning and introduce the theoretical underpinnings of reflective approaches to language learning and teaching. How these theoretical underpinnings collectively inform the reflective approach to computer-assisted language learning (CALL), emphasizing the active, experiential, and critical nature of language learning through technology, are then discussed. Next, a review of the empirical studies on incorporating various technological tools, for example, blogs, e-journals, e-portfolios, and reflective mobile applications, and their associated effects on creating reflective learning opportunities for students are provided. Finally, we delve into the discussion of emerging technologies that hold potential relevance to reflective CALL, exploring the possible effects and applications of these technologies in enhancing the reflective process, engagement, and learning outcomes.

Keywords

Reflective learning · Computer-assisted language learning · Technological tools · Emerging technologies

Introduction

Reflective learning is a process where learners deliberately reflect on their experiences to extract lessons that inform their future actions and behaviours. Reflective practices encourage learners to take responsibility for their own learning and development to promote learner autonomy and agency. Learners are thus actively involved in their learning process, leading to engaging and meaningful learning experiences.

The practice of reflective learning involves actively monitoring and assessing one's knowledge, abilities, and performance, facilitating continuous improvement in understanding and performance. Reflective learning employs a variety of tools, including journaling, discussions, peer feedback, and digital technologies, to develop greater self-awareness, critical thinking, and problem-solving skills. These practices are instrumental in forming dynamic learning environments where adaptability and ongoing personal development are emphasized.

Several theorists have formed the foundation of reflective learning. One of the most influential is Donald Schön. Schön (1983) introduced the idea of "reflection-in-action" and "reflection-on-

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_28-1

action". These concepts emphasize the importance of professionals thinking about their actions during and after the event to improve their practical capabilities and outcomes. Confusions are viewed as a prerequisite to learning and need to be tolerated and understood by students. In other words, students should accept that what they are being taught would make sense in time. Another key figure is John Dewey, who underlined the role of reflection in learning and critical thinking. In his work, *How We Think*, Dewey (1933) advocated for an approach where education centred around the experiences and interests of students, encouraging them to reflect on these experiences to build knowledge. David Kolb also contributed to reflective learning with his experiential learning theory. Kolb (1984) proposed a model outlining a four-stage cycle of learning, including concrete experience, reflective observation, abstract conceptualization, and active experimentation. The model emphasizes the role of reflection in transforming experience into learning.

Based on the theoretical conceptualization of reflective learning, reflective practices in language learning can be considered deliberate and purposeful acts of thinking that centre on ways of responding to problem situations in language learning. In language education, reflective practices have been applied to help learners take informed and self-regulated actions towards language learning outcomes (Chaika, 2023; Farrell, 2022). Reflective practice in learning a new language can provide inclusive and equitable environments for learners to be exposed to various cultural perspectives and challenge their cultural assumptions and biases. It has also been adopted to help language learners develop intercultural competence and prepare them for real-world communication in multicultural and intercultural settings.

In the context of language learning, technological tools such as interactive online learning environments and multimedia resources are commonly used for students to analyse their interactions with these tools, evaluate their responses, and consider how they can improve their language skills based on feedback.

Computer-Assisted Reflective Language Learning

Reflective learning in computer-assisted language learning (CALL) environments leverages interactive platforms, such as virtual classrooms and language learning apps, to promote active participation and hands-on engagement with learning materials. These technological tools facilitate a reflective cycle where students continuously review, assess, and adjust their language skills based on immediate feedback from their interactions. The reflective processes facilitate a more engaging and personalized learning experience and significantly enhance the effectiveness of language education by allowing learners to actively reflect on and improve their language abilities.

Numerous studies have been conducted, and findings have shown that when technology is integrated into language learning, it provides opportunities for learners to reflect on their progress and outcomes, enhancing learning experiences and language skills. Social media, such as blogs, are most reported to have successful computer-assisted reflective language learning attempts. Blogs have been used to support pedagogical approaches focusing on both individual reflection and collaborative learning. For example, research by Sun and Chang (2012) and Murray and Hourigan (2008) highlights how blogs facilitate academic writing and cognitive development through interactive dialogues and peer feedback. Studies by Gebhard et al. (2011) and Martínez (2012) demonstrate that blogs enhance emergent literacy and metacognitive skills by providing platforms for reflective writing and feedback and allowing learners to share and refine learning strategies effectively. This dual capability of blogs supports self-expression and socio-cognitive learning approaches, making them an invaluable resource in language education that promotes both personal and communal knowledge construction, enhancing overall language learning engagement.

Studies have demonstrated that integrating e-journals and e-portfolios could enable learners to engage deeply with their learning processes,

allowing for more detailed and critical reflection. Nonetheless, structured and unstructured journals seemed to have influenced the depth of reflection and learning (Alt & Raichel, 2020). When combined with mobile-assisted learning applications like Duolingo, e-journals helped learners reflect on various aspects of their language learning strategies, task engagement, and challenges, thereby improving metacognitive awareness (Kessler, 2023). Reflective portfolios, used in collaborative settings, allow for both individual and collective reflection, facilitating critical engagement with the learning content and enhancing problem-solving skills (Marden & Herrington, 2022).

Technology-mediated tasks, such as in Chen and Lin's (2024) study adopting a reading-based e-learning approach for EFL learners to improve reading comprehension, have also shown improvements due to the affordances of reflective thinking and behaviour engagements. In another study, Chong and Reinders (2020) designed the task of literature search on digital libraries and refereed journals, which has also helped students become more aware of their language errors and the use of words and phrases in communication. Both technology-mediated tasks improved learners' practical language skills and enhanced comprehension and reflective thinking about texts.

In addition to classroom settings, mobile technological tools allow students to engage in day-to-day practices as well as in short-term study abroad settings. For example, Ros i Solé et al. (2010) integrated MP3 so that the students could record their language tasks, questionnaires, classroom discussions, and English interviews with their teacher and the researcher. The exercises allowed the students to use technology for their day-to-day practices in a variety of situations and to rehearse and confront this newly found L2 self. In the study by Lomicka and Ducate (2021), Padlet was used for students on short-term study-abroad trips to promote students' deeper engagements with the target culture and reflection on the experience. The approach helped the students form an intellectual community in which the members were observant of the target culture.

Technology-enhanced reflective approaches have also been found to be effective in language teacher education. Gönen (2019) found that technology can become an effective teaching and learning tool by providing preservice teachers with training to integrate technology into their language teaching practices. A review study by Tuncer and Özkan (2021) found that reflective tools have been applied to teacher professional development, including videos, e-portfolios, and blogs, leading to positive effects, such as fostering self-awareness in individuals, promoting skills development, maintaining ongoing professional growth, integrating theoretical knowledge with practical application, and strengthening relationships among all participants in the educational environment. The edited volume by Hubbard and Ioannou-Georgiou (2017) introduced various approaches in which teachers used technology to teach English and other languages reflectively and offered guidelines for incorporating this into teaching practice.

Emerging Technologies for Reflective Language Learning

As technology advances, language educators and learners will have more tools available to facilitate language learning. Several emerging technologies hold the potential to enhance language learning through reflection, as they are designed to deepen learners' engagement with the language learning process. For example, artificial intelligence (AI) and adaptive learning systems can analyse learners' interactions and performances to deliver automated feedback and content, thereby supporting reflective learning by helping learners identify areas for improvement and encouraging personalized learning paths (Liu et al., 2023). In addition, learners can use AI Chatbots to practice language, reflect on their performance, and adjust their language use. Similarly, XR immersive language learning environments that simulate real-world scenarios enable learners to practice language skills in contextually rich environments, reflect on their performance, and adjust their language use in real time. Advanced analytics tools

integrated within language learning platforms are enabling educators and learners to track and reflect on detailed learning data. By understanding language acquisition and usage patterns, learners can make informed decisions about their learning strategies, fostering a reflective approach to language mastery.

Conclusion

Grounded in the theoretical frameworks established by scholars like Donald Schön, John Dewey, and David Kolb, reflective learning emphasizes continuous reflection during the learning process, fostering deeper understanding and skill development. The practice of reflective language learning incorporates a deliberate and structured approach to enhance learners' linguistic, cognitive, and intercultural competencies through the use of various technological tools.

Computer-assisted reflective language learning environments enable students to reflect on their language use, assess their progress, and interact with diverse cultural materials, enhancing their learning experience. Reflective practices supported by these technologies have shown significant benefits in making them essential components of language learning.

As this field continues to evolve, emerging technologies are set to facilitate a more nuanced engagement with the language, allowing learners to critically analyse their progress and adapt their learning strategies in real time. As such, the future of language education will increasingly take advantage of technological advancement to foster learning processes that are more reflective, inclusive, and adapted to the diverse learning styles and needs of language learners.

Cross-References

- ▶ Integrating Social Media in CALL
- ▶ Mobile-Assisted Language Learning (MALL)
- ▶ Task-Based Language Learning in CALL Environments
- ▶ The Impact of Artificial Intelligence (AI) on CALL Pedagogies
- ▶ Transforming Language Education: Exploring the Pedagogy of Adaptive Learning Systems
- ▶ Trends in Pedagogical Approaches to CALL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Alt, D., & Raichel, N. (2020). Reflective journaling and metacognitive awareness: Insights from a longitudinal study in higher education. *Reflective Practice, 21*(2), 145–158. https://doi.org/10.1080/14623943.2020.1716708

Chaika, O. (2023). Role of reflective practice in foreign language teaching and learning via multicultural education. *International Journal of Social Science And Human Research, 6*(2), 1343–1350. https://doi.org/10.47191/ijsshr/v6-i2-74

Chen, M. R. A., & Lin, Y. H. (2024). A reflective e-learning approach for reading, thinking, and behavioral engagement. *Language Learning & Technology, 28*(1), 1–20. https://hdl.handle.net/10125/73548

Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning and Technology, 24*(3), 70–86.

Dewey, J. (1933). *How we think*. Henry Regnery Company.

Farrell, T. S. C. (2022). *Reflective practice in language teaching*. Cambridge University Press.

Gebhard, M., Shin, D., & Seger, W. (2011). Blogging and emergent L2 literacy development in an urban elementary school: A functional perspective. *CALICO Journal, 28*(2), 278–307.

Gönen, S. İ. K. (2019). A qualitative study on a situated experience of technology integration: Reflections from pre-service teachers and students. *Computer Assisted Language Learning, 32*(3), 163–189. https://doi.org/10.1080/09588221.2018.1552974

Hubbard, P., & Ioannou-Georgiou, S. (2017). *An introduction to teaching languages reflectively with technology*. IATEFL. ISBN: 978-1-901095-86-9.

Kessler, M. (2023). Supplementing mobile-assisted language learning with reflective journal writing: A case study of Duolingo users' metacognitive awareness. *Computer Assisted Language Learning, 36*(5–6), 1040–1063. https://doi.org/10.1080/09588221.2021.1968914

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

Liu, C., Hou, J., Tu, Y. F., Wang, Y., & Hwang, G. J. (2023). Incorporating reflective thinking promoting

mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments, 31*(9), 5614–5632. https://doi.org/10.1080/10494820.2021.2012812

Lomicka, L., & Ducate, L. (2021). Using technology, reflection, and noticing to promote intercultural learning during short-term study abroad. *Computer Assisted Language Learning, 34*(1–2), 35–65. https://doi.org/10.1080/09588221.2019.1640746

Marden, M. P., & Herrington, J. (2022). Encouraging reflective practice through learning portfolios in an authentic online foreign language learning environment. *Reflective Practice, 23*(2), 177–189. https://doi.org/10.1080/14623943.2021.2001321

Martínez, C. d. A. (2012). Developing metacognition at a distance: Sharing students' learning strategies on a reflective blog. *Computer Assisted Language Learning, 25*(2), 199–212. https://doi.org/10.1080/09588221.2011.636056

Murray, S., & Hourigan, T. (2008). Blogs for specific purposes: Expressivist or socio- cognitivist approach? *ReCALL, 20*(1), 82–97.

Ros i Solé, C., Calic, J., & Neijmann, D. (2010). A social and self-reflective approach to MALL. *ReCALL, 22*(1), 39–52. https://doi.org/10.1017/S0958344009990188

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Sun, Y.-C., & Chang, Y.-J. (2012). Blogging to learn: Becoming EFL academic writers through collaborative dialogues. *Language Learning and Technology, 16*(1), 43–61.

Tuncer, H., & Özkan, Y. (2021). Reflective tools in an English language teaching context: Contributions, challenges and recommendations. *Kahramanmaraş Sütçü İmam University Journal of Social Sciences, 18*(1), 181–204. https://doi.org/10.33437/ksusbd.671482

C

Corpus-Informed Materials in Language Teaching

Eric Friginal[1] and Malila Prado[2]
[1]The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[2]BNU-HKBU United International College, Zhuhai, China

Abstract

This entry discusses the use of corpus-informed materials for language teaching and learning. It outlines how corpus linguistics employs large datasets of authentic texts to study language patterns and distributions, particularly for teaching applications. Educators and materials creators have utilized real-world and frequency-based language data that are ideal for materials design and classroom activities. Corpus tools and resources, including learner corpora have grown more accessible, especially online for both teachers and students. While barriers remain, corpus-based materials can enhance learner autonomy, exposure to authentic language samples, and data-driven discovery of patterns. Over the past two decades, teachers and materials developers who have incorporated corpus data, therefore, have produced teaching materials that reflect accurate language-in-use across pedagogical contexts.

Keywords

Corpus linguistics · Data-driven learning · Corpora in the classroom · Online corpora and databases

Introduction

Corpus linguistics is a research methodology focused on studying language, particularly the structure, patterns, and usage of discourse through the analysis of corpora (Biber et al., 2010). A corpus, then, serves as a dataset of "systematically collected, naturally-occurring registers of texts" (Friginal & Hardy, 2014, p. 20), allowing for in-depth examination of language use, the identification of linguistic distributions and trends, and the exploration of discourse features across different contexts and genres. The use of corpora has increasingly gained popularity as a method for quantitatively analysing the linguistic features of both written and spoken language, especially in the context of academic discourse. Not too long ago, it was relatively common to find English Language Teaching (ELT) materials presenting language in an artificial, unauthentic manner. Despite the Communicative Language Teaching approach advocating for real language, it was not until the advent of corpus linguistic studies that we began to witness the integration of authentic language data into teaching resources in ELT, enabling learners to engage with genuine

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_45-1

examples of language use and to better prepare for real-world communication (Friginal, 2018).

Following Leech (1997), Römer (2011) identifies two applications of corpora in ELT: *direct* and *indirect*. Direct applications involve the use of corpora or corpus tools by teachers and students in the classroom, either through worksheets or computer-based activities. In contrast, indirect applications refer to pedagogical materials such as learner dictionaries, textbooks, syllabi, and language tests, which are informed by corpus research. Both direct and indirect applications of corpora and corpus tools in educational settings support various theories and constructs related to successful ELT as well as second language acquisition. These include learner autonomy, motivation through interactions between learners and computers or among peers, and the explicit instruction of language features and patterns.

The presence and influence of corpus-based materials in the classroom increased exponentially from the 1990s to the present. Whereas corpus databases and concordancing software were previously limited, such resources are now readily available to teachers online. There are several corpora containing a variety of registers (also referred to as "text types") including classroom and learner English, formal and informal speeches, academic and newspaper articles, short stories and legal documents, or scientific and technical writing that have the potential for direct teaching applications. Additionally, several textbooks focused on corpus linguistics for language teaching have been published within the past two decades.

Data-driven learning (DDL), which is directly aligned with the corpus approach, adopts the inductive model of turning the learner into a researcher. The basic principle is that teachers provide students with authentic language data, such as concordance lines, and scaffold them until they can independently understand how a specific language phenomenon occurs through analysing patterns in the examples (Boulton & Tyne, 2013; Boulton & Vyatkina, 2024). In the early applications of DDL, students primarily analysed concordance lines manually to identify textual characteristics (Johns, 1991). However, DDL has since evolved into what Pérez-Paredes (2024) calls Broad DDL (BDDL), which includes not only more traditional DDL materials but also a wider array of online language learning resources such as online dictionaries, text analysis and text processing tools, vocabulary-oriented websites and apps, translation services, and artificial intelligence tools for language learning across various contexts and platforms, allowing for more self-directed study (Pérez-Paredes, 2024). This evolution is facilitated by specialized corpora designed specifically for DDL activities, such as CorpusMate (corpusmate.com) which provides students direct access to annotated corpora. Learners can then autonomously explore corpora to uncover patterns in usage, gaining a deeper understanding of a word or structure through analysing its various contexts of use (Crosthwaite & Baisa, 2024).

Corpus-Based Dictionaries

The use of corpora in language teaching saw slow adoption, with lexicography as perhaps the first area of study to enjoy fruitful benefits of corpus-based methods by relying on frequency, collocates, semantic preference, and prosody (Moon, 2007). Corpus-based dictionaries such as Oxford English Dictionary, Longman Dictionary of Contemporary English, or COBUILD Dictionary leverage corpus data to enhance the accuracy and relevance of their entries, benefitting learners of English. Word meanings and their frequency of use have been found to help determine what students needed to prioritize in the beginning of their studies (Sinclair & Renouf, 1988). In addition, word collocations could assist in the development of structural awareness (Carter & McCarthy, 1988), and vocabulary clusters may be helpful in identifying discourse and conversational markers (Burton, 2023). This emphasis on corpus-based insights not only enriches the lexicographical materials and resources available to learners but also significantly enhances their understanding and application of language in real-world contexts.

Reference Grammars and Textbooks

Historically, major initiatives such as the pioneering work by Biber and his colleagues, *The Longman Grammar of Spoken and Written English* (1999) demonstrated the value of accurate representations of language in textbooks and classroom settings, including register distinctions between British and American English. More specifically, *Real Grammar* (Conrad et al., 2009) was one example of a teacher-focused monograph that presented grammar and structural forms of English from corpus data to match a particular teaching context. To illustrate this, the book highlights a real-world preference for the use of adjectives instead of adverbs in colloquial conversations, such as *you could do it* ***quick*** and *look at how* ***slow*** *this car is going* (Conrad et al., 2009, p. 73). The book then guides students in understanding the contexts in which it is relatively acceptable to use adjectives rather than adverbs, contrary to traditional grammar rules. More recently, the project *Corpus for Schools* developed by Lancaster University (https://wp.lancs.ac.uk/corpusforschools/) involves the design of materials available online both for A-Level students, that is, English first language (L1) speakers, and second language (L2) students. *Corpus for Schools* materials address the preference of certain structures depending on the speakers and also the context in which they are uttered.

Databases and Open-Sourced Materials

Popular corpus databases such as the Corpus of Contemporary American English (COCA) developed by Davies (see other databases from Davies here: https://www.english-corpora.org) or those compiled with the help of the corpus tool Sketch Engine (https://www.sketchengine.eu/) have been used for teacher-created lessons and handouts. This trend suggests that not only are teachers making use of corpora to inform their lesson plans, but students are also directly engaging with corpus data, gaining hands-on experience with language analysis. The flexibility of corpus-informed materials is also particularly advantageous for learners of languages other than English. The dearth of specially designed materials for these learners encourages corpus use, though these resources are also still limited. For example, Forti (2024) discusses challenges in this field using a course for Chinese L1 speakers learning Italian as an example. Due to time constraints, corpora were not used in the classroom; instead, handouts with concordance lines and activities such as pattern hunting, matching, and gap-filling were used to focus on collocations. In an open-source initiative by Pinto et al. (2023), a lesson plan for learning Spanish was presented using Voyant Tools (https://voyant-tools.org/) to help students analyse a text for frequent words and verbs collocating with the pronoun *me*. The activities have taken into account multilingual students, enhancement of linguistic skills, and the goal of developing intercultural communication competence.

Open-sourced resources are a valuable means through which teachers can collaboratively design and implement corpus-informed classroom materials. Several initiatives have provided important resources for this purpose. Projects led by Le Foll (2021), Goulart and Veloso (2023), and Vyatkina (2020) offer accessible tools and data. Among the many available resources are lesson plans designed by language teachers for use in primary, secondary, and tertiary classrooms. Ma's *Corpus-Aided Platform for Language Teachers* (https://corpus.eduhk.hk/cap/) is active and up-to-date with tools and relevant corpora. A tab for "Teaching Activities" includes teaching vocabulary lessons developed by teachers and links to resources and publications are also provided (Ma et al., 2022).

Materials for ESP/EAP

Work in English for Specific/Academic Purposes (ESP/EAP) settings also tends to be more attuned to corpus-informed materials, perhaps due to an emphasis on needs analysis and realia. Studies by Charles (2018, 2024) recommend do-it-yourself

(DIY) corpora to engage students with language data in their own areas of expertise, enhancing understanding of domain-specific vocabulary and structures and empowering independent research. Similarly, Frankenberg-Garcia et al. (2022) propose workshops where university teachers and researchers learn to apply corpus tools to improve academic writing skills. The goal is to enhance participants' ability to use corpus data to inform writing practices in the disciplines, making output more precise and contextually appropriate.

Concluding Remarks

Corpus linguistics can make a positive contribution in various language teaching contexts through the use of corpora and the production of corpus-based materials. Learners stand to benefit from the practical applications of corpus data as they learn English (and other languages) in the classroom. Corpora have greatly contributed to the study of typical vocabulary, phraseological, and collocational patterns of language, demonstrating how learning and teaching can be informed by such patterns. Learners can more readily grasp that a word's significance and function reach even further to incorporate recurring lexical features. As Römer (2009) observed, "language uses highly patterned structures" (p. 140), and highlighting these patterns is often important in designing materials for language teaching.

However, despite the advantages of incorporating corpus-based materials, there are notable limitations that educators and learners may encounter. A significant challenge includes accessibility, which may require a certain level of technological proficiency that not all teachers and students possess. Additionally, Friginal (2018) reported teacher resistance to classroom use of corpus-informed materials as one key stakeholder concern. Teachers have also mentioned that corpus activities are time-consuming (McCarthy, 2008) and that indirect materials and resources are not generally user-friendly (Frankenberg-Garcia, 2012; Farr, 2008). Also, the quality and representativeness of the corpora themselves can vary, potentially leading to skewed or incomplete language models (Egbert et al., 2022). If the data is not diverse enough, it may fail to capture regional variations or the nuances of specific contexts, which can limit learners' exposure to the full spectrum of language use. Furthermore, reliance on corpus data may inadvertently prioritize certain language patterns at the expense of others, potentially constraining learners' linguistic creativity and flexibility.

Looking ahead, there are promising future directions for corpus-based materials in language teaching. Advances in technology, such as artificial intelligence and machine learning, offer opportunities to develop more sophisticated corpus tools that can provide personalized learning experiences tailored to individual learner needs. Additionally, the integration of corpora into mobile applications can make these resources more accessible and engaging for students. Future research could also explore the integration of contextualized corpora that reflect specific cultural and situational uses of language. By continuously adapting and expanding corpus-based materials, educators can foster an even richer and more effective language learning environment that prepares students for real-world communication.

Cross-References

- ▶ Corpus Literacy and Data-Driven Learning
- ▶ Data-Driven Learning and Languages Other than English
- ▶ Future Challenges and Opportunities for Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Usage-Based Approaches and Data-Driven Learning
- ▶ Writing Aids for Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (Eds.). (1999). *Longman grammar of spoken and written English* (10. impression ed.). Longman.

Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In R. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 548–570). Oxford University Press.

Boulton, A., & Tyne, H. (2013). *Corpus linguistics and data-driven learning: A critical overview*. https://api.semanticscholar.org/CorpusID:162048311

Boulton, A., & Vyatkina, N. (2024). Expanding methodological approaches in DDL research. *TESOL Quarterly, 58*(3), 1193–1204. https://doi.org/10.1002/tesq.3269

Burton, G. (2023). *Grammar in ELT and ELT materials*. Multilingual Matters. https://doi.org/10.21832/BURTON5270

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. Routledge. https://doi.org/10.4324/9781315835860

Charles, M. (2018). Using do-it-yourself corpora in EAP: A tailor-made resource for teachers and students. *Journal of Teaching English for Specific and Academic Purposes*, 217. https://doi.org/10.22190/JTESAP1802217C

Charles, M. (2024). Benefits and challenges of using do-it-yourself corpora for academic writing development. *TESOL Quarterly, 58*(3), 1205–1214. https://doi.org/10.1002/tesq.3276

Conrad, S., Biber, D., Daly, K., & Packer, S. (2009). *Real grammar: A corpus-based approach to English*. Pearson/Longman.

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*. https://doi.org/10.1075/ijcl.23056.cro

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Routledge.

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness, 17*(1), 25–43. https://doi.org/10.2167/la414.0

Forti, L. (2024). Data-driven learning, second language acquisition and languages other than English. In P. Crosthwaite (Ed.), *Corpora for language learning* (1st ed., pp. 59–70). Routledge. https://doi.org/10.4324/9781003413301-5

Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching, 45*(4), 475–489. https://doi.org/10.1017/S0261444810000480

Frankenberg-Garcia, A., Pinto, P. T., Bocorny, A., & Sarmento, S. (2022). Corpus-aided EAP writing workshops to support international scholarly publication. *Applied Corpus Linguistics, 2*(3), 100029. https://doi.org/10.1016/j.acorp.2022.100029

Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. Routledge.

Friginal, E., & Hardy, J. A. (2014). *Corpus-based sociolinguistics: A guide for students*. Routledge.

Goulart, L., & Veloso, I. (2023). *Corpora in English language teaching: Classroom activities for teachers new to corpus linguistics*. Open Educational Resource. https://pressbooks.pub/testbook123/

Johns, T. F. (1991). Should you be persuaded? Two samples of data-driven learning materials. https://api.semanticscholar.org/CorpusID:53988458

Le Foll, E. (2021). *Creating corpus-informed materials for the English as a foreign language classroom: A step-by-step guide for (trainee) teachers using online resources*. Zenodo. https://doi.org/10.5281/ZENODO.4992504

Leech, G. (1997). Teaching and language corpora: A convergence. In I. A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–24). Longman.

Ma, Q., Yuan, R., Cheung, L. M., & Yang, J. (2022). Teacher paths for developing corpus-based language pedagogy: A case study. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2040537

McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching, 41*(4), 563–574. https://doi.org/10.1017/S0261444808005247

Moon, R. (2007). Sinclair, lexicography, and the Cobuild project: The application of theory. *International Journal of Corpus Linguistics, 12*(2), 159–181. https://doi.org/10.1075/ijcl.12.2.05moo

Pérez-Paredes, P. (2024). Data-driven learning in informal contexts? Embracing broad data-driven learning (BDDL) research. In P. Crosthwaite (Ed.), *Corpora for language learning* (1st ed., pp. 211–226). Routledge. https://doi.org/10.4324/9781003413301-15

Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. (Eds.). (2023). *Using language data to learn about language: A teachers' guide to classroom corpus use*. The University of Queensland. https://doi.org/10.14264/3bbe92d

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*, 140–162.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225. https://doi.org/10.1017/S0267190511000055

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. In Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140–160). Routledge.

Vyatkina, N. (2020). *Incorporating corpora: Using corpora to teach German to English-speaking learners*. University of Kansas Open Language Resource Center. https://opentext.ku.edu/corpora

T

The Scope of Data-Driven Learning

Fiona Farr
University of Limerick, Limerick, Ireland

Abstract

This entry explores the scope of data-driven learning (DDL) from its original understanding presented to the academic world by Tim Johns in the 1990s to what it has emerged to become in a modern technological world at the frontier of the potential of artificial intelligence (AI). It discusses its original conceptualisation and how that might be adjusted and expanded in light of more modern and current data availability, configuration and digital manipulation tools and resources. The impact of the development and growth of the World Wide Web is discussed as both a data source and search tool, as well as the more ready availability of corpus-based pedagogic resources which were not present more than 30 years ago. A distinction between DDL and corpus linguistics, as these terms are now used, is made before a final note on the ever-changing face of such technologies with reference to the growing capabilities of AI.

Keywords

Data-driven learning · Corpus linguistics · Language learning

Introduction

In the late 1990s, as a then PhD student, when I first read Tim Johns' article introducing the notion of DDL (Johns, 1991), it was relatively clear and obvious to me what it meant and how it was operationalised in the language classroom. Many years later, I have to admit that I am not so sure, as I contemplate the same concept in an educational context which has been revolutionised by technology in the interim. In fact, I re-read the original article in preparation for writing this chapter to re-anchor my thoughts in his original thinking, trying to recall papers I saw him present at conferences and discussions had with him in the years following the original publication. The idea then could be easily described, not so any longer. In this short entry, I begin by revisiting the original concept as it was understood at the time; next I trace some changes over time, which have, in turn, led to the evolution of DDL. This is followed by a brief discussion of the criteria I consider to be important in distinguishing between DDL and corpus linguistics, and I conclude with a short note on the potential impact of AI in this pedagogic space.

DDL in a Changing World

When it was first articulated by Johns, DDL meant providing language in the form of:

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_46-1

> evidence needed to answer the learner's questions, and rely on the learner's intelligence to find answers. The most important computing tool for the data-driven approach is the concordancer, which is able to recover from text all the contexts for a particular item (morpheme, word or phrase) and to print them out in a way which facilitates rapid scanning and comparison. The most usual format is the keyword-in-context (KWIC) concordance in which the keywords are arranged one below the other down the centre of the page, with a fixed number of characters of context to the left and to the right. A useful refinement, particularly where one is concerned with regularities and patternings in large numbers of citations, is the ability to sort alphabetically the contexts to the left or right of the keyword so that similar contexts are grouped together. (Johns, 1991, p. 1).

In the original paper, Johns frames the corpus as a kind of language informant and illustrates with two examples of printed handouts of concordance lines used by his learners: the first to help them distinguish between the meanings and uses of the two words *convince* and *persuade* and the second to revise the modal verb *should*. One of the convince/persuade handouts from the original paper is included in Fig. 1.

He claimed, based on his observations of DDL in his own classrooms over a 3–4-year period, that, firstly, the use of a concordancer stimulates learners' sense of enquiry, justifying his assertion that 'research is too serious to be left to the researchers' (Johns, 1991, p. 2), secondly, the teacher's role changes to becoming a coordinator of student research and thirdly, that grammatical description emanates from the learner's engagement with the language rather than from descriptive accounts produced in grammar reference books. Indeed, since that time, empirical studies support assertions about the positive impact of DDL approaches on learners (for a systematic review, see Boulton & Cobb, 2017). In practical terms then, for some time in the 1990s and early 2000s, DDL meant generating wordlists or concordances from one of the few corpora available at the time (see Murphy & Riordan, 2016, for a discussion of corpus collections), or a DIY built corpus collection (O'Keeffe & Farr, 2003). Learners would then try to find the answers to their linguistic questions through reading, investigating and examining the corpus output. This could be done by learners through direct access to the corpus and computer with the associated software or through investigating corpus-based materials produced in advance by the teacher.

So what has changed since then? Two major developments have occurred. Firstly, in 1991, in exactly the same year that Johns' article was published, the first website was introduced to the world after the original idea was proposed in 1989 (Fischels, 2021). Since then, the Web, in all of its iterations, has become a massive repository/database of language, making it easy to search for language examples through general search tools such as Google or more purpose-designed interfaces such as WebCorp Live. This raises the question as to whether the Web can be considered a corpus and whether search engines can be considered concordancers, based on our definitions and understandings of what these are. Some accept that it is while others cast doubt (for a full discussion, see Kilgarriff & Grefenstette, 2003 and other papers in this special issue of the journal *Computational Linguistics*). This is still an open question, especially given that many corpus collections are now compiled/scraped from the Web, for example, the EnTenTen corpus available on SketchEngine, and many specialised corpora built by language teachers for use in their classrooms. And while the purist in me is inclined towards not accepting the Web as a corpus due to the lack of principled data collection, storage, metadata and the non-specificity and limitations of search engines as concordancers, I am reminded that in all of my introductory classes I teach on corpus linguistics, one of the ways through which I introduce the principles of corpus linguistics is with a demonstration of Google N-Gram Viewer. On that basis, in the context of DDL, I concede that the web-as-corpus is a useful, if not somewhat limited, language learning tool, compared with more traditional corpus approaches. Some have used it successfully but often with a need for significant teacher mediation (for example, Meyer, 2006). Whatever your stance, there is no denying that there is much validity in Gatto's assertion that the development of web-based 'tools and resources are decidedly redirecting the way we conceive of corpus work' (Gatto, 2011, p. 35), and that includes DDL.

CONVINCE - PERSUADE

Look carefully at the following examples of the verbs 'convince' and 'persuade' (and words derived from th
verbs such as 'convincing* and 'persuasion'/'persuasive'). What similarities and what differences can you
find between the two verbs?

1) ays that universities urgently need to convince academics that popularising research is re
2) rviews by Professor Ian Fells ought to convince producers elsewhere that talking heads are
3) produce literature detailed enough to convince the prospective buyer. Ivanov's major inte
4) hbouring system will find it harder to convince their own establishment that they need new
5) ggling sister or even the queen should convince us that behaviour can seem intelligent in
6) prising that New Scientist should have convinced itself that the nuclear weapons policy of
7) ed the Neolithic revolution and became convinced that it was not a matter of someone havin
8) given. But French nuclear experts are convinced that the "reference accident" (a core mel
9) ating how the British had determinedly convinced themselves they were enjoying themselves
10) s that lead you to think that they are convinced they are doing everyone else a favour, or
11) mpounded the crime by failing to build convincing alternative views of the relationship be
12) Lady Di look-alike. The results looked convincing. Equally important, the need for rigour
13) t ability to make and use maps provide convincing evidence of active intelligence? And if
14) ome of these interpretations. The most convincing example is that of the Davy lamp (New Sc
15) serve its own equilibrium, was never a convincing model of human behaviour. Many observed
16) ad been a success, we would still need convincing that a trial like this, with patients re
17) 1 spent, even if results are less than convincing. When it comes to marketing community in
18) ean areas. It is fascinating and it is convincing. Yet von Daniken is a millionaire and Ra
19) . Early on in both editions he writes, convincingly enough, of those who are unable to pre
20) t has been demolished so often, and so convincingly, that most readers will find further r

1) manager for remote sensing will try to persuade different parts of the government to spend
2) in of sense. Incidentally, how did you persuade Michael Heseltine to write it for you? Gal
3) n early stage. Second, it is trying to persuade researchers that it is a good thing to wor
4) t two years trying, unsuccessfully, to persuade the British government to make some contrl
5) ogy, is planning a mission to India to persuade the country to invest in British satellite
6) ely that only a big fire disaster will persuade the government to look harder at fire rese
7) entry to show farmers its work, and to persuade them to "get more out of muck". Scientists
8) likely to survive. So a female who can persuade two males that they each have a stake in t
9) nisation that focuses on that disease, persuaded both the House and Senate that arthritis
10) 's bad image among many scientists has persuaded it to woo the science community more expl
11) . At the end of last year, the FDA was persuaded that the Utah Medical Center had evolved
12) ncy services. In 1979, the Home Office persuaded the radio conference to allow it to conti
13) bolster Berlin's economy? Kewenig has persuaded the state of Berlin to give him OM 11 mil
14) nd by word of mouth; how people can be persuaded to adopt it by first influencing opinion
15) the state governments, who can also be persuaded to find fairly large experimental facilit
16) at Marly in France where a dragoon was persuaded to pick up a long brass wire inside a gla
17) literature. There is no one better at persuading people to do things they might not other
18) ablish the economic basis of Berlin by persuading researchers of industry to put their hea
19) ment to take action which, if friendly persuasion fails, end up ultimately in the magistra
20) how good they are at communication and persuasion, how resistant the population is to chan
21) he agricultural order. There have been persuasive arguments for the building of raised fie
22) ers that the Milwaukee Project "offers Persuasive evidence that mental retardation in the

The Scope of Data-Driven Learning, Fig. 1 Convince and persuade DDL hand-out (Johns, 1991, p. 16)

Secondly, two developments in corpus linguistics have also had a strong influence on the potential for integration as DDL. Firstly, corpus size and availability have exploded over the last 30 years, and the initial difficulties and costs associated with teachers trying to get access to or build their own corpora have virtually disappeared, opening the gateway even further for DDL. We now have access to spoken, written and multimodal corpora, parallel corpora and even simplified readers (for an account of corpus types and uses, see Murphy & Riordan, 2016). This has had two primary impacts. It has enhanced opportunities for DDL but has also ameliorated and refocused the role of the teacher as mediator for learners in the vast array of available options. However, aligned with this growth, corpus materials have also increasingly been repackaged and offered through pedagogically oriented interfaces such as SKELL from Sketch Engine, meaning that the backend of the operation is often hidden from the view of the learner. This, to me, is simplified

and more accessible DDL, which is likely to be more attractive to learners as it is less onerous on them both technically and conceptually, but the main drawback is that it makes serendipitous inquiry and learning less likely and could undermine, to an extent, Johns' notion of 'every learner a Sherlock Holmes'.

DDL and Corpus Linguistics: The Same but Different?

It strikes me as important that any discussion of DDL and corpus linguistics must necessarily include a strong distinction between these two terms, not only in terms of their meaning but in terms of their intended and likely outcomes. In contemporary narrative and usage, these terms, with DDL meaning the classroom application of corpus linguistics, seem to have drifted quite far apart. This is especially true if we accept all of the web and corpus-based digital tools and interfaces now used for pedagogic purposes, as discussed in the last section, as DDL tools. Pedagogic intent is key to this distinction and the related learner's perspective. It is possible to use/produce DDL materials, which rely on corpora, in contemporary learning contexts without directly consulting a (traditional) corpus. The learner will have access to language evidence which can act as informant for them to test their hypotheses and draw conclusions, as was intended by Johns. Sometimes this will mean accessing output in a form similar to a concordance and may or may not allow different configuration options such as alphabetical presentation. This wide range of resources, with a DDL flavour, are potentially useful and beneficial in language learning contexts. Therefore, it is reasonable to stretch the definition of DDL as the world of related resources stretches.

If consulting a corpus directly for DDL purposes, then teachers (and often students) need to consider questions such as how big does it need to be? The answer to this question is almost invariably; it depends on what one wants to search for and investigate and whether ample examples can be found in the size of corpus under scrutiny. High-frequency items, such as primary verbs, are easily retrievable from even small corpora, whereas more specialised and lower-frequency items may need a larger corpus to contain ample examples. This brings us to the related question of how much is ample for language learning purposes? Again, the answers is 'it depends' and it will vary from one learner to the next. I may need to see 20 examples to grasp meanings and use and associate patterns, while another learner may need just 10. And, of course, some examples may be more transparent, and telling than others so if a random selection of concordances is generated, it may be effective enough to get ample good examples, whereas the next random selection may present a number of examples that are not particularly useful because they are hard to interpret out of context or because the concordance format has cut them off at too early a point to be helpful to the learner. There are also added constraints associated with copying concordances onto a physical sheet of paper which is space-limited, although the use of paper seems to be quickly disappearing from many classroom contexts. Equally, under universal design for learning principles, we may need to consider whether a concordance-based presentation, which relies on word-heavy input for learners, is most suitable. Corpus software, and related pedagogic resources, now offer alternative presentation modes, such as word clouds, sketches and other modes of visualisation, which may be more appropriate or attractive to learners. This offers us unique opportunities to change the face of DDL in ways that were not possible 30 years ago.

Corpus linguistics, on the other hand is a related but different endeavour and while it might be implemented in language learning contexts, it is nowadays largely employed by language researchers or teachers with the intention to create materials for the classroom. To be considered as such, I hold that it should maintain the important characteristics of this approach by affording the following:

- A principled collection of texts with appropriate metadata, which can be exploited by recognised corpus analysis software
- Quantitative and qualitative outputs and dimensions

- Output format variability and flexibility (highlighting patterns and allowing reorganisation of the data)
- Attested language examples in their 'raw' state, rather than interpretations (i.e. unmediated or interpreted by intelligent technology).

This final bullet point brings us to a brief discussion of AI in the conclusion of this entry.

Conclusion

In his 1991 paper, Tim Johns interestingly made the case for DDL because of the inadequacies of AI at the time by saying:

> This failure [of AI] has been due not merely to the ignorance of linguistics and pedagogy of AI researchers, or to the ignorance of information technology of language teachers, but rather to the nature of the enterprise itself: the "rules of language" cannot with any plausibility be specified in computer terms to allow the machine to carry out as apparently simple an "informant" task as correcting the English of a piece of student writing. A spelling checker can make a reasonable stab at identifying and suggesting corrections for mis-spellings, while a so-called grammar checker can look for a limited range of "obvious" blunders (and may not be very good even at identifying those). Otherwise, a rule-based system is helpless: it cannot begin to come near the ability of the human informant/teacher to work out what the writer was trying to say or to find an appropriate way of expressing that meaning for the intended audience. (Johns, 1991 p. 1).

In effect, he argues that the weaknesses of AI as a language 'informant' could be compensated for by the potentials of DDL. So the question now arises as to whether the strengths and astounding affordances of AI in modern times negates or reduces the need for a more traditional corpus-based DDL approach? I think few would argue that the answer is not yes. For language learning purposes, AI tools such as ChatGPT or Copilot can relatively easily and accurately produce language information by teachers or learnings inputting the correct instruction (in other words, asking the right questions). And it can generally produce it in a more familiar reader-friendly format. Of course, it does this by using the Web as a corpus; in other words it is a DDL approach. Have we therefore come full circle?

Cross-References

- ▶ DIY Corpora and DDL
- ▶ Hands On vs Hands-Off Methods
- ▶ Non-Corpus DDL (DDL Lite)
- ▶ The History of DDL
- ▶ Tim Johns and DDL
- ▶ Web as Corpus for DDL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Fischels, J. (2021). *A look back at the very first website ever launched, 30 years later*. Retrieved November 7, 2024, from https://www.npr.org/2021/08/06/1025554426/a-look-back-at-the-very-first-website-ever-launched-30-years-later#:~:text=On%20August%206%2C%201991%2C%20the,about%20how%20to%20use%20it

Gatto, M. (2011). The 'body'and the 'web': The web as corpus ten years on. *ICAME Journal, 35*, 35–58. https://doi.org/10.1515/icame-2015-0008

Johns, T. (1991). Should you be persuaded – Two samples of data-driven learning materials. *English Language Research Journal, 4*, 1–16.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics, 29*(3), 333–347. https://doi.org/10.1162/089120103322711569

Meyer, C. F. (2006). Corpus linguistics, the World Wide Web, and English language teaching. *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos, 12*, 9–21.

Murphy, B., & Riordan, E. (2016). Corpus types and uses. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 388–403). Routledge.

O'Keeffe, A., & Farr, F. (2003). Using language corpora in language teacher education: Pedagogic, linguistic and cultural insights. *TESOL Quarterly, 37*(3), 389–418. https://doi.org/10.2307/3588397

D

Data-Driven Learning and Languages Other than English

Luciana Forti
University for Foreigners of Perugia, Perugia, Italy

Abstract

This entry describes the status of Languages other than English (LOTEs) in Data-driven learning (DDL) research and practice. First, it situates the topic within the broader domain of applied linguistics, with reference to the impact that Global English has had and continues to have on language pedagogy and scientific generalisability of language studies. Then, it provides an overview of how the focus on LOTEs in DDL has risen in the past few years, based on a range of publications. Next, specific attention is given to the status of LOTEs in DDL research with an empirical grounding, i.e. containing some form of evaluation of DDL. Finally, some key issues are identified and briefly discussed.

Keywords

Data-driven learning · Languages other than English · Language equality · Corpora · Language learning

Introduction

Data-driven learning (DDL) can be defined as the learning that results from the use of corpus-based resources by learners of a language. This kind of language learning is based on a combination of corpus linguistics techniques and a range of language learning pedagogies. Research and pedagogical practices in DDL are dominated by a focus on English as the target language. This is a reflection of both the broader research domain related to applied linguistics, as well as of global language learning practices. In the first case, the fact that an applied linguistics study deals with the English language and not with one of the other 7000+ languages in the world is not considered to be of specific significance, which is why this fact is generally not made explicit (Lei & Liu, 2019, pp. 544–545, see also Vyatkina, 2024, Chap. 1). In the second case, the vast majority of language learning happening around the world involves English, alongside the "Englishisation" of education systems, and in the context of what is considered to be a "language learning crisis, triggered by Global English" (Lanvers, 2024, p. 3). Yet, equality of all languages is part of the fundamental human rights formulated by UNESCO (UNESCO, n.d.) and at the core of the multilingual turn witnessed in second language acquisition (May, 2019).

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_48-1

Languages other than English (LOTEs) have been featured in DDL research from the very beginning of its development. However, their presence has been limited to certain geographical areas and remains too marginal with respect to what is desirable in terms of scientific generalisability and spread in pedagogical practices beyond academic domains. Furthermore, the scientific and pedagogical publications on DDL are also overwhelmingly in English, as far as the latest reviews allow us to gather, as we shall see. Nevertheless, interest in DDL for LOTEs is increasingly becoming more widespread, because it is an area that carries a number of theoretical, methodological, and pedagogical implications.

This entry illustrates how the focus on DDL for LOTEs has risen in recent years, what the status of LOTEs in empirical DDL research has been over time, and which issues may be considered key for the development of this area.

The Rising Focus on DDL for LOTEs

One way of determining how much research and pedagogical practices have developed in the context of DDL for LOTEs is to identify this specific area within syntheses of various sorts (meta-analyses, systematic reviews, overviews, etc.) on DDL overall. Another way is to retrieve publications that deal specifically with the issue of how much research of DDL for LOTEs there is and of what kind. This section of the entry will address both of these points.

First, we elicited proportions of DDL studies on English and on LOTEs from several research syntheses that looked at DDL either comprehensively, i.e. included all the studies published that fit certain inclusion/exclusion criteria, or selectively, i.e. focused on the most representative and significant studies over time. Chambers (2007) collected and examined DDL studies dealing with the emic data, i.e. data related to learner perceptions and attitudes towards DDL. In terms of target language distribution, 36% of the studies involved LOTEs, while 64% involved English. Boulton (2017) offered a research timeline of DDL, made of the studies considered to be most representative and significant over time. In this case, only 11% of the studies focused on LOTEs, while the rest dealt with English. Boulton and Cobb (2017) provided the first comprehensive meta-analysis on DDL, focused on etic studies, i.e. studies involving the measurement of changes in language competence as an effect of exposure to DDL, targeting various L2s. In this case, the studies on DDL and LOTEs covered only 2%. A second meta-analysis on DDL was published a year later. Lee et al. (2018) considered a single learning aim (i.e. vocabulary), while maintaining the outlook towards various L2s already presented in Boulton and Cobb (2017). In this meta-analysis, the studies on DDL for LOTEs covered 3% of the total. Finally, Pérez-Paredes (2019) analysed the uses and spread of DDL during the years 2011–2015. Here, the proportion of LOTEs was 5%. Boulton and Vyatkina (2021) will be examined more minutely in section "The Status of LOTEs in Empirical DDL Research" as the most recent and comprehensive synthesis of empirical studies on DDL.

One of the first publications dealing explicitly with the representation of LOTEs in DDL research is Jablonkai et al. (2020), who synthesised presentations given within the symposium "Data-driven learning for languages other than English", organised by the EuroCALL (European Association for Computer-Assisted Language Learning) CorpusCALL SIG (Special Interest Group) at the 2020 EuroCALL conference. They presented and briefly discussed the current situation and challenges for DDL and selected LOTEs: Spanish (Abad Castelló and Salengros Iguenane), Italian (Forti), Schaeffer-Lacroix (French), and German (Vyatkina).

Forti (2023), to the best of our knowledge, was the first research monograph devoted to DDL for a LOTE, namely, Italian. In this monograph, the author surveyed the area of corpus linguistics in Italy, along with the main corpus resources that may be used in Italian language pedagogy. She then reported on the results of an empirical study aimed at investigating DDL effects in relation to the development of phraseological competence in Italian L2, both from an etic and emic perspective.

In the same year, Forti et al. (2023) was published, a short paper stemming from the

symposium on DDL and LOTEs conducted at the 2023 EuroCALL conference and organised, again, by the EuroCALL CorpusCALL SIG. The symposium and related paper focused on three books: Forti (2023), Schaeffer-Lacroix (2019), Vyatkina (2024, in print at the time). In the paper, each author presented the key aspects of their book. Forti (2023) is summarised in the previous paragraph. Schaeffer-Lacroix (2019) presented five pedagogical scenarios in which German can be learnt/taught through corpora at varying levels of difficulty. Vyatkina (2024) also focused on DDL for German, providing a comprehensive presentation and discussion of three major divides: the English–LOTEs divide; the DDL–Instructed Second Language Acquisition research divide and the research–pedagogy divide. The book, then, presented an empirical study focused on learning German verb-noun collocations as well as corpus-based Open Educational Resources (OER) for German.

One of the latest efforts to move away from the dominance of English is Tyne and Spina (in press). This volume focuses on using corpora for teaching and learning Romance languages by showcasing studies presented at the 2020 Teaching and Language Corpora conference.

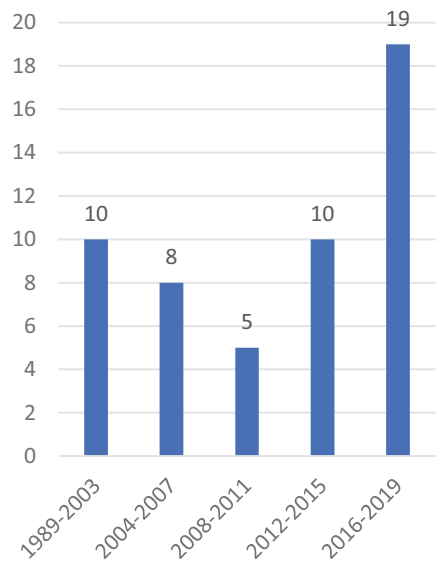
The focus on DDL for LOTEs is certainly rising within scholarly debate, and this shows promise for the future. Let us now zoom into the specific status of LOTEs in empirical DDL research over time.

The Status of LOTEs in Empirical DDL Research

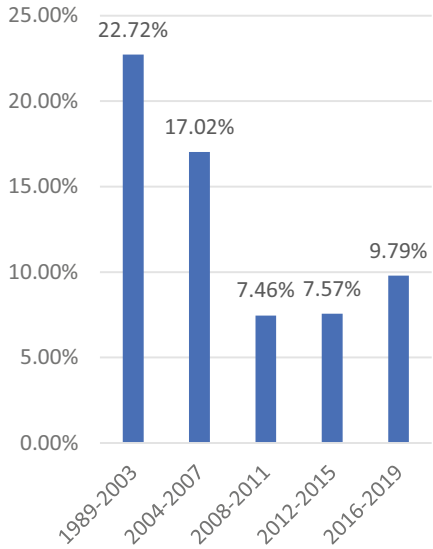
At the time of writing, the most comprehensive research synthesis available for DDL research is Boulton and Vyatkina (2021). The synthesis includes studies on DDL published at any point in time—thus the exhaustive nature of the study from a chronological perspective -, written in English, focused on any target language, and containing some form of empirical evaluation of the approach, whether etic or emic. The research synthesis uncovered a total of 489 studies over a 30-year timeframe from 1989 to 2019. While this number has almost doubled since 2020 (personal communication, May 17, 2024), for the purposes of the present entry we will consider the dataset published in the supplementary materials of the article.

When filtering through Boulton and Vyatkina's (2021) dataset supplement, one finds that English as a target language covers the vast majority of empirical studies on DDL: 435 out of 489 (88.89%). The remaining 11.04% are, thus, studies involving LOTEs. Among these, the most prominent language is German with 16 publications (3.27%), followed by French with 13 publications (2.66%), Italian with 9 publications (1.84%), Spanish with 4 publications (0.82%), Chinese with 3 publications (0.61%), Cantonese and studies involving mixed languages for translation purposes with 2 publications (0.41%), and Arabic, Catalan, Croatian, Czech, and Russian with 1 publication each (0.20%).

Figures 1 and 2 show how the publications on DDL for LOTEs have evolved over time, both in terms of raw quantities, and in terms of proportions related to the overall sample of publications considered. Both figures display a U-shaped pattern. However, Fig. 1 shows a marked upward trend in the quantity of publications, while Fig. 2



Data-Driven Learning and Languages Other than English, Fig. 1 DDL for LOTEs: Number of publications over time. (Adapted from Boulton & Forti, 2024)



Data-Driven Learning and Languages Other than English, Fig. 2 DDL for LOTEs: Percentage of publications over time

shows a much more modest upward trend in their proportion with respect to the total.

What are the implications of this situation? First, the research conducted so far on whether DDL can reach "the parts that other teaching can't reach" (Boulton, 2008) can hardly be generalised, as with the dominance of English, we can only refer to learning aims connected with the nature of this language. Languages of the world are highly diverse. Romance languages, for example, tend to exhibit a much more complex inflectional morphology than English and this has an inevitable impact on the identification of learning aims in these languages. But language diversity can go well beyond morphology, reaching all levels of linguistics analysis, and thus determining a wide range of language learning needs. Without an empirically founded knowledge base on what DDL can do for learning a wide range of languages, there is actually very limited scope for evaluating its effects and effectiveness on a scientific level.

Second, the little evidence on DDL for LOTEs that is available is very much skewed towards certain geographical areas, namely, WEIRD (Western, Educated, Industrialized, Rich, Democratic) countries (Henrich et al., 2010; Plonsky, 2023). This not only considerably extends the scientific problematicity deriving from the lack of target language diversity in DDL research, but also places it into wider sets of issues regarding representativeness and ethics. This implication is characteristic of applied linguistics studies as a whole, which DDL research is part of.

Key Issues in DDL for LOTEs

The research syntheses cited above have two major limitations: they are published in English and they typically focus on empirical studies. Considering only studies written and published in English is problematic in that it potentially overshadows publications of interest written in other languages. Furthermore, considering only studies with an empirical objective will not provide us with information about the actual spread of DDL in the research and practices within specific LOTEs traditions.

In reviewing, as exhaustively as possible, the existing literature on DDL for Italian, Forti (2024) considered publications in English as well as in Italian, and included papers with descriptive aims alongside those with empirical ones. A total of 45 studies were found, 22 published in Italian and 23 in English, 33 with descriptive aims and 12 with empirical ones. The status of DDL and Italian, based on Boulton and Vyatkina's (2021) synthesis, being limited to empirical studies published in English, could be described only on the basis of nine studies. By including papers written in Italian with a descriptive focus, the picture becomes much more complete and nuanced. Mapping the status of LOTEs in DDL research and pedagogical practices requires consideration of multiple publication languages and publication types.

Teacher training and OER are key topics in DDL as a whole and even more key in relation to the status of LOTEs in DDL. Most resources and materials that could be useful for language teachers to learn how to use corpora in the classroom are in English and refer to learning/teaching English (e.g. Le Foll, 2021), with some exceptions in relation to the target language (e.g. Vyatkina, 2020; Pinto et al., 2023). The book by Schaeffer-Lacroix (2019) represents a rarity and, at the same time, a

possible way forward. Written in French, it is aimed at French-speaking researchers and teachers of German, wishing to learn how to use corpora in their teaching. The book offers five practical pedagogical scenarios. English is absent both as the target language of learning/teaching and as the language of the publication. With books such as the one by Schaeffer-Lacroix, DDL can effectively move into non-English-dominated teaching contexts, develop more broadly as a language-independent approach to L2 learning, and foster multilingualism.

Summary

This entry surveyed the status of LOTEs in DDL research and pedagogical practices. While the research syntheses published so far on empirical DDL research show that LOTEs are considered only in a handful of cases, a series of other types of publications have put the topic into focus. This has created greater awareness of why we need more DDL studies on LOTEs: both to increase the generalisability of DDL research and to consider geographical areas that are under-represented. Methodologies for conducting DDL syntheses should evolve so as to include a broader range of publication languages and publication types, thus being able to provide us with a more complete perspective on how DDL for LOTEs is developing.

Cross-References

- ▶ Concordancers for DDL
- ▶ Evolving Paradigms in DDL
- ▶ Hands-On Versus Hands-Off Methods in DDL
- ▶ Pedagogical Guides for DDL
- ▶ Research Synthesis in Data-Driven Learning
- ▶ Teacher Training for DDL
- ▶ Tools for DDL

References

Boulton, A. (2008). DDL: Reaching the parts other teaching can't reach? In A. Frankenberg-Garcia (Ed.), *Proceedings of the 8th teaching and language corpora conference* (pp. 38–44). Associação de Estudos e de Investigação Cientifica do ISLA-Lisboa.

Boulton, A. (2017). Research timeline. Corpora in language teaching and learning. *Language Teaching, 50*(4), 483–506. https://doi.org/10.1017/S0261444817000167

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Forti, L. (2024). *Corpus linguistics and data-driven learning* [Manuscript under review].

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning and Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3–16). Rodopi.

Forti, L. (2023). *Corpus use in Italian language pedagogy. Exploring the effects of data-driven learning*. Routledge.

Forti, L. (2024). *Didattica dell'italiano L2 e corpora: una rassegna sistematica* [L2 Italian teaching and corpora: a systematic review] [Manuscript in preparation].

Forti, L., Vyatkina, N., & Schaeffer-Lacroix, E. (2023). Data-driven learning beyond English: Insights and implications from three monographs. In B. Bédi, Y. Choubsaz, K. Friðriksdóttir, A. Gimeno-Sanz, S. B. Vilhjálmsdóttir, & S. Zahova (Eds.), *CALL for all languages – EUROCALL 2023 short papers* (pp. 128–132). Editorial Universitat Politècnica de València. https://doi.org/10.4995/EuroCALL2023.2023.16865

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Jablonkai, R., Forti, L., Castelló, M. A., Iguenane, I. S., Schaeffer-Lacroix, E., & Vyatkina, N. (2020). Data-driven learning for languages other than English: The cases of French, German, Italian, and Spanish. In K.-M. Frederiksen, S. Larsen, L. Bradley, & S. Thouësny (Eds.), *CALL for widening participation: Short papers from EUROCALL 2020* (pp. 132–137). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.48.1177

Lanvers, U. (2024). *Language learning beyond English: Learner motivation in the twenty-first century*. Cambridge University Press.

Le Foll, E. (2021). *Creating corpus-informed materials for the English as a Foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. https://elenlefoll.pressbooks.com. CC-BY-NC 4.0. https://doi.org/10.5281/zenodo.4992504

Lee, H., Warschauer, M., & Lee, J. H. (2018). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Lei, L., & Liu, D. (2019). Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics, 40*(3), 540–561. https://doi.org/10.1093/applin/amy003

May, S. (2019). Negotiating the multilingual turn in SLA. *The Modern Language Journal, 103*(S1), 122–129. https://doi.org/10.1111/modl.12531

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. (Eds.). (2023). *Using language data to learn about language: A teachers' guide to classroom corpus use*. University of Queensland. https://doi.org/10.14264/3bbe92d

Plonsky, L. (2023). Sampling and generalizability in Lx research: A second-order synthesis. *Language, 8*, 75. https://doi.org/10.3390/languages8010075

Schaeffer-Lacroix, E. (2019). *Corpus et didactique de l'allemand – La langue à bras-le-corps* [Encounters with German language in use promoted by pedagogical corpora]. Lambert-Lucas.

Tyne, H., & Spina, S. (Eds.). (in press). *Applying corpora in teaching and learning Romance languages*. Benjamins.

UNESCO. (n.d.). *Universal declaration of humans rights* (UDHR). https://www.un.org/en/about-us/universal-declaration-of-human-rights

Vyatkina, N. (Ed.). (2020). *Incorporating corpora: Using corpora to teach German to English-speaking learners* [Online instructional materials]. University of Kansas Open Language Resource Center. https://opentext.ku.edu/corpora

Vyatkina, N. (2024). *Corpus applications in language teaching and research: The case of data-driven learning of German*. Routledge.

Y

Young Learners and Data-Driven Learning

Eri Hirata
Fukuoka Jo Gakuin University, Fukuoka, Japan

Abstract

Data-driven learning (DDL) refers to a pedagogy which involves language learning through the use of corpora. The mainstream of its application has been for learners at the tertiary level, and it has not been widely used with young learners. Though the number of empirical studies at the primary level remains relatively small compared to those with tertiary learners, studies on the use of DDL with young learners and secondary students have started to emerge and identified its potential for contributing to young learners' language learning, in terms of promoting their meta-linguistic awareness, enhancing their lexico-grammatical and phraseological awareness, as well as in improving their self-efficacy. For further implementation of DDL in young learners' language education, research suggests the incorporation of corpus linguistics in teacher training and further developments of suitable corpora and tools and related methodologies considering the local and contextual needs of learners.

Keywords

Data-driven learning (DDL) · Corpus-based language teaching · Young learners · Primary education · Pedagogy · Concordancing · Corpus linguistics

Introduction

With advances in corpus linguistics, pedagogical applications of corpus findings are increasingly found in language teaching contexts in recent years. One of the applications of corpora in pedagogy was termed "data-driven learning (DDL)" by Tim Johns (1991), which involves the use of "tools and techniques of corpus linguistics for pedagogical purposes" (Gilquin & Granger, 2010, p. 359).

Data-driven learning can be largely divided into two types: teacher-led DDL ("soft" or indirect version) and learner-led DDL ("hard" or direct version of DDL) (Gabrielatos, 2005; Mukherjee, 2006). Teacher-led DDL involves more controlled tasks in which the items to be searched are pre-decided by the teacher, such as gap-fill exercises, and conducted deductively. Learner-led DDL entails a more inductive task of "discovery learning" (Bernardini, 2004, p. 22) through the direct exploration of the corpus. While teacher-led DDL tends to be better suited for beginners, learner-led discovery learning is considered more appropriate for learners at an advanced level (Gilquin & Granger, 2010, p. 363).

Research on corpus-based language teaching suggests that incorporating DDL in language

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_49-1

teaching has several benefits, such as providing exposure to authentic language, developing learners' analytical and problem-solving skills and promoting learner autonomy through corpus consultation (Vyatkina & Boulton, 2017). The use of DDL in language teaching has been proven to be effective in contributing to the learners' linguistic development (Chambers, 2005, 2010), especially in promoting knowledge of vocabulary, lexico-grammatical and phraseological features of the language (Boulton & Cobb, 2017; Lee et al., 2019; Li, 2017) and in improving writing skills (Chambers, 2010).

Despite the fact that DDL has more than 30 years of history and that a number of empirical studies on DDL show positive results suggesting that incorporating corpus consultation in language teaching benefits learners (Boulton & Vyatkina, 2021; Chambers, 2019), the number of studies conducted on the use of DDL with pre-tertiary learners aged under 18 is very small (Boulton & Vyatkina, 2021), and the use of corpora in their language teaching context is an uncommon practice, especially for young learners who are in primary school aged under 12 (Ellis, 2014; Enever, 2016; Garton & Tekin, 2022).

One of the causes stems from the lack of adequate corpora or tools suitable for young learners. Since many of the corpora are created for linguistic research with adult users in mind, the language in concordances is often beyond the level of young learners. Even for tertiary level, these corpora often require "pedagogical mediation" (Widdowson, 2003) to suit the needs of learners. The software too can be daunting at this level. Moreover, as learner-led ("hard") DDL, in particular, requires inductive tasks on the part of learners, it is considered cognitively challenging especially for young L2 learners. Another possible reason is the lack of accessible information about the corpus-based approach to language teaching and related training for teachers. In teacher training courses for the primary level, teachers are rarely introduced to corpus linguistics or corpus-based materials, let alone DDL pedagogy.

Nevertheless, given the positive outcomes of DDL at tertiary level, there has been a recent rise of studies on the use of DDL with young learners exploring the possibility of implementing corpus-based or corpus-informed instruction with young learners. This entry firstly describes an overview of DDL studies with young learners in both L1 and L2 educational contexts and discusses the remaining challenges and the implications for future developments in this area.

Studies on DDL with Young Learners

While the majority of the empirical research on DDL is found within education at tertiary level, some studies have explored the use of DDL at the primary and secondary level. In the case of young learners in the L1 context, Sealey and Thompson (2004, 2006, 2007) explored the possibility of using DDL as part of a literacy classroom. The study conducted corpus-based teaching with L1 young learners at two primary schools in England, with six pupils from each school either in Year 4 or in Year 5 (i.e. aged between eight and ten). For the corpus-based activities, they created the CLLIP (Corpus-based Learning about Language In the Primary-school) corpus, which contained 40 texts written for a child audience, extracted from the British National Corpus (BNC). DDL activities were prepared using concordances from the prepared corpus, which were initially colour-coded according to word class, to promote the children's metalinguistic awareness of their L1 (i.e. English) through the discovery of frequencies and patterns in the concordance outputs. The items used in the activities were decided in accordance with the teaching objectives in the National Curriculum, National Literacy Strategies (NLS) so that the DDL tasks would contribute to the learners' knowledge and understanding of the materials taught at school. It was reported that they were able to identify lexico-grammatical features and describe the differences of certain word classes by making use of the coloured concordances in the DDL activity (Sealey & Thompson, 2004, 2007).

Another example of DDL in the L1 primary context is a study conducted by Crosthwaite and Stell (2020), who explored the use of DDL with two primary school students (i.e. aged 10) in private tutoring sessions for improving their argumentative essay writing. The pupils participated in

DDL by using Sketch Engine for Language Learning (SkELL) and the British Academic Corpus of Written English (BAWE) for writing and revising their argumentative essays. The tutor provided "scaffold help" on their queries and concordance results. The children reacted positively to the DDL revision tasks, and the experience of having access to corpora and making use of their problem-solving skills in exploring the usages in the corpus contributed to improving their self-efficacy. This study demonstrates how existing corpora or tools can be successfully used by young learners when appropriate guidance and support with the corpus consultation provided by the instructors.

One of the rare studies on the use of DDL in the L2 primary setting was conducted by Kim (2019) exploring the use of DDL in grammar instruction with a total of 18 Korean primary school pupils aged 12. The study used English Literature for Children (ELFC) corpus (Min et al., 2014) for creating paper-based DDL materials in teaching the use of prepositions (i.e. *in, at, into*). Kim employed paper-based DDL, considering the pupils' proficiency and their level of computer literacy. While identifying some challenges that pupils experienced (e.g. unknown words in the concordance, unfamiliarity with reading the concordance lines), the paper concluded that DDL activity was perceived positively by both pupils and teachers based on the post-interviews conducted with each participant.

While it is very rare to find empirical studies on the use of DDL in the L2 primary context, some studies can be found with L2 learners at the secondary level. For instance, Braun (2007) conducted a study which investigated the conditions and challenges integrating DDL with German L2 secondary learners aged 14 to 15 by making use of a corpus of video-based narrative interviews (i.e. ELISA—English Language Interview Corpus as a Second Language Application). Though the participants generally perceived the corpus-based activities positively, some methodological and logistical challenges were identified for the integration of DDL, such as a lack of learners' autonomy and skills in corpus consultation. Braun points out the need for modifying traditional DDL for secondary learners, suggesting a shift from traditional corpus-based "data-driven" to "needs-driven corpora" along with related activities and methodologies in order to suit the pedagogical requirements of each context.

A recent collection of papers on DDL with younger learners can be found in *Data-driven learning for the next generations: Corpora and DDL for pre-tertiary learners*, edited by Crosthwaite (2020), which includes some empirical research with secondary L2 learners. For instance, Szudarski (2020) explored the effects of paper-based DDL on the acquisition of L2 phraseology with 22 Polish students aged 16–18 in secondary education and compared the effectiveness of paper-based DDL and dictionary-based instruction. The results revealed that paper-based DDL was more effective in enhancing the learners' phraseological competence in terms of their meaning recognition and meaning recall, in comparison to the dictionary-based instruction. While this showed that paper-based DDL contributed more to the acquisition of target phrases, the qualitative analysis from a questionnaire also revealed that the students' limited proficiency and history of receiving deductive instruction influenced their attitudes towards the new form of teaching. Szudarski (2020, p. 147) argues the need for "didactic processing of DDL" in the secondary context, whereby teachers contextualise the approach to correspond to the local context and learning needs of the learner.

Liontou (2020) conducted a study with secondary school students aged 12–15, and explored the effectiveness of DDL activities on the EFL learners' understanding and use of idiomatic expressions, by comparing the performance of the experimental group using DDL, with that of the control group using textbooks and activities. For the DDL group, Liontou (2020) made use of both deductive teacher-led DDL with prepared worksheets and activities created based on using concordances, and inductive learner-led DDL for students to explore the idiomatic phrases directly using British National Corpus (BNC) and Corpus of Contemporary American Corpus (COCA). The post-test results indicated that the DDL group showed a significant improvement in both comprehending idiomatic meaning and using idiomatic expressions in writing.

The above empirical studies concerning the use of DDL in both primary and secondary education report that DDL contributed to learning. In all of these studies, the pedagogical mediation by the teachers played an essential role in the success, in providing appropriate scaffolding during corpus consultation, and in preparing paper-based DDL tasks or other suitable tasks suitable for the needs and local context of the learners.

Prospects for DDL with Young Learners

Boulton and Cobb (2017, p. 386) note that "DDL works pretty well in almost any context where it has been extensively tried". Although the number of empirical studies on DDL with younger learners is still relatively small compared to those aimed at tertiary learners and it might be far from being claimed as "extensive", the existing studies at the primary and secondary levels have indicated its strengths and potential to enhance learners' understanding of their target language. Nevertheless, there is clearly a need for more research on DDL in contemporary young learners' classrooms, especially in the L2 primary context. Given the results and identified challenges, further research and considerations are needed in the following areas before realising the successful use of DDL with young learners.

Child-Friendly Corpora and Tools

One of the recurring causes documented about the limited DDL studies with young learners is the lack of suitable corpora and tools for their use. Since the majority of available corpora are usually created for linguistic research with adult users in mind, such corpora could contain words which are pedagogically irrelevant or even educationally inappropriate for young learners. As Whistle (1999, p. 78) emphasises "[t]he value and usefulness of the concordance will be largely determined by the corpus". For the successful integration of DDL for young learners especially in the L2 context, further development of appropriate corpora or tools is required. The importance of contextualising DDL within young learners' classrooms has been emphasised, which involves the choice of corpora or texts to be included in a corpus so that they fit the local contexts and learning needs of the young learners (Braun, 2007). The notion of "pedagogic corpus" (J. Willis, 1998, p. 46; D. Willis, 2003, p. 163), which consists of texts or materials used in the classroom, can be useful in contextualising DDL and making the corpus consultation relevant for learners (Granger, 2011). Since a pedagogic corpus includes materials children encounter in their learning, DDL can be integrated within the learning process, providing opportunities to focus on the kinds of patterns in language which learners need to familiarise themselves with. Moreover, when conducting DDL with lower-proficiency learners, using corpora of graded readers may help them feel less overwhelmed as the texts included in those graded corpora would be better suited to their proficiency level while maintaining authenticity in the data (Allen, 2009; Hadley & Charles, 2017).

Most of the DDL tend to make use of textual corpora, which can also present some challenges for L2 young learners, as their language classroom often have a primary focus on communicative tasks with an emphasis on developing oral skills. Several proposals have been made for making the content of the corpus relevant for younger learners in order to contextualise DDL, considering the pedagogical needs of the learners (Braun, 2007; Hirata, 2020; Pérez-Paredes, 2020). For instance, Pérez-Paredes (2020) suggests the use of SACODEYL-English multimedia pedagogic corpus of teenage speakers of English for L2 secondary education to provide learners with opportunities to be exposed to authentic oral language use by speakers of similar age. For the L2 primary context, Hirata (2020) argues the need for reflecting the multi-modal features in the corpus for primary learners of English as a foreign language (EFL), in order to make learning through DDL more relevant for young learners who experience various modes of interactions through the increasing engagement of digital media in their daily lives both at home and in educational contexts. Children discern meaning not just by language alone, but through the interaction of other modes such as image, gesture, movement, sounds and music (Jewitt & Kress, 2003), and studies suggest that multimodal stimuli contribute to the

young learners' language learning (e.g. Fleta, 2017; Kaminski, 2019; Wolf et al., 2017). Multimodal corpora (aligning text, audio and/or video) are likely to assist learners in comprehending the meaning better, rather than through the text-only corpus, especially for young L2 learners who may be building their L2 literacy skills simultaneously.

There is certainly a need for more child-friendly corpora which contain relevant materials, and also cognitively appropriate tools that provide additional support such as layout, colour-coding, multimedia (Pérez-Paredes, 2020) and multimodal information (Meunier, 2020; Hirata, 2020) to make DDL more accessible and relevant to young learners' reality. Moreover, suggestions have been made for the use of existing tools or applications which are not originally created for traditional DDL but still provide learners opportunities to explore language in a way similar to DDL for young learners (Meunier, 2020; Gatto, 2020). Such a proposal is considered useful in familiarising young learners with the idea of "discovery learning", and the experience may also help them to transition easily to other forms of DDL.

Methodological Consideration

As Sinclair (2004, p. 7) notes, "[c]orpus evidence is essentially indirect, which means that it cannot be taken at face value but must go through a process of interpretation". This process can be time-consuming (Yoon & Hirvela, 2004) because of the inductive nature of the task, which would also be overwhelming for children if the learner-led "hard" version of DDL were to be directly employed, using existing corpora and tools created for linguistic research. For lower-level L2 learners, it has been suggested that "hard DDL may be less helpful than softer versions or traditional teaching methods" (Vyatkina & Boulton, 2017, p. 6). Studies have revealed that this softer version of paper-based DDL, using printouts of concordances instead of directly accessing corpora on computer (Boulton, 2010; Mizumoto & Chujo, 2016; Smart, 2014) is also effective with medium effect sizes (Boulton & Cobb, 2017). With secondary learners, the use of paper-based DDL, together with the teacher's mediation or manipulation of concordance lines, has proved effective (Braun, 2007; Szudarski, 2020) in assisting L2 secondary learners to interpret concordance lines and learn from the DDL activity.

It is not realistic to use DDL as the only method of language teaching, and in adopting DDL in young learners' classrooms, methodological consideration of aligning with the teaching approach is necessary (Boulton, 2010; Meunier, 2020; Wicher, 2020) together with appropriate pedagogical mediation and extended scaffolding to support corpus consultation (Crosthwaite & Stell, 2020; Wicher, 2020). It is necessary for teachers to have enough corpus literacy to successfully mediate the concordance materials to suit the needs and levels of learners in each local context, and provide scaffolding help in corpus consultation accordingly. This point has a bearing on the need for teacher training which will be discussed below.

Teacher Education

Another important development needed for the successful implementation of DDL with young learners is the integration of corpus linguistics, including DDL, in teacher education. Research suggests that such integration is necessary within in-service or pre-service teacher training (Chambers, 2019; Mukherjee, 2006; O'Keeffe & Farr, 2003; Römer, 2009) in order for teachers to be introduced to the corpus analysis for "pedagogical needs" (Braun, 2007, p. 326) and convinced of the benefits of DDL in the young learners' classroom (Hirata, 2020; Schaeffer-Lacroix, 2020). The use of a corpus can serve as an awareness-raising device for teachers themselves (J. Willis, 1998) and "supplement teachers with their intuition" (Sealey, 2009, p. 57). The corpus can also serve as a valuable resource for teachers in preparing teaching materials and making pedagogically informed decisions about the kinds of language to be presented to learners, including the kinds of necessary mediations suitable for the leaners' learning needs (Braun, 2007; Sealey & Thompson, 2006).

Summary

DDL is an approach which encourages learners to explore language through corpus analysis. Existing corpora and tools are often intended for linguistic research, and due to the lack of corpora

and tools suitable for young learners' learning, and because of the possible cognitive burden, DDL tended to be excluded from mainstream language education for young learners.

Although the number of empirical studies is still limited compared to those at the tertiary level, research on DDL with young learners in L1 and L2 contexts at the secondary level has demonstrated the potential for such incorporation. It has been reported that corpus-based exploratory learning contributes to an increase in learners' meta-linguistic knowledge, understanding of lexico-grammatical and phraseological features of the target language, as well as improving their self-efficacy through corpus consultation.

Attempts to introduce DDL with L2 secondary learners have started to emerge, but still remain few in L2 primary education contexts. As primary teachers know the value of authentic experience in children's learning (Sealey & Thompson, 2006), given the appropriate corpora and tools suitable for young learners, they may be willing to provide children with the authentic experiences of exploratory learning about language through DDL. As Boulton (2020: xviii) notes, "DDL might even be more appropriate for children than for older learners, as their expectations are less fixed and the approach lends itself to the inquisitive".

The successful integration of DDL with young learners requires local and contextual considerations in terms of having suitable corpora and tools, applying a methodology adopted to the local context and pedagogical needs of young learners' classrooms, so that corpus-based or corpus-informed DDL tasks can be relevant and integrated into the language learning process. To achieve such integration, the incorporation of DDL in teacher education is vital, as none of the initiatives can be achieved in young learners' classrooms without the contribution of teachers.

Cross-References

- ▶ Pedagogical Corpora and DDL
- ▶ Pedagogical Guides for DDL
- ▶ Teacher Training for DDL
- ▶ Techniques for DDL Beginners
- ▶ Tools for DDL

References

Allen, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal, 63*(1), 23–32. https://doi.org/10.1093/elt/ccn011

Bernardini, S. (2004). Corpora in the classroom. In J. McH & Sinclair (Eds.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.12

Boulton, A. (2010). Data-driven learning: Taking the computer out of equation. *Language Learning, 60*(3), 534–572. https://doi-org.bases-doc.univ-lorraine.fr/10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2020). Data-driven learning for younger learners: Obstacles and optimism. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. xiv–xx). Routledge.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology, 25*(3), 66–89. https://doi.org/10.10125/73450

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*(3), 307–328. https://doi.org/10.1017/S0958344007000535

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology, 9*(2), 111–125. https://doi.org/10.10125/44022

Chambers, A. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 345–358). Routledge. https://doi.org/10.4324/9780203856949-25

Chambers, A. (2019). Towards the corpus revolution? Bridging the research – Practice gap. *Language Teaching, 52*, 460–475. https://doi.org/10.1017/S0261444819000089

Crosthwaite, P. (Ed.). (2020). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge. https://doi.org/10.4324/9780429425899

Crosthwaite, P., & Stell, A. (2020). "It helps me get ideas on how to use my words": Primary school students' initial reactions to corpus use in a private tutoring setting. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 150–170). Routledge. https://doi.org/10.4324/9780429425899-9

de Wolf, S., Smit, N., & Lowie, W. (2017). Influences of early English language teaching on oral fluency. *ELT Journal, 71*(3), 341–353. https://doi.org/10.1093/elt/ccw115

Ellis, G. (2014). 'Young learners': Clarifying our terms. *ELT Journal, 68*(1), 75–78. https://doi.org/10.1093/elt/cct062

Enever, J. (2016). Primary ELT. In G. Hall (Ed.), *The Routledge handbook of English language teaching*

(pp. 353–266). Routledge. https://doi.org/10.4324/9781315676203-30

Fleta, T. (2017). The sounds of picturebooks for English language learning. *Children's Literature in English Language Education Journal, 5*(1), 21–43. http://clelejournal.org

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ, 8*(4), 1–35.

Garton, S., & Tekin, S. (2022). Teaching English to young learners. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp. 83–96). Routledge. https://doi.org/10.4324/9781003106609-7

Gatto, M. (2020). Query complexity and query refinement: Using web search from a corpus perspectives with digital natives. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 106–129). Routledge. https://doi.org/10.4324/9780429425899-7

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge. http://hdl.handle.net/2078.1/75724

Granger, S. (2011). From phraseology to pedagogy: Challenges and prospects. In T. Herbst, S. Faulhaber, & P. Uhrig (Eds.), *The phraseological view of language: A tribute to John Sinclair* (pp. 123–146). de Gruyter. https://doi.org/10.1515/9783110257014.123

Hadley, G., & Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Language Learning & Technology, 21*(3), 131–152. https://doi.org/10.10125/44624

Hirata, E. (2020). The development of a multi-modal corpus tool for young EFL learners: A case study on the integration of DDL in teacher education. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 88–104). Routledge. https://doi.org/10.4324/9780429425899-6

Jewitt, C., & Kress, G. (2003). *Multimodal literacy.* Peter Lang.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), Classroom concordancing (pp. 1–13). *English Language Research Journal*, 4.

Kaminski, A. (2019). Young learners' engagement with multimodal texts. *ELT Journal, 73*(2), 175–185. https://doi.org/10.1093/elt/ccy060

Kim, H. (2019). The perception of teachers and learners towards an exploratory corpus-based grammar instruction in a Korean EFL primary context. *Primary English Education, 25*(1), 123–152. https://doi.org/10.25231/pee.2019.25.1.123

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Li, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology, 21*(3), 153–171. https://doi.org/10.10125/44625

Liontou, T. (2020). The effect of data-driven learning activities on young EFL learners' processing of English idioms. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 208–227). Routledge. https://doi.org/10.4324/9780429425899-12

Meunier, F. (2020). A case for constructive alignment in DDL: Rethinking outcomes, practices, and assessment in (data-driven) language learning. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 13–30). Routledge. https://doi.org/10.4324/9780429425899-2

Min, D. G., Lee, S. M., & Kim, H. J. (2014). A study on uses of adjectives in English corpora for elementary school student. *Primary English Education, 20*(4), 213–244.

Mizumoto, A., & Chujo, K. (2016). Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System, 61*, 55–64. https://doi.org/10.1016/j.system.2016.07.010

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Peter Lang.

O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly, 37*(3), 389–418. https://doi.org/10.2307/3588397

Pérez-Paredes, P. (2020). The pedagogic advantage of teenage corpora for secondary school learners. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 67–87). Routledge. https://doi.org/10.4324/9780429425899-5

Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language learning* (pp. 83–98). John Benjamins. https://doi.org/10.1075/scl.33.09rom

Schaeffer-Lacroix, E. (2020). Barriers to trainee teachers' corpus use. In P. Crosthwaite (Ed.), *Data- driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 47–64). Routledge. https://doi.org/10.4324/9780429425899-4

Sealey, A. (2009). Exploring vocabulary with young L1 learners: The contribution of a corpus. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 39–58). Palgrave Macmillan. https://doi.org/10.1057/9780230242258_3

Sealey, A., & Thompson, P. (2004). "What do you call the dull words?" Primary school children using corpus-based approaches to learn about language. *English in Education, 38*(1), 80–91. https://doi.org/10.1111/j.1754-8845.2004.tb00774.x

Sealey, A., & Thompson, P. (2006). "Nice things get said": Corpus evidence and the National Literacy Strategy. *Literacy, 40*(1), 22–28. https://doi.org/10.1111/j.1467-9345.2006.415.x

Sealey, A., & Thompson, P. (2007). Corpus, concordance, classification: Young learners in the L1 classroom. *Language Awareness, 16*(3), 208–223. https://doi.org/10.2167/la453.0

Sinclair, J. McH. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins. https://doi.org/10.1075/scl.12

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL, 26*(2), 184–201. https://doi.org/10.1017/S0958344014000081

Szudarski, P. (2020). Effects on data-driven learning on enhancing the phraseological knowledge of secondary school learners of L2 English. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 133–149). Routledge. https://doi.org/10.4324/9780429425899-8

Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology, 21*(3), 1–8. https://doi.org/10.10125/44621

Whistle, J. (1999). Concordancing with the with students using and "off-the-web" corpus. *ReCALL, 11*(2), 74–80. https://doi.org/10.1017/S0958344000004985

Wicher, O. (2020). Data-driven learning in the secondary classroom: A critical evaluation from the perspective of foreign language didactics. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 31–46). Routledge. https://doi.org/10.4324/9780429425899-3

Widdowson, H. (2003). *Defining issues in English language teaching*. Oxford University Press.

Willis, J. (1998). Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 44–66). Cambridge University Press.

Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511733000

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

L

Language for General Purposes and Data-Driven Learning

Mariangela Picciuolo
Department of Modern Languages, Literatures and Cultures (LILEC), University of Bologna, Bologna, Italy

Abstract

Data-Driven Learning (DDL) is a pedagogical approach to foreign or second language learning and teaching which employs authentic language data from corpora to engage learners in guided discovery tasks, fostering deeper understanding of linguistic patterns and usage. Historically focused on error-correction and short-term gains, recent advances expand DDL's application to discourse, pragmatics, and rhetorical functions, enhancing its effectiveness in language for general purposes (LGP) contexts. Integrated with LGP, DDL draws on key theoretical frameworks like constructivism and sociocultural theory, supporting learner autonomy and critical thinking. Methodologically, DDL employs tools such as concordance lines, frequency analysis, and collocations to analyse authentic language data. Challenges include data complexity and teacher preparation, necessitating pedagogic transfer for educational adaptation. Indirect and direct DDL approaches offer distinct benefits. Indirect DDL simplifies learning through corpus-informed materials without direct corpus visibility, maintaining a traditional teaching approach. Direct DDL involves learners directly with corpus extracts or activities like concordance use, promoting autonomy. Empirical evidence supports DDL's efficacy, illustrated by recent case studies. Future trends in DDL for LGP include integrating large language models (LLMs) and mobile apps for enhanced accessibility and real-time linguistic analysis. While LLMs represent a paradigm shift, traditional DDL methods remain valuable, particularly in indirect approaches where linguistic analysis informs learning materials. As these technologies evolve, they will complement DDL, enabling learners to explore and understand target languages more effectively. Future research should apply DDL in diverse educational settings, develop accessible tools, and provide educator training for optimized use across learner profiles and contexts.

Keywords

Data-Driven Learning (DDL) · Language for General Purposes (LGP) · Corpus linguistics · Language learning · Language teaching

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_50-1

Introduction

Data-driven learning (DDL) is an educational approach to language learning, teaching, and use of foreign or second languages (L2) (Boulton & Vyatkina, 2021). This approach shifts from traditional teacher- and textbook-guided methods to a research-oriented model by leveraging corpora—"carefully compiled samples of language intended to be representative of a particular variety" (Boulton, 2012a, p. 23)—to enhance language acquisition, particularly focusing on phraseology and grammar. Using real-life language data, learners gain an authentic and practical understanding of how language is used in various contexts.

The connection between DDL and Language for General Purposes (LGP) lies in the systematic approach that DDL offers. Through the analysis of corpora, learners observe natural language patterns, including collocations, idiomatic expressions, and grammatical structures. This empirical evidence provides a concrete basis for understanding how language functions in real-world scenarios, which is crucial for achieving accuracy and fluency in both spoken and written communication.

This entry explores the historical context, theoretical foundations, and methodologies of DDL in LGP learning. It covers corpus types, technological tools, benefits, and practical applications of DDL for LGP. Challenges and criticisms are also addressed, offering a balanced perspective. Finally, future directions and research needs are discussed, highlighting potential developments and areas for further exploration.

Historical Context and Development

The term "DDL" was coined by Tim Johns in the 1990s during his tenure at the English for Overseas Students Unit of Birmingham University. He introduced an inductive approach to language learning, challenging the traditional "presentation, practice, production" paradigm prevalent in early language teaching. Johns (1997, p. 101) portrayed learners as "language detectives", emphasizing their direct access to data and empowerment through engagement with authentic language materials. He initially focused on error-correction in DDL: in one-on-one advisory sessions, he collaboratively explored errors in students' written work using concordances and corpus data, akin to his earlier informal "kibbitzing" sessions. He also expanded the applications of DDL in various innovative ways. Emphasizing the importance of collocations and phrases, Johns advocated for the use of corpora to explore common word combinations, such as *make a decision* or *take a risk* (Johns, 2002). He also introduced genre and register analysis through corpus comparison, helping students distinguish linguistic features specific to different types of texts (Johns, 1997). Recognizing cultural variations in language use, Johns used corpora to compare British and American English, providing insights into vocabulary and usage differences (Johns, 1994). Finally, Johns contributed to the development of new corpus-based tools and resources, enhancing the sophistication of language analysis available to learners and educators (Johns, 2002).

Notably, Johns was involved in the development of the Collins Birmingham University International Language Database (COBUILD), an electronic corpus of contemporary texts in English, which constitutes the earlier development of the DDL approach for LGP. That project extensively used keyword-in-context (KWIC) concordance data from a 20-million-word corpus of modern English to create various reference and teaching materials for English language learners. This initiative, initially overseen by Sinclair (1987), evolved into the renowned Bank of English (BoE). The BoE comprised a diverse collection of English texts sourced from various mediums, including newspapers, magazines, and books, as well as spoken data from radio, television, and informal conversations. This corpus laid the foundation for the creation of significant corpus-based dictionaries and grammars of English for learners, most notably the Collins COBUILD English Language Dictionary, which debuted in 1987. The primary innovation of the

COBUILD English Language Dictionary lies in its phraseological description of entry words through KWIC concordance lines. For example, *brink* is defined as being *on the brink of* something significant, reflecting Sinclair's methodology of using real language examples for whole-sentence definitions, unlike traditional dictionaries. The entry also summarizes common colligational patterns, such as "N-SING: usu. on/to/from the N of n", with prepositions listed by frequency. Additionally, the dictionary highlights frequently co-occurring words like *collapse* and *war*, indicating a generally negative connotation for *brink* (Flowerdew, 2024). Early DDL efforts often faced technical and resource limitations, but advances in corpus linguistics and computational tools have significantly expanded the scope and application of DDL in language education (Boulton, 2012a) such that DDL's principles have been increasingly adopted in LGP contexts to provide learners with a deeper, data-driven understanding of general language use.

DDL has been overwhelmingly employed for general purposes even within highly specialized disciplines, particularly involving university students with intermediate levels of proficiency or above, usually aiming at short-term learning outcomes in vocabulary and lexicogrammar, as well as error-correction (Boulton & Vyatkina, 2021). English remains predominantly favoured as the target language, yet there has been a notable surge in studies investigating DDL within the realm of other languages (Vyatkina, 2020).

Theoretical Foundations of DDL in LGP

The integration of DDL with LGP is supported by several key theories (cf. Flowerdew, 2015). Constructivism (see Cobb, 1999) encourages learners to build their understanding through exploration and discovery, which aligns well with DDL's inductive approach. This method fosters critical thinking and learner autonomy by allowing students to formulate their own hypotheses and rules based on empirical data. The associative link to constructivism is thus seen as a key pedagogical feature of DDL.

Sociocultural Theory (SCT) emphasizes the importance of social interaction and cultural context in learning. Essentially, learning occurs through mediation expressed via dialogue, as this collaborative process involves the co-construction of knowledge (see Huang, 2011). DDL supports this theory by exposing learners to authentic language data, highlighting the role of collaborative learning and the use of language as a social tool. Thus, while SCT and constructivist principles share some similarities, an SCT approach to DDL emphasizes the nature and support—or *scaffolding*—of learner engagement with data, focusing on enhancing learning through self-regulation or mediation by peers or a teacher, rather than solely independent discovery.

The Noticing Hypothesis suggests that learners must notice language features to effectively learn them. The use of computer screens as a central tool in most forms of DDL enhances the frequency and quality of exposure to language structures, helping learners notice and internalize them, thereby enhancing their overall language proficiency (see Papp, 2007).

However, several scholars (e.g. Flowerdew, 2015; O'Keeffe, 2021) observe the absence of robust theoretical foundations in DDL and advocate a more comprehensive research perspective, calling for an expanded examination beyond constructivism and sociocultural theory, to firmly anchor DDL within the realm of Second Language Acquisition (SLA).

Methodologies in DDL for LGP

When applied to LGP, DDL employs various tools and methodologies to analyse and use authentic language data. By using concordance lines, learners can observe how everyday vocabulary and phrases are employed in authentic conversations, thus gaining a better understanding of how language functions in real-world interactions. Frequency analysis in LGP helps learners prioritize learning high-frequency words and structures that are most relevant to daily communication, thereby enhancing their communicative competence more efficiently. Studying collocations and patterns

within LGP contexts allows learners to develop more natural and fluent speech and writing, essential for effective general language use.

Historically, DDL research has predominantly been experimental, with a concentrated focus on short-term gains in vocabulary and lexicogrammar, particularly within writing. Initial learner-technology interactions centred on rudimentary tools for generating concordances, evolving to encompass features like frequency lists, collocations, clusters, keywords, and distributions. More recently, DDL has moved into wider areas of language, such as discourse, pragmatics and rhetorical functions, discourse markers, and cohesion (Boulton & Vyatkina, 2021). In the context of LGP, this evolution is significant. By expanding beyond vocabulary and grammar, DDL now addresses how language is used to achieve various communicative purposes in everyday situations. This includes understanding the nuances of discourse markers, which help in managing conversations, and rhetorical functions, which are key in persuasive or explanatory contexts common in general communication.

Despite these advances, scholars like Boulton and Vyatkina (2021) note the enduring reliance on fundamental procedures such as examining concordances and frequency information. Similarly, more recent studies have called for a broader investigation into the long-term benefits and higher-level cognitive skills developed through DDL, such as critical thinking and independent learning (Boulton, 2012b).

Corpora for DDL in LGP

The choice of corpus is pivotal in DDL for LGP, influencing the effectiveness of concordance analysis (Whistle, 1999). Various corpus types—written, spoken, multimodal, monolingual, bilingual, general, specialized, native, or non-native—serve different purposes in DDL. In LGP contexts, large corpora like the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) offer extensive datasets ideal for studying broad language trends, including frequency, collocations, and usage across diverse contexts and genres. However, these large corpora may overwhelm beginners and require advanced tools for analysis. Conversely, smaller, specialized corpora offer focused datasets that are more manageable and relevant for pedagogical applications, particularly in understanding context-specific language use (Cheng, 2010).

The authenticity of corpora is critical; while inherently authentic, scholars like Widdowson (2000) argue about their cultural relevance to learners. Pedagogic and local learner corpora address this by providing samples from teaching materials or learners' own language use, offering tailored insights into language learning challenges. For instance, the TeMa corpus (Meunier & Gouverneur, 2009) compiles popular English for General Purposes coursebooks, while the Louvain International Database of Spoken English Interlanguage (LINDSEI—Gilquin et al., 2010) features non-native speaker recordings, annotated to highlight language errors and fluency issues.

Annotation of corpora, whether through part-of-speech (POS) tagging, syntactic parsing, or error-tagging in learner corpora, enhances their utility for language exploration. While raw corpora offer flexibility, they require extensive editing to remove irrelevant data. In contrast, annotated corpora refine search queries and facilitate the identification of linguistic patterns and errors, thereby supporting more focused language learning.

In summary, while new, larger, and more specialized corpora are emerging, there remains a lack of substantially original corpus types, e.g. learner corpora and graded corpora, as well as speech corpora or ones suitable for mobile consultation.

Tools in DDL for LGP

Software for DDL in LGP encompasses both downloadable tools and online interfaces, each offering their own advantages for language learners and educators.

Downloadable tools like AntConc (Anthony, 2023) are invaluable for DDL by enabling the creation of concordances, which facilitate the

exploration of lexical patterns, collocations, and word frequencies. These tools support customization of search patterns using wildcards and regular expressions, allowing for focused educational interventions. For instance, AntConc's flexibility in search criteria empowers users to delve deep into textual data to uncover subtle language nuances and usage patterns. Such tools aid in tasks crucial to LGP, including identifying significant words, patterns of usage, and relationships between words within texts. BootCaT (https://bootcat.dipintra.it) automates corpus creation from web content, facilitating the rapid compilation of small- to medium-sized corpora. This tool streamlines the process of gathering and extracting text from the web based on user-specified search parameters. By automating data collection and outputting results in plain text format, BootCaT simplifies the task of compiling diverse textual samples for linguistic analysis. This capability is particularly useful in LGP contexts where access to a wide range of textual data is essential for exploring language usage across different genres, registers, and contexts. However, while downloadable tools offer robust functionalities for in-depth linguistic analysis, they may present challenges for learners new to corpus linguistics due to their interface complexity and the need for careful corpus selection. Balancing these complexities with their pedagogical benefits is crucial to effectively support language learning objectives without overwhelming learners.

Online interfaces such as the English-Corpora.org website, SkELL (Sketch Engine for Language Learning; https://skell.sketchengine.eu/#home?lang¼en), Voyant Tools (http://www.voyant-tools.org), BNCweb (http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php), and LexTutor (https://www.lextutor.ca/) provide convenient access to a wealth of linguistic resources without the need for software installation. These platforms offer intuitive user interfaces with features like concordance search, word sketches, and collocation analysis, making them accessible across various devices and operating systems. This accessibility supports flexible learning environments where learners can engage with language data from diverse sources and contexts.

For instance, English-Corpora.org stands out by offering insights into the diversity within English across different genres, historical periods, and regional varieties. Such platforms enable users to explore how language evolves over time and varies across different social and cultural contexts. SkELL, designed specifically for language learners and educators, simplifies the exploration of language usage patterns with user-friendly tools like simplified word sketches and integrated dictionary features. Voyant Tools, known for its visual approach to text analysis, enhances comprehension and engagement by visually representing textual data.

While online interfaces provide ease of access and collaboration opportunities, they may have limitations in terms of depth of analysis and the availability and quality of corpora compared to downloadable tools. Additionally, reliance on Internet connectivity and potential subscription costs for premium features can impact their usability in certain educational settings.

In summary, both downloadable tools and online interfaces play crucial roles in supporting DDL in LGP by offering diverse functionalities that cater to different learning needs and preferences. Effective integration of these tools into language learning environments requires consideration of their respective strengths and limitations to optimize educational outcomes.

Benefits of DDL in LGP

The benefits of DDL for LGP are substantial. One of the key benefits is the emphasis on phraseology which includes mastering common phrases and expressions crucial for fluency. It emphasizes lexicogrammatical, functional, and prosodic aspects, enabling learners to improve overall language skills. DDL aids in identifying collocations —words that commonly co-occur—promoting fluency in comprehension and production. It promotes learner autonomy by treating language as data, engaging students as researchers in guided tasks to explore linguistic patterns and corpus tools independently. This approach supports vocabulary acquisition, grammar understanding,

and proficiency development, aligning with practical communication needs. By using corpora for writing and revision, learners verify language usage, improving naturalness and accuracy. Furthermore, multimodal corpora, integrating written text with audio or video, allow the study of prosodic features like intonation and stress patterns, enriching learning. DDL thus enriches LGP education through structured, research-oriented approaches that foster autonomy, critical thinking, and real-world language application, preparing learners for diverse language learning contexts and specialized language needs.

Challenges and Criticisms

DDL for LGP faces several challenges. The complexity of authentic data can be overwhelming for beginners, who might struggle with the sheer volume and intricacy of the datasets without proper guidance. Additionally, the availability of suitable corpora and user-friendly tools is not universal, particularly for less commonly taught languages, which can hinder the widespread adoption of DDL. Building local corpora (see Abuhakema et al., 2008), though not easy, might help overcome this scarcity. Using local corpora in DDL, potentially built by students themselves, offers language data directly relevant to everyday communication, enhancing learning relevance and applicability. Student involvement in corpus creation promotes active participation and ownership, deepening language analysis and understanding. However, constructing a local corpus is complex, requiring significant effort in data collection, cleaning, and organization, often constrained by limited resources and technical support. Teacher preparation is another significant hurdle, as educators require specific training to effectively integrate DDL into LGP curricula. Without adequate support and training, teachers may find it challenging to implement DDL effectively. Crucially, there is a necessity for "pedagogic transfer" (Pérez-Paredes, 2010, p. 3) which is vital in adapting methods and corpora originally designed for language research for educational purposes.

Applications and Approaches of DDL in LGP

The integration DDL into LGP teaching has led to ongoing debate on its role alongside traditional approaches. Should DDL seamlessly blend in or stand as a distinct transformative method?

Leech (1997) distinguished between two main approaches to incorporating corpora into language teaching. In indirect DDL, teachers use corpus data to design materials and activities without requiring students to directly interact with the concordance. This approach simplifies the learning process by curating and presenting relevant data to learners, though it may not fully embody DDL if teaching remains largely traditional. In indirect corpus-driven teaching methods, learners typically review corpus excerpts prepared by the teacher, with the corpus itself remaining invisible. In contrast, in direct DDL students engage directly with the concordance (hands-on or "inductive," as in Cobb, 1997), enabling them to explore, discover, and analyse language patterns independently. This approach empowers learners with autonomy and deeper engagement, as described by Vyatkina (2016) as the "strong version" of DDL (see Bernardini, 2004). Alternatively, students may interact with pre-prepared corpus-informed materials (hands-off), where they engage with educational materials that incorporate insights from corpus analysis. This method allows students to access corpus-derived data directly through curated materials, although they do not engage in direct interaction with the raw corpus data or a concordance. Furthermore, Seidlhofer (2002) proposed "learning-driven data" to parallel Tim Johns' *data-driven learning*. This involves systematically storing learner-generated content for comparison with native speaker corpora. The aim is to help learners understand their position within their L2 learning environment and their potential learning goals.

The integration of corpora in language teaching and learning has predominantly taken an indirect approach rather than a direct one (McEnery & Xiao, 2011). This preference may stem from various constraints that limit the direct use of corpora

in language pedagogy. These factors include the proficiency levels and backgrounds of learners, time limitations within curricula, the expertise necessary for corpus analysis and instructional adaptation by teachers, and access to essential resources such as computers, suitable software tools, and relevant corpora. To exemplify how these approaches can be used to teach LGP, two case studies are provided below: *Case study 1* exemplifies hands-on DDL, with a language focus on lexicogrammar, particularly accuracy in writing, where participants were given additional instruction on performing independent concordance searches; *case study 2* provides an example of hands-off DDL, with participants being instructed to explore extracts from an integrated bilingual learner corpus.

Case Study 1: Hands-On DDL

Gaskell and Cobb (2004) present a learning system using concordance information to provide feedback for sentence-level written errors made by lower-intermediate English L2 writers. The study involved adult Chinese EFL learners enrolled at a university in Montreal, Canada, in a course spanning 15 weeks and including ten assignments. To improve the usability of data-driven writing feedback, the authors developed a concordance integrated into a learning platform, comprising an assignment upload program, a corpus, a concordance, and a form submitter for learners. The concordance was linked to the 1-million-word Brown corpus. However, since the Virtual Language Centre's concordance is no longer available, alternatives like BNCweb can be used. Errors were addressed by providing correct examples through concordance searches, with URLs pasted into learners' texts for revision. Learners submitted forms detailing corrections based on concordance information. A needs analysis identified common errors, guiding concordance feedback throughout the course. Initial in-class training sessions taught students to retrieve and interpret concordances. Subsequently, the authors switched to a more hands-on approach to DDL with participants receiving further guidance on conducting independent concordance searches. Participant feedback via questionnaire was positive, expressing satisfaction with the methodology and their writing progress. However, the authors noted higher participation when pre-cast URL links were provided in revision assignments, suggesting a need for greater confidence with technology, and a longer training period.

Case Study 2: Hands-Off DDL

Belz and Vyatkina (2008) report on instructional modules they developed for teaching German modal particles and pronominal *da*-compounds using a Developmental *Learner* Corpus (DLC). This bilingual corpus included chat and email interactions between English native speakers learning German and German native speakers learning English. Learners' conversations were stored in online folders and in the DLC. Participants explored these conversations both as collaborative online documents and as DLC excerpts to gain familiarity with the technology. The first six-stage module focused on modal particles (MPs). Initially, learners' usage of MPs was assessed (pre-intervention stage). Stage 1 (1 week) taught the expressive force of MPs, while stages 2 and 3 (2 and 1 week, respectively) focused on MP usage rules, meanings, and syntax by analysing DLC excerpts. Stage 4 (1 week) involved discussing correct and incorrect MP usage with additional examples from the DLC, connecting these to native speaker patterns. Learners were then encouraged to use MPs in their writing. The study assessed learners' MP competence post-intervention and collected ethnographic data, including surveys and feedback, documenting micro-changes in their language use.

Research indicates that both hands-off and hands-on DDL approaches can be effective in LGP contexts. Hands-off DDL simplifies the learning process, while hands-on DDL empowers learners with autonomy and deeper engagement. Boulton and Cobb's (2017) meta-analysis affirms the overall efficacy of DDL across diverse learning environments, asserting that it performs well across a spectrum of contexts. However, further research is warranted to determine the optimal utilization of DDL with varying learner profiles and within different educational settings, catering to specific learning objectives.

Future Directions and Research Needs

DDL offers a powerful approach to general language education. By using authentic language data, DDL enhances the practicality, relevance, and depth of LGP instruction. Despite some challenges, the benefits of this integration are clear, paving the way for more effective and engaging language learning experiences. As the field continues to grow, further research and development will help to optimize DDL's application across different learning contexts and levels. Future research should explore DDL's application in diverse educational settings and with various proficiency levels. Developing accessible DDL tools and resources, and providing more training for educators, will be crucial for broader implementation. Additionally, more studies are needed to investigate the use of DDL with younger learners and in non-university contexts, as the majority of current research focuses on university students, often for specific or academic purposes, though a surprising number are for LGP. In this respect, Boulton and Vyatkina (2021, p. 77) suggest that "this may seem unintuitive: since general language can be found in generic resources (dictionaries, textbooks, etc.), it might be thought that DDL is most appropriate for specific needs in narrow fields". Therefore, it might be assumed that while applications of DDL in specialized fields are valuable and expanding, research primarily focuses on LGP as a starting point. Mastering general language patterns and structures through DDL lays a crucial groundwork for learners before they progress to more specialized language studies in specific fields or domains.

DDL is poised to evolve with the continuous integration of cutting-edge technologies. The development of mobile apps dedicated to DDL will make these resources more accessible and convenient, allowing learners to engage with language data anytime and anywhere. One significant trend is the use of large language models (LLMs) to enhance language learning experiences. These models can provide sophisticated, real-time linguistic analysis and personalized feedback, offering learners deeper insights into language patterns and usage. However, several authors (Flowerdew, 2024; Crosthwaite & Baisa, 2023; Lin, 2023) claim that the current paradigm shift with LLMs still leaves room for traditional DDL. This is especially true for indirect approaches, where deterministic results from linguistic analysis are integrated into learning materials, as they have been since the COBUILD dictionary. Additionally, in DDL pedagogy, the goal remains for every learner to act like "Sherlock Holmes" (Johns, 1997, p. 101), guided by the teacher, to discover for themselves how the target language functions and to incorporate that knowledge into their linguistic repertoire. As these technologies advance, they will complement traditional DDL methods, enabling learners to independently explore and understand the intricacies of the target language with greater ease and effectiveness.

Cross-References

- ▶ Hands-on vs Hands-off Methods in Data-Driven Learning
- ▶ Learner Corpora in Data-Driven Learning
- ▶ Mall and Data-Driven Learning
- ▶ Multimodal Tools and Corpora for Data-Driven Learning
- ▶ Techniques for Beginner-Level Data-Driven Learning
- ▶ Tools for Data-Driven learning

References

Abuhakema, G., Faraj, R., Feldman, A., & Fitzpatrick, E. (2008). Annotating an Arabic learner corpus for error analysis. In *Proceedings of the sixth international conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA). Retrieved July, 2024 from http://www.lrec-conf.org/proceedings/lrec2008/pdf/343_paper.pdf

Anthony, L. (2023). *AntConc* (version 4.2.4) [computer software]. Waseda University. Retrieved May, 2024, from https://www.laurenceanthony.net/software

Belz, J. A., & Vyatkina, N. (2008). The pedagogical mediation of a developmental learner corpus for classroom-based language instruction. *Language Learning & Technology, 12*(3), 35–52. https://doi.org/10125/44154

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In

J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.12.05ber

Boulton, A. (2012a). What data for data-driven learning? *The EUROCALL Review, 20*(1), 23–27. https://doi.org/10.4995/eurocall.2012.16038

Boulton, A. (2012b). Computer corpora in language learning: DST approaches to research. *Mélanges Crapel, 33*, 79–91. https://www.atilf.fr/wp-content/uploads/publications/MelangesCrapel/file-34-6-1.pdf

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Cheng, W. (2010). What can a corpus tell us about language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 319–331). Routledge. https://doi.org/10.4324/9780367076399

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*(3), 301–315. https://doi.org/10.1016/S0346-251X(97)00024-9

Cobb, T. (1999). Applying constructivism: A test for the learner as scientist. *Educational Technology Research & Development, 47*(3), 15–31. https://doi.org/10.1007/BF02299631

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast. *Applied Corpus Linguistics, 3*(3), 100066. https://doi.org/10.1016/j.acorp.2023.100066

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.69.02flo

Flowerdew, J. (2024). Data-driven learning: From Collins Cobuild Dictionary to ChatGPT. *Language Teaching*, 1–18. https://doi.org/10.1017/S0261444824000144

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://doi.org/10.1016/j.system.2004.04.001

Gilquin, G., De Cock, S., & Granger, S. (2010). *The Louvain International Database of Spoken English Interlanguage (LINDSEI)*. Presses Universitaires de Louvain.

Huang, L. (2011). Corpus-aided language learning. *ELT Journal, 65*(4), 481–484. https://doi.org/10.1093/elt/ccr031

Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge University Press. https://doi.org/10.1017/CBO9781139524605.014

Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100–115). Longman. https://doi.org/10.4324/9781315842677-9

Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107–117). Rodopi. https://doi.org/10.1163/9789004334236_010

Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–23). Addison Wesley Longman. https://doi.org/10.4324/9781315842677-1

Lin, P. (2023). ChatGPT: Friend or foe (to corpus linguists)? *Applied Corpus Linguistics, 3*(3), 100065. https://doi.org/10.1016/j.acorp.2023.100065

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 364–380). Routledge. https://doi.org/10.4324/9780203836507

Meunier, F., & Gouverneur, C. (2009). New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 179–201). John Benjamins Publishing. https://doi.org/10.1075/scl.33.16meu

O'Keeffe, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching, 54*, 259–272. https://doi.org/10.1017/S0261444820000245

Papp, S. (2007). Inductive learning and self-correction with the use of learner and reference corpora. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 207–220). Rodopi. https://doi.org/10.1163/9789401203906_014

Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: Appropriation and the possibilities scenario. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 53–73). Peter Lang.

Seidlhofer, B. (2002). Pedagogy and local learner corpora. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 213–234). John Benjamins. https://doi.org/10.1075/lllt.6.14sei

Sinclair, J. (1987). *Collins COBUILD English language dictionary*. Collins.

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL, 28*(2), 207–226. https://doi.org/10.1017/S0958344015000269

Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals, 53*(2), 359–370. https://doi.org/10.1111/flan.12464

Whistle, J. (1999). Concordancing with students using an 'off-the-Web' corpus. *ReCALL, 11*(2), 74–80. https://doi.org/10.1017/S0958344000004985

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 3–25. https://doi.org/10.1093/applin/21.1.3

L

Language for Academic Purposes and Data-Driven Learning

Reka R. Jablonkai
University of Bath, Bath, UK

Abstract

Data-driven learning (DDL) has significantly advanced since its early focus on academic writing, exemplified by Tim Johns's Kibbitzers. In recent years, a considerable proportion of DDL research has addressed specific aspects of language for academic purposes. A number of corpora aiming to represent academic language use in different contexts and levels have been compiled and used in DDL over the years. This entry discusses the key research areas where DDL has been successfully used in language for academic purposes to demonstrate effective academic writing and speaking and to address specific academic needs. The most widely studied areas include disciplinary variation, rhetorical functions, lexicogrammatical patterns, error correction and compiling personal corpora. Areas that have received little attention so far include spoken academic language and computer-mediated academic discourse. These areas are showcased in this entry with example studies and corpus resources designed for academic purposes.

Keywords

Disciplinary variation · Spoken academic language · Written academic language · Rhetorical functions · Error correction · Academic writing · Academic listening · Specialised corpora

Introduction

Data-driven learning (DDL) started with activities that focused on academic writing. Many of the first set of DDL activities, the Kibbitzers by Tim Johns, were designed to support academic writing development of university students (Johns, 1996–1997). Ever since language (especially English) for academic purposes has been a flourishing area in DDL. Lusta et al. (2023), in a recent systematic review of DDL, found that 19 (21%) of the 89 reviewed papers published between 1997 and 2022 dealt with aspects of language for academic purposes (LAP). Tool-centred research characterised DDL in the period between 1998 and 2011 (Dong et al., 2023) which resulted in corpus platforms and corpora specifically designed for academic purposes. This was the time when several of the widely used academic corpora were also compiled. These include the Michigan Corpus of Academic Spoken English (MICASE) (Simpson et al., 2002), the British Academic Written English corpus (BAWE; Nesi & Gardner, 2012), the British Academic Spoken corpus (BASE; Thompson & Nesi,

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_51-1

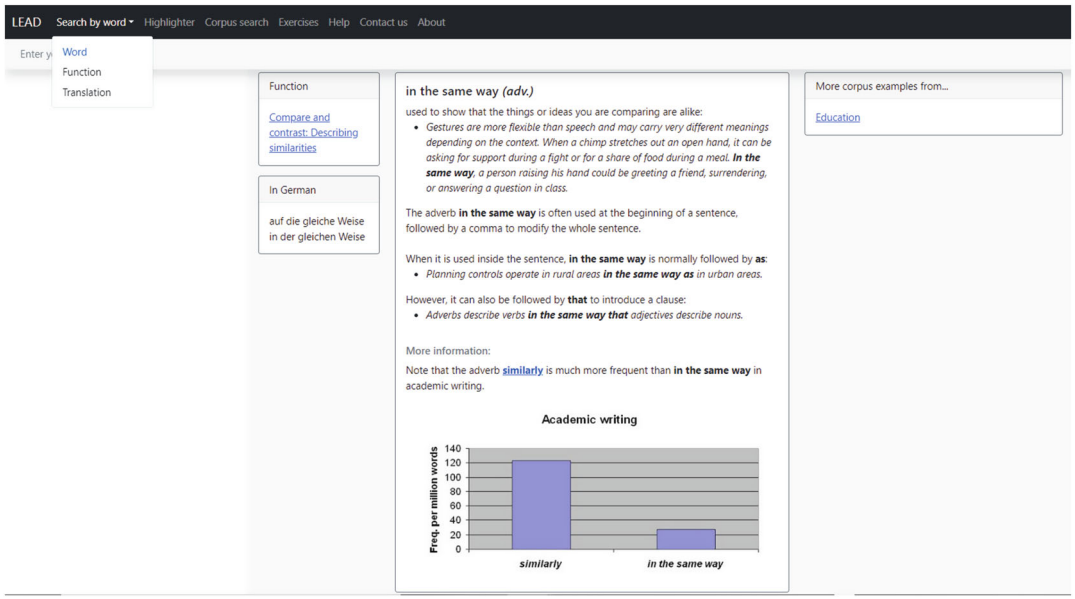
2001) and the Michigan Corpus of Upper-Level Student Papers (MICUSP) (Römer & O'Donnell, 2011). Several new corpora, web-based corpus platforms and tools (e.g. ColloCaid; Frankenberg-Garcia et al., 2019; LEAD; Paquot, 2012) have been developed since (Cangır, 2022; Chen & Flowerdew, 2018; Flowerdew, 2022). As regards the effectiveness of DDL for LAP, recent meta-analyses concluded that corpus use for learning purposes yielded large effect sizes (Boulton & Cobb, 2017) and that DDL's impact on L2 learning for LAP can enhance both short-term and long-term retention (Ueno & Takeuchi, 2023).

Written Academic Language

Many of the LAP and DDL studies apply a combination of a genre-based and corpus-based approach (Chen & Flowerdew, 2018). Such studies draw on the large body of corpus-based disciplinary variation literature that has examined a wide variety of linguistic features (e.g. the use of first-person pronouns, self-mentions, reporting verbs in research articles) and rhetorical structures and functions (e.g. stance in research articles, citations in research articles and theses) in academic discourse (Thompson, 2022). Students in the majority of these DDL studies consulted specialised corpora (e.g. BAWE or MICUSP) or bespoke corpora compiled for the specific purposes of their university courses. This allowed a more effective response to learner needs. Chen and Flowerdew's review also revealed that studies that combined genre- and corpus-based approaches suggested that this combination helps with scaffolding the learning process. Studies in general reported that students often felt overwhelmed by corpus consultation initially; however, they appreciated the benefits and believed that corpus consultation helped them understand academic texts better and allowed them to access a wide range of examples.

A further strand in LAP and DDL studies is when a DDL approach is applied to teaching particular sub-sections and rhetorical functions of research papers, dissertations and theses (Chen & Flowerdew, 2018; Karpenko-Seccombe, 2021). Studies frequently focus on sub-sections such as abstracts, discussion sections of dissertations and introductions of research articles, or rhetorical functions such as criticising previous studies, making and countering arguments, and comparing and contrasting ideas. A user-friendly corpus tool that allows learners to search for academic words according to their discourse function is The Louvain EAP dictionary (LEAD, https://leaddico.uclouvain.be/about, Granger & Paquot, 2022; Paquot, 2012). In addition to discipline-specific corpus search, the web-based LEAD platform includes a number of exercises to enhance learners' understanding and provide ample practice of the phraseology in specific disciplines. The tool can also be customised to match the learner's L1 background using data from multinational learner English corpora. A screenshot to show the functionalities of the tool can be seen in Fig. 1.

Corpus techniques have been found very powerful in demonstrating lexicogrammatical patterns (collocation and colligation) in academic texts. Therefore, DDL studies often focus on teaching academic vocabulary and use corpus techniques to explore such patterns (Chen & Flowerdew, 2018). DDL can be used to focus on collocation and colligation (e.g. prepositional patterns) of single- and multi-word units. For example, Huang (2014), in an experimental study, demonstrated that a DDL approach with paper-based concordances of single-word abstract nouns (e.g. *controversy, criticism, effect*) can be more effective in teaching collocational patterns than dictionaries. Multi-word units are often understood as collocational patterns, phrases and lexical bundles in DDL studies. Cortes (2022, p. 230), for example, proposes a sequence of activities to teach lexical bundles that are frequently used in the Swalesian Move 2 *Preparing present research* (Swales, 2004) in research article introductions. She suggests the following steps for instruction: first, students are shown several examples of lexical bundles in a relevant academic corpus; next, they identify these lexical bundles in their reading materials using corpus tools; and lastly, they can use the lexical bundles such as *In order to . . . it is necessary to . . ., It should be noted that . . ., There*

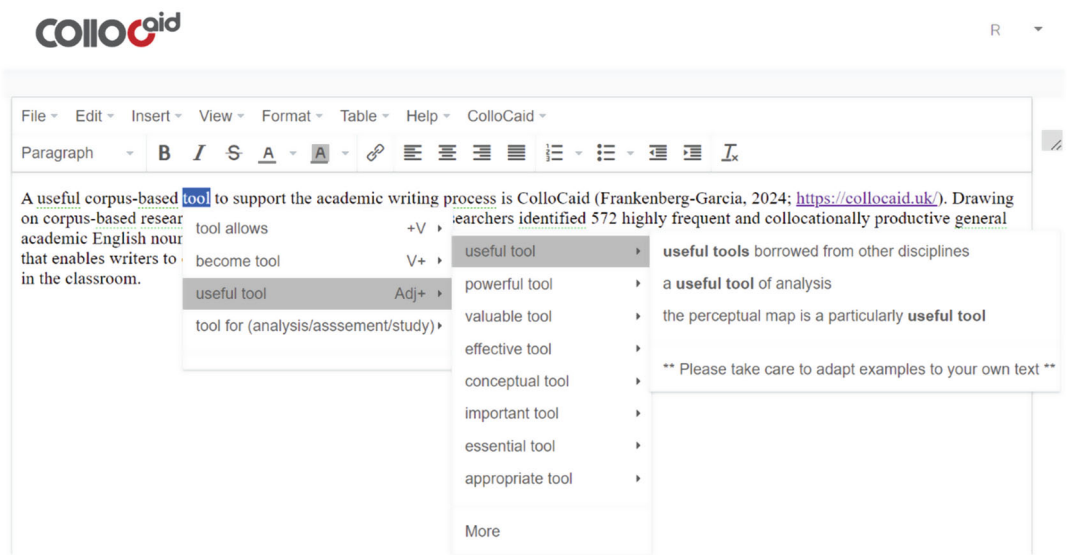


Language for Academic Purposes and Data-Driven Learning, Fig. 1 The entry of *in the same way* in LEAD

is a need to . . ., as frameworks to scaffold their own writing. Fewer DDL studies have focused on teaching cohesive devices such as linking adverbials (Chen & Flowerdew, 2018). A useful corpus-based tool to support the academic writing process is ColloCaid (Frankenberg-Garcia, 2024; https://collocaid.uk/). Drawing on corpus-based research into English academic texts, the researchers identified 572 highly frequent and collocationally productive general academic English nouns, verbs and adjectives and compiled the ColloCaid database. This is integrated into an open-source online text editor that enables writers to consult ColloCaid while writing, as shown in Fig. 2. ColloCaid supports autonomous academic writing and learning and can also be used in the classroom.

Chen and Flowerdew (2018) found that several DDL studies in LAP focus on students' error-correction. These studies often use available large general corpora (e.g. COCA, Davies, 2020; BNC, 2007, 2014) against which to compare and evaluate learner writing. Studies that compared two different types of corpus consultation (e.g. Yoon & Jo, 2014), referred to as indirect and direct (Leech, 1997) or hands-on and hands-off (Boulton, 2012) approaches, revealed that learners often make more corrections with a hands-off approach when they are provided with print-outs of ready-made and potentially edited or curated concordance lines for the targeted errors. However, learners can find direct corpus consultation more interesting. Other studies have also explored different ways and amounts of teacher intervention and guidance for error correction. Although findings regarding the effectiveness of the various types of teacher guidance are not conclusive, a combination of teacher guidance and self-directed corpus consultation—when teachers identify the errors and learners correct them with the help of corpora—has been found highly effective (e.g. Crosthwaite, 2017).

Self-directed DDL can also be taken a step further by encouraging learners to compile, annotate and consult personal corpora, also called do-it-yourself (DIY) corpora, for their own specific courses or writing purposes such as for assignment, dissertation and thesis writing (e.g. Charles, 2012; Lee & Swales, 2006). Such studies have found that the design and compilation process enhances learners understanding of the nature and limitations of corpus data in general thus promoting corpus literacy (e.g. Jablonkai & Čebron, 2021). A useful tool to support the



Language for Academic Purposes and Data-Driven Learning, Fig. 2 Interactive menus of ColloCaid in In-Text View

compilation of DIY corpora for academic purposes is AntCorGen (Anthony, 2024). AntCorGen can compile specialised academic corpora using research articles from the Open Access journal *PLOS ONE* based on relevant search criteria (e.g. subject category, date of publication, keywords). The tool automatically divides the texts into relevant sections (e.g. abstract, introduction and methods) thus allowing learners to explore the language use in these specific parts of academic texts. There is still little research into learners' long-term autonomous corpus use for academic purposes. An exception is Charles and Hadley's (2022) study that followed up learners' corpus use 1 year after their corpus-based courses finished; the course had run 50 times between 2009 and 2017. Responses showed that around 62% of the learners continued using their DIY corpora after taking the DDL academic writing course.

Spoken Academic Language

A less widely researched aspect of LAP and DDL is how DDL can be used when it comes to spoken academic language. There are corpora that represent spoken academic language use, such as BASE and MICASE. These corpora comprise written transcripts of academic lectures and seminars. A recent study by Zare and Delavar (2023) demonstrated the effectiveness of DDL activities in providing learners with concordance lines of importance markers in lectures (e.g. *I think the first point we need to make is that . . ., . . . so that's the main point for today's lecture*) for lecture comprehension. The authors concluded that incorporating DDL activities into language learning tasks enhances academic English lecture comprehension by improving learners' noticing and discovery learning, and heightening their metalinguistic awareness.

In recent years, more and more multimodal corpora, which integrate textual, video and audio recordings, are available for DDL (Crawford Camiciottoli, 2022). ELAN (2020) and ANVIL (Kipp, 2012), for example, are corpus tools that allow multimodal annotation of such corpora and can be used for pedagogic purposes. Crawford Camiciottoli (2021) proposed a lesson plan for using video clips from the Pisa Audiovisual Corpus (Crawford Camiciottoli & Bonsignori, 2015) for academic listening skill development.

In addition to written and spoken discourse in higher education contexts (e.g. assignments, lectures, PhD theses) and in academic publications (e.g. research articles), recent years have witnessed the emergence and increasing relevance of computer-mediated discourse within LAP. Drawing on corpus-based research for EAP, Viana and O'Boyle (2022) discuss genres such as e-mails, discussion forums and blogs, which have become important in computer-mediated academic discourse (CMAD). While they do not provide specific DDL activities for teaching these academic genres, the findings of such corpus-based studies can inform future DDL tasks and lesson plans in LAP.

Summary

LAP and DDL have a long and fruitful history. Since the first DDL activities that focused primarily on academic writing, several corpus resources such as academic corpora, corpus tools and web-based platforms have been developed that can be used in DDL for academic purposes. DDL draws on the substantial corpus-based research into academic discourse. The areas that have received considerable attention in LAP and DDL research focus on disciplinary variation, rhetorical functions and lexicogrammatical patterns, with studies often combining genre- and corpus-based methods to address specific needs in academic writing. A small number of DDL studies have focused on spoken academic language and listening skills. Emerging trends also show the growing importance of computer-mediated academic discourse. As a result, DDL has been enriched with multimodal corpora and tools to enhance academic language instruction.

Cross-References

- ▶ Genre Awareness and Data-Driven Learning
- ▶ Language for Specific Purposes and Data-Driven Learning
- ▶ Writing Aids for Data-Driven Learning

References

Anthony, L. (2024). *AntCorGen (Version 1.3.0)* [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Boulton, A. (2012). Hands-on/hands-off: Alternative approaches to data-driven learning. In J. Thomas & A. Boulton (Dir.), *Input, process and product: Developments in teaching and language corpora* (pp. 152–168). Masaryk University Press.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

British National Corpus (BNC). (2007). http://www.natcorp.ox.ac.uk/corpus/index.xml

British National Corpus (BNC). (2014). http://corpora.lancs.ac.uk/bnc2014/

Cangır, H. (2022). Using corpus tools for academic writing in EFL settings: A data-driven learning approach. In G. Yangın-Ekşi et al. (Eds.), *New directions in technology for writing instruction* (English language education 30) (pp. 171–191). Springer Nature. https://doi.org/10.1007/978-3-031-13540-8_9

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes, 31*(2), 93–102. https://doi.org/10.1016/j.esp.2011.12.003

Charles, M., & Hadley, G. (2022). Autonomous corpus use by graduate students: A long-term trend study (2009–2017). *Journal of English for Academic Purposes, 56*, 101095. https://doi.org/10.1016/j.jeap.2022.101095

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics, 23*(3), 335–369. https://doi.org/10.1075/ijcl.16130.che

Cortes, V. (2022). Lexical bundles in EAP. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 220–233). Routledge. https://doi.org/10.4324/9781003002901-18

Crawford Camiciottoli, B. (2021). Teaching lecture comprehension skills through OpenCourseWare video-recorded lectures: A research-informed classroom application. *Journal of English for Academic Purposes, 54*, 101055. https://doi.org/10.1016/j.jeap.2021.101055

Crawford Camiciottoli, B. (2022). Corpora and EAP listening comprehension. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 248–263). Routledge. https://doi.org/10.4324/9781003002901-20

Crawford Camiciottoli, B., & Bonsignori, V. (2015). The Pisa Audiovisual Corpus project: A multimodal approach to ESP research and teaching. *ESP Today, 3*(2), 139–159.

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning, 30*(6), 447–473. https://doi.org/10.1080/09588221.2017.1312462

Davies, M. (2020). *The COCA corpus*. https://www.english-corpora.org/coca/help/coca2020_overview.pdf

Dong, J., Zhao, Y., & Buckingham, L. (2023). Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL, 35*(3), 339–355. https://doi.org/10.1017/S0958344022000222

ELAN (Version 5.9) [Computer software]. (2020). Max Planck Institute for Psycholinguistics. https://archive.mpi.nl/tla/elan

Flowerdew, L. (2022). Corpora for EAP writing. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 234–247). Routledge. https://doi.org/10.4324/9781003002901-19

Frankenberg-Garcia, A. (2024). DDL support for academic English collocations. In P. Crosthwaite (Ed.), *Corpora for language learning* (pp. 71–84). Routledge. https://doi.org/10.4324/9781003413301-6

Frankenberg-Garcia, A., Lew, R., Rees, G., Roberts, J., Sharma, N., & Butcher, P. (2019). Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL, 31*(10), 23–39. https://doi.org/10.1017/S0958344018000150

Granger, S., & Paquot, M. (2022). The Louvain English for academic purposes dictionary manual. CECL Papers 5. Centre for English Corpus Linguistics/Université catholique de Louvain. https://cdn.uclouvain.be/groups/cms-editors-cecl/cecl-papers/LEAD_HELP_march2022_0.pdf

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL, 26*(2), 163–183. https://doi.org/10.1017/S0958344014000020

Jablonkai, R. R. & Čebron, N. (2021). Undergraduate students' response to a corpus-based ESP course with DIY corpora. In M. Charles & A. Frankenberg-Garcia (Eds.), *Corpora in ESP/EAP writing instruction* (pp. 100–120). Routledge.

Johns, T. (1996–1997). Kibbitzers. University of Birmingham. https://lexically.net/TimJohns/#gsc.tab=0

Karpenko-Seccombe, T. (2021). *Academic writing with corpora: A resource book for data-driven learning*. Routledge. https://doi.org/10.4324/9780429059926

Kipp, M. (2012). Multimedia annotation, querying, and analysis in ANVIL. In M. Maybury (Ed.), *Multimedia information extraction. Advances in video, audio, and imagery analysis for search, data mining, surveillance and authoring* (pp. 351–368). Wiley-IEEE Computer Society. https://doi.org/10.1002/9781118219546.ch21

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75. https://doi.org/10.1016/j.esp.2005.02.010

Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 11–23). Addison Wesley Longman.

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data-driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*(12), e22731. https://doi.org/10.1016/j.heliyon.2023.e22731

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge University Press. https://doi.org/10.1017/9781009030199

Paquot, M. (2012). The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool. In S. Granger & M. Paquot (Eds.), *Electronic lexicography* (pp. 163–185). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654864.003.0009

Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora, 6*(2), 159–177. https://doi.org/10.3366/corp.2011.0011

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. The Regents of the University of Michigan.

Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524827

Thompson, P. (2022). Corpus analysis of disciplinary variation and the teaching of ESP/EAP. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 177–192). Routledge. https://doi.org/10.4324/9781003002901-15

Thompson, P., & Nesi, H. (2001). The British Academic Spoken English (BASE) corpus project. *Language Teaching Research, 5*(3), 263–264. https://doi.org/10.1177/136216880100500305

Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics, 3*(3), 100076. https://doi.org/10.1016/j.acorp.2023.100076

Viana, V., & O'Boyle, A. (2022). *Corpus linguistics for English for academic purposes*. Routledge. https://doi.org/10.4324/9781003245988

Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology, 18*(1), 96–117. https://doi.org/10.10125/44356

Zare, J., & Delavar, K. (2023). A data-driven learning focus-on-form approach to academic English lecture comprehension. *Applied Linguistics, 44*(3), 485–504. https://doi.org/10.1093/applin/amac021

C

Constructivist Approaches to Data-Driven Learning

Lynne Flowerdew
Birkbeck, London, UK

Abstract

Constructivism is a theory within the field of Second Language Acquisition. It is associated with an inductive, discovery-based approach to learning, in which students apply their existing background knowledge and cognitive problem-solving skills to understand new data. Data-driven learning (DDL) relies on the use of corpora and concordancing software to retrieve, for example, multiple uses of vocabulary or phrases from a corpus. DDL also promotes an inductive, discovery-based approach to language learning as students are expected to extrapolate rules or probabilistic tendencies of language from examples in the concordance output. DDL is a methodology which is thus seen as conducive to a constructivist approach. The teacher's role is seen as a facilitator, ensuring that the learning process is effective. This can be done through providing scaffolding input, which can take the form of co-construction of knowledge through collaborative dialogue and negotiation with guidance and support mediated by the teacher or other students. This entry discusses key DDL pedagogic initiatives which take a constructivist stance. It also reviews empirical studies which aim to investigate the link between DDL and constructivist approaches. Recommendations for research to establish a closer association between constructivist theory and DDL methodological practice are given.

Keywords

Constructivism · Data-driven-learning · DDL · Inductive · Discovery-based · Socio-cultural theory · SCT · Scaffolding · Tom Cobb · Tim Johns

Introduction

Cogent accounts of the theoretical background of constructivism and its position within the field of language education and, more specifically, educational technology, are provided in Cobb (1999, 2005, 2024), a leading figure in the field:

> Constructivism, the notion that knowledge must be assembled from pieces rather than assimilated whole, has been a principle learning theory in psychology for about 20 years and in psycholinguistics for 10. The theory is now making a strong entry into educational thinking. . . a more explicit understanding of the constructivist approach is useful within language education, particularly in providing a framework for exploiting information technologies. (Cobb, 2005, p. 1)

In the language education field, the pedagogic application of constructivism is seen as an

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_53-1

alternative to the deductive Presentation-Practice-Production model, in which the rules of grammar, for example, "can be imparted to learners wholesale through telling reinforced by practice, repetition, and application" (Cobb 2024, p. 2). Constructivism, on the other hand, is associated with an inductive 'discovery-based' approach to learning that "involves learners grappling in some sense with (more or less) 'raw' data and moving it in the direction of abstraction or generality, rather than simply being told the generalisation about the data and asked to seek out and remember instances of it" (Cobb 2024, p. 2). Students thus apply their existing background knowledge and cognitive problem-solving skills to understand new data. Additionally, Collentine (2000, p. 47) argues that in this approach students can be given multiple entry points for a targeted item of relevance to them "since features lacking salience in one context might be more salient in anther".

Cobb (2024) points out that constructivism is not to be confused with Krashen's (1989) Input Hypothesis model for language acquisition, now largely discredited on account of its lack of empirical validation, in which the teacher played a minimal role, and learners were expected to learn by themselves in a somewhat uncontrolled environment. While Cobb concedes a superficial resemblance to the "just-let-it-happen aspects of Krashenism" (p. 7), the teacher's role in constructivist learning is of paramount importance in ensuring the learning process is effective. This can be achieved, if necessary, by drawing on Lantolf and Thorne's (2006) Socio-cultural Theory (SCT), anchored in the work of Vygotsky (1934/1986), and scaffolding input for the learner. This can take the form of co-construction of knowledge through collaborative dialogue and negotiation with guidance and support mediated by the teacher or student. While many constructivists accept that some form of scaffolding or guidance can be accommodated within the constructivist approach, as pointed out by Cobb (1999), the learner is still considered to have agency and can choose their own learning pathways and whether to accept or reject feedback and suggestions offered. Constructivism also has some affinity with usage-based approaches to language acquisition (see Ellis et al., 2016). As noted by Cobb (2024, p. 5), the empirical research by Ellis (2002) demonstrates "that L2s are regularly constructed through general cognitive processes, such as generalising from exemplars, as supported by normal cognitive sensitivity to features of language like frequency, context and analogy".

The above overview of constructivism has been positioned within the general field of Second Language Acquisition (SLA) to illustrate its complementarity with other theories, especially SCT. Against this theoretical backdrop, the following section discusses how a constructivist approach can be seen to underpin the field of data-driven learning (DDL) pedagogy, a branch of computer-assisted language learning (CALL) within the wider field of educational technology.

How Do Constructivist Approaches Relate to DDL Pedagogy?

Cobb (2024, p. 7) underscores how the tools for implementing DDL facilitate a constructivist approach: "language patterns have more chance of being noticed and incorporated if the exemplars are assembled and sorted in various ways by text-processing computer software such as concordancers". As Boulton (2021) notes, it was the methodology of DDL which was conducive to a constructivist approach with the theorising following on from this. Three overview studies in the DDL literature discuss and evaluate the types of second language learning theories that can inform corpus pedagogy, one of which is constructivism (see Flowerdew, 2015; O'Keeffe, 2021; Mitchell, 2021). Building on this previous literature, this section examines how some key DDL initiatives reflect the main tenets of constructivism, but with different emphases.

It was Tim Johns (1991), generally regarded as the father of DDL, who first coined the term DDL to describe a methodology making use of corpora and concordancing tools, to foster an inductive, discovery-based approach to learning. This innovative paradigm of Identify-Classify-Generalise challenged the traditional Presentation-Practice-

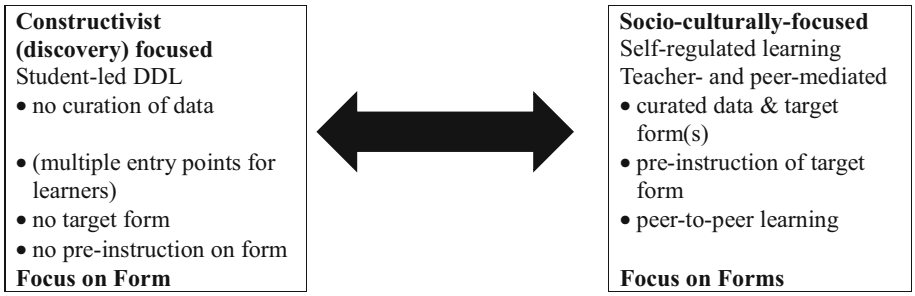
Production model, which still prevailed at the time. Johns' (1994, p. 29) groundbreaking work in DDL "attempts to cut out the middleman as far as possible" by giving learners direct access to the corpus data. He proposes that the learner takes on a role as a linguistic researcher, referring to the learner as a "language detective". His pity epithet "Every student a Sherlock Holmes!" aptly conveys the constructivist orientation underpinning his pedagogy. By way of example, Johns (1991) gave his students a set of printed concordance output with *convince* and *persuade* and asked them to analyse the data. Students working together on the task were able to verbalise the different ways in which these two verbs were used and their different constructions (see Gavioli, 1996 for an application of this methodology using corpora of small, specialised corpora from the medical field).

However, Bernardini (2002, 2004) takes issue with the *learner as researcher* favoured by Johns (1991) and Gavioli (1996) on the grounds that their approach "implicitly assumes that learners share the same interests, competences and capacities". Instead, she suggests that a more appropriate metaphor would be "the learner as traveller", invested with a high degree of learner agency. For Bernardini this serendipitous, incidental type of learning "encourages learners to follow their own interests whilst providing them with opportunities to develop their capacities and competences so that their searches become better focused, their interpretation of results more precise" (Bernardini, 2004, p. 23). Bernardini aligns her concept of DDL with schema theory, a branch of cognitive science, in which mental frameworks based on past experiences are used to process the corpus data. Both Johns and Bernardini acknowledge the importance of teacher guidance and scaffolding when necessary—Johns advocates cutting out the middleman but only "as far as possible" with Bernardini proposing guidance to help students progress from more convergent, guided activities to divergent autonomous browsing of the corpus (see Cobb & Boulton, 2015, for a discussion on scaffolding).

Römer (2011) and Gabrielatos (2005) have proposed similar paradigms for DDL. In Römer's (2011) model corpus interactions are viewed as either teacher- or learner-initiated. Teacher-initiated tasks involve careful guidance. Teacher-curated handouts can be provided to familiarise students with inductive reasoning and reading concordance lines before having them manipulate corpus tools. Learner corpus interactions, on the other hand, involve student autonomy and self-directed corpus consultation. Römer's (2011) teacher-initiated and learner-initiated corpus interactions correspond with what Gabrielatos (2005) refers to as 'hard' and 'soft' approaches, ranging from teacher- to learner-controlledness.

Karpenko-Seccombe (2018) makes use of a revised version of Bloom's taxonomy for DDL tasks. The taxonomy—Remembering, Understanding, Applying, Analysing, Evaluating and Creating—progresses from lower-order thinking skills to higher-order thinking skills. Her DDL tasks are graded according to their level of cognitive complexity with the teacher providing scaffolding in the form of prompts to guide students towards completing a complex task. By way of example, using the freely available LexTutor (https://www.lextutor.ca) program containing corpora of expert academic writing, Karpenko-Seccombe asked students to find out how the word *hypothesis* is used in academic writing. Students were prompted to look for verb collocations and prepositional patterns and articulate their findings, e.g., when *hypothesis* occurs as an object it is found with the following verbs: *X adopts, supports, presents, confirms*. The prompts become less detailed in line with students' growing competence and independence. The teacher's role thus changes from being the source of knowledge to acting as a facilitator or consultant through to the final Creating stage where students conduct independent searches.

Meanwhile, O'Keeffe (2021) presents a model for DDL which links both constructivist and socio-cultural approaches to learning (see Fig. 1 below). O'Keeffe also incorporates the concepts of Focus on Form (FonF) and Focus on Forms (FonFs) within these two learning paradigms. While FonFs would involve explicit language instruction, in a constructivist DDL approach the focus is on meaning with learners' attention drawn to forms as they arise incidentally.



Constructivist Approaches to Data-Driven Learning, Fig. 1 DDL cline of learning from constructivist to socio-culturalist stances on learning (O'Keeffe, 2021, p. 264)

While the above DDL initiatives embody constructivist principles (e.g., inductive learning, analytical processes, higher-order critical thinking skills), they differ in respect to where they can be positioned on the continuum in Fig. 1. The above task from Karpenko-Seccombe (2018), while intrinsically following an inductive, discovery-based approach, commences more at the socio-culturally focused end of the cline with her focus on forms, whereas Bernardini's approach takes a more constructivist approach at the outset.

What Evidence Is There that That DDL Fosters a Constructivist Approach to Learning?

Empirical studies which shed light on constructivist approaches to DDL are few and far between. Bernardini's (2000) study with six students of translation is situated within a constructivist framework, although this is not explicitly stated. The large-scale corpus in her study is used not as a reference source but rather as a browsing resource for exploratory activities allowing students to find their own entry points. The activities are "largely unstructured and open-ended. These activities required learners to participate actively in the learning process, and to a large extent shape it according to their cognitive preferences, interests and learning styles" (p. 226), with learners in the driving seat and teachers taking on the role of expert advisor. Follow-up interviews revealed that while students designed complex queries, they seldom used the browse mode and thought up alternative strategies. Also, it was found that students were sensitive to position/function of words in the text but that they seldom made lexicogrammatical observations. A small-scale study by Chang (2012), involving seven doctoral students, set out to investigate the types of cognitive skills students drew on in DDL activities to learn about effective authorial stance-taking, using a web-based corpus interface platform. However, it was found that this interface did not necessarily foster a constructive learning environment for attaining higher-level cognitive skills with learners infrequently using higher-order skills such as inferencing and relying on lower-level skills. Cobb (1997, 1999), meanwhile, designed a platform integrating a corpus, concordance software and a vocabulary list to aid lower-level proficiency students. The delayed post-tests indicated that students had drawn on more complex knowledge required for recognising new words in extended contexts rather than relying on the more superficial knowledge found in dictionaries. Lee and Lin's (2019) two-week study found no significant difference between inductive and deductive learning approaches, suggesting "the need to look beyond constructivism as the sole explanation of how learning takes place during DDL" (Crosthwaite & Boulton, in press), as elaborated below.

Future Directions

Various meta-analyses and surveys on DDL signal that the approach is under-theorised (Boulton & Cobb, 2017; Pérez-Paredes, 2019; Boulton &

Vyatkina, 2021; Crosthwaite & Boulton, in press), with Mitchell (2021) also noting the paucity of extended discussions and empirical studies such as those above. In this respect, Boulton (2012) observed that most studies focus on targets that are easy to measure such as short-term learning outcomes and error correction, often using pre- and post-tests, while other areas such as cognitivism are under-researched, a situation that prevails today. Both Mitchell (2021) and O'Keeffe (2021) have called for studies which investigate the *nature* of learning, i.e., how learners process corpus data, which can be captured through eye-tracking and screen and voice-recording. Another factor to consider is the degree of mediation involved depending, for example, on the learner's level of competence, previous learning experiences and curation of the data, as noted by O'Keeffe (2021). Students' self-regulated learning, such as their ability to set goals and make adjustments to their learning, etc., is another factor to consider as this involves cognitive and metacognitive strategies. The association between constructivism and DDL presents a complex picture; more research is thus needed to shed light on this phenomenon to inform pedagogy.

Cross-References

- ▶ Complexity, Accuracy and Fluency and Data-driven Learning
- ▶ Learner Autonomy and Data-Driven Learning
- ▶ Techniques for Beginner-Level Data-Driven Learning
- ▶ Tim Johns and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy form a corpus perspective* (pp. 225–234). Peter Lang.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Rodopi. https://doi.org/10.1163/9789004334236_015

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future development. In J. McH. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.12.05ber

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261–291). John Benjamins. https://doi.org/10.1075/scl.52.11bou

Boulton, A. (2021). Research in data-driven learning. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 9–34). John Benjamins. https://doi.org/10.1075/scl.102.01bou

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Chang, P. (2012). Using a stance corpus to learn about effective authorial stance-taking: A textlinguistic approach. *ReCALL, 24*(2), 209–236.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*(3), 301–315.

Cobb, T. (1999). Applying constructivism: A test for the learner as scientist. *Educational Technology Research & Development, 47*(3), 15–31. https://doi.org/10.1007/BF02299631

Cobb, T. (2005). Constructivism, applied linguistics, and language education. In *Encyclopedia of language and linguistics* (pp. 5–88). Elsevier.

Cobb, T. (2024). Constructivism in language learning and applied linguistics (pp. 1–8). In *Reference collection in social sciences*. Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00020-X

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *Cambridge handbook of corpus linguistics* (pp. 478–497). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.027

Collentine, J. (2000). Insights into the construction of grammatical knowledge provided by user-behaviour tracking technologies. *Language Learning & Technology, 36*, 45–60. https://doi.org/10125/25072

Crosthwaite, P., & Boulton, A. (in press). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In H. Tyne, M. Bilger, L. Buscail, M. Leray, N. Curry, & C. Pérez-Sabater (Dir.), *Discovering language: Learning and affordance*. Peter Lang.

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N., Römer, U., & O'Donnell, M. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigation of construction grammar*. Wiley-Blackwell.

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.69.02flo

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ, 8*(4), 1–35.

Gavioli, L. (1996). *Exploring corpora for ESP learning*. John Benjamins.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing*. *ELR Journal, 4*, 1–16.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogic grammar* (pp. 27–45). Cambridge University Press. https://doi.org/10.1017/CBO9781139524605.014

Karpenko-Seccombe, T. (2018). Practical concordancing for upper-intermediate and advanced academic writing: Ready-to-use teaching and learning materials. *Journal of English for Academic Purposes, 36*, 135–141. https://doi.org/10.1016/j.jeap.2018.10.001

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440–464. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

Lantolf, J., & Thorne, S. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Lee, P., & Lin, H. (2019). The effect of the inductive and data-driven learning (DDL) on vocabulary acquisition and retention. *System, 81*, 14–25. https://doi.org/10.1016/j.system.2018.12.011

Mitchell, R. (2021). Corpora and instructed second language acquisition. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 252–264). Routledge. https://doi.org/10.4324/9781351137904-22

O'Keeffe, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching, 54*, 259–272. https://doi.org/10.1017/S0261444820000245

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225. https://doi.org/10.1017/S0267190511000055

U

Usage-Based Approaches and Data-Driven Learning

Ute Römer-Barron
Georgia State University, Atlanta, GA, USA

Abstract

Usage-based approaches to language and language acquisition work from the assumption that language structure results from natural language use and that language is learned through a person's exposure to and experience with language in communicative situations. Data-driven learning (DDL) as a pedagogical method involves the use of a corpus or of corpus-derived materials in contexts of language teaching and learning, as well as learner engagement with authentic language. This entry summarises the main tenets of usage-based approaches to second-language acquisition (SLA) and discusses the relevance and significance of usage-based SLA for DDL.

Keywords

Usage-based linguistics · Second language acquisition · Corpus linguistics · Data-driven learning · Lexicogrammar · Constructions

Introduction

The past two decades have seen an increase in the popularity of usage-based approaches to language in applied linguistics, particularly in the field of second-language acquisition (SLA) or, more broadly, second−/additional-language (L2) development (see, for example, Cadierno & Eskildsen, 2015; Ellis et al., 2016; Tyler et al., 2018). The present entry explores in what ways data-driven learning (DDL) as a corpus-based approach to language teaching and learning is supported by usage-based approaches to language and language acquisition and establishes connections between usage-based SLA and DDL. The entry is largely based on Römer (2024) which provides a more detailed discussion of usage-based approaches vis-à-vis DDL.

As implied by their name, usage-based approaches take language use and language in use seriously. Their focus lies on 'usage events', defined as 'actual instances of language use, in their full phonetic detail and contextual understanding' (Langacker, 2008, p. 81). Since usage affects linguistic structure (Bybee & Beckner, 2010, p. 827) and provides the basis of language development, these usage events are of key importance. In the field of linguistics, the term 'usage-based' was first introduced by Langacker (1987) with reference to models of language. Usage-based language models, such as cognitive grammar, are further described by Langacker (1988) as maximalist, nonreductive and bottom-up and

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_54-1

contrasted to the minimalist, reductive, top-down orientation of generative grammar theories. According to usage-based models, abstractions of patterns from specific instances of language use shape the grammar of a language. There is no clear separation between lexis and grammar: 'lexicon and grammar grade into one another so that any specific line of demarcation would be arbitrary' (Langacker, 2000, p. 33; for additional details on usage-based language models, see Kemmer & Barlow, 2000).

Usage-Based Approaches to SLA

Usage events do not only affect and shape the system of language; they also affect and shape language acquisition. In the words of Ellis et al. (2016), '[l]earning language involves determining structure from usage' (p. 23). This applies to both L1 (first language) and L2 acquisition (see, for example, for L1, Behrens (2009), Lieven (2016) and Tomasello (2003), and for L2, Bybee (2008), Ellis & Ferreira-Junior (2009), Eskildsen & Cadierno (2015) and Römer & Berger (2019)). Accordingly, a key tenet of usage-based approaches to language acquisition is that language emerges from use (Tomasello, 2003) or, put differently, that language use shapes and provides the foundation for learning (Tyler & Ortega, 2018). Exposure to the target language is hence essential for learners, and the type and amount of input they receive matter. A second major tenet of usage-based approaches is that, from the language input they receive, learners derive rules and regularities by employing domain-general cognitive mechanisms such as perception, generalisation and pattern identification. In this context, frequencies of items and their coselection patterns in the input play an important role. Items that occur frequently in the usage events a learner is engaged in or exposed to get entrenched more strongly in the learner's mind than low-frequency items (Ellis, 2002). Related to this, a key assumption of usage-based approaches is that language learning is exemplar-based. Learners do not only implicitly register how frequent a specific language item is as they are exposed to it; they also register how it is distributed in a text and what other words occur around it (Ellis & Wulff, 2020). We learn from exposure to items presented in context.

Another key feature of usage-based approaches to language and language acquisition is that they highlight the importance of meaningful interaction. Language learning takes place while speakers engage in communication and is, like language, meaning-based. This means that, in pedagogical contexts, language 'must always be taught in the service of meaning making and communication' (Tyler & Ortega, 2018, p. 7). When they are connected to discourse functions, linguistic forms become learnable and constitute 'constructions' or recurring form-meaning mappings (Bybee, 2010; Goldberg, 2006). Language consists of constructions, and language learning is essentially construction learning. When we learn new constructions, we learn associations between forms and meanings, for example, that between the morpheme *dis-* in *disappear* and the meaning 'opposite of' or that between the phrase *under the weather* and the meaning 'sick'/'unwell'. The traditional separation of form and meaning is one of the dichotomies that is rejected in usage-based approaches to SLA, together with that of lexis and grammar. Lexical items have a preference for certain grammatical constructions (and vice versa), an assumption for which corpus linguistics research has provided ample evidence (e.g. Hoey, 2005; Hunston & Francis, 2000; Römer, 2009; Sinclair, 1991, 2004).

A final important concept in usage-based SLA is attention. As Ellis et al. (2016) state, '[w]hat is attended to becomes the focus of learning, and so attention controls the acquisition of language itself' (p. 23; see also Ellis, 2019). In instruction, it is therefore important to help language learners focus their attention on the language items they wish to acquire. This could mean directing particular attention to those items they find particularly challenging or that are not particularly salient in the input they receive.

Connections Between Usage-Based SLA and DDL

The core tenets of usage-based approaches to language and language acquisition provide strong support for data-driven learning as a pedagogical approach. DDL is founded on the Firthian linguistic tradition (Firth, 1957; Sinclair, 1966, 1991), listed by Kemmer and Barlow (2000) as an important early usage-based tradition. With its focus on context and collocation, it inspired the work of DDL pioneer Tim Johns, who worked at the University of Birmingham as a colleague of Firth student John Sinclair in the 1980s. In DDL, learners work with corpus-derived materials, often concordances, which display multiple instances of a word or phrase in context and allow the learner to explore usage patterns in an inductive and autonomous way.

Besides the heavy reliance on corpus data on both sides, usage-based SLA and DDL both highlight domain-general cognitive processes, such as generalisation and pattern identification, in language learning. According to Johns (1991), DDL is highly beneficial to learners as it helps them 'develop the ability to see patterning in the target language and to form generalisations to account for that patterning' (p. 2). In DDL settings, learners are required to apply the kinds of cognitive mechanisms that facilitate acquisition. Very much in line with how language learning within usage-based frameworks 'involves determining structure from usage' (Ellis et al., 2016, p. 23), DDL exposes learners to often substantial amounts of rich, authentic usage data from which they can derive regularities. It hence appears to be a suitable method for language learning from a usage-based perspective. As we know from usage-based SLA research, input matters and high input frequencies facilitate entrenchment. Concordances, used in DDL, provide the learner with repeated instances of the language item of interest.

As they carry out DDL activities, learners are encouraged to actively identify recurring sequences of words and phrases which, through their combinations, create meanings. As another central attribute of DDL, this focus on pattern detection and meaningful item discovery in concordance exploration is also very much in line with one of the tenets of usage-based SLA discussed above, namely, the importance of meaningful input in the acquisition process. DDL allows learners to engage with natural language data in the form of sets of concordance lines or even longer extracts from corpora. While certainly beneficial from a usage-based SLA perspective, this type of engagement is of course not as rich and meaningful as exposure to real communicative situations. This problem could be remedied at least partially if learners were given access to video-text-aligned corpora that provide links between transcripts and videos of lectures or interactions and make it easy to go directly from a concordance example to a video clip which captures additional aspects of spoken interaction, such as intonation and gestures. An example of such a resource is the TED Corpus Search Engine (TCSE; https://yohasebe.com/tcse/; Hasebe, 2015). Tools like the TCSE could be used to provide even more meaningful input in DDL settings. An additional connection between usage-based SLA and DDL lies in their mutual acceptance of the inseparability of lexis and grammar. In usage-based approaches, constructions transcend any traditional boundaries between the two; in DDL, lexis and grammar are explored simultaneously under the umbrella of 'lexicogrammar' (Boulton, 2017; Flowerdew, 2015). Concordances provide easy access to lexicogrammatical patterns (and to exemplars of constructions), especially when the context of the search term is sorted alphabetically. Through the exploration of concordances, learners discover in a serendipitous fashion how strongly lexis and grammar are linked and are exposed to typical exemplars—which, as we know from usage-based SLA research, promotes learning. DDL therefore appears to be a promising pedagogical intervention.

A final aspect of usage-based SLA which provides support for the use of DDL is its observation that attention controls language acquisition. In DDL, we use concordance displays that are designed to draw the learner's attention to the

search term and to the context words around it. DDL can hence be described as an excellent method to help learners notice relevant aspects of lexicogrammar by focusing their attention on a specific language feature, perhaps one that they find particularly difficult. As Flowerdew (2015, p. 31) points out, 'the inductive approach usually associated with DDL is underpinned by 'noticing' of rules and patterns from the concordance data' (see also Boulton, 2017). Like language models in usage-based linguistics, learning from corpora and retrieving answers to language questions from concordances is inductive and 'bottom-up'.

Conclusion

As the discussion in the previous section has shown, DDL appears to be a suitable method to support learners in the process of acquiring a new language as it is highly compatible with a usage-based SLA perspective. There are such clear connections between usage-based SLA and DDL (see also O'Keeffe, 2021, p. 45), and DDL practices are so aligned with the major tenets of usage-based approaches to language and language acquisition that it could even be argued that DDL constitutes a practical application of such approaches. It is likely that the benefits of DDL and its success in promoting L2 acquisition are derived from the approach's close alignment with usage-based acquisition principles. These DDL benefits have been illustrated in numerous publications (see Pérez-Paredes, 2022; Boulton & Cobb, 2017, for relevant overviews).

To further capitalise on this alignment, future work in DDL could focus more on the creation of materials that are guided by usage-based SLA principles. For example, we know from usage-based SLA research that input enhancements may facilitate the fast learning of new constructions (e.g. Goldberg & Casenhiser, 2008). This known sensitivity of learners to input frequencies could be made good use of as we prepare corpus-based teaching materials which contain what Madlener (2015) calls 'structured input floods'. In such input floods, learner input is enhanced or flooded 'with task-essential exemplars of the target structure in a principled way' (Madlener, 2015, p. 92). Since, via concordance displays, DDL exposes learners to multiple instances of target language items in context, it already incorporates a form of input flooding, but it could take additional cues from the SLA literature on the specific types of input enhancements that are ideal for learning to take place (see also O'Keeffe, 2023).

A stronger engagement with research in usage-based SLA could benefit DDL researchers and practitioners and help them gain a better understanding of how L2 learners acquire language. As suggested by Pérez-Paredes (2022), an improved understanding of popular DDL topics, such as collocation or formulaic language, could be brought about by theoretical framing. Perhaps even more importantly, a stronger theoretical grounding and visibility in the field of SLA might help DDL as a proven-to-be-effective teaching approach move more into the L2 education mainstream. This may allow us to support more L2 learners in achieving their language acquisition goals.

Cross-References

- ▶ Concordancers for DDL
- ▶ Corpus-Informed Materials in Language Teaching
- ▶ First-Language Acquisition and DDL
- ▶ Future Challenges and Opportunities for DDL
- ▶ Input Enhancement and Input Flood in DDL
- ▶ The Scope of Data-Driven Learning
- ▶ Tim Johns and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*(2), 383–411.

Boulton, A. (2017). Data-driven learning and language pedagogy. In S. Thorne & S. May (Eds.), *Language,*

education and technology. Encyclopedia of language and education (pp. 181–192). Springer.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). Routledge.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Bybee, J., & Beckner, C. (2010). Usage-based theory. In H. Narrog & B. Heine (Eds.), *Handbook of linguistic analysis* (pp. 827–855). Oxford University Press.

Cadierno, T., & Eskildsen, S. W. (Eds.). (2015). *Usage-based perspectives on second language learning*. De Gruyter.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188.

Ellis, N. C. (2019). Essentials of a theory of language cognition. *The Modern Language Journal, 103*, 39–60.

Ellis, N. C., & Ferreira-Junior, F. (2009). Constructions and their acquisition. Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics, 7*(1), 187–220.

Ellis, N. C., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In B. VanPatten, G. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (pp. 63–82). Routledge.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Wiley.

Eskildsen, S. W., & Cadierno, T. (2015). Advancing usage-based approaches to L2 studies. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 1–15). De Gruyter.

Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins.

Goldberg, A. E. (2006). *Constructions at work. The nature of generalization in language*. Oxford University Press.

Goldberg, A. E., & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 197–215). Routledge.

Hasebe, Y. (2015). Design and implementation of an online corpus of presentation transcripts of TED talks. *Procedia – Social and Behavioral Sciences, 198*, 174–182.

Hoey, M. P. (2005). *Lexical priming. A new theory of words and language*. Routledge.

Hunston, S., & Francis, G. (2000). *Pattern grammar. A corpus-driven approach to the lexical grammar of English*. John Benjamins.

Johns, T. F. (1991). Should you be persuaded – Two samples of data-driven learning materials. In T. F. Johns & P. King (Eds.), *Classroom Concordancing*. *ELR Journal, 4*, 1–16.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. vii–xxviii). CSLI Publications.

Langacker, R. W. (1987). *Foundations of cognitive grammar. Vol. I: Theoretical prerequisites*. Stanford University Press.

Langacker, R. W. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 127–161). John Benjamins.

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–64). CSLI Publications.

Langacker, R. W. (2008). Cognitive grammar as a basis for language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 66–88). Routledge.

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition, 8*, 346–368.

Madlener, K. (2015). *Frequency effects in instructed second language acquisition*. De Gruyter.

O'Keeffe, A. (2021). Data-driven learning, theories of learning and second language acquisition: In search of intersections. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 35–55). John Benjamins.

O'Keeffe, A. (2023). A theoretical rationale for the importance of patterning in language acquisition and the implications for data-driven learning. *Nordic Journal of English Studies, 22*(1), 16–41.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *CALL, 35*(1–2), 36–61.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*(1), 140–162.

Römer, U. (2024). Usage-based approaches to second language acquisition vis-à-vis data-driven learning. *TESOL Quarterly, 58*(3), 1235–1245.

Römer, U., & Berger, C. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition, 41*(5), 1089–1110.

Sinclair, J. M. H. (1966). Beginning the study of lexis. In C. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 410–430). Longman.

Sinclair, J. M. H. (1991). *Corpus concordance collocation*. Oxford University Press.

Sinclair, J. M. H. (2004). *Trust the text. Language, corpus and discourse*. Routledge.

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Harvard University Press.

Tyler, A. E., & Ortega, L. (2018). Usage-inspired L2 instruction: An emergent, researched pedagogy. In A. E. Tyler, L. Ortega, M. Uno, & H. I. Park (Eds.), *Usage-inspired L2 instruction: Researched pedagogy* (pp. 3–26). John Benjamins.

Tyler, A. E., Ortega, L., Uno, M., & Park, H. I. (Eds.). (2018). *Usage-inspired L2 instruction: Researched pedagogy.* John Benjamins.

L

Learner Autonomy and Data-Driven Learning

Maggie Charles
University of Oxford, Oxford, UK

Abstract

Learner autonomy has generally been considered to play a central role in data-driven learning (DDL), as it enables students to take responsibility throughout the learning process. This implies that learners provide input at all stages, from initial decisions as to what and how to learn, through implementation of their chosen strategies, to evaluation of the outcomes. The social context of a DDL intervention plays an important role in developing individual autonomy through teacher input and mediation alongside interactive collaboration with other learners. In class-based DDL, teachers often have overall responsibility for macro-level decisions affecting the whole class, while learners take decisions on a micro-level that affect their own learning. Although autonomy is often claimed as a positive outcome of DDL, there is little research that sets out to test this assumption or attempts to quantify the effect of DDL on the achievement of learner autonomy. Encouraging developments include improved tools to aid autonomous learning, especially for academic writing, and the spread of open educational resources.

Keywords

Learner autonomy and data-driven learning · Autonomous data-driven learning · Independent data-driven learning · Self-regulated data-driven learning · Autonomous corpus-based learning · Autonomous corpus use · Corpora and autonomous language learning · Self-directed corpus-based learning

Introduction

Attention to learner autonomy in language learning can be traced back to the work of Holec (1979, p. 3), who defined autonomy as "the ability to take charge of one's own learning". This definition was later refined by Benson (2013, p. 839) as "the capacity to control important aspects of one's language learning", while Little (2022, p. 64) further explained that autonomy is "a teaching/learning dynamic in which learners plan, implement, monitor and evaluate their own learning". Thus, the fully autonomous learner takes responsibility for the whole of the learning process. Holec set out two prerequisites for the achievement of learner autonomy: first students should know how to take decisions, and second they should have the possibility of exercising autonomy. These two conditions reflect a duality within the notion of autonomy: the first implies that autonomy is an individual learner characteristic, while the second suggests that the social nature of the learning context also exercises an important

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_55-1

function. This in turn indicates that some measure of institutional and teacher autonomy is necessary for learners to exercise their autonomy (Yi, 2017). As Godwin-Jones (2019, p. 8) points out, however, the two aspects of the learner and the social context of learning, should not be seen as dichotomous; rather they exist in a relationship of "mutual dependence and interaction".

The individual and the social-contextual aspects of autonomy play out to varying degrees in applications of data-driven learning (DDL). This term originated with Johns, who viewed the learner as "a research worker whose learning needs to be driven by access to linguistic data" (1991, p. 2). Later, DDL was broadly defined as "using the tools and techniques of corpus linguistics for pedagogical purposes" (Gilquin & Granger, 2010, p. 359). In Johns' work, learners decided on issues to examine, used inductive strategies to interrogate the corpus data, and formulated their own conclusions, an approach which assigns learners' autonomy a central role within the learning process. However, the social context of the classroom also played an important part, since the teacher acted as facilitator of students' learning, selecting and providing the data on which they worked, while students reached their conclusions through a process of collaboration and interaction with their classmates. Thus, Johns' practice of DDL fostered learner autonomy through social processes that supported individual development.

Learner Autonomy and DDL Approaches

Numerous reports of DDL are based on accounts of its application in the classroom and are characterised by the use of group interaction and teacher input to underpin learner autonomy at the individual level. This approach is illustrated by Carter and McCarthy's (1995, p. 155) three-part DDL procedure: *Illustration* (students examine the data)—*Interaction* (students discuss findings)—*Induction* (students formulate a rule), while the optional addition of an *Intervention* stage highlights the potential for teacher input (Flowerdew, 2009, p. 407). DDL approaches can be situated on a cline from *hard* to *soft* DDL (Gabrielatos, 2005), with minimal teacher input at the harder end of the spectrum where students are often engaged in inductive tasks in which they formulate rules from the corpus data. By contrast, softer approaches tend to rely on deductive strategies, with students seeking corpus evidence for given rules and here there is greater potential for teacher mediation. However, the greater the teacher's role, the less scope there is for learner autonomy; thus, it is the harder approaches that are more likely to align more closely with autonomous learning.

Occupying a middle ground between hard and soft DDL, a "guided inductive approach" has been proposed, which combines inductive and deductive procedures to provide a level of pedagogical scaffolding that is appropriate to the individual teaching/learning context (Johansson, 2009, p. 42). While the scope for learner autonomy is somewhat curtailed by such scaffolding, it is likely that particularly at lower levels and with less mature learners, some measure of teacher mediation will be necessary (e.g. various chapters in Crosthwaite, 2020). Even for advanced-level students, Bernardini (2002) argues that self-directed learning should take place within a structured environment, and DDL courses often start with guided tasks, moving on later to more autonomous work. For example, learners of Italian were seen as apprentices and trained over time to adopt an "observe and borrow chunks mentality" (Kennedy & Miceli, 2017, p. 93), in which they were encouraged not only to scrutinise specific language items, but to collect useful chunks for re-purposing in their own writing. Learner autonomy in DDL can thus be seen as a process of gradual development, supported and facilitated by the teacher in an interactive classroom setting.

Learner Autonomy and DDL in the Classroom

DDL is not specifically mentioned by Reinders (2010), but his framework for the adoption of autonomy in the classroom offers a useful means of assessing the areas in which DDL facilitates learners' autonomy. The framework consists of

eight stages as follows: (1) *Identifying needs*; (2) *Setting goals*; (3) *Planning learning*; (4) *Selecting resources*; (5) *Selecting learning strategies*; (6) *Practice*; (7) *Monitoring progress*; (8) *Assessment and revision* (Reinders, 2010, p. 46). Reinders argues that learners need to be actively involved in all stages of the process. However, in many DDL classes, responsibility for the learning process is shared between teachers and students, with teachers taking decisions at the macro-level for the whole class, while students exercise autonomy at the micro-level on decisions affecting their own learning (Charles, 2023). Accounts of this shared approach are seen, for example, in studies on learners' use of DDL to correct their own texts; teachers may carry out Reinders' stages 1–4, while students revise their texts using self-chosen learning strategies, as evidenced in studies with learners of French (O'Sullivan & Chambers, 2006) and English (Bridle, 2019).

Opportunities for autonomous learning may also be built into DDL courses at an earlier stage by having students construct their own do-it-yourself (DIY) corpora from material relevant to their work. This enables learners to take responsibility for creating personalised resources as shown for English (Charles, 2014; Lee & Swales, 2006) and French (Tyne, 2009). Although directed by the teacher, this practice entails needs identification, goal setting and planning by the student. At the practice stage, aspects of learner autonomy may be incorporated by asking students to work on self-chosen projects or presentations (Geluso & Yamaguchi, 2014), or by the use of homework tasks, as seen in work for learning pragmatic routines (Bardovi-Harlig et al., 2017) and Italian phraseology (Forti, 2023). However, although such approaches show evidence of learner autonomy, it is exercised at the individual micro-level; there is little evidence of macro-level involvement, for example in setting class goals or planning and revising courses, although students may suggest improvements, e.g. to research writing workshops (Chen & Flowerdew, 2018).

If learner autonomy were a prime goal of DDL (Forti, 2023) rather than a by-product of the approach, we would expect it to be closely monitored and assessed following Reinders' (2010) stages 7 and 8, potentially leading to revised and improved future courses. Instead, the evaluation of DDL has primarily focused on whether a given linguistic feature has been learnt or on student attitudes to the approach. In particular, students' intentions to use DDL autonomously in future have been widely reported including for learning German collocations (Vyatkina, 2016) and for English reading comprehension (Curado Fuentes, 2015). However, few studies attempt to quantify how far DDL has been instrumental in achieving learner autonomy. Notable exceptions are Lin (2016), who found greater self-efficacy in students using DDL for English grammar learning in comparison to those who used traditional methods, and Mizumoto et al. (2015), who developed a psychometric scale to measure students' perceived benefits (including autonomy) of using DDL.

Learner Autonomy and DDL Outside the Classroom

After some initial training in corpus use, learners can access many resources autonomously to carry out DDL according to their individual language needs, and the improvement of learners' academic written work in English has been a particular focus of research. Self-report data on out-of-class corpus use by post-graduates is given by Karpenko-Seccombe (2018), who found a high level of take-up by 30 students (93%) immediately after a corpus course. However, such autonomous use is likely to diminish over time, as shown by Charles and Hadley (2022), whose data for a similar population of over 200 students show a drop from 87% immediate post-course use to 62% a year later. Tracking data provide a more reliable measure of actual corpus use and are provided by Crosthwaite et al. (2019, p. 18), who found autonomous use by a "large number" of their 327 students. Such figures support the contention that autonomous DDL is a viable option for such students. Several DDL tools have also been developed specifically for learners' autonomous use. Cotos (2016) describes how students learn genre conventions through

revising their texts according to suggestions given by the Research Writing Tutor tool. Mizumoto et al.'s (2017) tool for writing research articles was piloted with eight users who reported increased autonomy in open-ended comments on the experience. ColloCaid, which helps learners with collocation suggestions for incorporation in research writing, is another tool rated highly by users for promoting autonomy (Frankenberg-Garcia et al., 2022).

Recently, Meunier (2023) has argued for a wider view of DDL, pointing out the wealth of digital resources that can be used both in-class and outside that offer multiple opportunities for autonomous learning (see also Godwin-Jones, 2019). However, although it may be easier to secure students' input at the initial stages of planning and selection of resources, their monitoring and evaluation of outcomes may be harder to achieve. Other highly promising approaches still to be fully explored include mobile-assisted technology (Pérez-Paredes et al., 2019) and the use of open educational resources such as that devised by Vyatkina (2024) for German.

Summary

The development of learner autonomy has long been considered a key benefit of data-driven learning. Thus, with access to corpora of the target language, students can take control of their learning by acting as researchers, choosing the corpus (or creating their own), deciding what linguistic features they need to learn, selecting suitable strategies and resources and carrying out their own investigations. However, several researchers have also stressed the important role of the teacher and the social context in facilitating the development of autonomous corpus use (Carter & McCarthy, 1995; Flowerdew, 2009). Given that most accounts of DDL are based on classroom interventions, the use of teacher mediation and collaborative work with fellow students are widespread. The teacher generally has overall control of macro-level decisions affecting the whole class, while the learner exercises autonomy at the micro-level over decisions that address their individual needs. Although learner autonomy may be somewhat curtailed as a result, this approach does allow a measure of autonomy to be practised at several stages of the learning process. Research has shown not only how opportunities for autonomy can be offered in class-based corpus work, but also that the resulting autonomy is highly appreciated by learners (Vyatkina, 2016). Outside class, an increasing number of tools for autonomous data-driven learning are being developed, especially for helping students with academic writing, and future promising avenues include the use of open educational resources and mobile-assisted learning.

Cross-References

- ▶ DIY Corpora and DDL
- ▶ Future Challenges and Opportunities for DDL
- ▶ Inductive and Deductive Approaches to DDL
- ▶ Language for Academic Purposes and DDL
- ▶ Learner Attitudes, Styles, Strategies and Behaviours in DDL
- ▶ Learning Outcomes from DDL
- ▶ MALL and DDL
- ▶ Tim Johns and Data-Driven Learning
- ▶ Tools for DDL
- ▶ Writing Aids for DDL

References

Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines. *Language Learning & Technology, 21*(3), 76–103.

Benson, P. (2013). Learner autonomy. *TESOL Quarterly, 47*(4), 839–843. https://doi.org/10.1002/tesq.134

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Rodopi. https://doi.org/10.1163/9789004334236_015

Bridle, M. (2019). Learner use of a corpus as a reference tool in error correction: Factors influencing consultation and success. *Journal of English for Academic Purposes, 37*, 52–69. https://doi.org/10.1016/j.jeap.2018.11.003

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–158. https://doi.org/10.1093/applin/16.2.141

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for*

Specific Purposes, 35, 30–40. https://doi.org/10.1016/j.esp.2013.11.004

Charles, M. (2023). Corpora and autonomous language learning. In R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 406–419). Routledge. https://doi.org/10.4324/9781003002901-32

Charles, M., & Hadley, G. (2022). Autonomous corpus use by graduate students: A long-term trend study (2009–2017). *Journal of English for Academic Purposes, 56*, 101095. https://doi.org/10.1016/j.jeap.2022.101095

Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes, 50*, 97–112. https://doi.org/10.1016/j.esp.2017.11.004

Cotos, E. (2016). Computer-assisted research writing in the disciplines. In S. Crossley & D. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 198–210). Routledge. https://doi.org/10.4324/9781315647500-15

Crosthwaite, P. (Ed.). (2020). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge. https://doi.org/10.4324/9780429425899.

Crosthwaite, P., Wong, L., & Cheung, J. (2019). Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. *ReCALL, 31*(3), 255–2275. https://doi.org/10.1017/S0958344019000077

Curado Fuentes, A. (2015). Exploiting keywords in a DDL approach to the comprehension of news texts by lower-level students. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 177–198). John Benjamins. https://doi.org/10.1075/scl.69.09cur

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy. *International Journal of Corpus Linguistics, 14*(3), 393–417. https://doi.org/10.1075/ijcl.14.3.05flo

Forti, L. (2023). *Corpus use in Italian language pedagogy: Exploring the effects of data-driven learning*. Routledge. https://doi.org/10.4324/9781003137320.

Frankenberg-Garcia, A., Pinto, P., Bocorny, A., & Sarmento, S. (2022). Corpus-aided EAP writing workshops to support international scholarly publication. *Applied Corpus Linguistics, 2*(3), 100029. https://doi.org/10.1016/j.acorp.2022.100029

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ, 8*(4), 1–35.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*(2), 225–242. https://doi.org/10.1017/S0958344014000044

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge. https://doi.org/10.4324/9780203856949-26

Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language Learning & Technology, 23*(1), 8–25.

Holec, H. (1979). *Autonomy and foreign language learning*. Pergamon Press.

Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 34–44). John Benjamins. https://doi.org/10.1075/scl.33.05joh

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal, 4*, 1–16.

Karpenko-Seccombe, T. (2018). Practical concordancing for upper-intermediate and advanced academic writing: Ready-to-use teaching and learning materials. *Journal of English for Academic Purposes, 36*, 135–141. https://doi.org/10.1016/j.jeap.2018.10.001

Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning, 30*(1–2), 91–114. https://doi.org/10.1080/09588221.2016.1264427

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75. https://doi.org/10.1016/j.esp.2005.02.010

Lin, M. (2016). Effects of corpus-aided language learning in the EFL grammar classroom: A case study of students' learning attitudes and teachers' perceptions in Taiwan. *TESOL Quarterly, 50*(4), 871–893. https://doi.org/10.1002/tesq.250

Little, D. (2022). Language learner autonomy: Rethinking language teaching. *Language Teaching, 55*(1), 64–73. https://doi.org/10.1017/S0261444820000488

Meunier, F. (2023). Revamping DDL: Affordances of digital technology. In R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 344–360). Routledge. https://doi.org/10.4324/9781003002901-28

Mizumoto, A., Chujo, K., & Yokota, K. (2015). Development of a scale to measure learners' perceived preferences and benefits of data-driven learning. *ReCALL, 28*(2), 227–246. https://doi.org/10.1017/S0958344015000208

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the bundle-move connection approach to the development of an online writing support tool for research articles. *Language Learning, 67*(4), 885–921. https://doi.org/10.1111/lang.12250

O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing, 15*(1), 49–68. https://doi.org/10.1016/j.jslw.2006.01.002

Pérez-Paredes, P., Ordoñana Guillamón, C., Van De Vyver, J., Meurice, A., Aguado Jiménez, P., Conole, G., & Sánchez Hernández, P. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System, 84*, 145–159. https://doi.org/10.1016/j.system.2019.06.009

Reinders, H. (2010). Towards a classroom pedagogy for learner autonomy: A framework of independent language learning skills. *Australian Journal of Teacher Education, 35*(5), 39–55. https://doi.org/10.14221/ajte.2010v35n5.4

Tyne, H. (2009). Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL, 31*, 91–111.

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159–179.

Vyatkina, N. (2024). *Corpus applications in language teaching and research: The case of data-driven learning of German*. Routledge. https://doi.org/10.4324/9781003223627.

Yi, W. (2017). When teacher autonomy meets management autonomy to enhance learner autonomy. *Chinese Journal of Applied Linguistics, 40*(4), 392–409. https://doi.org/10.1515/cjal-2017-0023

T

Techniques for Beginner-Level Data-Driven Learning

Jihua Dong [1] and Louisa Buckingham [2]
[1]Shandong University, Shandong, China
[2]School of Cultures, Languages and Linguistics, University of Auckland, Auckland, New Zealand

Abstract

Data-driven learning (DDL) is an instructional approach that uses corpora and corpus analytic tools to learn about the target language. This entry introduces the core techniques of DDL within the context of Computer-Assisted Language Learning (CALL), particularly targeting learners with little DDL experience. This entry introduces the core techniques of data-driven learning (DDL), its relevant theories and characteristics, commonly used corpora and software, and offers practical examples to illustrate how this approach can be implemented with learners. Through a brief discussion of a range of topics that include hands-on and hands-off DDL, teacher-led and student-led DDL activities, and the selection of general or self-built corpora, we show that the DDL approach can support the development of language skills in both the classroom environment and in autonomous language learning contexts. We show that the DDL approach can both respond to learners' momentary language-related queries and contribute to longer-term language learning and teaching objectives. By possessing even rudimentary DDL skills, language learners and teachers can independently access and process ever-expanding resources for language analysis.

Keywords

Data-driven learning · CALL · Inductive/deductive learning · Discovery learning · Educational technology · Corpora · Corpus linguistics

Introduction

DDL is an instructional approach to language learning that draws on the principles of corpus linguistics, which is broadly defined as "the use of the tools and techniques of corpus linguistics for second language (L2) learning" (Boulton & Cobb, 2017, p. 352).

According to Johns (1991, p. 2), a pioneer in the field, the DDL approach transforms the learner into "a research worker whose learning needs to be driven by access to linguistic data". In the DDL approach, learners analyse corpora (or corpus-derived language data) to identify patterns of language use (whether grammatical or lexical). This approach to learning accords with the principles of constructivism (Cobb, 2024) in that learners take responsibility for their own learning by actively participating in discovering language features. Previous research has demonstrated that

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_56-1

DDL not only enhances the effectiveness of the teaching and learning of lexicogrammatical features (Lee & Lin, 2019) but it also favours higher-order skills such as learner autonomy and lifelong learning (Cobb & Boulton, 2015).

DDL has been shown to have a positive impact on learning outcomes in specific areas such as vocabulary acquisition (e.g. Sun & Hu, 2023), collocations (e.g. Saeedakhtar et al., 2020; Vyatkina, 2016a), academic writing (see reviews in Chen & Flowerdew, 2018), speaking (e.g. Hirschi & Kang, 2024), and comprehension (e.g. Zare & Aqajani Delavar, 2023). DDL is also believed to support the development of learner autonomy through (guided) explorations of corpus data (Boulton & Vyatkina, 2021). With some adjustment to the tasks used and consideration to the corpus employed, DDL can be suitable for both adult learners (e.g. tertiary students or professionals) and young (i.e. school-age) learners. The following features are central to DDL:

- DDL enables learners and teachers to access **a wealth of language data** in corpora, which they can explore to identify patterns and derive regularities.
- DDL is **discovery- and evidence-based**. In DDL activities, learners' exploration of corpus data has a research element. Accompanying tasks require learners to engage with the data; they encourage learners to observe and generalize about language features, and allow them to test their understandings.
- DDL emphasizes learners' **active participation** in their own learning process. As corpora and corpus analysis tools are available online, the analytical approaches practised in the classroom can be transferred to independent learning contexts.

Previous research has supported the efficacy of employing DDL for language learning, as detailed in previous meta-analyses (e.g. Boulton & Cobb, 2017), bibliometric analyses (e.g. Dong et al., 2023), and other systematic reviews (e.g. Pérez-Paredes, 2022). However, although previous studies have emphasized the positive reported results of DDL on language learning, learner autonomy, awareness, and classroom engagement, and the lack of integration of DDL into mainstream English as a second (or foreign) language (ESL/EFL) courses has also been observed (Crosthwaite, 2024; Römer, 2023; Zare & Aqajani Delavar, 2023).

Core Considerations: Corpora and Corpus Analysis Tools

The choice of corpus is an important initial consideration when designing DDL activities. Well-known corpora that have been extensively used for teaching, learning, and research purposes over the years are listed in Table 1. The right-hand column lists examples of DDL research that used the respective corpus. An important consideration for EFL teachers is the corpus type (i.e. general or specialized corpora; expert or learner corpora). General corpora, such as the British National Corpus (BNC; Aston & Burnard, 1998) and the Corpus of Contemporary American English (COCA; Davies, 2009), usually contain a large, structured collection of texts (often including both written and spoken language), which provide comprehensive coverage of different (language-specific) text types, disciplines and genres. Due to the large size (often millions or even billions of words) and the requirement for texts to reflect an advanced level of language proficiency (usually educated "native-speaker" level), this type of corpus is considered a "credible" authentic resource for EFL/ESL teachers (Boulton, 2017). For instance, the BNC (100 million) roughly represents British English of the 1990s, while COCA (over one billion words) represents contemporary American English. These "general" corpora typically contain collections of texts from different domains of use, such as academia, news media, or fiction, and these sub-corpora can be analysed independently in studies on specialized language use. General corpora can be used to teach vocabulary (by utilizing the frequency information, collocations, clusters and concordances) and grammar (by presenting corpus examples with specific grammatical features), and to identify collocational and colligational information (by using the

Techniques for Beginner-Level Data-Driven Learning, Table 1 Common corpora used in DDL research

Corpus	Type	Size (words)	Focus/Description	Suitable for learners of	Relevant research
British National Corpus (BNC)	General	100 million	British English in the 1990s; various registers; written and spoken texts	Intermediate to advanced learners	Şahin Kızıl (2023), Liu and Gablasova (2023)
Brown University Standard Corpus of Present-Day American English (Brown Corpus)	General	One million	American English; from 1961 and revised and amplified in 1979	Intermediate to advanced learners	Gaskell and Cobb (2004), Saeedakhtar et al. (2020)
Collins COBUILD corpus	General	4.5 billion	Contemporary English; Texts from the Birmingham Collection of English texts	Beginner to advanced learners	Boulton (2011), Quinn (2015)
Corpus of Contemporary American English (COCA)	General	One billion	Contemporary American English; various registers; written and spoken texts	Intermediate to advanced learners	Satake (2020), Tsai (2019)
Michigan Corpus of Upper-Level Student Papers (MICUSP)	Specialized; written	2.6 million	Academic writing by university students in the USA; various disciplines	Advanced learners	Crosthwaite (2017), Sun and Hu (2023)
British Academic Written English (BAWE) Corpus	Specialized; written	6.5 million	British academic writing by university students; various disciplines	Advanced learners	Lay and Yavuz (2020), Crosthwaite and Steeples (2022)
Michigan Corpus of Academic Spoken English (MICASE)	Specialized; spoken	1.8 million	Transcriptions of academic spoken English; various genres (e.g. lectures, seminars, consultations)	Intermediate to advanced learners	Swales and Feak (2012)
Penguin Readers	Specialized; written	80 thousand (all levels added)	Fiction; historical and contemporary	Beginner to advanced learners	Allan (2009)

collocation functions or by sorting concordances). In recent years, corpus analysis tools integrated with online corpora have been made available; these include CorpusMate (Crosthwaite & Baisa, 2024), YouGlish (Barhen, 2019), BNCweb (Hoffmann & Evert, 2006), and LexTutor (Cobb, 2002), among others. The website www.english-corpora.org presents numerous large corpora (specialized and general, contemporary and historic), which are available to for use both online and offline.

Small tailor-made (often self-built) corpora may be used when learners have specialized disciplinary or genre-related needs, as this allows targeted exposure to specific language features, rhetorical structures and writing styles. Based on learners' needs or the goals of teaching, corpora can be built to represent certain registers, genres, disciplines, and language varieties. The underlying rationale is that students then become familiar with a very specific style of language use relevant to their learning priorities. Small corpora should still have an adequate number of texts or tokens to enable sufficient breadth of exposure and a meaningful analysis of textual and linguistic features.

Another consideration for language teachers concerns the (target language) proficiency level of the writers (or speakers) of the texts contained in

the corpus. Here the distinction is made between expert and learner corpora. An "expert corpus" contains texts authored by experts in a certain genre or discipline and who employ written (or spoken) English at the level of an educated native speaker. For example, with the learning goal of English for research publication purposes, expert texts would most typically be published research articles and book chapters (two different genres) which have undergone peer-review. Many expert corpora used in previous studies on disciplinary-specific language features were self-built by researchers, and expert corpora used in classroom teaching can be created by either instructors or students. Learner corpora, defined as "searchable collections of the written or spoken production of language learners" (Crosthwaite, 2012, p. 164), can provide both teachers and learners with typical L2 features (e.g. language use inappropriate for a specific genre or discipline, non-standard language use, or errors). This is typically undertaken by contrasting learner writing to expert or native-speaker writing. Such features include, for example, the underuse of passive voice by Chinese learners of English (Xiao, 2007), the misuse of V-N collocations by German learners of English (Nesselhauf, 2005) and temporal morphology by Vietnamese and Somali learners of Norwegian (Helland Gujord, 2017). Learner corpora that are intended to support language teaching and learning are usually created prior to the start of a course, but they can also be created mid-course using texts that students have produced to date (see Staples et al., 2023). Such corpora provide opportunities for a "situated" or localized approach to lesson content and teaching materials (Götz & Granger, 2024). An example of how sample sentences from a learner corpus can be integrated into a learning activity appears in Exercise 3.

The corpus-building process may involve both teachers and learners. The involvement of learners can help foster a sense of ownership and agency in the learning process. In most cases, however, teachers independently create a learner corpus using texts from the previous courses; this ensures that the text type (or genre) and the text content (in terms of topics and level of language) are appropriate for a particular course. For the teacher, the process of building a learner corpus typically entails the following steps: (a) compiling texts produced by learners; (b) annotating the texts with desired features such as non-standard language use or errors, or alternatively, the successful use of specific language (or discourse) features relevant to a particular course curriculum; (c) annotating the texts with relevant metadata (e.g. student L1 background, student level, text genre, assignment type); (d) organizing the texts into an electronically searchable database; (e) creating a user guide to explain to students how to search the corpus (or different sub-corpora). The files can be analysed with software available online such as AntConc (Anthony, 2023) and LancsBox (Brezina et al., 2020).

Implementing DDL in Classrooms: Instructional Considerations

Teachers who are interested in using DDL in their courses have a number of options available to them, and decisions should be taken based on resource availability (e.g. time, staffing, equipment, and Internet connectivity) and teaching and learning objectives. We provide here a brief overview of issues to consider when implementing DDL approaches.

The first major decision when conducting DDL is whether learners explore corpus data using a computer (hands-on) or hardcopy materials (hands-off). This decision depends on factors such as the availability of hardware in classrooms, learners' age and language proficiency level, and teaching objectives. Beginner-level learners (and particularly young learners) usually respond better to DDL explorations using print resources (handouts or textbook-based exercises) (Boulton, 2010). This avoids the problem of presenting learners with content that is unsuited to their age or proficiency level, which can easily occur with hands-on corpus searches. The hands-off approach is also advantageous due to its (in-class) time-saving nature. Additionally, learners who do not ultimately intend to continue with corpus-based analyses in the future (e.g. language learning for travel, general interest, or vocational needs) may be better served by print materials.

The hands-on approach involves learners directly using concordancing software, whether online or downloaded onto a computer. This approach is considered useful in promoting learner autonomy, as learners are usually encouraged to discover linguistic patterns themselves (or with some measure of guidance) (Saeedakhtar et al., 2020). Generating and studying concordances is usually the first step in acquiring corpus exploration skills, and DDL beginners usually find this task to be manageable. The website Sketch Engine for Language Learning (SkELL, Kilgarriff et al., 2014) provides a very simple interface through which users can generate example sentences extracted from the in-built corpus (several different language options are available). This can be useful as an introduction to concordance lines and also collocations (through the function Word Sketch). Simple exercises can be created from these examples such as Exercise 1, which directs learners to distinguish between the different meanings of a polysemic word (*bold*), using examples from SkELL. Exercise 2 asks learners to consider the contexts in which two synonyms can be used. These types of exercise can be used in hands-on and hands-off learning contexts. While some learners (and teachers) may find the analysis of concordance lines sufficient, others may choose to investigate collocations or clusters, or they may be interested in comparing particular words or phrases in different corpora. As shown in Buckingham (2008), the lexicogrammatical focus of DDL activities will vary, depending on the language being learned, learners' proficiency level and their L1. The possible focus might include article or prepositional use, adjectival modification, nouns in plural/singular, tenses (e.g. simple past and present perfect), voice (passive or active), or collocational patterns.

Exercise 1

The adjective *bold* has two main meanings.

(a) To be courageous, or very confident;
(b) Something appears clearly and strongly and is easy to notice.

Look at the follow examples of the adjective *bold* used in a sentence. Which meaning is expressed in each sentence? For example, sentence 1 expresses meaning (b).

1. The pages had *bold* headlines and large lettering. *(b)*
2. The budget is remarkably *bold* on pensions.
3. This early attempt was *bold* and fundamentally sound.
4. *Bold* numbers represent the top 11 mentioned items.
5. The effort to slow climate change requires *bold* action.
6. The message is *bold* and honestly worded.
7. There was nothing shameful in making *bold* plans.
8. He was very *bold* and preferred large movements
9. Her solution was characteristically impulsive and *bold*.
10. *Bold* indicates players who are still active.

Exercise 2

In some contexts, the adjectives *bold* and *courageous* can be used interchangeably, as the meaning expressed is very similar, and both can be used to modify a particular noun. In other contexts, only one of these words would be suitable. Look at the examples with the adjective *courageous* below. In which of these sentences could *bold* be used without a noticeable change in meaning? For instance, in (1), *bold* would also be possible.

1. It was a *courageous* and truly democratic endeavor. (*Yes*)
2. He died after a *courageous* battle with cancer.
3. She was brave, determined and *courageous.*
4. The world is watching these *courageous* students.
5. He has made an extremely difficult and *courageous* decision.

In hands-off DDL, teachers manually select the concordance lines to be included in teaching materials, and may edit them slightly to ensure that they are better understood by the specific learner group. For example, the teacher may choose to substitute certain words that would cause undue difficulties. Concordance lines may

contain acronyms, terminology, or unusual words that are not suited to the lesson's learning objectives. For instance, the examples generated by SkELL for Exercise 1 included this sentence: *ICI is therefore taking a very bold leap*. If the teacher decided to include this sentence, the teacher might decide to substitute the acronym *ICI* for a noun such as *the bank*, or *the organisation*. Corpus examples may also contain errors (whether grammatical or orthographic); the teacher may decide to amend these before including them in lesson material. For example, the sentences returned in SkELL for *bold* include: *Key features are often presenting in bold text.* In this case, the teacher may amend the verb to *presented*. Alternatively, if the error corresponds to a language point the students can be expected to know, the teacher could also ask learners to identify the example containing an error themselves.

Hands-off DDL has been viewed as providing a more restrictive learning experience, as learners do not have the opportunity to explore on their own. Nevertheless, queries using a general corpus (usually) deliver the learner with more examples than they are able to process, and confront them with language that surpasses their proficiency level. This lack of control over the language that learners are exposed to can mean much of the language generated from learners' corpus explorations is not available as input, as the learner is unable to process it.

DDL activities can be conducted in teacher-led or student-led environments, although typically the difference between teacher and student-led activities is one of degree. According to Mukherjee (2006, p. 12), "DDL activities can be plotted on a cline of learner autonomy, ranging from instructor-led and relatively closed concordance-based to entirely learner-centred corpus-browsing projects". For teacher-led DDL activities, the search queries have a higher degree of teacher control, either through spoken instructions or the provision of worksheets. Learner-led activities typically involve discovery tasks that rely on learners' noticing and intuition, such as "browsing large and varied text collections in open-ended, exploratory ways" (Bernardini, 2004, p. 22). Student-led activities typically require a higher level of corpus competence.

A further distinction in the delivery of DDL concerns inductive and deductive approaches. Inductive DDL, which draws from the constructivist view of learning, encourages the learner to formulate (and perhaps also test) hypotheses related to the targeted linguistic features, based on the available evidence, usually in the form of concordance lines. In deductive DDL activities, learners are provided with explanatory information first (for instance, in the form of rules or typical patterns) and they subsequently test their understanding of these by completing exercises which require the analysis of relevant corpus data, typically concordance lines. Exercises 1 and 2 are examples of deductive DDL activities. The choice of approach largely depends on the goal of the class, available time, learners' proficiency and preferred learning styles. For instance, DDL activities that are primarily teacher-led and deductive may be more suitable for low-level learners. Inductive DDL activities may require more time to complete, as time should be allowed for learners to sift through corpus data (usually concordance lines) and build hunches, or formulate or test hypotheses about language use, whether independently or with guidance.

DDL in Practice: A Sample Classroom Activity

Exercise 3 demonstrates an example of a (hands-off) guided inductive DDL activity for students with little DDL experience. The activity is intended to illustrate the specific steps in conducting DDL activities for in-class use. It can be adapted for use as a hands-on activity.

Exercise 3
Teaching lexicogrammatical patterns in academic writing with paper-based materials

- Step 1: The teacher identifies the lexicogrammatical pattern(s) to be learned (e.g., *one of the most*, *a number of*, *found*, etc.).

- Step 2: The teacher searches for the target form (s) in the corpus, selects appropriate concordance lines, and copies them onto a handout (see the example in Fig. 1; these results for *found* were extracted from BAWE corpus using Sketch Engine).
- Step 3: The teacher guides students in exploring the usage patterns of *found* in academic writing. For example, the following instructions can be given to students:
  1. Find all instances of *found* that appear in the passive voice on the handout.
  2. What grammatical structure(s) often follows *found* in the passive voice in these examples?
  3. Do we know who/what the subject (i.e., the agent) is in these examples of passive voice? Explain the reason.
  4. Identify examples of *found* used in the active voice with the meaning of *identify* or *discover*. Choose three of these examples that could be used in the passive voice without much loss or change of meaning. Explain why you chose these examples.
  5. In what section(s) of a study (e.g., an article or a thesis) would these sentences typically appear?
- Step 4: The teacher asks students in pairs to summarise what they know about *found* in the passive voice in two sentences, using one (shortened) corpus example to illustrate their explanation.
- Step 5: The teacher asks several pairs to share their explanation with the class.
- Step 6: The teacher asks students to look at the five examples of the use of *found* in the passive voice in academic texts written by learners (see below). In each example, there are errors. The teacher asks students to reformulate the sentences (individually or in pairs) so that they are correct.

Left context	KWIC	Right context
Yorkshire, Scotland and Humberside were the worst affected, and Ashton et al	found	that a young man's chance of being unemployed in Sunderland was one in thr
led maintenance of the status quo.</s> <s>A further source of frustration was	found	in the fact that, despite repeated pledges for reform from the party leadership,
er, mitochondria only generate energy in aerobic conditions, while the latter is	found	in obligate or aerotolerant anaerobes, such as Trichomonas.</s> <s>Chloropla
s (15)<s>The results of such experimentation have been varied.</s> <s>It was	found	that if the experimental conditions were not correctly set, the holoclone-gener
the absence of opsonization, the Mip protein (macrophages invasion protein)	found	on the L. pneumophila outer membrane, potentiates the invasion of human m
even men and women attempting to keep within their official social positions	found	it an impossible task. I have looked at some of the factors that made this impo
ch should tax limited resources for parallel semantic processing.</s> <s>They	found	a significant effect of noise overall on the recall task but no significant differen
ncient times.</s> <s>'The Middle Ages reversed Augustus's boast that he had	found	Rome a city of brick and left it one of marble'.</s> <s>Materials from ancient r
ger an eating disorder.</s> <s>Lieberman, Gauvin, Bukowski and White (2001)	found	significant causes to be peer pressure to be thin, social reinforcement and bod
ts on caregivers offered in this essay (e.g. social role and gender), research has	found	that there are no significant differences in coping strategies between caregiver
s suggestibility to leading questions and post event information has also been	found	to increase (Dywan & Bowers, 1983).</s> <s>State theorists would claim t
ess is only an adequate strategy if the suspect is present.</s> <s>Wells (1993)	found	that when a suspect is removed from a set of distractors and not replaced, tho
facilitate their search even though they attended to them (as invalid cues were	found	to increase their reaction times).</s> <s>This coupled with the findings that in
extreme argument indicates that anticlericalism in Germany was probably not	found	on such a wide scale as other historians have claimed and that it varied region
ationship between the cell standard deviations and means of the raw data was	found	and thus a square-root transformation was applied.</s> <s>With the transforr
-looks with positive characteristics.</s> <s>Dion (1973; cited by Durkin, 1995)	found	that preschoolers predicted attractive peers to be better friends and frequently
ency and adult criminality.</s> <s>Amongst upper-class boys, popularity was	found	to be a predictor of delinquency (Roff et al, 1972).</s> <s>Additionally, militar
guity codes.</s> <s>Due to different programming techniques, databases are	found	in a few different file formats.</s> <s>Fortunately, most of them are interconv
cording to the assumption that the best information on gap placement will be	found	among the most similar sequence, once-a-gap rule is considered as a 'golden
lves performing a Fourier transform on the location of the maximal intensities	found	- from this the structure of the crystal and be inferred and thus the shape of th
e inner membrane is far less permeable and has similarities to the membranes	found	surrounding prokaryotes.</s> <s>This phenomenon can be explained by the e
s.</s> <s>Both of these have implications for health 6, 7.</s> <s>Studies have	found	apparent relationships between income and health, where low-income groups
on Coefficients<s>Two values for the attenuation coefficient of lead (Pb) were	found	:</s> <s>&</s> <s>Both values are in agreement with the theoretical valu
d, because he has become a marchant, is afraid of making mistakes and being	found	out of being a paysan - 'devant les gens qui parlaient bien il se taisait,...toujour
2 of VH, yet its composition and structure differ from the biantennary complex	found	in the CH2 domain, since it is of the high mannose type.</s> <s>The following
jested by Maltz and Borker (1982, as in Coates and Cameron 1988 124) can be	founded	and men have been found to favour public speech (Tannen, 1990: 76/7) does t
as women do in private conversations.</s> <s>The men in this transcript were	found	to use no minimal responses, whilst women used them 9 times.</s> <s>Men i
significant to this investigation more than men involved.</s> <s>Women were	found	to use more interruptions than men in their conversations, with men demonstr

Techniques for Beginner-Level Data-Driven Learning, Fig. 1 Concordance lines of *found* in BAWE

- Step 7. The teacher asks students to work in pairs and to formulate a brief explanation (in their L1 or L2) to help them understand how to use *found* in the passive voice. Different explanations may be needed for each sentence.
- Step 8. The teacher checks the corrected sentences and the explanations in class.

Below are some examples of *found* in the passive voice written by learners (from the ArabCC—Learner Corpus of English Essays, in Sketch Engine).

1. There are several Arab African dialects, for instance Algerian, Tunisian, and Moroccan; however, the latter is to be *found* the hardest one to understand.
2. According to they *found* that the amount of time spent editing photos for Instagram is considerable.
3. According to *found* that school students enjoyed using tablets more than print textbooks but learned better when they used print textbooks.
4. It was *found* in research that women or girls are getting less computer science degrees than they got in the past.
5. Furthermore, according to he "*found* that people who followed many celebrities on Instagram started experiencing a decline in life satisfaction".

Conclusion

DDL is an instructional approach to language learning that draws on the principles of corpus linguistics, and advocates greater attention to authentic, evidence-based language use. As corpus analysis tools have become easier to use, more accessible, and some even come equipped with an inbuilt corpus (for instance AntConc or Sketch Engine), DDL techniques can be easily integrated into everyday learning and teaching. This entry discussed core considerations for the integration of DDL in instructed learning contexts, provided research-informed recommendations, and presented practical activities to model how DDL can be integrated into a lesson. Areas of research that warrant further research include the use of DDL with learning languages other than English (see Boulton & Vyatkina, 2021; Buckingham, 2008), the implementation of DDL in resource-poor contexts, the integration of DDL approaches in materials development, and the use of DDL methods to explore discourse in the digital humanities, that is, disciplines other than applied linguistics and language teaching such as media studies, history or politics. In these disciplines, the principles of data-driven language learning can be applied to study how language is used to present and describe topics, individuals or social groups, events across different time periods, text types, or to investigate language use by different speakers/writers.

References

Allan, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal, 63*(1), 23–32. https://doi.org/10.1093/elt/ccn011

Anthony, L. (2023). *AntConc* (Version 4.2.4) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Aston, G., & Burnard, L. (1998). *The BNC handbook: Exploring the British national corpus with SARA*. Edinburgh University Press.

Barhen, D. (2019). YouGlish. *TESL-EJ*, *23*(2). https://www.tesl-ej.org/wordpress/issues/volume23/ej90/ej90m1

Bernardini, S. (2004). Corpora in the classroom. An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.12.05ber

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60*(3), 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2011). Bringing corpora to the masses: Free and easy tools for language learning. In N. Kubler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 69–95). Peter Lang.

Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching, 50*(4), 483–506. https://doi.org/10.1017/S0261444817000167

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. https://doi.org/10.10125/73450

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *#LancsBox v. 5.x.* [software]. http://corpora.lancs.ac.uk/lancsbox

Buckingham, L. (2008). Spanish verb support constructions from a learner perspective. *Estudios de Lingüística Inglesa Aplicada, 8*, 151–179.

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics, 23*(3), 335–369. https://doi.org/10.1075/ijcl.16130.che

Cobb, T. (2002). Web vocabprofile. http://www.lextutor.ca/vp

Cobb, T. (2024). Constructivism in language learning and applied linguistics. In *Reference module in social sciences*. Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00020-X

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478–497). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.027

Crosthwaite, P. (2012). Learner corpus linguistics in the EFL classroom. *PASAA Journal, 44*, 133–147. https://doi.org/10.58837/CHULA.PASAA.44.1.6

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning, 30*(6), 447–473. https://doi.org/10.1080/09588221.2017.1312462

Crosthwaite, P. (2024). *Corpora for language learning: Bridging the research-practice divide*. Routledge. https://doi.org/10.4324/9781003413301

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*. https://doi.org/10.1075/ijcl.23056.cro

Crosthwaite, P., & Steeples, B. (2022). Data-driven learning with younger learners: Exploring corpus-assisted development of the passive voice for science writing with female secondary school students. *Computer Assisted Language Learning*, advance access. https://doi.org/10.1080/09588221.2022.2068615

Davies, M. (2009). The 385+ million-word corpus of contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(2), 159–190.

Dong, J., Zhao, Y., & Buckingham, L. (2023). Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL, 35*(3), 339–355. https://doi.org/10.1017/S0958344022000222

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*, 301–319. https://doi.org/10.1016/j.system.2004.04.001

Götz, S., & Granger, S. (2024). Learner corpus research for pedagogical purposes: An overview and some research perspectives. *International Journal of Learner Corpus Research, 10*(1), 1–38. https://doi.org/10.1075/ijlcr.00039.got

Helland Gujord, A.-K. (2017). The "perfect candidate" for transfer: A discussion of L1 influence in L2 acquisition of tense-aspect morphology. In A. Golden, S. Jarvis, & K. Tenfjord (Eds.), *Crosslinguistic influence and distinctive patterns of language learning: Findings and insights from a learner corpus* (pp. 29–63). Multilingual Matters. https://doi.org/10.21832/9781783098774-005

Hirschi, K., & Kang, O. (2024). Data-driven learning for pronunciation: Perception and production of lexical stress and prominence in academic English. *TESOL Quarterly*, advance access. https://doi.org/10.1002/tesq.3302

Hoffmann, S., & Evert, S. (2006). BNCweb (CQP-edition): The marriage of two corpus tools. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 177–195). Peter Lang.

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria, 10*, 14–34.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography, 1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Lay, K. J., & Yavuz, M. A. (2020). Data-driven learning of academic lexical bundles below the C1 level. *Language Learning and Technology, 24*(3), 176–193. https://doi.org/10.10125/44741

Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System, 81*, 14–25. https://doi.org/10.1016/j.system.2018.12.011

Liu, T., & Gablasova, D. (2023). Data-driven learning of collocations by Chinese learners of English: A longitudinal perspective. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2023.2214605

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Peter Lang.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins. https://doi.org/10.1075/scl.14

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal, 69*(2), 165–177. https://doi.org/10.1093/elt/ccu062

Römer, U. (2023). Usage-based approaches to second language acquisition vis-à-vis data-driven learning. *TESOL Quarterly, 58*, 1235–1245. https://doi.org/10.1002/tesq.3278

Saeedakhtar, A., Bagerin, M., & Abdi, R. (2020). The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System, 91*, 102268. https://doi.org/10.1016/j.system.2020.102268

Şahin Kızıl, A. (2023). Data-driven learning: English as a foreign language writing and complexity, accuracy and fluency measures. *Journal of Computer Assisted Learning, 39*(4), 1382–1395. https://doi.org/10.1111/jcal.12807

Satake, Y. (2020). How error types affect the accuracy of L2 error correction with corpus use. *Journal of Second Language Writing, 50*, 100757. https://doi.org/10.1016/j.jslw.2020.100757

Staples, S., Dang, A., & Wang, H. (2023). Learner corpora in corpus-informed instruction: Moving toward an asset- and genre-based model. *TESOL Quarterly,* advance access. https://doi.org/10.1002/tesq.3293

Sun, X., & Hu, G. (2023). Direct and indirect data-driven learning: An experimental study of hedging in an EFL writing class. *Language Teaching Research, 27*(3), 660–688. https://doi.org/10.1177/1362168820954459

Swales, J., & Feak, C. B. (2012). *Academic writing for graduate students* (3rd ed.). Michigan University Press. https://doi.org/10.3998/mpub.2173936

Tsai, K. J. (2019). Corpora and dictionaries as learning aids: Inductive versus deductive approaches to constructing vocabulary knowledge. *Computer Assisted Language Learning, 32*(8), 805–826. https://doi.org/10.1080/09588221.2018.1527366

Vyatkina, N. (2016a). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159–179. https://doi.org/10.10125/44487

Xiao, R. (2007). What can SLA learn from contrastive corpus linguistics: The case of passive constructions in Chinese learner English. *Indonesian Journal of English Language Teaching, 3*(1), 1–19.

Zare, J., & Aqajani Delavar, K. (2023). A data-driven learning focus on form approach to academic English lecture comprehension. *Applied Linguistics, 44*(3), 485–504. https://doi.org/10.1093/applin/amac021

Input Enhancement and Input Flood in Data-Driven Learning

Reka R. Jablonkai
University of Bath, Bath, UK

Abstract

Effective language instruction aims to convert input into intake through input flooding and input enhancement. Input flooding involves exposing learners to numerous instances of a target language feature, while input enhancement uses strategies to make language features more salient and noticeable. This entry discusses techniques of input flood and input enhancement that are often used in DDL. These techniques—such as using concordance lines, textual highlights, frequency information, comparison across languages and varieties, collocations, visualisation, auditory highlights, focusing on pragmatic functions, and corrective feedback—are showcased with examples from research and practical collections of DDL activities.

Keywords

Input flood · Input enhancement · Inductive learning · Discovery learning · Noticing · Focus-on-form (FonF) · Form-focused instruction (FFI) · Usage-based instruction · Input enhancement techniques

Introduction

Research has established that using corpora in tandem with a data-driven approach to teaching and learning (DDL) benefits language learners (Boulton & Vyatkina, 2021; Boulton & Cobb, 2017). DDL has often been linked with a focus-on-form approach (Long, 1991) and a usage-based model of language acquisition (Ellis et al., 2016; Römer, 2023). The primary links between usage-based second-language acquisition (SLA) and DDL have been proposed as follows (e.g. O'Keeffe, 2023; Römer, 2023).

Bottom-up approach: DDL is typically linked with inductive learning, supported by Schmidt's Noticing Hypothesis (1990), which posits that attention and noticing of patterns govern language learning. This 'bottom-up' approach emphasises the importance of learners actively identifying and understanding language patterns which aligns with usage-based SLA research.

Easy access to multiple examples: Concordance lines used in DDL provide multiple examples of use of words or phrases in context and highlight these by centring them in the middle of the page or screen. This way learners can be exposed to rich, naturally occurring, authentic language use.

Relevance of frequent patterns: DDL emphasises the interconnectedness of vocabulary and grammar, challenging this traditional distinction. Research in corpus linguistics underscores this inseparability, revealing that words and word

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_57-1

forms often align with specific grammatical constructions. Therefore, corpus-based approaches often focus on exploring and identifying lexicogrammatical patterns in language use. With the help of concordance lines and frequency-based searches, DDL is well-suited to demonstrate such frequent and typical patterns and provides a useful way to present learners with multiple examples of these. Usage-based SLA research revealed that such repeated exposure to typical, salient examples enhances learning.

Input Enhancement, Input Flood and Intake in DDL

Input in second language acquisition (SLA) encompasses the language environment or primary language data that learners encounter during language instruction (e.g. Schwartz, 1993). This input is crucial as it forms the basis for language learning, but not all input is necessarily internalised by learners. The distinction between input and intake, proposed by Corder (1967) highlights that while input refers to what learners are exposed to, intake refers to what is actually processed and internalised. Effective language instruction involves facilitating the transformation of input into intake through pedagogic techniques and strategies such as input flooding and input enhancement. Input flooding is a focus-on-form intervention where learners are inundated with numerous examples of a target language feature. This method aims to enhance learners' exposure to and awareness of specific language forms, thereby promoting intake creation from the input provided (Hernández, 2011). Szudarski and Carter (2016), for example, provided learners with stories which included the target collocation several times in different grammatical forms without any highlights. Input enhancement, proposed by Sharwood Smith (1993), however, involves pedagogic strategies designed to make relevant features of input more noticeable to learners. Examples of input enhancement techniques include consciousness-raising tasks, textual enhancement (e.g. highlighting, underlining, and annotation), contrastive analysis, aural enhancement (e.g. auditory highlighting), explicit instruction, and corrective feedback (Simard, 2018).

Techniques of Input Flooding and Input Enhancement in DDL

All DDL inherently involves some degree of input flood and enhancement as it highlights language features for learners, provides multiple instances of these language features, and emphasises noticing and consciousness raising. While DDL studies may not explicitly concentrate on input flood or enhancement (Dong et al., 2023; Lusta et al., 2023), they employ various techniques related to these concepts. These techniques are discussed and showcased here with examples from research and recently published collections of practical resources, DDL activities and lesson plans (e.g. Le Foll, 2021; Pinto et al., 2023; Viana, 2023).

Concordance Lines

Probably the most frequently used technique for input flooding and enhancement in DDL is to provide learners with concordance lines of instances of specific language features, for example, words, word forms, multi-word units, and grammatical structures. Concordance lines, also called KWIC (key-word-in-context) lines, present the language feature in focus in the middle of the line in their original textual context (See an example in Fig. 1). Jones (2023), for example, demonstrates how concordance lines can provide learners with examples of grammatical structures, such as *going to*, by flooding them with examples like *What are you going to do?* and *What are you going to be doing?* These examples highlight tense variations and illustrate how these patterns function in different textual contexts. In another example, Dünnhaupt and Haase (2021, Fig. 10) use concordance lines from the BNC using the COCA platform (Davies, 2020a) corpus tool to help learners explore the grammatical structures of the word *tell*.

Textual Highlights

Typical forms of textual enhancement include typographical modification of texts by, for example, using bold, italics, capitals, underlining, colour coding, and changing the font or the size of letters (Simard, 2018). All of these forms can be used in DDL when using printouts of concordance lines or other types of corpus data. An example is Johns'

(1996–1997) Kibbitzer 34, where the different verbs (*take part in/carry out/engage/undertake*) that can be used together with the word *activity* are highlighted in blue in the relevant concordance lines on the handout. When applying corpus consultation, that is, learners explore corpora directly with the help of corpus tools, the options are limited by the functionalities of the specific corpus tool. Some tools use default colour coding in concordance lines for part-of-speech as shown in Fig. 2 (e.g. CorpusMate—Crosthwaite & Baisa, 2024; COCA—Davies, 2008–2019, 2020b), others use colours for sort highlighting and colour code the words occurring at different distances from the word in focus (e.g. AntConc—Anthony, 2024a). Another useful way to highlight relevant language features and patterns is to sort concordance lines alphabetically or by frequency (Römer, 2023). This way learners can explore likely good examples and discover typical patterns which often results in converting such input into intake.

Teachers might also find corpus tool functions such as highlighting words in specific frequency bands by colour coding in teaching materials useful. AntWordProfiler (Anthony, 2024b), for example, can be used with English word lists. An example using the COCA platform (Davies, 2020a) can be seen in Fig. 9 in Hörnschemeyer and Janzen (2021) which used a text on climate change. MultilingProfiler (Finlayson et al., 2022) is a similar tool for vocabulary profiling for German, French, and Spanish. The tool highlights cognates with English in blue and colour codes words that are not used elements of the selected word list in orange. See an example in German in Fig. 3.

Frequency Information

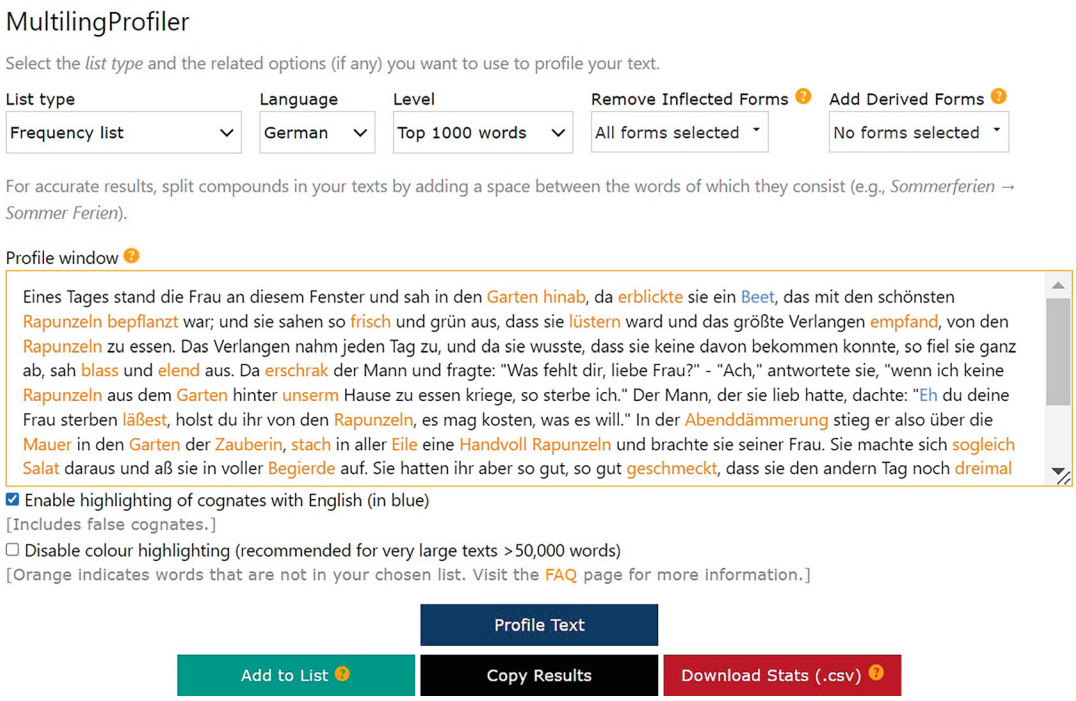
One specific form of input enhancement in DDL involves presenting corpus-based frequency information on the language feature in focus. This technique can be used to draw learners' attention to typical use of particular grammatical structures as well as individual words and expressions. For instance, Friginal (2018) provides an example DDL activity where learners are given the frequency of perfect and progressive aspects in British and American conversations and news texts as part of a lesson on tenses in English

what it's like slightly orange slightly orange tasty orange I **must** be half an hour in to cook mm rice is good
they had this like giant communal hot tub thing and like there **must** 've been about fifty odd people in there
I just yeah once oh my word that **must** well our housemate like played it continuously
done this what controller and button combination **must** be used? I've done this as well I know this one ah
oh no they weren't there for that it **must** 've been the one the year before oh it was the curry
he seems to be the man who the Traceman Group it so people **must** work for him but he's not exposed to something else
oh well I don't see how you can do that by accident do you? **must** well most people must have really complicated programming
it just felt really really nice mm just erm like you you **must** be a northern European at heart I must be I just really

Input Enhancement and Input Flood in Data-Driven Learning, Fig. 1 Concordance lines demonstrating the use of the modal verb *must* in the BNC 2014 Spoken corpus

29 nly a few that have done so at a pace consistent with dealing with the **climate** crisis. So nuclear seems like a pretty good option, but there's this big 3x
30 :eserve Valle Nuevo (Reserva Científica de Valle Nuevo). Climate The **climate** of the province is a tropical climate, hot most of the year, but it is 3x
31 de Valle Nuevo). Climate The climate of the province is a tropical **climate** , hot most of the year, but it is cooler on the mountains. Municipalities 3x
32 Borgen from The Socialist Left Party. Oslo has a humid continental **climate** (Dfb in the Koeppen climate classification). History The Saga of 2x
33 Cook collapses. 2019 Russia joins the Paris Climate Agreement. 2019 **Climate** change activist Greta Thunberg speaks at a United Nations Climate 2x
34 he Feed-In Tariff Handbook, Earthscan. Benjamin K. Sovacool (2010). **Climate** Change and Energy Security: A Global Overview of Technology and P 2x
35 change. Political leaders ought to give a "funk" because the impact of **climate** change and extreme poverty comes right to our shore. Now, global 2x
36 that are expected to affect the climate More scientists believe future **climate** change can be predicted. Climate prediction has improved but not 2x
37 Youth Climate Network. She took part in the "2021 United Nations **Climate** Change Conference" (COP26) in Glasgow. She also represented the 2x
38 27, 1924 – April 6, 2020) was an Austrian-born American physicist and **climate** change denier. He was an emeritus professor of environmental scienc 2x
39 . You see, the world is getting warmer, and climate change is a reality. **Climate** change is affecting my country. Our glaciers are melting, causing 2x

Input Enhancement and Input Flood in Data-Driven Learning, Fig. 2 Concordance lines of *climate* by CorpusMate using colour codes for part-of-speech: noun-red, verb-blue, adverb-green, adjective-pink, article-orange



Input Enhancement and Input Flood in Data-Driven Learning, Fig. 3 Vocabulary profile of a short extract of the Grimm tale Rapunzel in German by Multilingual profiler

(p. 259). In another example, Johns' (1996–1997) Kibbitzer 6 gives frequency information on the usage of the word *data* with singular or plural verb forms in different journals.

Comparison Across Languages and Varieties

Contrastive analysis is another technique to enhance input (Simard, 2018). This technique draws learners' attention to similarities and differences between languages, often between learners' L1 and L2. In DDL, parallel corpora provide an obvious way to provide examples in two or more languages. Le Foll (2023), for instance, provides a simple way to use the free multilingual interface KonText (Čermák & Rosen, 2012) to create worksheets for available language pairs.

DDL also offers straightforward ways to compare the language use and the use of specific language features in different varieties such as specific genres, registers, and disciplines. In this case, specialised corpora or sub-corpora of larger corpora can be used to highlight specific differences to enhance language input. Karpenko-Seccombe (2024), for example, provides an overview of the frequency distribution of the words *problem*, *hypothesis* and *argument* in the 16 discipline-specific sub-corpora in MICUSP (pp. 125–126) to demonstrate disciplinary differences for teaching written argumentation.

The usage of specific words can also be compared in the same, often large general corpora. Johns' (1996–1997) Kibbitzer 11 provides an example of comparing the use of *predict* and *forecast* to help learners understand the difference between the meaning of these words.

Show the Company Words Keep (Collocations)

An often-cited benefit of DDL is that it provides a simple way to present words in their immediate and wider context which helps learners to develop their productive vocabulary. Some corpus tools provide an overview of collocates grouped according to the grammatical relation they form. An example in SkELL (Basia & Suchomel, 2014) shows the collocates of *be* grouped into the categories of subject

of *be*, object of *be*, phrasal, and phrasal with object (Garcia, 2023, Fig. 1). Another example created by Hörnschemeyer and Janzen (2021, Fig. 14) shows the nouns, verbs, adjectives, and adverbs co-occurring with the noun *climate* created with COCA (Davies, 2020b) for their teaching materials on climate change.

In addition to grammatical relations, pedagogic collocational profiles (Jablonkai, 2024, p. 96) can also give information about the semantic fields a word is frequently associated with. As can be seen in an example in Fig. 4, learners studying the noun *criterion* can see its collocates in the *verb + criterion* construction such as *fulfil*, *meet*, *satisfy*, and *conform to*. To highlight the semantic fields the collocates are associated with, verbs are grouped into semantic categories such as MEET (e.g. *fulfil, satisfy, conform to*), SET (e.g. *agree, establish*), or LIST (e.g. *specify, give*). This categorisation can aid in understanding nuances of meaning and learning similar, potentially synonymous words.

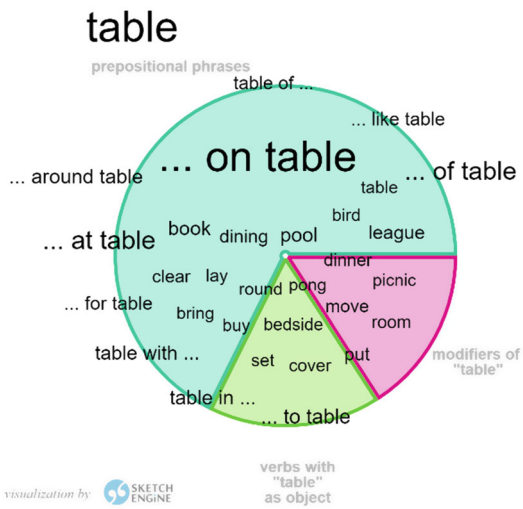
Visualisation

Visualisation of corpus data is an effective way to enhance input as it can draw learners' attention to different aspects of the language data and adds variety to the primarily written textual input. Corpus tools offer a range of ways to visualise language data from bar charts or histograms to more complex, interactive collocation maps and word clouds (Friginal, 2018).

Busman (2021, Fig. 12), for example, presents the use of *freak out* in different registers and time periods using COCA (Davies, 2008–2019). Learners can easily see that this verb is most frequently used in the TV sub-corpus and least frequently in academic registers. Similarly, Sperling and Winkelmann (2021) show the appearance of the word *Brexit* by showing bar charts representing the frequency of this word in different years from 2010 to 2021 (Fig. 12). Figure 8 in the same material gives an example of visualising the frequency list of the collocates of *glamping*. Fuchs and Liebelt (2021, Fig. 29) demonstrate how the visualisation function of Sketch Engine (Kilgarriff et al., 2014) can be used to show collocates of the word *bus*. This function allows the selection of specific grammar relations to visualise, thus providing useful ways to enhance input by only presenting the relevant relations. An example of the visualisation of selected collocates of the noun *table* by Sketch Engine can be seen in Fig. 5.

CRITERION	
grammatical relation	semantic field
verb + CRITERION	1 MEET *conform to, fulfil, meet, satisfy* subject to the necessary **criteria** being **met** 2 SET *set, agree, establish, lay* Have any additional eligibility **criteria** been **set**? 3 RESPECT *respect* Banks worldwide aim at **respecting** sustainability **criteria** in investing policies. 4 LIST *list, specify, give* Paragraph 3 **specifies** the **criteria** that may be used by the courts. 5 APPLY *apply* They stress the need to **apply** quality **criteria** for health-related websites.

Input Enhancement and Input Flood in Data-Driven Learning, Fig. 4 Example of pedagogic collocational profile. (Adapted from Jablonkai, 2024, p. 95)



Input Enhancement and Input Flood in Data-Driven Learning, Fig. 5 Visualisation of selected collocates of the noun *table* using SketchEngine

Liu and Gablasova (2023) offer another example of input flooding and enhancement by using #LancsBox to develop learners' collocation knowledge. They flooded learners with numerous examples of collocates for selected words. The GraphColl tool in #LancsBox (Brezina et al., 2015) visualises collocations by showing the strength of the association between words (indicated by the length of the line) and the frequency of collocates (indicated by the colour density of blue circles). This visualisation serves as input enhancement by directing learners' attention to word relationships, helping them actively analyse and identify collocational patterns (Brezina, 2018; Liu, 2021).

Auditory Highlights

Multimodal corpora can be used to provide auditory enhancement in DDL (Coccetta, 2022). Hirschi and Kang (2024), for example, demonstrate DDL and input flooding in teaching pronunciation. Their online intervention included lessons on lexical stress, where learners listened to examples of correct stress patterns from a corpus and categorised various audio samples. This approach combined input flooding with input enhancement techniques, such as audio examples of target words and self-recording options for learners to assess their own pronunciation accuracy. These structured activities helped learners internalise correct stress patterns and improve their pronunciation skills through repeated exposure and practice.

Focus on Form and Pragmatic Function

DDL proves pivotal in enhancing pragmatic competence and awareness of discourse markers and their usage through corpus-based methodologies. Bardovi-Harlig and Mossman (2022), for example, demonstrate DDL's effectiveness with MICASE (Michigan Corpus of Academic Spoken English) to search for instances of expressions such as *I mean* and *apologize* for heightening awareness of nuanced pragmatic details often overlooked in traditional language instruction. Extracting metapragmatic insights from corpus data can enhance input by illustrating contextual pragmatic functions of words and expressions.

Corrective Feedback

DDL can also enhance input by providing corrective feedback through examples that illustrate the correct usage of specific language features. This can be done by consulting corpora in the classroom or by using a hyperlink to selected concordance lines that include instances of correct usage of the target language structure. A recent study that compared different types of feedback found that concordance feedback provided on an online quiz of advanced English grammar featured among the most preferred forms of feedback by learners (Kılıçkaya, 2022).

The BAWE Quicklinks project (Vincent et al., 2021) is an excellent example of how to create and present such feedback to learners. This resource includes a directory of ready-made hyperlinks to concordance lines through Sketch Engine that can be used as feedback to address common errors and issues in academic writing. In addition to re-usable links, there are guidelines to how to create such links with relevant concordance lines.

Summary

The primary links between usage-based SLA and DDL include a bottom-up approach, emphasising

inductive learning and the importance of noticing language patterns. DDL offers a straightforward way to provide access to multiple examples through concordance lines and such repeated exposure to typical examples enhances language input and promotes learning. DDL inherently integrates elements of input flooding and enhancement. The often-used specific input enhancement techniques in DDL include employing concordance lines, applying textual highlights, presenting frequency data, conducting cross-linguistic and variety comparisons, utilising visualisation, incorporating auditory enhancement, demonstrating collocations, and exploring the relationship between form and pragmatic function, as well as providing corrective feedback. These techniques can be employed to heighten learners' awareness and make specific aspects of language input more salient, thereby facilitating the learning process.

Cross-References

- ▶ Concordancers for Data-Driven Learning
- ▶ Genre Awareness and Data-Driven Learning
- ▶ Hands-On Versus Hands-Off Methods in Data-Driven Learning
- ▶ Inductive and Deductive Approaches to Data-Driven Learning
- ▶ Integrating Corpora and Data-Driven Learning into Lesson Planning
- ▶ Lexicogrammar and Grammar and Data-Driven Learning
- ▶ Listening and Data-Driven Learning
- ▶ Multimodal Tools and Corpora for Data-Driven Learning
- ▶ Pedagogical Corpora and Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Pragmatics and Data-Driven Learning
- ▶ Usage-Based Approaches and Data-Driven Learning
- ▶ Vocabulary and Data-Driven Learning

References

Anthony, L. (2024a). *AntConc (Version 4.3.0)* [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2024b). *AntWordProfiler (Version 2.2.1)* [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Bardovi-Harlig, K., & Mossman, S. (2022). Corpora in instructed second language pragmatics. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 71–88). Routledge.

Basia, V., & Suchomel, V. (2014). SkELL—Web interface for English language learning. In *Eighth workshop on recent advances in Slavonic natural language processing* (pp. 63–70). Tribun EU. ISSN 2336-4289.

BAWE Quicklinks. (n.d.). https://bawequicklinks.coventry.domains/creating-quicklinks/

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89.

Brezina, V. (2018). Collocation graphs and networks: Selected applications. In P. Cantos-Gómez & M. Almela-Sánchez (Eds.), *Lexical collocation analysis: Advances and applications* (pp. 59–83). Springer.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139–173. https://doi.org/10.1075/ijcl.20.2.01bre. https://lancsbox.lancs.ac.uk/

Busman, K. (2021). How to do data-driven learning (DDL) with secondary school pupils to improve their writing. In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3ed ed.). Open Educational Resource. https://pressbooks.pub/elenlefoll. CC-BY-NC 4.0. https://doi.org/10.5281/zenodo.4992504. https://pressbooks.pub/elenlefoll/chapter/busmann/

Čermák, F., & Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics, 17*(3), 411–427.

Coccetta, F. (2022). Multimodal corpora and concordancing in DDL. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 361–376). Routledge.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161–170.

CorpusMate. (https://corpusmate.com/info)

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*. https://doi.org/10.1075/ijcl.23056.cro

Davies, M. (2008–2019). The Corpus of Contemporary American English (COCA): 600 million words, 1990–present. https://www.english-corpora.org/coca/

Davies, M. (2020a). English corpora. https://www.english-corpora.org/. BNC. https://www.english-corpora.org/bnc/

Davies, M. (2020b). The COCA corpus. https://www.english-corpora.org/coca/help/coca2020_overview.pdf

Dong, J., Zhao, Y., & Buckingham, L. (2023). Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL, 35*(3), 339–355. https://doi.org/10.1017/S0958344022000222

Dünnhaupt, G., & Haase, L. (2021). Differences in the grammatical structures of 'tell' and 'say'. In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. https://pressbooks.pub/elenlefoll. CC-BY-NC 4.0. https://doi.org/10.5281/zenodo.4992504. https://pressbooks.pub/elenlefoll/chapter/duennhaupt-haase/

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Wiley.

Finlayson, N., Marsden, E., & Anthony, L. (2022). MultilingProfiler (Version 3) [Computer software]. University of York. https://www.multilingprofiler.net/. Accessed 14 July 2024.

Friginal, E. (2018). *Corpus linguistics for English teachers*. Routledge.

Fuchs, M., & Liebelt, A. (2021). "How do I get there?": Travelling by various means of transport. In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. https://pressbooks.pub/elenlefoll. CC-BY-NC 4.0. https://doi.org/10.5281/zenodo.4992504. https://pressbooks.pub/elenlefoll/chapter/fuchs_liebelt/

Garcia, W. (2023). So many things in a day. In P. T. Pinto, P. Crosthwaite, C. T. de Carvalho, F. Spinelli, T. Serpa, W. Garcia, & A. O. Ottaiano (Eds.), *Using language data to learn about language: A teachers' guide to classroom corpus use*. University of Queensland. https://doi.org/10.14264/3bbe92d. https://uq.pressbooks.pub/using-language-data/chapter/lesson-description-2/#lesson2

Hernández, T. A. (2011). Re-examining the role of explicit instruction and input flood in the acquisition of Spanish discourse markers. *Language Teaching Research, 15*, 159–182.

Hirschi, K., & Kang, O. (2024). Data-driven learning for pronunciation: Perception and production of lexical stress and prominence in academic English. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3302

Hörnschemeyer, L., & Janzen, L. (2021). In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. https://pressbooks.pub/elenlefoll. CC-BY-NC 4.0. https://doi.org/10.5281/zenodo.4992504. https://pressbooks.pub/elenlefoll/chapter/janzen-hoernschemeyer/

Jablonkai, R. R. (2024). Transforming language teaching with corpora—Bridging research and practice. In P. Crosthwaite (Ed.), *Corpora for language learning* (pp. 93–111). Routledge.

Johns, T. (1996–1997). Kibbitzers. University of Birmingham. https://lexically.net/TimJohns/#gsc.tab=0

Jones, C. (2023). Corpus linguistics and grammar teaching. In K. Harrington & P. Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 119–138). Palgrave Macmillan.

Karpenko-Seccombe, T. (2024). Practical use of corpora in teaching argumentation. In P. Crosthwaite (Ed.), *Corpora for language learning* (pp. 111–130). Routledge.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, S. (2014). The sketch engine: Ten years on. *Lexicography, 1*, 7–36. http://www.sketchengine.eu

Kılıçkaya, F. (2022). Pre-service language teachers' online written corrective feedback preferences and timing of feedback in computer-supported L2 grammar instruction. *Computer Assisted Language Learning, 35*(1–2), 62–87. https://doi.org/10.1080/09588221.2019.1668811

KonText: Advanced and flexible corpus query interface. (n.d.). In N. Calzolari, et al. (Eds.), *Proceedings of the 12th conference on language re-sources and evaluation* (pp. 7003–7009). European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020. lrec-1.865.pdf. https://www.korpus.cz/kontext

#Lancsbox tool. (n.d.). https://lancsbox.lancs.ac.uk/

Le Foll, E. (Ed.). (2021). *Creating corpus-informed materials for the English as a Foreign Language classroom*. https://pressbooks.pub/elenlefoll/front-matter/how-to/

Le Foll, E. (2023). Raising awareness of first-language interference using parallel corpora of subtitles. In V. Viana (Ed.), *Teaching English with corpora* (pp. 49–54). Routledge.

Liu, T. (2021). Data-driven learning: Using #LancsBox in academic collocation learning. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Applications of corpora in language education* (pp. 177–206). John Benjamins.

Liu, T., & Gablasova, D. (2023). Data-driven learning of collocations by Chinese learners of English: A longitudinal perspective. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2023.2214605

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. DeBot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in crosscultural perspective* (pp. 39–52). John Benjamins.

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data-driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9* (12), e22731. https://doi.org/10.1016/j.heliyon.2023.e22731

MICASE Michigan Corpus of American Spoken English. (n.d.). https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple

O'Keeffe, A. (2023). A theoretical rationale for the importance of patterning in language acquisition and the implications for data-driven learning. *Nordic Journal of English Studies, 22*(1), 16–41.

Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. (Eds.). (2023). Using Language Data to Learn About Language: A Teachers' Guide to Classroom Corpus Use. *University of Queensland*. https://doi.org/10.14264/3bbe92d

Römer, U. (2023). Usage-based approaches to second language acquisition Vis-à-Vis data-driven learning. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3278

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*, 143–146.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*, 165–179.

Simard, D. (2018). *Input enhancement. The TESOL Encyclopedia of English Language Teaching*. Wiley.

SkELL. (n.d.). https://www.sketchengine.eu/skell/

Sperling, N., & Winkelmann, S. (2021). In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. https://pressbooks.pub/elenlefoll. CC-BY-NC 4.0. https://doi.org/10.5281/zenodo.4992504. https://pressbooks.pub/elenlefoll/chapter/sperling_winkelmann/

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245–265.

Viana, V. (2023). Teaching English with corpora. *A resource book. Routledge*.

Vincent, B., Nesi, H., & Quinn, D. (2021). Exploiting corpora to provide for academic writing: The BAWE Quicklinks project. In M. Charles & A. Frankenberg-Garcia (Eds.), *Corpora in ESP/EAP writing instruction: Preparation, exploitation, analysis* (pp. 13–31). Routledge.

Inductive and Deductive Approaches to Data-Driven Learning

Sonia Di Vito
University of Tuscia, Viterbo, Italy

Abstract

Data-driven learning (DDL) is a pedagogical approach that assumes the use of large amounts of authentic language data to facilitate foreign language teaching and learning. The approach involves students discovering linguistic phenomena through data analysis, acting as researchers and developing their ability to observe and infer. There are two main approaches: inductive and deductive DDL. Inductive DDL is based on the observation of data in order to reach the formulation and generalisation of a rule, whereas deductive DDL centres on the statement of a rule, which is then observed within a corpus. The inductive approach encourages student autonomy and develops their curiosity, acting on their motivation; the teacher is an 'instructor' (Smart, ReCALL 9:5–14, 1997, p. 186) who simply acts if necessary. The deductive approach, on the other hand, is more structured and guided by the teacher, who keeps their role as the holder of knowledge; the student observes the application of the rule stated in the materials provided by the teacher. The choice of one approach over the other depends on numerous factors including the educational context, learning objectives, students' skills and their learning styles.

Keywords

DDL · Deductive · Inductive · Learning styles · Advantages · Disadvantages · Concordances · Bottom-up · Top-down

Introduction

Data-driven learning (DDL) is a pedagogical approach that uses large amounts of authentic data as a primary resource to facilitate learning. The term DDL was first coined by Tim Johns (1990) revised in 1994 and 1991 in his seminal articles, as a synonym for 'classroom concordancing' (Johns, 1994, p. 296), that is, a teaching method that makes use of 'computer-generated concordances' (Johns & King, 1991, p. iii) in the classroom for the learning of foreign languages, which is an approach that entails the application of computers to language learning.

This approach was influenced by the development of computer science (Boulton, 2011) in the 1980s and 1990s. However, the first application of corpora to language teaching occurred as early as 1969 when Peter Roe began using LSP corpora during his courses at Aston University (McEnery & Wilson, 1997). In the 1980s, as pointed out by

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_58-1

Boulton (2011), other scholars in different universities used data retrieved from corpora and concordances during their courses, to teach, for example, the use of verbs or to give learners access to numerous examples of linguistic phenomena.

Access to linguistic data as the main source of information regarding the functioning of the language being learned entails a reflection on the role of the two 'protagonists' involved in learning, namely, the teacher and the learner. Johns (1991) describes the tasks of both: the learner is in charge of discovering the foreign language while the teacher has to create a learning context that fosters discovery by students, thus helping them to develop strategies for learning 'how to learn' (Johns, 1991, p. 1).

As McEnery and Wilson (1997) note, the use of corpora did not come as a surprise in an environment (that of the UK in the late 1990s) where the focus was to see teaching as mediating learning and not just as a way of imparting knowledge. Indeed, the use of corpora 'encourages the student to act as the producer of research, rather than its passive receptacle' (McEnery & Wilson, 1997, p. 6). In the early 1980s, Seliger (1983) studied the hypothesis that the workings of a language learning mind are similar to those of a scientific mind, so since then the language learner has been described as a 'research worker' (Johns, 1991, p. 3) or 'language detectives' (Johns, 1997, p. 101). This characteristic is considered by Boulton and Vyatkina (2021, p. 66) to be the 'prototypical core' of DDL-related activities; over the years, this primordial definition of DDL has been joined by a broader and more nuanced definition involving various ways of using corpora, corpus linguistics methods and concordances in foreign language teaching/learning activities.

Over the past 30 years, DDL activities, both inductive and deductive, have mainly been for university students as many researchers in the field have embraced this teaching practice (Boulton, 2020). In general, studies describing language learning experiences with corpora highlight the methodology used with students. Already Johns (1991, p. 4), in describing the activities he prepared for overseas students in the Spring Term of 1989, set out in detail what type of data he used, how long he had to spend collecting them, how long the activity with the students lasted and what questions he prepared on the delivered material for students. In 1999, Conrad gave an overview of the type of examples of language learning through corpus data presented by scholars and pointed out that they presented lexical or lexicogrammatical analyses (1999, p. 2). She argued that this type of examples may not make teachers adequately appreciate the potential of language learning through corpora. In her study, she highlighted how corpus analysis also allows one to focus on phenomena such as frequency, semantic categories and grammatical structures, emphasising how important register variation is in the study of language. She stated that teachers must not only know structural information about the language but must also know how native speakers use these structures and under what conditions. In 2004, Tsui argued that the relevance of corpus linguistics is underexplored in the training of language teachers, who were not at all aware of the potential of corpus linguistics in foreign language teaching. Indeed, she argued that it is necessary to contribute to teachers' linguistic awareness by proposing 'self-discovery' (Tsui, 2004, p. 59) paths so that they too can experience formulating rules from observations of linguistic data and can master grammar like a language researcher. Bennet, in 2010, provided further ideas and actual examples for using this methodology in the classroom in order to make the DDL approach accessible to teachers. For example, she provided a detailed description of an activity that can be proposed to discover the patterns of the articles *a/an* and *the* in TV programmes presenting the news: type of students (language level and general theme of the course attended), tools and procedures (Bennet, 2010, pp. 49–57). Unfortunately, the worlds of primary and secondary education and that of applied linguistics do not often interact, and research on DDL is rarely if ever included in training programmes for future teachers (Boulton, 2020). Furthermore, there is a tendency to think that young pupils do not have a sufficient level of

Inductive and Deductive Approaches to Data-Driven Learning, Table 1 Scholars' designations of the stages of a DDL activity

Stages of a typical DDL activity				
1.	Illustration (Carter & McCarthy, 1995; McEnery et al., 2006)	Identification (Johns, 1991)	Research task (Johns, 1997)	
2.	Interaction	Classification	Practice task(s)	
3.				Intervention (Flowerdew, 2009)
4.	Induction	Generalisation	Improvisation activities	

foreign language proficiency to allow the use of corpora and concordances in teaching. However, in his pioneering work, Crosthwaite (2020) collected contributions in which DDL experiments with young learners are described, highlighting how DDL, if properly implemented, can be an efficient and effective method even with younger learners.

Although the term 'data-driven learning' suggests an inductive approach in that one starts from the observation of data in order to identify the regularity of a linguistic phenomenon, many teaching practices using corpora actually follow a deductive approach, as highlighted by Flowerdew (2012), which starts from the teacher's enunciation of the rule, which is then later observed in concordances and eventually refined or corrected.

The Inductive Approach

As the term itself suggests, 'data-driven language' presupposes an inductive approach, that is, a bottom-up method that starts from observations of specific cases and draws general conclusions. According to this approach, the learner is like a 'researcher' (Johns, 1988, pp. 13–14; Gavioli, 2001, p. 108) of linguistic phenomena in medium or small corpora or like a 'traveller' (Bernardini, 2001, p. 220) in a large corpus. The inductive process concerns the observation of phenomena, the search for patterns or regularities and the formulation of hypotheses or theories that can explain these patterns. The inductive method is widely used in the empirical sciences, where observation and experimentation are fundamental. It consists of various phases that scholars have in turn given different names to (see Table 1) in which the learner is the protagonist of their own learning, to the point that the DDL is considered a learner-centred approach. In the first phase, students are exposed to data, that is, they are asked to observe authentic texts (collected in a corpus, which may be newspaper articles, transcripts of conversations, literary works, etc.). This phase can be supported by corpus analysis tools provided by concordance programmes to better observe the occurrences of keywords or to study the frequency of particular collocations. In the second phase, students share their observations and opinions on the so-called linguistic patterns, that is, the regularities of lexical, grammatical and syntactic uses that they noticed in the previous phase. We then move on to an optional phase in which the teacher can provide the students with suggestions to allow them to move on to the next phase, which is that of the formulation of hypotheses and of rules that explain the patterns observed. The hypotheses, after being discussed in class, also with the teacher's help, can be generalised and thus become rules concerning the functioning of the language. Finally, the last step in this discovery/research process coincides with the application of the formulated rules in new contexts, verifying their validity and effectiveness through practical exercises.

Inductive DDL has three main features: the first is the use of linguistic corpora to explore and analyse data, which therefore serve as a 'medium' through which learning takes place (McEnery & Wilson, 1997). Students can use concordance software to search for specific words or phrases and observe how these are used in different contexts. The second is autonomous discovery, as students are encouraged to formulate hypotheses and test them through data analysis. This discovery process helps develop critical thinking and problem-solving skills and allows students to actively engage in the learning process, which can increase their motivation and involvement (Boulton, 2010; O'Sullivan & Chambers, 2006; Hunston, 2002, among others). The third is contextualisation, as the use of authentic data allows students to see how language is used in the real world.

Inductive DDL may give rise to concerns related to some of its characteristics. The first is the question of whether students should be allowed to make their own observations and produce their own reasoning; this could lead them to feel that they are not up to the task assigned to them, thus generating a sense of frustration. As highlighted by Aston (1995, 1996), students may also arrive at wrong or approximate conclusions. Moreover, as suggested by Leech (1997, p. 5), the corpus does not give the learner the direction of learning, that is, to use the words of Aston (2001, p. 42), 'learners need to acquire criteria and methods which will enable them to choose and follow suitable learning paths'. Learners should then get used to abandon their first assumption and constantly interrogate the corpus in search of forms and meanings that might be even more appropriate with respect to the context and even better attested. Another difficulty that could arise is related to the use of concordance programmes that require a certain degree of technical knowledge (Boulton & Vyatkina, 2021): students would have to familiarise themselves with the new technologies for enough time to be able to use them confidently. This could be a valid alternative to, for example, the mode that Flowerdew (2012, p. 197) calls 'pen-and-paper based', that is, handing students the concordances on paper and leaving them free to make their own assumptions. This method has been proposed since the beginning of DDL practice in language classes. Johns himself (1991, p. 298) used computer-extracted concordances that were both printed on the paper and saved on the PC. Today, there are many ways by which concordances can be presented with authentic data, either through ppt slides or by using online tools that save the searches performed and learners need only to click on the link and see the search and its results (Vincent & Nesi, 2018). Pérez-Paredes et al. (2019) propose another way of learning through DDL, 'mobile data-driven language learning' (Pérez-Paredes et al., 2019, p. 145): language learning is based on the use of an app powered by free NPL tools. The authors state that the first results of the study suggest on the one hand a positive evaluation of the new learning mode and on the other hand the need to improve the types of tasks and especially the learners' need for specialised training. A further drawback concerns the time it takes on the one hand for teachers to prepare the materials on which students will work (from the collection of texts to the preparation of any materials with paper concordances) and, on the other hand, for students to learn how to use computer tools, to spend time observing concordances, to formulate hypotheses and rules and finally to apply them. A fourth disadvantage (Kennedy & Miceli, 2010) might be related to the students' need to know the foreign language in a very in-depth manner and have a high capacity for observation and inductive reasoning. One final objection is linked to teacher training in DDL and teachers' possible perception of a loss of power and control (Boulton, 2009, p. 88) as they would need to strip themselves of the role of knowledge holders and take on that of a guide or research organiser (Johns, 1991, p. 297) or facilitator (Bernardini, 2004, p. 16; Cresswell, 2007, p. 270) or coordinator of research (Bernardini, 2004, p. 16), who leaves students free to explore, observe and formulate hypotheses. Moreover, as stated in Aston (1997, p. 52), teachers do not know in advance what the results of students' research will be: 'One student told me that the best thing about the BNC was that she felt able to contradict her teacher'. To overcome some

of these possible impediments to the application of inductive DDL, gradual, 'gentle introduction' could be proposed (Boulton, 2009, p. 87) at different times of the new way of observing language (the inductive one), new materials to observe (corpora and concordance lines) and new technologies to use (concordancers).

The Deductive Approach

The deductive approach in DDL is based on the principle that a particular inference derives from more general assumptions, and it consists of a top-down method in which general rules are stated first and then observed through the examples in the corpus. This is a corpus-based approach as opposed to the corpus-driven one, that is, a methodology that makes use of the data contained in a corpus mainly to 'expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study' (Tognini-Bonelli, 2001, p. 65). This type of approach was first considered by Johns himself (Johns, 1986) and was first referred to as 'deductive DDL' by Cresswell (2007, p. 270) who defines it as an approach based on 'less autonomous activities in which student discovery is guided by exercises designed by the teacher'. During a deductive DDL session, the instructional approach of the three Ps (presentation, practice, production) is applied, which is widely known in the field of foreign language learning and consists of three phases: The initial phase, known as the *presentation* phase, is a teacher-centred stage, during which instructors introduce new elements of the language being studied through the use of corpus-informed material or from a reference grammar text; the teachers ask questions about the rules just presented to ensure that everything is clear. The second phase is called *practice* and involves structured activities, aimed at consolidating the language points just acquired, in exercises designed using sentences taken from the corpora. It is a learner-centred phase in which the learner is encouraged to use the language point in context. During the third phase, called *production*, the learner puts the linguistic phenomenon just studied into practice by carrying out short activities involving written, oral and receptive skills.

A study of some experiments conducted using this approach reveals two distinct ways of applying deductive DDL. For example, corpus-informed material can be prepared (as in Smart (2014) or Lee and Lin (2019)), that is, derived from concordance analysis, and specially crafted for the objective to be achieved, with which to present students with grammatical rules or linguistic phenomena. Cresswell (2007), on the other hand, presents activities in which students can choose a deductive-inductive path, which has as its starting point the description of the use of a given linguistic structure in a reference book; in a second phase, students will search the corpus for examples that correspond to that description and practise applying the studied rules through purposely prepared exercises.

The deductive approach is thus more structured and guided by the teacher who explicitly presents the linguistic phenomenon being taught. For this reason, we can outline three main features of this type of approach, which constitute its strengths. Firstly, it is, as anticipated, explicit teaching, in which linguistic rules and structures are clearly presented to students before they are exposed to linguistic data. Secondly, the analysis is not carried out independently by the individual student (or group of students), but the teacher's task is to guide the students and help them focus on specific aspects of the phenomenon being studied. A third feature of deductive DDL concerns rule checking, the latest aim of which is to verify and consolidate students' understanding of the rules through the study of authentic data and allow students to learn and then apply them. It could be argued that its main features may also be its weaknesses in that it does not stimulate student curiosity and thus undermines student motivation. Secondly, a reduced need to observe data due to the prior presentation of the rule means that students are exposed to the linguistic phenomenon fewer times and acquire a less in-depth understanding of the phenomenon (Lee & Lin, 2019). However, the findings of their study lead on the fact that 'deductive DDL is

equally effective and more efficient than inductive DDL for vocabulary acquisition' (Lee & Lin, 2019, p. 24), circumventing some of the disadvantages of inductive DDL.

Summary

Inductive and deductive DDL represent two poles of an educational continuum. While inductive DDL promotes autonomy and deep learning through discovery, deductive DDL offers clarity and structure, facilitating rapid and systematic acquisition of language knowledge. The choice between the two approaches often depends on the educational context, learning objectives and learners' skills. Preferring deductive to inductive DDL may also depend on the learning styles of students 'who prefer to accommodate rather than analyse' (Cresswell, 2007, p. 277).

In a gradual and guided learning path, considering the advantages and disadvantages of the two types of approaches as well as all the variables presented (learning styles, level of language proficiency, logistical aspects related to the availability of computers and concordance programmes, etc.), an integration of deductive and inductive DDL could be envisaged. Indeed, the deductive DDL could allow students to get in touch with concordances, and teacher-led activities could teach students what to observe in concordance lines and how to observe. Only after learners have reached a certain degree of familiarity with DDL-related tools and methods could they proceed with inductive DDL activities that presuppose more independent work in discovering the language to be learned. An integration of the two approaches, combining inductive and deductive elements, can provide a balanced and flexible pathway to language learning and can maximise the benefits of both approaches by giving students both the necessary guidance and the opportunity to develop autonomous and critical skills (Boulton, 2010). Furthermore, in a gradual and scaffolded learning pathway, deductive DDL can be considered as a first step towards inductive DDL.

Cross-References

- ▶ Integrating Corpora and Data-Driven Learning into Lesson Planning
- ▶ Learner Autonomy and Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Quantitative Designs in Data-Driven Learning Research
- ▶ Techniques for Beginner-Level Data-Driven Learning
- ▶ Tim Johns and Data-Driven Learning
- ▶ Young Learners and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson* (pp. 257–270). Oxford University Press.

Aston, G. (1996). The British National Corpus as a language learner resource. In S. Botley, J. Glass, A. McEnery, & A. Wilson (Eds.), *Proceedings of TALC 1996* (UCREL technical papers, Vol. 9, pp. 178–191).

Aston, G. (1997). Small and large corpora in language learning. In B. Lewandowska-Tomaszczyk & J. Melia (Eds.), *Practical applications in language corpora* (pp. 51–62). Lodz: Lodz University Press.

Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 7–45). CLUEB.

Bennet, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press. https://doi.org/10.3998/mpub.371534

Bernardini, S. (2001). 'Spoilt for choice': A learner explores general language corpora. In G. Aston (Ed.), *Learning with corpora* (pp. 220–249). CLUEB.

Bernardini, S. (2004). Corpora in the classroom. In J. M. H. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins.

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics, 35*(1), 81–106.

Boulton, A. (2010). Consultation de corpus et styles d'apprentissage. *Les Cahiers de l'APLIUT, 29*(1), 98–115. https://doi.org/10.4000/apliut.967

Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Roszkowski & B. Lewandowska-

Tomaszczyk (Eds.), *Explorations across languages and corpora* (pp. 563–580). Peter Lang.

Boulton, A. (2020). Data-driven learning for younger learners: Obstacles and optimism. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. ix–xviii). Routledge, xiv–xx.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. 10125/73450.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*, 141–158. https://doi.org/10.1093/applin/16.2.141

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System, 27*, 1–18.

Cresswell, A. (2007). Getting to "know" connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom*. Rodopi. https://doi.org/10.1163/9789401203906_018

Crosthwaite, P. (Ed.). (2020). *Data-driven learning for the next generation. Corpora and DDL for pre-tertiary learners*. Routledge. https://doi.org/10.4324/9780429425899

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*(3), 393–417.

Flowerdew, L. (2012). *Corpora and language education*. Palgrave, Macmillan. https://doi.org/10.1057/9780230355569

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108–137). CLUEB.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524773

Johns, T. (1986). Micro-concord: A language learner's research tool. *System, 14*(2), 151–162. https://doi.org/10.1016/0346-251X(86)90004-7

Johns, T. (1988). Whence and whither classroom concordancing? In P. Bongaerts, P. de Haan, S. Lobbe, & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9–27). Foris.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing. ELR journal* (Vol. 4, pp. 1–16).

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge University Press.

Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100–115). Harlow.

Johns, T., & King, P. (1991). Editors' preface. In T. Johns & P. King (Eds.), *Classroom concordancing. ELR journal* (pp. iii–iv).

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language, Learning and Technology, 14*(1), 28–44.

Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System, 81*, 14–25. https://doi.org/10.1016/j.system.2018.12.011

Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–23). Harlow.

McEnery, T., & Wilson, A. (1997). Teaching and language corpora. *ReCALL, 9*(1), 5–14. https://doi.org/10.1017/S0958344000004572

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

O'Sullivan, I., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing, 15*, 49–68.

Pérez-Paredes, P., Ordoñana Guillamón, C., Van de Vyver, J., Meurice, A., Aguado Jiménez, P., Conole, G., & Sánchez Hernández, P. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System, 84*, 145–159. https://doi.org/10.1016/j.system.2019.06.009

Seliger, H. (1983). The language learner as linguist: Of metaphors and realities. *Applied Linguistics, 4*(3), 179–191. https://doi.org/10.1093/applin/4.3.179

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL, 26*(2), 184–201. https://doi.org/10.1017/S0958344014000081

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins. https://doi.org/10.1075/scl.6

Tsui, A. B. M. (2004). What teachers have always wanted to know – And how corpora can help. In J. M. H. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39–61). John Benjamins.

Vincent, B., & Nesi, H. (2018). The BAWE quicklinks project: A new DDL resource for university students. *Lidil, L'enseignement et l'apprentissage de l'écrit académique à l'aide de corpus numériques, 58*. https://doi.org/10.4000/lidil.5306

Integrating Corpora and Data-Driven Learning into Lesson Planning

Larissa Goulart
Montclair State University, Montclair, NJ, USA

Abstract

Despite challenges such as computer power and instructors' technological knowledge (Yoon and Hirvela, J Second Lang Writ 13: 257–283, 2004; Boulton, Lang Lear 60: 534–572, 2010), the development of user-friendly online corpus tools and ready-made lesson plans has made corpus-based teaching more accessible for language teachers. These resources enable educators to create dynamic and interactive lessons, enriching the learning experience and preparing students for real-world language use. This chapter explores the integration of corpus tools and Data-Driven Learning (DDL) into lesson planning. The first part of the chapter addresses the benefits of incorporating corpus tools and techniques in lesson planning. The second part focuses on the indirect applications of corpus resources in the classroom, while the third part examines the direct applications.

Keywords

Corpora in the classroom · Corpus tools · English as a Second Language · English as a Foreign Language · Language teaching

Introduction

Starting in the 1980s, the availability of personal computers revolutionized linguistic research by enabling the analysis of electronic text collections (or corpora). This shift allowed researchers to describe language use based on empirical evidence rather than their intuitions. In time, corpus research made its way into language classrooms, shaping materials, curricula, and lesson plans. Initially, corpus linguistics had 'indirect' applications in the classroom. These applications took the form of dictionaries, wordlists, reference grammars, and textbooks, which were *informed* by the results of corpus research. These resources contrast to 'direct' applications of corpora, or data-driven learning (DDL; Johns, 1991), which proposes that students interact with corpus data directly in the classroom. DDL encourages students to act as 'language detectives' (Johns, 1997), analysing corpus data themselves in order to draw conclusions about language use. While DDL is typically understood as the hands-on use of authentic corpus data, such as concordance lines by advanced English as a Second Language (ESL) or English as a Foreign Language (EFL) learners (Boulton, 2011), this chapter adopts a

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_60-1

broader perspective proposed by Gilquin and Granger (2010), which suggests that any use of corpus tools and techniques for pedagogical purposes falls within the scope of DDL.

Three decades ago, Rundell and Stock (1992) envisioned a future where corpus would revolutionize language teaching, however, recently Chambers (2019) pointed out that most practitioners doing DDL in the classroom are still corpus linguists themselves who have access to a classroom. Ädel (2010) also suggests that DDL curriculums are primarily developed by corpus linguists themselves. One of the reasons for this is the scarcity of corpus linguistics training in teacher education programmes. In sum, corpora are slowly being integrated into educational contexts, but the agents of change are primarily corpus researchers, rather than language teachers. Recently, researchers have sought to promote the use of corpus in the classroom by (1) developing corpus tools that are more user-friendly (e.g. Crosthwaite & Baisa, 2024), and (2) creating lesson plans using corpus resources that are readily available for teachers (e.g. Le Foll, 2021; Goulart & Veloso, 2023; Pinto et al., 2023). The goal of these initiatives is to make the integration of corpora in the classroom more accessible for language teachers. This chapter will focus specifically on these two aspects of corpus linguistics in lesson planning.

The Advantages of Incorporating Corpora in Classroom Materials

Numerous publications have emphasized the benefits of incorporating corpus linguistics into classroom materials (e.g. Friginal, 2018; Reppen, 2010). One of the primary advantages of DDL is the inclusion of authentic examples of language use. Research suggests that to acquire a language, learners must be exposed to real examples of the lexical and lexico-grammatical features that they are studying (Friedman, 2009; Mukherjee & Rohrbach, 2006; Römer, 2006). DDL enables learners to understand how language is used in natural contexts. Additionally, DDL fosters learner's autonomy by allowing students to make informed decisions about specific linguistic features based on real language use. Further, corpora enable learners to develop their numerical literacy by analysing the frequency of linguistic features. In DDL, learners get to examine the co-occurrence of linguistic features, and make informed decisions about the appropriateness of specific linguistic features in various contexts (Crosthwaite, 2020).

Studies investigating the language gains of students exposed to corpus tools generally find a positive impact on the use of DDL in the classroom. A meta-analysis by Boulton and Cobb (2017) examined 64 studies on the effectiveness of using corpus linguistics tools in language learning contexts. Their findings showed that DDL yielded promising results, with large effect sizes for both studies comparing control and experimental groups ($d = 0.95$) and those with pre- and post-test designs ($d = 1.50$).

Empirical research has shown that the integration of corpus linguistics and DDL in language classrooms not only provides authentic language examples and fosters learner autonomy but also demonstrates significant positive impacts on language acquisition. Therefore, it is important to consider how to include corpus tools and techniques in the classroom.

Indirect Applications of Corpora in the Classroom

The 'corpus revolution' mentioned by Rundell and Stock (1992) began with lexicographers, who analyzed words and their usage to create dictionaries. Consequently, some of the first classroom applications of corpora were dictionaries, such as the Collins COBUILD Dictionary of English (Sinclair, 1987). Nowadays, most dictionaries are informed by corpus data and provide examples of collocations and authentic sentences.

Another indirect application of corpora in the classroom involves wordlists. These were developed to help teachers select the most important English words for their students in general and specialized English classes. Examples of such applications date back to the 1950s with West's

(1953) General Service List and continue to the present with lists like the Academic Word List (AWL; Coxhead, 2000) and the Contracts Word List (Hanks et al., 2024). These lists inform the educators' decisions about what words should be included in classroom materials and guide the development of textbooks.

Other indirect applications of corpora in the classroom involve corpus-based or corpus-informed textbooks. These textbooks, developed from extensive corpus research, include works like Grammar and Beyond (Reppen, 2012) and Touchstone (McCarten et al., 2006). These textbooks aim to facilitate the integration of corpora into the classroom by condensing corpus research findings into teaching units.

While these applications expand the range of corpus resources available to teachers, they do not involve students and teachers directly engaging with the corpus to explore their language questions. Instead, these resources are hands-off and might lack the flexibility of teachers developing their own lesson plans using corpora.

Direct Applications of Corpora in Lesson Planning

Similar to indirect applications, where the researcher or material developer is the mediator between the student and the data, direct applications of corpora in the classroom require that the teachers mediate students' interaction with corpus data. This might involve selecting specific concordance lines or helping students interpret the results of an analysis (Charles, 2007).

Two of the main challenges of incorporating corpus tools and techniques in the classroom are: (1) access to corpora and corpus tools, and (2) pedagogical knowledge of how to include corpora in the language classroom. In relation to the first point, teachers and instructors need user-friendly and accessible corpus resources. This often involves developing online interfaces that provide direct access to corpora in combination with analysis techniques. COCA (Davies, 2008-) and Sketch Engine (Kilgarriff et al., 2014) are examples of web applications that combine both sides of DDL in one interface. However, these tools are often behind paywalls—COCA allows a limited number of free searches per day, while Sketch Engine offers a 30-day free trial—making them less reliable for consistent use by teachers. Recognizing these challenges, corpus researchers have collaborated with developers to create more accessible tools for educators.

Table 1 presents a series of such tools and summarizes their functionalities. Two notable examples are Corpus Mate (Crosthwaite & Baisa, 2024) and Sketch Engine for Language Learning (SKELL). Corpus Mate is an online interface that combines six corpora (e.g. BASE, BAWE, BBC, TedCorpus) covering over 20 different topics and both spoken and written language. Designed for teaching young learners, it also includes a profanity filter. SKELL, a simplified version of Sketch Engine, offers three main functions: concordance lines (examples), collocations (word sketches), and a thesaurus function. However, users have limited control over the source corpora used in these searches.

The development of the online tools described in Table 1 addresses one of the challenges in including corpus in the classroom: access to corpus resources. A second challenge is the technical pedagogical knowledge required to effectively integrate corpora into lesson plans. While the ideal solution is to incorporate more corpus training into teacher education and professional development programmes, this is not always feasible. In response, researchers have developed ready-made lesson plans that incorporate corpora. These lesson plans serve a dual purpose: (1) they are ready-to-use resources that teachers can implement in their classrooms, and (2) they provide a model for teachers on how to integrate corpora into their lessons. Many published lesson plans include recommendations for adaptation to different topics or proficiency levels.

Many efforts to produce corpus-based or corpus-informed lessons are the by-product of corpus linguistics classes at the Masters and/or undergraduate level. This is the case with Le Foll (2021) and Goulart and Veloso (2023), who present numerous lesson plans, complete with handouts and corpus instructions in an online

Integrating Corpora and Data-Driven Learning into Lesson Planning, Table 1 Online corpus tools

Tool	Description
AWL Gapmaker	This website creates gap-fill exercises based on the Academic Word List. Users input a text and select the sublist to be blanked in the exercise. (https://www.nottingham.ac.uk/alzsh3/acvocab/awlgapmaker.htm)
AWSuM	The Academic Word Suggestion Machine suggests lexical bundles using move analysis results from a user-inputted text. (https://langtest.jp/awsum/)
CLiC	This web application, part of a larger stylistics project, allows users to search a corpus for concordance lines, clusters, keywords, frequency counts, and more. (https://clic.bham.ac.uk/)
COCA	The Corpus of Contemporary American English, part of the English-Corpora suite, enables users to search for collocations, clusters, concordances, and more across different years, registers, and modes. (https://www.english-corpora.org/coca/)
Compleat LexTutor	This tool offers options for researchers and teachers, including checking academic word coverage in a text, assessing text word levels, and searching for collocates. (https://www.lextutor.ca/)
CorpusMate	A simplified corpus interface for teachers and students, focusing on young learners. It allows searches by mode and topic and includes a profanity filter. (https://corpusmate.com/)
FreeTxt	Originally developed for survey visualizations, this tool allows teachers to import spreadsheets and create ESL classroom materials. (https://corcencc.org/freetxt/)
HyperCollocation	This search engine finds the frequency and collocations of a given word across research articles on arXiv. (https://hypcol.marutank.net/)
Just the word	This tool enables language learners to search for collocates and grammatical patterns of a given word. (http://www.just-the-word.com/)
Kontext	A corpus interface for searching parallel corpora, creating wordlists, and finding collocates and n-grams, useful for both research and teaching. (www.korpus.cz/kontext/)
Linggle	Developed for language learners, it uses Google Books as a corpus for searching word combinations and frequency lists. (https://search.linggle.com/)
MICASE	The Michigan Corpus of Academic Spoken English offers an online interface for searching words or strings with filters for discipline or lecture type. (https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple)
MICUSP	The written counterpart to MICASE, the Michigan Corpus of Upper-Level Student Papers, allows searches for words and word combinations across disciplines and text types. (https://micusp.elicorpora.info/main)
Netspeak	This tool allows searches for words and word combinations using regex, with a corpus sourced from Google Books. (https://netspeak.org/)
SKELL	Sketch Engine for Language Learning is a simplified version of Sketch Engine, offering examples, word sketches, and similar words. (https://skell.sketchengine.eu/)
String Net	Allows users to search for patterns or collocates of a given word or string, with options for advanced filtering by frequency and n-grams. (http://nav.stringnet.org/)
TED Corpus	A search engine for TED transcripts, offering functionalities like concordancing, n-grams, and frequency lists, useful for creating vocabulary exercises. (https://yohasebe.com/tcse)

(continued)

Integrating Corpora and Data-Driven Learning into Lesson Planning, Table 1 (continued)

Tool	Description
Versatext	A web application for English learners to explore language patterns in a text by creating word clouds, analysing text profiles, and viewing concordance lines. (https://versatext.versatile.pub/)
Visual Thesaurus	A collocations network that visualizes word connections and allows users to create wordlists from a text. (https://www.visualthesaurus.com/)
Voyant Tools	A web application for analysing user-inputted texts online, offering features like word clouds, vocabulary density checks, and concordance lines. (https://voyant-tools.org/)
WebCorp	This tool allows the user to turn the web into a corpus. Users can create wordlists from URLs or search for concordance lines for a given word or string. (https://www.webcorp.org.uk/live/)

format hosted by Pressbook. Le Foll (2021) offers 16 lesson plans divided into 4 educational contexts, while Goulart and Veloso (2023) provide 14 lesson plans and 2 tutorials on using corpus tools (Sketch Engine and COCA). Other publications, such as Pinto et al. (2023), present lesson plans that incorporate corpus tools into language classes for English, Portuguese, and Spanish. Additional resources include Viana's (2022) book, which features 70 chapters with lesson plans written by corpus researchers and English teachers. Projects like the University of Lancaster's Corpus for Schools (https://wp.lancs.ac.uk/corpusforschools/) and the Corpus and Repository of Writing (CROW), CROW for Teachers (https://writecrow.org/crow-for-teachers/pedagogical-materials/) have also produced lesson plans. These projects aim to showcase the diverse applications of corpora in the classroom and encourage teachers to adapt these ready-made plans to their specific contexts.

Given the abundance of resources available for teachers, it is worth considering how corpora have been integrated into these materials (e.g. Le Foll, 2021; Viana, 2022; Goulart & Veloso, 2023; Pinto et al., 2023) at different stages of the lesson. The following table is not exhaustive but provides examples from these materials of how and where corpus was integrated into lesson plans, following a presentation-production-practice (PPP; see DeKeyser, 2007) model.

As Table 2 shows, corpus resources can be included in different stages of the lesson plan. In addition, Table 2 highlights how DDL can be used to teach different skills, from grammar to listening and speaking.

Conclusion

Integrating Data-Driven Learning (DDL) into lesson plans involves a multi-step approach that begins with selecting relevant corpora and designing tasks that encourage data exploration. Ready-made lesson plans can serve as examples for teachers on how to include corpus tools into their own classroom. By scaffolding student learning, educators can ensure that students effectively interpret and apply the insights gained from these activities. This approach not only enhances language proficiency but also fosters critical thinking and analytical skills, making lessons more dynamic and interactive.

Despite the numerous benefits, implementing corpora in the classroom is not without challenges. Computer power and computer skills have been cited as significant barriers (Boulton, 2010; Yoon & Hirvela, 2004). However, the development of user-friendly corpus tools and the availability of ready-made lesson plans have facilitated the inclusion of these resources into teaching.

Integrating Corpora and Data-Driven Learning into Lesson Planning, Table 2 Examples of corpus integration in different stages of a lesson plan

Warm-up	A corpus can be used to select vocabulary to pre-teach before a reading/listening task, or to introduce the topic to students. Example: In Step 6 of Tombrägel (2021), the teacher selects concordance lines from a soap opera corpus, where characters make invitations, and then searches for these specific scenes online to share with the class.
Presentation	A corpus can be used to help students notice language patterns. Example: In Task 2 of Batista (2023), students examine the difference between the collocates of 'transgender' and 'transsexual' on COCA to understand the difference in connotations of these words. In Worksheet 1 of Füchs and Liebelt (2021), students search for verbs that collocate with modes of transportation and then analyse the collocates.
Practice	Teachers can use concordance lines, frequency information, collocates and other corpus techniques to give students a chance to practice the topic being taught. Example Andrade et al. (2023) created a gap-fill exercise based on concordance lines for the prepositions "through" and "toward," which are the focus of the class. In Activity 2 of Garcia (2023), students identify the type of collocation (verb-preposition, noun-preposition, adjective preposition) in collocates extracted from Linggle.
Production	Students can use corpus resources to create language more freely In Activity 4 of Garcia (2023), students write sentences using the collocations with prepositions found in Linggle. In Activity 3 of Pesenti et al. (2023), students complete a crossword puzzle using concordance lines from SKELL and use the words in the crossword in new sentences. In Task 4 of Nottmeier et al. (2021), the student sees a text excerpt from a corpus and is asked to correct the use of commas, which is the topic of the lesson plan.

Cross-References

- ▶ Corpus Literacy
- ▶ Emerging Technologies and New Methodologies for DDL
- ▶ Learner Corpora in DDL
- ▶ Pedagogical Corpora and DDL
- ▶ Teacher Training for DDL
- ▶ Wordlists and DDL

References

Ädel, A. (2010). Just to give you kind of a map of where we are going: A taxonomy of metadiscourse in spoken and written academic English. *Nordic Journal of English Studies, 9*(2), 69–97.

Andrade, B., Britto, G., De Leeuw, R., & Silva, R. (2023). Using COCA to explore the differences between 'through' and 'toward'. In L. Goulart & I. Veloso (Eds.), *Corpora in English language teaching: Classroom activities for teachers new to corpus linguistics* [Open Educational Resource]. Montclair State University. https://pressbooks.pub/testbook123/chapter/using-coca-to-explore-the-differences-between-through-and-toward/

Batista, L. (2023). Using gender-related vocabulary. In L. Goulart & I. Veloso (Eds.), *Corpora in English language teaching: Classroom activities for teachers new to corpus linguistics* [Open Educational Resource]. Montclair State University. https://pressbooks.pub/testbook123/chapter/using-gender-related-vocabulary/

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60*(3), 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563–580). Peter Lang.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching, 52*(4), 460–475. https://doi.org/10.1017/S0261444819000089

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic*

Purposes, 6(4), 289–302. https://doi.org/10.1016/j.jeap.2007.09.009

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Crosthwaite, P. (2020). Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics, 43*(2), 169–195. https://doi.org/10.1075/aral.00031.cro

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*. Advance online publication. https://doi.org/10.1075/ijcl.23056.cro

Davies, M. (2008–). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

DeKeyser, R. M. (2007). *Practice in a Second Language: Perspectives from applied linguistics and cognitive psychology.* Cambridge University Press.

Friedman, G. L. (2009). Learner-created lexical databases using web-based source material. *ELT Journal, 63*(2), 126–136. https://doi.org/10.1093/elt/ccn022

Friginal, E. (2018). *Corpus linguistics for English teachers*. Routledge.

Füchs, M., & Liebelt, A. (2021). "How do I get there?": Travelling by various means of transport. In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom*. https://pressbooks.pub/elenlefoll/chapter/fuchs_liebelt/

Garcia, W. (2023). Lesson 1: It depends on the situation. In P. Pinto, P. Crosthwaite, C. de Carvalho, F. Spinelli, T. Serpa, W. Garcia, & A. Ottaiano (Eds.), *Using language data to learn about language: A teachers' guide to classroom corpus use*. University of Queensland. https://doi.org/10.14264/3bbe92d

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge.

Goulart, L., & Veloso, I. (2023). Introduction. In L. Goulart & I. Veloso (Eds.), *Corpora in English language teaching: Classroom activities for teachers new to corpus linguistics* [Open Educational Resource]. Montclair State University. https://doi.org/10.5281/zenodo.7900420

Hanks, E., Hashimoto, B., & Egbert, J. (2024). The contracts word list: Integral vocabulary for reading and writing English contracts. *English for Specific Purposes, 75*, 37–48. https://doi.org/10.1016/j.esp.2024.03.002

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal, 4*, 27–45.

Johns, T. F. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora*. London: Longman.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, M., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*, 7–36. https://doi.org/10.1007/s40607-014-0009-9

Le Foll, E. (Ed.). (2021). *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. https://elenlefoll.pressbooks.com

McCarten, J., McCarthy, M., & Sandiford, H. (2006). *Colombo touchstone 2 student's book/workbook with audio CD/CD-ROM*. Cambridge University Press.

Mukherjee, J., & Rohrbach, J. M. (2006). Rethinking applied corpus linguistics from a language pedagogical perspective: New departures in learner corpus research. In B. Kettemann & G. Marko (Eds.), *Planning, gluing and painting Corpora: Inside the applied corpus linguist's workshop* (pp. 205–232). Peter Lang.

Nottmeier, M., Peters, A., & Warnecke, L. (2021). Commas in argumentative writing. In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom*. https://pressbooks.pub/elenlefoll/chapter/nottmeier_peters_warnecke/

Pesenti, M., Ribeiro, J., Tumux, J., & Zimny, R. (2023). What can we learn about "take", "have", and "make" from COCA? In L. Goulart, & I. Veloso (Eds.), *Corpora in English language teaching: Classroom activities for teachers new to corpus linguistics* [Open Educational Resource]. Montclair State University. https://pressbooks.pub/testbook123/chapter/what-can-we-learn-about-take-have-and-make-from-coca/

Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. (Eds.). (2023). *Using language data to learn about language: A teachers' guide to classroom corpus use*. University of Queensland. https://doi.org/10.14264/3bbe92d

Reppen, R. (2010). *Using corpora in the language classroom* (Vol. 6). Cambridge University Press.

Reppen, R. (2012). *Grammar and beyond*. Cambridge University Press.

Römer, U. (2006). *Progressives, patterns, pedagogy. A corpus-driven approach to English progressive forms, functions, contexts and didactics*. John Benjamins.

Rundell, M., & Stock, P. (1992). The corpus revolution. *English Today, 8*(3), 21–32. https://doi.org/10.1017/S0266078400006520

Sinclair, J. M. (1987). Introduction. In J. Sinclair (Ed.), *Collins COBUILD English language Dictionary* (pp. vii–xxi). Collins.

Tombrägel, S. (2021). "Let's have a party!": Teaching students how to express, accept and refuse invitations. In E. Le Foll (Ed.), *Creating corpus-informed materials for the English as a foreign language classroom*. https://pressbooks.pub/elenlefoll/chapter/tombraegel/

Viana, V. (Ed.). (2022). *Teaching English with corpora: A resource book*. Routledge.

West, M. (1953). *A general service list of English words*. Longman, Green.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257–283.

P

Pedagogical Guides for Data-Driven Learning

Elen Le Foll
Department of Romance Languages, University of Cologne, Cologne, Germany

Abstract

Pedagogical guides for data-driven learning (DDL) can function as resources for both language learners and educators. Existing guides cover a range of topics, from selecting an appropriate corpus and corpus tool to strategies for querying copora and interpreting the results. Some of the guides targeted at teachers additionally discuss the design of DDL tasks and explain how DDL can be integrated into everyday teaching and learning practices. Pedagogical guides come in diverse formats, ranging from printed textbooks published by academic presses to multimodal webbooks released under Creative Commons licences, and from individual paper-based worksheets to interactive online courses featuring instructional videos and interactive tasks with automatic feedback. The affordances of pedagogical guides can be examined in terms of their accessibility. Some formats suffer from limited physical, intellectual, and/or social accessibility. In recent years, the emergence of Open Educational Resources (OERs) and of guides coauthored with (trainee) teachers has contributed to the growth and increased accessibility of pedagogical guides for DDL. While acknowledging the current dominance of resources focusing on DDL for English L2 in tertiary education, recent developments point to an enduring commitment to improving the accessibility of pedagogical guides for DDL in a variety of teaching and learning contexts.

Keywords

Lesson planning · Open education resources (OERs) · Corpora · Teaching · Task design · Design · Textbook · Corpus linguistics · Teacher education · Accessibility

Introduction

Data-driven learning (DDL) aims to empower learners to derive linguistic insights independently by exploiting corpora that represent 'real-life' language use. Within computer-assisted language learning (CALL), it represents a learner-centred approach that fosters learner autonomy and critical data literacy. The term 'data-driven learning' (DDL) may give the impression that it is an approach that is entirely 'driven' or guided by the data itself. However, merely having access to corpora and corpus tools does not guarantee effective DDL. Adequate guidance, thoughtful task design, and meaningful integration into the regular teaching and learning practices are critical to the success of DDL. The importance of pedagogical guides

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_61-1

tailored to the needs of both educators and learners, as well as of teacher training (see entry ▶ 'Teacher Training for Data-Driven'), cannot be overstated.

Pedagogical guides for DDL are instructional manuals or tools that guide teachers and/or learners through the process of using corpus data for various language learning activities. They may include guidance on corpus selection, querying, analysis, and interpretation, as well as strategies for integrating DDL into everyday learning activities and suggestions for DDL task design and lesson planning. The provision of high-quality, ready-to-use teaching and learning resources can boost the effectiveness of DDL in the classroom and help to bridge the 'corpus research–practice gap' (Chambers, 2019; Crosthwaite, 2024) by facilitating the uptake of DDL outside of academia. Making available complete lesson plans with worksheets and instructions for teachers can also help (pre-service) teachers to develop corpus literacy skills (Le Foll, 2023; Ma et al., 2022; and entry on ▶ 'Corpus Literacy and Data-Driven Learning'), allowing them to progressively develop a greater degree of creativity and independence in integrating DDL into their regular teaching practice (Farr & Leńko-Szymańska, 2023a, pp. 8–9).

This entry provides an overview of existing pedagogical guides for DDL. It first surveys the various aspects of DDL that current guides cover, before considering their target audiences. It then reviews the different types and formats of presently available guides and discusses their affordances in terms of their physical, intellectual, and social accessibility for their respective target users. The entry concludes with a summary of the *status quo* and an outlook on future advances in the development and dissemination of pedagogical guides for DDL.

Contents

Pedagogical guides can inform teachers and learners about one or more aspects relevant to DDL. These contents can range from corpus selection and corpus queries to the interpretation of corpus results and the design and integration of DDL activities in lesson plans.

In DDL activities, the choice of corpus and of corpus tool impacts the learning outcome. Some pedagogical guides therefore include discussions on the choice of an appropriate corpus and/or corpus tool in line with an activity's learning objectives and instructional context (e.g. Anderson & Corbett, 2017; Poole, 2018). Some guides even include a chapter on the use of custom 'DIY' corpora (see ▶ 'DIY Corpora and Data-Driven Learning') and therefore also cover corpus building (e.g. Friginal, 2018; Reppen, 2010). By contrast, pedagogical resources that make use of corpus tools that rely exclusively on a single corpus (e.g. SKELL, MICUSP Simple, TED Corpus Search Engine; see ▶ 'Tools for Data-Driven Learning' entry) tend to focus more on the technical aspects of using the tool than on the pedagogical implications of choosing one corpus over another.

Numerous pedagogical guides for DDL include user manual-style instructions that guide language teachers and/or learners through the technical aspects of using one or more corpus tools. Depending on the target audience of the guide, the steps are described in greater or lesser detail, sometimes with annotated screenshots (e.g. Karpenko-Seccombe, 2020; Le Foll, 2021) and/or screencast videos (e.g. Anthony, 2023; Le Foll, 2020). However, not all DDL activities proposed in existing pedagogical guides require access to corpus tools during the learning process. Many guides feature paper-based DDL activities (e.g. Johns, n.d.; Le Foll, 2021; Pinto et al., 2023; Viana, 2023), thus allowing for the integration of DDL activities in contexts where learners do not have access to the necessary devices and/or the Internet (see entry ▶ 'Hands-On Versus Hands-Off Methods in Data-Driven Learning').

Pedagogical guides which are primarily targeted at teachers usually place a strong emphasis on DDL task design (e.g. Friginal, 2018; Le Foll, 2021; Reppen, 2010). However, relatively few of these guides devote a great deal of attention to the issue of integrating DDL activities into regular teaching and learning activities (see ▶ 'Integrating Corpora and Data-Driven Learning into Lesson Planning' entry). In this respect, the Corpus-Aided Platform for Language Teachers (Ma, 2018) constitutes an exception; among other resources, it features video

recordings of demonstration lessons involving DDL activities.

Well-designed DDL tasks exploit the potential of corpora and corpus tools for language learning by guiding learners to reach data-driven conclusions on language phenomena. Typically, pedagogical guides for DDL recommend beginning with relatively controlled exercises. These may include detailed instructions aimed at supporting the development of the technical skills necessary for querying corpora in DDL. When it comes to interpreting corpus results, initial tasks typically direct learners' attention to specific language patterns to scaffold the DDL process. As learners progress, DDL tasks become more exploratory, encouraging independent discovery learning and hypothesis testing. This task progression can also be found in materials targeted directly at learners (in the context of EAP, see, e.g. Karpenko-Seccombe, 2020; Viana & O'Boyle, 2021). As learners' (L1 and L2) language awareness and metacognitive abilities are likely to significantly influence their ability to draw insightful conclusions from corpus data (see entry on ▶ 'Metacognition and Data-Driven Learning'), these DDL guides frequently encourage in-depth reflection during learners' interaction with corpus tools and in the process of interpreting the results (e.g. Anderson & Corbett, 2017; Friginal, 2018; Poole, 2018).

Target Users

In instructed language learning settings, the role of the teacher is crucial in providing the scaffolding necessary for learners to derive meaningful knowledge from corpus data. Although the edited volume 'How to Use Corpora in Language Teaching' (Sinclair, 2004) was published over two decades ago, studies continue to conclude that one of the major barriers to incorporating corpora and DDL into language instruction 'originate[s] in teachers themselves who frequently complain about a lack of guidance on how to take advantage of advances in corpus-based technologies' (Leńko-Szymańska & Boulton, 2015, p. 5) (see also Farr & Leńko-Szymańska, 2023b; Leńko-Szymańska, 2014, 2015).

Responding to this need, a number of existing pedagogical guides for DDL are explicitly aimed at teachers (e.g. Bennett, 2010; Friginal, 2018; Le Foll, 2021; Reppen, 2010; Viana, 2023). Conversely, some guides are primarily designed with language learners in mind, either for independent study or as supplementary resources to a taught course (e.g. Anderson & Corbett, 2017; Karpenko-Seccombe, 2020; Poole, 2018; The University of Queensland, 2019). An emerging category of guides caters for both audiences: resources such as those provided by the Corpus-Aided Platform for Language Teachers (Ma, 2018) and Corpus for Schools (Gablasova et al., 2018) exemplify this approach, offering DDL activities accompanied by solutions, detailed teaching notes and lesson plans.

To date, most published pedagogical guides focus on teaching and learning English as a foreign/second language (L2). Notable exceptions include Pinto et al. (2023) which features resources for teaching Spanish and Portuguese L2, and Schaeffer-Lacroix (2019) and Vyatkina (2020) who propose a wide range of DDL activities for German L2. Moreover, most materials are written in English (though Schaeffer-Lacroix [2019] is in French and a few are in German, e.g. Mukherjee, 2002, 2009; Siepmann, n.d.). Finally, the DDL activities described in such guides are typically targeted at specific learner age groups, language proficiencies, and instructional settings. In particular, a number focus on learners of English as an Academic Language (EAP) at university level (e.g. Friginal, 2018; Karpenko-Seccombe, 2020; Poole, 2018; Viana & O'Boyle, 2021) (see also entry on ▶ 'Language for Academic Purposes and Data-Driven Learning'). Far fewer resources guide teachers in the use of DDL with younger, pre-tertiary language learners (though Crosthwaite, 2020a; Ma, 2018; Le Foll, 2021; and Viana, 2023 present teaching resources for primary and lower-secondary school EFL learners, see also ▶ 'Young Learners and Data-Driven Learning').

Formats

Pedagogical guides for DDL have traditionally taken the form of books written by distinguished

corpus linguists and published by prestigious academic presses. To mention but one, 'From corpus to classroom: Language use and language teaching' (O'Keeffe et al., 2007) was instrumental in attempting to bridge the gap between applied corpus linguistics research and language teaching and learning practice. However, as the authors themselves acknowledge, this highly influential research-oriented textbook 'stop[s] at the classroom door' (p. 248). Subsequent academic book publications have continued to popularise corpus linguistics and DDL among language students and (future) language teachers. Although still firmly rooted in research, these publications (e.g. Bennett, 2010; Crawford & Csomay, 2016; Reppen, 2010; Timmis, 2015) exemplify how corpus linguistics can be used in ELT by presenting a range of example DDL activities. As such, they can be said to have taken further steps towards 'entering the classroom'.

A more recent development has seen the publication of 'resource books' that provide materials that can readily be used in the classroom and thus aim to contribute to integrating DDL in day-to-day learning and teaching practices. 'Academic writing with corpora: A resource book for data-driven learning' (Karpenko-Seccombe, 2020), for example, fosters autonomous learning among intermediate and advanced EAP learners by demonstrating how to solve everyday difficulties in word choice, grammar, and text production using freely accessible online corpus tools. The largest such resource book to date is 'Teaching English with Corpora' (Viana, 2023). As the title implies, it is targeted at teachers rather than learners and brings together 70 tested corpus-based/driven lesson plans and teaching suggestions for all proficiency levels, as well as for EAP and ESP contexts. These lessons were contributed by a mixture of corpus linguists and practicing teachers from diverse backgrounds and do not assume any prior knowledge of corpus linguistics.

In parallel, some researcher-educators have begun to publish banks of resources for DDL as Open Educational Resources (OERs). Open, online formats offer abundant space and thus allow for multimodal content, interactive elements, instructional videos, detailed teaching notes, and (editable) classroom handouts in convenient formats. The Corpus-Aided Platform for Language Teachers (CAP; Ma, 2018) hosts teaching activities for English L2 classes at different proficiency levels and English-Chinese translation classes, as well multimodal teacher training materials, and instructional videos contextualising DDL in classroom practice. Being hosted online, its bank of lessons plans and worksheets can be regularly updated and continue to grow, e.g. with the recent addition of student-contributed 'AI-enhanced' DDL lesson plans.

Worksheets are another popular format for pedagogical guides for DDL. In addition to CAP, several collections of standalone worksheets have been published. The oldest are likely those created by the 'founding father of DDL' (see ▶ 'Tim Johns and Data-Driven Learning' entry): Tim Johns' collection of 'Kibbitzers' is freely accessible as downloadable PDF worksheets (see https://lexically.net/TimJohns and https://web.archive.org/web/20090425093656/http:/www.eisu2.bham.ac.uk/Webmaterials/kibbitzers/indexes/indexchron.htm). Each worksheet encourages learners to identify language patterns from a curated list of concordance lines to solve a specific language issue (e.g. *by* vs. *using*). Tom Cobb's (n.d.; Gaskell & Cobb, 2004) online Corpus Grammar tool encourages English learners to infer information from concordance lines in a very similar way but via a web interface that provides automated feedback. Many collections of DDL worksheets are associated with specific tools. For instance, the Corpus for Schools platform features resources for teaching English as an L1 and L2 using the BNCLab (Brezina et al., 2018) alongside additional pedagogical guidance for teachers (Gablasova et al., 2018).

Other freely accessible pedagogical guides take the form of instructional videos (e.g. Anthony, 2023; Crosthwaite, 2020b; Le Foll, 2020). Tutorial videos typically also introduce DDL via specific tools (e.g. AntConc, Sketch Engine, english-corpora.org). They tend to focus more on the technical component of learning to use corpora for language learning, rather than on pedagogical aspects.

Webbooks are hybrid formats that combine the familiar structure of traditional textbooks with the

affordances of online publication formats. Integrating Corpora (Vyatkina, 2020) is an example of an OER targeted at learners of German as a second/foreign language. It is an interactive online workbook that guides readers through the process of DDL using a suite of online German corpora and many short DDL activities for which the webbook provides automatic corrections. In contrast, the OER webbook 'Creating Corpus-Informed Materials for the English as a Foreign Language Classroom' (Le Foll, 2021) is aimed at (trainee) teachers of English L2. Each chapter provides a fully fledged corpus-based lesson plan, including DDL activities and worksheets, as well as richly illustrated instructions for creating new corpus-informed materials for primary, secondary, and tertiary education (see entry on ▶ 'Corpus-Informed Materials in Language Teaching'). Open webbooks appear to be growing in popularity with several recent book-length pedagogical guides for DDL being published as OERs (Goulart & Veloso, 2023; Pinto et al., 2023).

Finally, online courses can provide guidance for DDL by combining textbook-like explanations and instructions, instructional videos, and practical tasks and activities. In the spirit of Open Education, Leńko-Szymańska (2014) published a Moodle course which exemplifies applications of corpora in language teaching and learning with a variety of receptive and productive tasks. It was originally developed for a one-semester pre-service teacher training course. Targeted at learners rather than (student) teachers, 'Improving writing through corpora' (The University of Queensland, 2019) is an example of a free self-study course. It provides a hands-on introduction to DDL for EAP in five short modules. This course is also available as an 8-hour self-study course in Mandarin Chinese (University of Queensland, n.d.). Although it is not specifically about DDL, Lancaster University's (2014) massive open online course (MOOC) on corpus linguistics also deserves a mention. Running annually since 2014, it has introduced tens of thousands of students and educators to the basic principles of corpus linguistics and its applications, including for language teaching and learning.

Accessibility

While pedagogical guides for DDL can undoubtedly contribute to bridging the well-known corpus research-practice research gap, their potential is often hampered by their limited accessibility. Accessibility, in this context, is a multidimensional construct (Burnett et al., 2008). Below we consider how three dimensions of accessibility contribute to determining the success of pedagogical guides for DDL: their physical, intellectual, and social accessibility (see also Le Foll, 2023).

Regarding the physical accessibility of pedagogical guides for DDL, it is important to first acknowledge that most book-length publications on DDL are commercial publications that come at a cost. While student teachers in the Global North may have access to these resources through their institutional libraries, this only applies to a minority of teachers and learners worldwide. In contrast, Open Educational Resources (OERs) and other freely available resources such as online courses, video tutorials, and downloadable worksheets have the potential to reach a much wider range of language teachers and learners. In addition, physical accessibility implies a focus on freely available, user-friendly corpus tools. The latter, in particular, is not always the case as many corpus tools and interfaces have been primarily designed for academic research rather than for teaching and learning (see entry on ▶ 'Tools for Data-Driven Learning'). Another issue affecting the physical accessibility of pedagogical guides arises from the fact that corpus tools and platforms are constantly evolving, frequently rendering instructions published in formats that cannot easily be updated (e.g. traditionally published books as opposed to dynamic OERs) obsolete. Having to adapt to updated graphical user interfaces (GUIs) and new features without the corresponding instructions can represent an additional barrier to access for less technically versed users (Le Foll, 2023, p. 173).

Second, effective pedagogical guides for DDL must be 'intellectually accessible'. To date, however, most guides are the work of dedicated corpus linguists who primarily work as researchers rather

than as educators and who rarely have first-hand experience of teaching in pre-tertiary instructional contexts. This can lead to overly complex materials that presuppose too much linguistic, technical, and, in the case of resources targeted at (pre-service) teachers, pedagogical knowledge. Several factors can influence the intellectual accessibility of a resource for any individual, ranging from the language in which it is written and its general legibility, to its degree of technical and linguistic complexity. Some recent publications can be said to address the intellectual accessibility of pedagogical guides for DDL by featuring DDL lesson plans and activities designed by pre- and/or in-service teachers in collaboration with researchers (e.g. Friginal, 2018; Le Foll, 2021; Ma, 2018).

Third, for teachers and learners to value and make use of a DDL guide, it must have high 'social accessibility'. This third dimension of accessibility is closely linked to the concept of 'affordances'. In the introduction to the edited volume 'Multiple Affordances of Language Corpora for Data-Driven Learning', Leńko-Szymańska and Boulton (2015, p. 5) emphasise that 'an affordance of an object (a corpus in this case) is not inherent in the object itself, but depends on the perception and ability of the user'. In other words, users' willingness to engage with materials is inextricably shaped by societal norms and attitudes within a particular community. Teacher education and professional development courses for in-service teachers are crucial for the social accessibility of DDL and here, too, introducing accessible pedagogical guides can contribute to these courses' effectiveness and sustainability (see entry on ► 'Teacher Training for Data-Driven Learning'). Guides that broach the issue of meaningfully integrating DDL in everyday teaching and learning activities can contribute to making DDL more socially accessible.

Summary and Outlook

This entry on existing pedagogical guides for data-driven learning (DDL) has brought to light a growing wealth of resources designed to support both learners and teachers. DDL guides serve not only language learners but their teachers too, with some resources specifically developed to address both user groups. The present survey of the content of existing DDL guides has underlined the importance not only of mastering the technical aspects of DDL, but also of efficient task design and DDL task integration, together with the judicious selection of appropriate corpora and tools.

Beyond the traditional research-oriented textbooks, DDL guides come in a range of formats, including banks of lesson plans and DDL activities, multimodal webbooks released under Creative Commons licences as Open Educational Resources (OERs), worksheets, tutorial videos, and interactive online courses. The ease of publishing OERs offering easily updatable, multimodal content, and the possibility of tailoring contents for specific instructional contexts and proficiency levels offers hope for the mainstreaming of DDL in everyday education (see also Le Foll, 2024). A growing trend towards a more hands-on approach is evident in more recent pedagogical guides, thereby broadening their potential use among increasingly diverse teaching and learning communities. Such guides can boost the incorporation of DDL in day-to-day teaching and learning practices, especially at pre-tertiary levels.

However, accessibility remains a significant challenge. To bridge the gap between the potential of these guides and their actual use, emphasis needs to be placed on increasing their physical, intellectual, and social accessibility. While the basic principles of DDL are theoretically applicable to a range of languages, the majority of existing pedagogical guides are in English and focus exclusively on English language teaching and learning. There is an urgent need to increase the number of resources available for a wider range of languages. The proliferation of resources published as OERs makes this possible, as their licences explicitly state that they can be freely updated, translated, adapted, and extended. Increased collaboration between researchers and teachers can further enhance the social accessibility and thus the practical use of these guides, thereby increasing their uptake by educators. Future possibilities also include the integration of AI tools that may allow for more sophisticated

yet intuitive user interfaces (see recent developments of AntConc [Anthony, 2024] and the entry on ▶ 'Emerging Technologies and New Methodologies in Data-Driven Learning'). In addition, although much corpus research suggests that learner corpora have an important role to play in language teaching and learning (see ▶ 'Learner Corpora in Data-Driven Learning'), pedagogical guides for DDL rely almost exclusively on corpora of native and/or expert language users (but see Karpenko-Seccombe, 2020, who includes DDL activities based on the Michigan Corpus of Upper-level Student Papers).

In conclusion, the landscape of pedagogical guides for DDL is evolving. While barriers such as cost, intelligibility, complexity, and social accessibility persist, there are promising developments, largely driven by the emergence of free, user-friendly OERs, and other innovative web-based resources and initiatives.

Cross-References

- ▶ Corpus Literacy and Data-Driven Learning
- ▶ Corpus-Informed Materials in Language Teaching
- ▶ DIY Corpora and Data-Driven Learning
- ▶ Emerging Technologies and New Methodologies in Data-Driven Learning
- ▶ Future Challenges and Opportunities for Data-Driven Learning
- ▶ Hands-On Versus Hands-Off Methods in Data-Driven Learning
- ▶ Integrating Corpora and Data-Driven Learning into Lesson Planning
- ▶ Learner Corpora in Data-Driven Learning
- ▶ Metacognition and Data-Driven Learning
- ▶ Teacher Training for Data-Driven Learning
- ▶ Tim Johns and Data-Driven Learning
- ▶ Tools for Data-Driven Learning
- ▶ Young Learners and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anderson, W., & Corbett, J. (2017). *Exploring English with online corpora*. Bloomsbury Publishing.

Anthony, L. (2023). *AntConc 4 (ver. 4.2)—Getting started* [Tutorial videos]. YouTube. https://www.youtube.com/watch?v=_GSlwIO5QZE

Anthony, Laurence. 2024. Enhancing Corpus Analysis through the Integration of Large Language Models (LLMs). Presented at the 11th Inter-Varietal Applied Corpus Studies (IVACS), University of Cambridge. https://osf.io/qcu3z/. (23 January, 2025)

Bennett, G. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press. https://doi.org/10.3998/mpub.371534

Brezina, V., Gablasova, D., & Reichelt, S. (2018). *BNClab* [Electronic resource]. http://corpora.lancs.ac.uk/bnclab/search

Burnett, G., Jaeger, P. T., & Thompson, K. M. (2008). Normative behavior and information: The social aspects of information access. *Library & Information Science Research, 30*(1), 56–66. https://doi.org/10.1016/j.lisr.2007.07.003

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching, 52*(4), 460–475. https://doi.org/10.1017/S0261444819000089

Cobb, T. (n.d.). *Corpus grammar*. Compleat Lexical Tutor. Retrieved 17 July 2024, from https://www.lextutor.ca/conc/gram/

Crawford, W. J., & Csomay, E. (2016). *Doing corpus linguistics*. Routledge.

Crosthwaite, P. (Ed.). (2020a). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge.

Crosthwaite, P. (2020b, May 1). *Corpora and data-driven learning for ELT lesson planning* [Recording of an online workshop]. YouTube. https://www.youtube.com/watch?v=Igo6iVb1m0M

Crosthwaite, P. (Ed.). (2024). *Corpora for language learning: Bridging the research-practice divide*. Routledge. https://doi.org/10.4324/9781003413301

Farr, F., & Leńko-Szymańska, A. (2023a). Corpora in English language teacher education: Research, integration, and resources. *TESOL Quarterly, Early View, 58*, 1181–1192. https://doi.org/10.1002/tesq.3281

Farr, F., & Leńko-Szymańska, A. (2023b). Corpus linguistics in second language teacher education. *Second Language Teacher Education, 2*(2), 117–132. https://doi.org/10.1558/slte.28536

Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. Routledge. https://doi.org/10.4324/9781315649054

Gablasova, D., Brezina, V., McEnery, T., Meyerhoff, M., Reichelt, S., Arnold, T., & Cheung, C. (2018). *Corpus for schools: Teaching English language with corpus linguistics*. http://wp.lancs.ac.uk/corpusforschools/

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://doi.org/10.1016/j.system.2004.04.001

Goulart, L., & Veloso, I. (Eds.). (2023). *Corpora in English language teaching*. https://pressbooks.pub/testbook123/

Johns, T. (n.d.). *Tim Johns' Kibbitzers*. Retrieved 17 July 2024, from https://lexically.net/TimJohns/

Karpenko-Seccombe, T. (2020). *Academic writing with corpora: A resource book for data-driven learning*. Routledge. https://doi.org/10.4324/9780429059926

Lancaster University, & Tony, M. (2014). *Corpus Linguistics Analysis—Online Course*. FutureLearn. https://www.futurelearn.com/courses/corpus-linguistics

Le Foll, E. (2020). *Short tutorials on using corpus tools for (trainee) teachers, students, linguists and anyone interested in corpus linguistics and corpus methods* [Tutorial videos]. YouTube. http://www.youtube.com/playlist?list=PLAq6uhS_0brxTW99jeZjxdslQ5BRy1Eb5

Le Foll, E. (Ed.). (2021). *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom: A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). https://pressbooks.pub/elenlefoll

Le Foll, E. (2023). 'Opening up' corpus linguistics: An open education approach to developing corpus literacy among pre-service language teachers. *Journal of Second Language Teacher Education, 2*(2), 161–186. https://doi.org/10.1558/slte.25371

Le Foll, E. (2024). Why we need Open Science and Open Education to bridge the corpus research–practice gap. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 142–156). Routledge.

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL, 26*(2), 260–278. https://doi.org/10.1017/S095834401400010X

Leńko-Szymańska, A. (2015). A teacher-training course on the use of corpora in language education: Perspectives of the students. In A. Turula, B. Mikolajewska, & D. Stanulewicz (Eds.), *Insights into technology enhanced language pedagogy* (pp. 129–144). Peter Lang. https://doi.org/10.3726/978-3-653-04995-4

Leńko-Szymańska, A., & Boulton, A. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning*. John Benjamins.

Ma, Q. (2018). *The corpus-aided platform for language teachers (CAP)*. https://corpus.eduhk.hk/cap/

Ma, Q., Yuan, R. (Eric), Cheung, L. M. E., & Yang, J. (2022). Teacher paths for developing corpus-based language pedagogy: A case study. *Computer Assisted Language Learning*, *37*(3), 461–492. https://doi.org/10.1080/09588221.2022.2040537.

Mukherjee, J. (2002). *Korpuslinguistik und Englischunterricht: Eine Einführung*. Peter Lang.

Mukherjee, J. (2009). *Anglistische Korpuslinguistik: Eine Einführung*. Erich Schmidt Verlag.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Pinto, P. T., Crosthwaite, P., Tavares de Carvalho, C., Spinelli, F., Serpa, T., Garcia, W., & Orenha Ottaiano, A. (2023). *Using language data to learn about language: A teachers' guide to classroom corpus use*. The University of Queensland. https://doi.org/10.14264/3bbe92d

Poole, R. (2018). *A guide to using corpora for English language learners*. Edinburgh University Press. https://doi.org/10.13136/2281-4582/2020.i15.644

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.

Schaeffer-Lacroix, E. (2019). *Corpus et didactique de l'allemand. La langue à bras le corps*. Lambert Lucas.

Siepmann, D. (n.d.). *Digitalisierung im Englischunterricht – Data-driven learning*. Klett Akademie. Retrieved 17 July 2024, from https://www.klett.de/inhalt/digitalisierung-im-englischunterricht/data-driven-learning/9965

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins.

The University of Queensland. (2019). *Improving writing through corpora: Data-driven learning?* [Online course]. https://edge.edx.org/courses/course-v1:UQx+SLATx+2019/

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge. https://doi.org/10.4324/9781315715537

University of Queensland. (n.d.). 使用语料库提升英语写作:数据驱动学习 [Online course]. Retrieved 17 July 2024, from https://edge.edx.org/courses/course-v1:UQx+HASS0653N_7120_SLATx+2021_S1/about

Viana, V. (Ed.). (2023). *Teaching English with corpora: A resource book*. Routledge.

Viana, V., & O'Boyle, A. (2021). *Corpus linguistics for English for academic purposes*. Routledge.

Vyatkina, N. (Ed.). (2020). *Incorporating corpora: Using corpora to teach German to English-speaking learners*. University of Kansas Open Language Resource Center. https://opentext.ku.edu/corpora

M

MALL and Data-Driven Learning

Pascual Pérez-Paredes
Universidad de Murcia, Murcia, Spain

Abstract

This entry examines the contributions of data-driven learning (DDL) and broad data-driven learning (BDDL) to mobile-assisted language learning (MALL), and vice versa. The entry reviews the emergent research in this field and proposes critical ways to integrate DDL skills and corpus literacy in a new ecology of digital learning that takes advantage of mobile devices to situate language learning and language learners' needs across both formal and informal learning practices. (In this entry, informal learning is conceptualized as "everyday life [learning] activities whose main objective is not educational, [and] non-intentional and not leading to certification. Non-formal learning is [however] embedded in planned activities not always explicitly designated as learning (in terms of learning objectives, learning time or learning support), but which contain an important learning element. Non-formal learning is intentional from the learner's point of view. In particular, adult non-formal learning refers to the self-directed intentional learning that takes place in people's leisure time outside of both the workplace and the curriculum of formal learning activities". (Conole & Pérez-Paredes, 2017, p. 46). In this entry, the term *self-directed uses* is preferred when describing non-formal learning.

Keywords

Mobile-assisted language learning · DDL · Language data · Mobile devices · Apps · MALL

Introduction

Mobile-assisted language learning (MALL) is one of the areas in computer-assisted language learning (CALL) that has attracted increased interest over the last 15 years. This is due to a multitude of factors, including the rise and spread of mobile devices worldwide and the impact of apps on language learning, extramural language learning and self-directed uses (Zhang & Pérez-Paredes, 2021). This entry examines the contribution of different forms of DDL (prototypical DDL vs. Broad DDL, after Pérez-Paredes, 2024) in MALL contexts, presenting a critical scope of the field and its potential contributions to language learners' development of digital literacies.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_63-1

Mobile-Assisted Language Learning

Mobile-assisted language learning (MALL) examines the use of mobile technologies and devices to facilitate language learning. MALL encompasses a range of applications and tools that support learners in accessing language resources, engaging in interactive activities and practicing language skills anytime and anywhere. According to Stockwell (2022), MALL leverages context awareness and personalization to tailor learning experiences to individual needs.

The popularity of MALL, and particularly the widespread use of apps for language learning, is situated in a context where CALL has been gradually moving away from desktop-only environments to hybrid contexts where mobile devices, portability and connectivity play a key role (Gillespie, 2020; Pérez-Paredes & Zhang, forthcoming). This aspect deserves attention as much DDL research depends on desktop/laptop software-mediated classroom experiences (Pérez-Paredes, 2022) where the size of the screens is not a major constraint.

MALL research shows that using mobile devices seems to be effective for language learning. However, this finding needs to be contextualized. First, some meta-analyses have found that the research methodology used in some of the published papers on MALL (as indeed in other areas of CALL) displays unsophisticated designs that depend on quantitative methods with problematic samplings. Second, research shows that a mobile device per se does not seem to be enough to warrant language learning. Chen et al.'s (2020) meta-analysis of MALL research from 2008 to 2018 revealed a medium-to-high effect size of 0.722 for the effectiveness of mobile devices in language learning. However, other meta-analyses have questioned the validity of the research in the field. Sung et al. (2015) demonstrated that 70.7% of learners using mobile devices performed significantly better than those without. However, their meta-analysis of 44 studies from 1993 to 2013 found that the portability of mobile devices did not consistently enhance language learning outside the classroom. This suggests that most of the studies surveyed were conducted in institutionalized contexts with structured, curriculum-driven MALL (Pérez-Paredes & Zhang, 2022). Burston (2015) highlighted that many MALL studies suffer from short treatment durations, small sample sizes, lack of tracking of device use, poor control group descriptions, confounding variables and inadequate statistical analyses. Of the nearly 300 studies reviewed, only 35 met basic criteria for duration and sample size. If only 12% of MALL research is methodologically sound, we need a cautious and critical interpretation of findings to better understand the actual impact of MALL. Burston and Giannakou (2022) carried out a meta-analysis of research published between 1994 and 2019 and showed that MALL learning outcomes are overwhelmingly positive. The methodological issues detected in Burston (2015) remain. Their meta-analysis highlights that a large proportion of the experimental studies reported favourable results regarding language acquisition, particularly in vocabulary learning. Most of the research analysed, however, looks at the English language and is localized in Asian countries.

Apps and the Commercialization of Language Learning

MALL apps have become popular for different reasons: they are either free or affordable, they are easy to download and install and attractive in gamified contexts, which appeals to the tastes and expectations of younger generations of users (Godwin-Jones, 2011, 2021, 2023). Apps such as Duolingo have massively favoured informal learning of, largely, a reduced number of major world languages such as English, Chinese or Spanish (Wagner, 2019). However, the widespread availability of apps on the online app stores in mobile devices has led the way to an increasing popularization and widespread adoption of a few popular language learning solutions and the commercialization of language learning. This fact may risk equal access to language education, its quality, as well as the influence of profit-driven motives on educational practices (Godwin-Jones, 2021; Kern, 2021). At the time of writing, none of the major language learning apps integrate DDL in their

design (Li, 2024). This is a major obstacle to promoting critical skills in language learning as one of the challenges in MALL is promoting individualization through critical thinking and data literacy. It is in this area where DDL can contribute to empowering language learners with critical skills that can favour further lifelong formal and informal language learning, and a language analysis mindset that can enhance language learning experiences.

Mobile (Broad) Data-Driven Learning

Is it possible to bring corpus analysis and corpus literacy to a mobile phone by designing a DDL app? Some of the skills that DDL promotes are likely to be transportable to mobile devices (Pérez-Paredes et al., 2019), facilitating a critical approach to noticing patterns in language and evaluating statistical properties of language such as frequency, distribution or saliency (Pérez-Paredes, 2024). Besides, DDL works well with the emphasis on vocabulary learning in MALL (Lin & Lin, 2019). However, taking corpora to mobile devices such as smartphones has not been explored. One of the reasons lies in the fact that corpus consultation for language learning has been conceptualized mainly as a classroom activity in universities (Pérez-Paredes, 2022); informal, extramural language learning has received less attention.

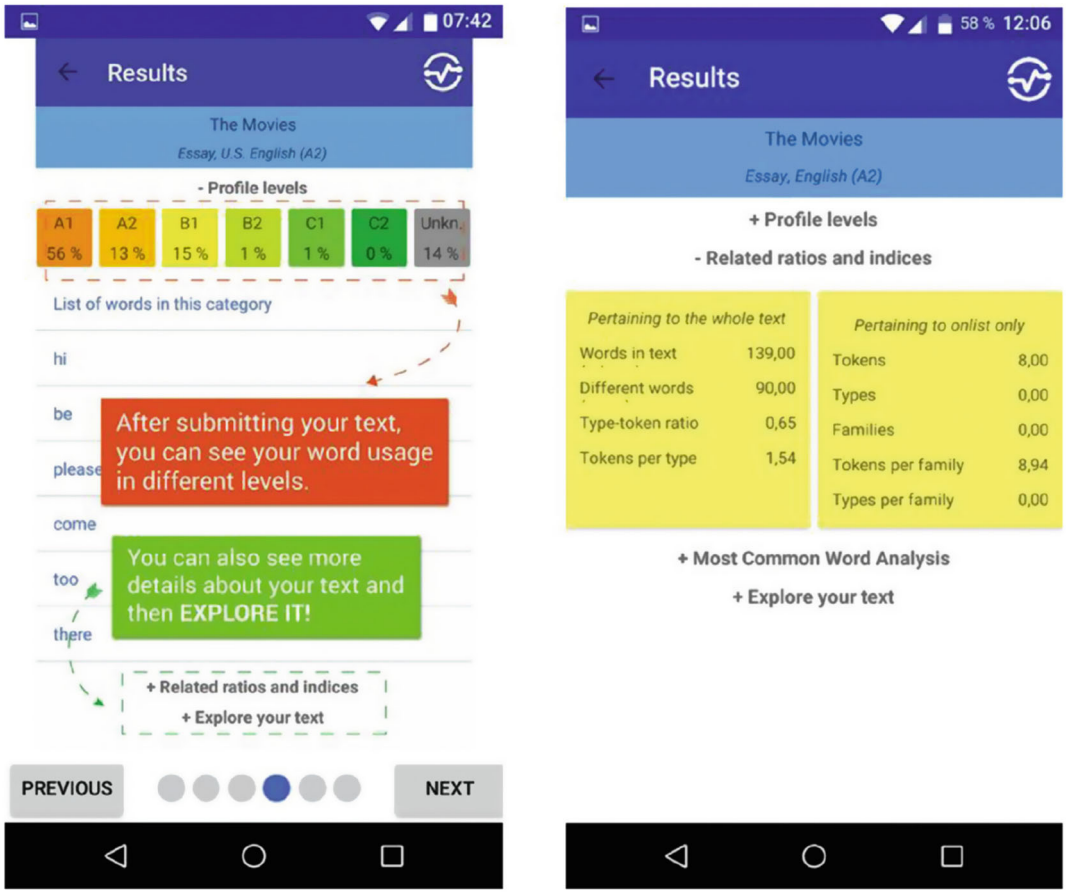
Data-driven learning can be defined as the use of corpus linguistics tools and techniques for learning or using an additional language (L2) (Boulton & Pérez-Paredes, 2024). A broader definition that includes resources that cannot be exclusively conceptualized, strictly speaking, as corpora fits in current approaches to MALL (Stockwell, 2022). Broad data-driven learning (BDDL) (Pérez-Paredes, 2024) encompasses learning practices that take place both in and outside the classroom, bringing together corpora and other resources to the learning experience. In this context, engaging with language data involves learners analysing and making sense of the linguistic patterns and structures found in the data, identifying recurring word combinations or collocations (Latif, 2021). Language data interpretation is essential for extracting meaningful information from corpora and other sources and applying it to language learning.

At the time of writing, web services such as Sketch Engine (http://www.sketchengine.eu) or Mark Davies' English Corpora (http://www.english-corpora.org) can be searched on tablets and mobile phones via browser, showing that at least web-based corpus consultation is available on mobile devices. However important this is, navigation and consultation in these two resources have not been tailored to the specifics of mobile access—working with your own data on a tablet continues to be challenging.

So far, mobile DDL has adapted resources and interfaces to the reduced screen of mobile devices and, to a lesser extent, has approached DDL in a way that exploits learners' mobility and ubiquitous learning. The still-limited research in this area is characterized by (1) increasing learners' language awareness about how lexicogrammatical items are used in authentic contexts; (2) personalizing the learning experience; and (3) favouring user-generated content and evaluation of such content.

Quan's (2016) early mobile DDL research on academic English involved the use of COCA with an app which displayed limited concordance lines and allowed broader searches with dictionary links. The option to display a reduced number of lines is driven by the limited space available on mobile screens and by the need to guide the user's attention to a manageable number of concordance lines. Pérez-Paredes et al. (2019) developed a mobile app that sought to shed light on how personalized and autonomous DDL could be conceptualized and tested by real language learners across different contexts and countries. The TELL-OP app (www.tellop.eu) helped learners evaluate word frequency in texts, enhancing their vocabulary and grammar awareness. The app provided access to collocations, synonyms and dictionary-like information, with the user clicking on any word in their own text. These were analysed automatically and each word classified into CEFR levels. Figure 1 offers two screenshots of the app tutorial where users are walked through the results screen.

While the app increased user sensitivity to vocabulary frequency, learners felt that more guidance was needed and suggested that



MALL and Data-Driven Learning, Fig. 1 TELL-OP app result screens

gamification could boost app engagement. This is precisely the area that Özbay and Özer (2017) explored in their research. The authors gamified DDL using Kahoot.com to improve EFL students' use of phrasal verbs, showing that gamification can benefit language learning with tablets or computers. The authors suggest that Web 2.0 applications can greatly help prototypical DDL to become popular for students who expect engaging web interfaces and the use of mobile devices in their language learning.

Contributions of DDL to MALL

By focusing on language data analysis, DDL can potentially maximize some of the benefits of MALL and complement desktop DDL. Mobile BDDL practices can contribute to making corpus literacy available and accessible to language learners who use their devices as an extension of their classroom learning experience. For those engaging in informal learning and extramural language learning, mobile BDDL can enhance language skills such as reading and listening by facilitating learners' implicit and inductive learning that exploits language data-driven analysis. Examples of BDDL resources include the following (Pérez-Paredes, 2024):

- Use of online dictionaries: monolingual, bilingual or bilingualized dictionaries (Zhang & Pérez-Paredes, 2024), collocation dictionaries, thesauri
- Frequency information of lexical items

- Tools for visual representation of lexical relations
- Google searches
- Large-language model (LLM) applications such as ChatGPT or Copilot
- A wide range of natural language processing technologies (NLPTs) (Pérez-Paredes et al., 2018) such as spellcheckers, vocabulary profiling tools or, among many others, readability indexes

Some of these resources have already been explored in Pérez-Paredes et al. (2019), but further research in this area is urgent to understand how DDL and BDDL can benefit from the affordances of MALL. Interacting and working with these resources will require that language learners enhance their digital literacy skills. Working with corpora and using DDL activities can help students reframe how language data is generated and presented and how it can be interpreted in language education. In this context, learners will have to understand data concepts such as types of data, data sources and formats; interpret data, e.g. frequency, and its impact on language use; analyse data using basic concepts drawn from corpus linguistics and talk about language data (Pérez-Paredes & Boulton, 2025). Some of the spaces where mobile BDDL can positively impact language learning are summarized in Table 1. They are drawn from Stockwell's (2022) conceptualization of MALL.

Prototypical DDL has explored just a fraction of the four spaces outlined in Table 1, perhaps because most DDL has taken place in university classrooms (Boulton & Cobb, 2017; Pérez-Paredes, 2022) under the supervision of lecturers in the context of one-term modules. There is room for DDL to inform MALL and for MALL to inform DDL. These conversations are urgent if we want to contribute to the next generation of language learning apps that will for sure integrate generative AI and text generation.

Mobile devices offer opportunities to exploit multimodal internet capabilities, integrating audio, video and text, which contributes to a new ecology of language learning where user-generated input along with mobile-friendly concordancing can collaborate to deliver authentic language learning experiences and connectivity with L2 user communities. One of the most challenging areas of analysis will be the role that traditional corpora will play in mobile DDL. As corpora begin to offer access through APIs and integration with LLMs, they will hopefully become available in apps that simplify and integrate searches in the new digital learning ecologies of language learners (Gee & Hayes, 2011).

According to Crosthwaite (2019), technology has played and will continue to play a crucial role in enabling access to corpora across educational

MALL and Data-Driven Learning, Table 1 Contributions of mobile DDL to language learning

MALL areas	DDL contribution
Flexibility and accessibility	Mobile devices offer learners the flexibility to engage in language learning anytime, anywhere, allowing for out-of-class practice and reinforcement of language skills
Personalization and adaptivity	Mobile DDL can potentially provide personalized learning experiences based on individual learner needs. By integrating adaptive learning, content, activities and feedback can be tailored to each learner's proficiency level and progress
Integration of multimedia	Video, audio, visual aids and enriched learning material can enhance language learning experiences. Bringing multimedia to DDL activities may increase the interest of younger learners
Collaborative learning	Thanks to MALL features for communication, peer interaction and project-based language learning, language educators can integrate social learning experiences and foster a sense of community among language learners, enhancing language practice. Working on DDL tasks collaboratively requires further attention
Continuous learning support	The integration of BDDL and MALL can provide learners with further opportunities for language practice and an increasing sense of expanding digital literacy skills

contexts. He suggests that a key advance would be a killer app for DDL that is multimodal and accessible on mobile devices, incorporating standardized corpora for different proficiency levels, genres and users. Unfortunately, language learners are still a long way from downloading this killer app and using it on their phones.

Conclusion

This entry has addressed both the impact of MALL on DDL and present and future uses of DDL in MALL. Although some steps have been taken in this area, it is necessary to advance our understanding of how mobile BDDL can complement what has already been accomplished in the last decades of research and classroom DDL practices (Boulton & Cobb, 2017; Pérez-Paredes, 2022). A mobile BDDL that integrates the advantages of MALL, namely the ease and ubiquity of access to language learning resources, the personalization of learning episodes as well as the maximization of in-class and out-of-class incidental learning and the integration of multimodal resources is possible. However, apart from the development of MALL-friendly applications, the main challenge ahead will be to inform digital literacy research and practices with existing and future insights from corpus and data literacy that illustrate how learners' analysis of statistical properties of language such as frequency, distribution and saliency can be integrated in mobile BDDL.

Cross-References

- ▶ Concordancers for DDL
- ▶ Corpus Literacy and DDL
- ▶ Dictionary Use in DDL
- ▶ Emerging Technologies and New Methodologies in DDL
- ▶ Future Challenges and Opportunities for DDL
- ▶ Learner Autonomy and DDL
- ▶ Multimodal Tools and Corpora for DDL
- ▶ Pedagogical Corpora and DDL
- ▶ Web as Corpus for DDL
- ▶ Young Learners and DDL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Pérez-Paredes, P. (2024). Data-driven language learning. In R. Hampel & U. Stickler (Eds.), *The Bloomsbury handbook of language learning and technology* (pp. 212–225). Bloomsbury.

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL, 27*(1), 4–20. https://doi.org/10.1017/S0958344014000159

Burston, J., & Giannakou, K. (2022). MALL language learning outcomes: A comprehensive meta-analysis 1994–2019. *ReCALL, 34*(2), 147–168.

Chen, Z., Chen, W., Jia, J., & An, H. (2020). The effects of using mobile devices on language learning: A meta-analysis. *Educational Technology Research and Development, 68*(4), 1769–1789. https://doi.org/10.1007/s11423-020-09801-5

Conole, G. & Pérez-Paredes, P. (2017). Adult language learning in informal settings and the role of mobile learning. In Yu, S., Ally, M., & Tsinakos, A. (Eds.) Mobile and ubiquitous learning. An international handbook. Springer, 45–58

Crosthwaite, P. (2019). Afterword. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 228–229). Routledge. https://doi.org/10.4324/9780429425899-13

Gee, J., & Hayes, E. (2011). *Language and learning in the digital age*. Routledge. https://doi.org/10.4324/9780203830918

Gillespie, J. (2020). CALL research: Where are we now? *ReCALL, 32*(2), 127–144. https://doi.org/10.1017/S0958344020000051

Godwin-Jones, R. (2011). Mobile apps for language learning. *Language Learning & Technology, 15*, 2–11. https://doi.org/10125/44244

Godwin-Jones, R. (2021). Evolving technologies for language learning. *Language Learning & Technology, 25*(3), 6–26. https://doi.org/10125/73443

Godwin-Jones, R. (2023). AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology, 27*(2), 6–27. https://doi.org/10125/73501

Kern, R. (2021). Twenty-five years of digital literacies in CALL. *Language Learning & Technology, 25*(3), 132–150. https://doi.org/10125/73453

Latif, M. M. M. (2021). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL, 33*(1), 34–48. https://doi.org/10.1017/S0958344020000129

Li, C. (2024). A review of theories, pedagogies and vocabulary learning tasks of English vocabulary learning apps for Chinese EFL learners. *Journal of China Computer-Assisted Language Learning*, advance online publication. https://doi.org/10.1515/jccall-2023-0026.

Lin, J. J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning, 32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Özbay, A. Ş., & Özer, M. (2017). Effects of data-driven learning on collocational competence of EFL learners in a Turkish context. *The International Journal of Research in Teacher Education, 8*(3), 27–39.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pérez-Paredes, P. (2024). Data-driven learning in informal contexts? Embracing Broad Data-driven learning (BDDL) research. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 210–226). Routledge. https://doi.org/10.4324/9781003413301

Pérez-Paredes, P., & Boulton, A. (2025). *Data-driven learning in and out of the language classroom*. Cambridge University Press.

Pérez-Paredes, P., & Zhang, D. (2022). Mobile assisted language learning: Scope, praxis and theory. *Porta Linguarum, 4*, 11–25. https://doi.org/10.30827/portalin.vi.21424

Pérez-Paredes, P., & Zhang, D. (forthcoming). Mobile dictionaries and apps for vocabulary learning. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (2nd ed.). Routledge.

Pérez-Paredes, P., Ordoñana, C., & Aguado, P. (2018). Language teachers' perceptions on the use of OER language processing technologies in MALL. *Computer Assisted Language Learning, 31*(5–6), 522–545. https://doi.org/10.1080/09588221.2017.1418754

Pérez-Paredes, P., et al. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System, 84*, 145–159. https://doi.org/10.1016/j.system.2019.06.009

Quan, Z. (2016). Introducing "mobile DDL (data-driven learning)" for vocabulary learning: an experiment for academic English. *Journal of Computers in Education, 3*(3), 273–287.

Stockwell, G. (2022). *Mobile-assisted language learning*. Cambridge University Press. https://doi.org/10.1017/9781108652087

Sung, Y. T., Chang, K. E., & Yang, J. M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review, 16*, 68–84. https://doi.org/10.1016/j.edurev.2015.09.001

Wagner, J. (2019). Towards an epistemology of second language learning in the wild. In J. Hellermann, S. W. Eskildsen, S. Pekarek Doehler, & A. Piirainen-Marsh (Eds.), *Conversation analytic research on learning-in-action* (pp. 251–271). Springer. https://doi.org/10.1007/978-3-030-22165-2_10

Zhang, D., & Pérez-Paredes, P. (2021). Chinese postgraduate EFL learners' self-directed use of mobile English learning resources. *Computer Assisted Language Learning, 34*(8), 1128–1153. https://doi.org/10.1080/09588221.2019.1662455

Zhang, D., & Pérez-Paredes, P. (2024). Chinese EFL learners' use of mobile dictionaries in reading comprehension tasks. *System, 121*, 103221. https://doi.org/10.1016/j.system.2024.103221

R

Reading and Data-Driven Learning

Gregory Hadley
Niigata University, Niigata, Japan
Kellogg College, Oxford, UK
University of Oxford, Oxford, UK

Abstract

This entry examines the application of Data-Driven Learning (DDL) in second language reading instruction, with a particular focus on the diverse methodologies and outcomes associated with this approach. It critically reviews existing research, highlighting the moderate success of DDL in improving writing, vocabulary, and grammar. However, less emphasis has been placed on its efficacy in developing reading skills. By surveying recent studies, this entry identifies various strengths and limitations of DDL in intensive, extensive, and task-based reading contexts. Practical suggestions for effectively implementing DDL in second language classrooms are provided, emphasizing the importance of tailoring approaches to learner needs and educational environments.

Keywords

Second language reading · Inductive thinking · Critical survey of DDL · Pedagogic strategies

Introduction

Data-driven learning (DDL) is a form of computer-assisted language learning (CALL) that was created and developed during the late 1980s and 1990s by the British applied linguist Tim Johns at the University of Birmingham. DDL builds upon the ability of computers to quickly find and organize discrete data points in a cogently structured, meaningful manner using a software programme called a concordancer. Johns would search for specific examples of lexical items and grammatical structures in large databases of pertinent English language materials (called corpora) and provide this information to non-native English speaking graduate students, and encourage them to inductively notice patterns from the corpus data, and to then generalize their insights towards improving their academic writing (Johns, 1991). From its early start as a teaching tool in one-to-one teaching contexts, the use of DDL in second language classrooms has since spread around the world, though it is primarily found in higher education institutional environments. Meta-analytical studies on the past 30 years of DDL research suggest that it has been used with moderate success in the enhancement of second language writing, the expansion of lexical knowledge, and the acquisition of new grammatical forms (Boulton & Cobb, 2017; Lusta et al., 2023).

However, less attention has been given to the effectiveness of DDL with receptive skills such as reading, even though some researchers have

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_64-1

started to focus on this largely uncharted area. This entry will survey a sample of these studies in an effort to highlight what is currently known about the efficacy of DDL with reading in the second language. It will also consider, based on the findings of current research, what is entailed for successful implementation of DDL with second language reading. To begin, however, it is important first to address the notion of "reading" within the context of ELT.

Clarifying the Notion of Reading

Reading in the realm of second language instruction is a rather general term for a range of approaches to the practice. For example, there is intensive reading, in which learners read short texts with the goal of learning vocabulary, grammatical structures, and understanding specific information. Students are taught skills such as skimming, scanning, and summarizing. Comprehension and critical thinking about the content are also frequently highlighted (Morris, 1972; Zhang & Yuan, 2022). Extensive reading, on the other hand, allows learners to read large amounts of text that are both comprehensible for their level of proficiency and which complement their personal interests. The focus here is on fostering overall language proficiency, greater speed and fluency, and vocabulary enhancement within the context of reading for enjoyment in the second language (Day & Bamford, 2004; Waring & McLean, 2015).

Other ways of approaching reading in the second language link it to other aspects of language instruction, such as task-based reading, which integrates the skills needed for completing various sets of short projects or activities in order to improve the learners' overall second language proficiency (Xu, 2021). Interactive reading is similarly related, as readers engage with texts as a springboard for group and pairwork discussions, and also as a stimulus for reflective writing (Mera Rivas, 1999). The purpose here is to enhance comprehension and retention through active collaboration.

This list of reading types in ELT is not comprehensive, as one could also consider other types of reading within the context of digital platforms as opposed to the medium of paper. Various gradations as well as combinations of the above approaches are also possible, such as using a task-based approach to extensive reading (Green, 2005). For the purposes of this entry, however, the conceptual shorthand of intensive, extensive, interactive, and task-based should be sufficient, and whenever possible, the type of reading that was linked with DDL will be noted in the following survey of studies.

A Critical Survey of Research on Reading and DDL

As noted earlier, papers exploring the effects of DDL in second language reading instruction are relatively few. The following is an overview of some of the major work that has been carried out thus far. In this survey, the type of reading studied, main findings, relative strengths, shortcomings, and implications for each of the studies are considered.

Research by Johns et al. (2008) explored the integration of CALL programmes in teaching English through children's literature, focusing on task-based reading. The study utilized various CALL programmes together with a DDL approach in an elective course for Taiwanese high school students, aimed at enhancing their English proficiency through engaging with the novel *Swallows and Amazons*. The findings indicated that the experimental group, which used CALL and DDL activities, showed significant improvements in their term exam scores, reading speeds, and reading comprehension compared to the control group. The strengths of the study include the use of innovative CALL programmes to enhance language learning and the integration of students' native language to aid comprehension. However, the limitations of the study are in the small sample size and the short duration of the study, which might have limited the generalizability of the findings. Additionally, it was difficult to isolate the direct impact of CALL activities due to potential external variables influencing students' performance. Overall, the study suggests that integrating DDL and CALL with children's literature

has some potential for improved learning through second language reading.

Rasikawati's (2020) doctoral thesis explored the use of DDL in an English for Academic Purposes (EAP) intensive reading course, and sought to determine the effectiveness of corpus-based DDL compared to regular vocabulary and grammar instruction among 153 freshmen in an urban Indonesian university. The findings revealed that while DDL significantly improved syntax knowledge, traditional instruction was more effective for vocabulary acquisition. The study's strengths include its large sample size and randomized experimental design, which aided in providing robust data on DDL's mixed results, but as this study considered not only reading comprehension, but also listening skills, syntax, and lexical learning, the study may have attempted to explore too many receptive features. A significant shortcoming was that the four classes of research participants participating were taught by teaching assistants, and it was unclear how skilled they were in teaching DDL, thereby calling into question whether the presentation of traditional instruction and DDL were given comparably equal levels of expertise and enthusiasm.

In Japan, Miura and Satake (2024) investigated the implementation of DDL in a university English reading and writing course, focusing on intensive reading. The research applied various DDL tasks to enhance vocabulary, collocations, and grammar comprehension among first-year students. Their findings suggest that DDL is particularly effective for lexical comprehension and verb-noun collocations, but only when students are given specific guidance, as inductive thinking seemed to be too challenging for most. The time-intensive nature of DDL tasks and their inconsistent impact on what they described as practical language use also presented themselves significant challenges to using DDL with reading. The strengths of the study include its detailed qualitative analysis and the practical applications of DDL tasks in real classroom settings. The limitations lie in the lack of quantitative validation beyond that of counting the times when students used collocations, and also the time-consuming nature of setting up and then doing the DDL activities. The authors conclude this last point is an indication of the difficulties that arise when using DDL for prolonged periods in the classroom. Overall, the study sidesteps directly addressing the efficacy of DDL for reading, and instead focuses on the potential of DDL in fostering independent learning, enhancing the recognition of collocations, and increasing vocabulary recognition, all of which could contribute to better reading skills.

Al-Mahbashi et al. (2015) examined the impact of DDL on the receptive vocabulary knowledge of second-year Yemeni university learners in an English for Academic Purposes (EAP) reading course, and focused on intensive reading. Their study compared the effectiveness of DDL with traditional dictionary use for learning word meanings and collocations among 60 female second-year English students. Findings indicate that DDL significantly enhanced vocabulary knowledge and retention more than dictionary use, with notable improvements observed in both posttest and delayed posttest results. While this study also involved a small and homogeneous sample size with a short period of time for the actual study, all of which place limits on the generalizability of the reported results, the strengths of this study lay in its empirical approach, robust design, and clear evidence arguing for of DDL's effectiveness in second language vocabulary acquisition, which is vital for success in intensive reading.

From Spain, a study by Curado Fuentes (2015) investigated the impact of DDL on the intensive reading comprehension of news texts among 50 lower-level business English students. The research divided the students into control and experimental groups and compared the effectiveness of DDL, using keyword in context (KWIC) concordances, with traditional paper-based reading comprehension activities. The findings indicate that DDL activities significantly improved students' comprehension of news keywords and their subsequent decoding of information in the texts. Similar to the study by Al-Mahbashi et al. (2015) Curado Fuente's research also featured a rigorous empirical approach. It discusses the practical applications of using DDL with reading in real classroom settings, and the detailed feedback

from students through logs and questionnaires provided further insight into their learning processes and attitudes. The study, however, also faced certain limitations, such as the short duration and the initial difficulties students faced in using the concordance software, which may have affected the results. While keeping these points in mind, the findings of this study are nevertheless promising.

Within the area of extensive reading, Hadley and Charles (2017) used a corpus created from the Oxford Bookworms Graded Readers series, and implemented DDL at a Japanese university with pre-intermediate level English reading classes that featured both Japanese and international university undergraduates. The experimental group used DDL materials from the Oxford Bookworms Graded Readers corpus to aid in the comprehension of vocabulary and to increase their reading speeds. The control group read their graded readers without any DDL input. While extensive reading was the locus of instruction, there was significant time given in both classes to interactive reading, where learners shared with each other and discussed the books they were reading. Posttest findings from this study showed significant improvements for both groups, but it was the control group that achieved greater improvements in reading speed and lexical learning. Qualitative data suggested that the experimental group struggled with engagement due to the perceived complexity and time demands of DDL tasks. The strengths of this study were in the use of DDL with diverse groups of research participants and its robust mixed-methods design that gave a fuller view of how students interacted with DDL while reading. The relatively small sample sizes, difficulty that students had in engaging with DDL, and the relative lack of direct access that students in the experimental class had to the corpus (due to limitations imposed by Oxford University Press in order to protect their copyrighted material) presented significant barriers which may have skewed the results in a manner similar to the study by Curado Fuentes (2015). Hadley and Charles (2017) conclude that with learners at lower levels of second language proficiency, a softer, more teacher-directed approach to DDL, emphasizing affective engagement, may be more effective for lower proficiency learners.

Gordani (2013) examined the effects of integrating online corpora in intensive reading comprehension classrooms on the vocabulary development of 70 first-year university Iranian EFL students majoring in mathematics. Using a randomized pretest-posttest control group design, it was found that the experimental group, which received corpus-based instruction, significantly outperformed the control group in vocabulary acquisition as measured by the posttest. Gordani's research, which features a rigorous experimental design and the practical application of corpus-based tools, provides compelling empirical evidence for the effectiveness of DDL in vocabulary development as it relates to intensive reading. Like other studies in this survey, the homogeneous sample (this study also featured female-dominated classes) and short time period during which the study was conducted, should temper any broad-based claims from the findings. Nevertheless, this study provides another example of how corpora in EFL classrooms might be used to enhance vocabulary learning, which in turn can positively affect the process of second language reading.

Hadley and Hadley (2021), reported on the results of a 4-year mixed-methods investigation of the impact of DDL on extensive reading proficiency among second language learners at a Japanese national university. Building upon the findings of Hadley and Charles (2017), and using the same corpus of graded readers created from their study, this project sought to discover the parameters that could make DDL more amenable to learners in the Japanese university context. It was carried out in three stages and included both quantitative and qualitative research methodologies. The quantitative aspect involved pretest-posttest designs, while the qualitative part comprised classroom observations, repertory grids, and follow-up interviews. They found that forms of DDL that used more explicit features of the pedagogy such as studying concordance lines with a key word in context or inductive thinking together with the teacher, were most successful in small groups that allowed for one-on-one

instruction. The personal feedback and less pressure afforded by small groups played important roles in the students' acceptance of DDL, which resulted in improved reading speeds and significant levels of lexical learning. The authors noted that the small group conditions in which explicit DDL was used were similar to those of Tim Johns when he used DDL in tutorials with individual students.

The authors contrasted explicit DDL with what they called implicit DDL, which involved the creation of teaching materials designed to enhance their students' extensive reading with lexical learning and the presentation of collocational knowledge hidden within formats looking more like what can be seen in traditional second language textbooks. They found that implicit DDL worked best in larger classes, and resulted in faster reading speeds and a greater volume of material read compared to control groups who studied extensive reading without DDL. The study's main findings highlighted the effectiveness of DDL in improving lexicogrammatical proficiency and reading speeds, particularly with implicit DDL. However, explicit DDL was often viewed unfavourably by students, and seen as a barrier to reading enjoyment and interaction. The strengths of the study include its longitudinal design, mixed-methods approach, and practical classroom applications that showed examples of the types of teaching materials created for learners. Shortcomings included the small sample sizes in some stages, variability in classroom conditions, and as with other studies, the challenges students had in engaging with the DDL materials. The implications for DDL and reading suggest that while DDL can enhance reading proficiency, its implementation needs to be carefully tailored to the learners' context, with consideration for balancing explicit instruction and reading enjoyment.

Implementing Data-Driven Learning in Second Language Reading

Based on the findings of this empirical research, some tentative suggestions can be made with regard to how DDL might be implemented in reading courses, both from the perspectives of the teachers and the students.

For teachers, it is crucial to assess students' willingness and ability to access specialist software and corpus data. Teachers need to evaluate their students' readiness to engage in inductive thinking, which is essential for effectively integrating DDL with second language reading. Depending on the learners' target language proficiency and IT literacy, teachers may need to design activities that provide explicit guidance in noticing, interpreting, and using new collocations and lexical items. Ensuring the availability of a suitable corpus that complements the students' reading materials is also essential. In larger classes, implicit presentations of DDL may be more effective, especially in extensive reading contexts. In contrast, explicit DDL may be more suitable for intensive reading contexts. The approach to DDL should be tailored to the students, their learning environment, and the type of reading they are engaged in. A tailored approach can lead to more effective language learning, as students will be less likely to feel overwhelmed by the often concordance-heavy presentation of the target language in DDL.

From the second language learner's perspective, engaging with DDL while reading requires an openness to new tools and approaches. Students need to recognize and appreciate how corpora provide authentic, real-life examples that can deepen their linguistic awareness. While some may not have been previously encouraged to make inferences, a willingness to reflect upon new language data and to collaborate with both peers and teachers can enhance their DDL experience. An interactive approach can then help in more learners attaining faster reading speeds, better-targeted lexical and grammatical development, and deeper levels of second language comprehension.

Conclusion

It must be noted that the critical review of pertinent studies on reading and DDL has been carried out within the context of small, homogeneous group samples. In addition, numerous variables

related to the quality of instruction, and significant variations in classroom conditions are issues that we must keep in mind when considering the impact of DDL on second language reading. This, of course, is a common feature of classroom research. Nevertheless, it can be tentatively concluded that if learners are appropriately scaffolded in their use of DDL, and if instruction is provided in a consistent manner, then DDL can be effective in improving reading skills for at least some students some of the time. However, it is uncertain at this point as to whether it will be effective with most students most of the time. More studies, both small and large scale, with longer research duration periods, will be needed in order to gain a clearer picture of how and when DDL can positively impact second language reading. Achieving this would require more language teachers to include a DDL element in their reading classes. In doing so, they will discover new ways of making the implementation of DDL more effective for a larger body of second language learners.

Cross-References

- ▶ Inductive and Deductive Teaching, Guided Induction in DDL
- ▶ Language for Academic Purposes and DDL
- ▶ Learner Attitudes, Styles, Strategies and Behaviours in DDL
- ▶ Outcomes: DDL for Learning
- ▶ Pattern-Hunting and Pattern-Defining in DDL
- ▶ Techniques for DDL Beginners
- ▶ Tim Johns

References

Al-Mahbashi, A., Noor, N., & Amir, Z. (2015). The effect of data driven learning on receptive vocabulary knowledge of Yemeni University learners. *3L: Language, Linguistics and Literature—The Southeast Asian Journal of English Language Studies, 21*, 13–20.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Curado Fuentes, A. (2015). Exploiting keywords in a DDL approach to the comprehension of news texts by lower-level students. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 177–197). John Benjamins.

Day, R., & Bamford, J. (2004). *Extensive reading in the second language classroom*. Cambridge University Press.

Gordani, Y. (2013). The effect of the integration of corpora in reading comprehension classrooms on English as a foreign language learners' vocabulary development. *Computer Assisted Language Learning, 26*(5), 430–445. https://doi.org/10.1080/09588221.2012.685078

Green, C. (2005). Integrating extensive reading in the task-based curriculum. *ELT Journal, 59*(4), 306–311. https://doi.org/10.1093/elt/cci059

Hadley, G., & Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Language Learning & Technology, 21*(3), 131–152.

Hadley, G., & Hadley, H. (2021). Exploring the impact of data-driven learning in extensive reading. In P. Perez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 149–176). John Benjamins.

Johns, T. (1991). Should you be persuaded – Two examples of data-driven learning materials. *English Language Research Journal, 4*, 1–16.

Johns, T., Lee, H. C., & Wang, L. X. (2008). Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning, 21*(5), 483–506. https://doi.org/10.1080/09588220802448006

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*(12), 1–19. https://doi.org/10.1016/j.heliyon.2023.e22731

Mera Rivas, R. M. (1999). Reading in recent ELT coursebooks. *ELT Journal, 53*(1), 12–21. https://doi.org/10.1093/elt/53.1.12

Miura, A., & Satake, Y. (2024). Implementation of data-driven learning (DDL) in a course for English reading and writing. *多言語教育実践ジャーナル (Journal of Multilingual Pedagogy and Practice), 4*, 1–17. www.rikkyo.repo.nii.ac.jp

Morris, J. (1972). Intensive reading. *ELT Journal, 27*(1), 38–47. https://doi.org/10.1093/elt/XXVII.1.38

Rasikawati, I. (2020). *Corpus-based data-driven learning in reading English for academic purposes* [Doctoral thesis, Seattle Pacific University]. Seattle, OR (USA). https://digitalcommons.spu.edu/soe_etd/55/

Waring, R., & McLean, S. (2015). Defining extensive reading. *Extensive Reading in Japan*, *8.1*(May), 7.

Xu, Q. (2021). Incorporating reading circles into a task-based EAP reading scheme. *ELT Journal, 75*(3), 341–350. https://doi.org/10.1093/elt/ccab012

Zhang, H., & Yuan, R. (2022). Rethinking the infusion approach to teaching critical thinking in reading in Chinese tertiary English classes. *RELC Journal*, 1–15. https://doi.org/10.1177/00336882221087466

T

Types of Corpora in Data-Driven Learning

Pascual Pérez-Paredes and
Carlos Ordoñana-Guillamón
Universidad de Murcia, Murcia, Spain

Abstract

This entry explores various types of corpora used in data-driven learning (DDL). It categorises corpora based on size (macro- and micro-corpora), nature (written, spoken and multimodal) and thematic cohesion. The entry highlights the importance of representativeness and practical considerations in corpus design, particularly for educational purposes. The document also discusses the advantages of different corpus types and their applications in enhancing vocabulary, grammar and contextual language understanding. It underscores the flexibility and adaptability of DDL to meet diverse learner needs and teaching goals.

Keywords

DDL · Corpora · Corpus linguistics · Corpus-based instruction · Language teaching

Introduction: Corpora for Data-Driven Learning

Data-driven learning (DDL) promotes inductive learning by fostering the autonomous exploration of authentic linguistic data. DDL situates the learner as a sort of researcher (Bernardini, 2004) who discovers linguistic patterns in attested uses of language. This approach to foreign language instruction has often been implemented using corpora selected or provided by the teacher in the form of either hands-on activities or printed concordance lines.

Corpora are the most often used language resource in DDL practices, as they facilitate the direct observation of frequently occurring lexicogrammatical items across different contexts of use. The emergence and widespread adoption of increasingly sophisticated computing devices has streamlined this process through the use of corpus management tools such as Sketch Engine, BNCWeb or AntConc (Boulton & Vyatkina, 2021).

Across the years, different corpora have been used in DDL research and classroom practices. The most widely utilised due to their size, their representativeness and their availability are the Brown Corpus, the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), the Michigan Corpus of Academic Spoken English (MICASE), the Michigan Corpus of Upper-level Student Papers (MICUSP) and the British Academic Written English Corpus

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_68-1

(BAWE) (Boulton & Cobb, 2017; Boulton & Vyatkina, 2021; Pérez-Paredes, 2022). Some studies even point to the *Google* search engine as a 'corpus' consultation tool (Conroy, 2010; Geluso, 2013; Pérez-Paredes et al., 2012). While corpora such as the BNC or the BAWE are compiled by a team of experts, usually funded by research bodies, there are tools nowadays that facilitate that individual language teachers or small teams of teachers put together their own corpora for DDL. Sketch Engine offers tools to find texts and download them using target words or specific domains or websites. If interested in research articles, AntCorGen can find and download open science research papers that can be grouped into domains and used as a corpus for DDL. (This can be downloaded from https://www.laurenceanthony.net/software/antcorgen/) A more flexible option is to use a web scraping Python script. This requires some basic understanding of coding and the use of a Python notebook such as Google Colab. (Google Colab is a free online Python development environment where scripts can be run ant tested. There are many Python libraries that will allow you to run a script that can collect the data for you. Also, there are R scripts that find and download texts from the Web if the user is more familiar with R language than with Python.)

There is, however, a wealth of different corpora at the disposal of the user beyond those mentioned above. This entry provides a comprehensive description of the different types of corpora available in DDL. Most of the research and practices discussed in this entry draw on the use of English as a foreign language (EFL) literature. For a discussion on corpora for French and Spanish, see Pérez-Paredes and Chambers (in press).

Types of Corpora for DDL

A corpus is 'a large and principled sample of texts designed to represent a target domain of language use (e.g., a language, dialect, or register)' (Egbert et al., 2022, p. 7). In this definition, 'text' involves naturally occurring, recognizably self-contained and functional language, either written or spoken.

There are different ways to conceptualise corpora. Most of the conceptualisations draw on corpus typologies in the corpus linguistics literature. Davies (2015) provided a categorisation of corpora based on their overall size, such as small corpora (up to 1–5 million words, citing the Brown corpus as an example), moderately sized corpora (up to 100 million words, such as the BNC) or large corpora (the 450-million word COCA—now 1 billion words), referring as well as the whole World Wide Web as an extremely large corpus likely to be explored through Google searches. For Boulton (2017), corpora for DDL can be classified as written corpora (e.g. books, research articles or essays), spoken corpora (e.g. conversations, interviews or radio shows), multimedia corpora (e.g. video interviews with transcripts), learner corpora (texts produced by language learners), specialised corpora (covering specific fields or genres) and general corpora (usually large collections of texts that represent a wide range of language use). Each type of corpus offers unique advantages and can be selected by language teachers based on specific learning objectives and contexts of the language teaching programme.

Boulton and Cobb (2017) distinguished between public (publicly available, often monolingual corpora), local (usually smaller corpora created and compiled to serve specific language learning contexts, also typically monolingual) and parallel corpora (consisting of collections of texts in two or more languages aligned at sentence level to provide translation or contrastive language analysis). Their meta-analysis shows that in within-group designs, all three types of corpora yielded strong effect sizes, while in between-group designs, the use of public monolingual corpora offered lower mean effect sizes. Large corpora above 100 million words yielded medium effect sizes in between-group designs. Overall, the choice of corpora does not seem to affect the main trend observed in empirical DDL studies that language gains are expected.

Ballance (2017) identified different types of corpora for DDL, such as general language corpora (very large collections covering a wide variety of texts and genres), specialised corpora

(focusing on particular topics such as medical, legal or academic language), learner corpora (language data produced by language learners), corpora of error-annotated data (designed to highlight language errors produced by learners), corpora of learner language interlanguage (reflecting learner language at different transitional stages of proficiency) and parallel corpora. Aston (1999) and Pérez-Paredes (2022), in turn, have acknowledged the existence of ad hoc corpora, which are often smaller and compiled to serve specific purposes in the context of DDL projects.

The following subsections provide a rationale for the categorisation of corpora for DDL while aiming to avoid unnecessary overlap by placing corpora in a three-dimensional, interrelated space: one related to the *size* of the corpus, another related to the *nature* of the corpus and the other dimension related to its *thematic cohesion.*

Corpus by Size

A corpus can range from a few texts to several billion words. Given this wide range, the question of representativeness often arises in the literature. It has been suggested that the process of building a corpus should not just focus on the number of words and texts but also on a clear definition of the target population and careful sampling methods to include the full variability of the target population (Biber, 1993). Sampling should be based, therefore, on criteria like demographics or sociocultural factors, and variability refers to the interaction between linguistic and situational parameters (Raineri & Debras, 2019). Additionally, the quality of selected texts, that is, the reliability of their sources, is crucial for assessing a corpus' representativeness, particularly when the sampling relies on online environments (Corpas Pastor & Seghiri Domínguez, 2010). Representativeness should be evaluated in terms of how well the corpus allows for generalisations about a specific domain and the distribution of linguistic variables within that domain (Egbert et al., 2022). Here, 'domain' refers to the language phenomena being studied, such as register or dialect, and 'distribution' refers to how frequently a linguistic feature appears in the texts of the corpus. Practicality is also important, especially when corpus designers face resource or time constraints (Reppen, 2010). This is particularly relevant for teachers compiling a corpus for classroom use. Such considerations must be pondered when selecting corpora for educational purposes, as the appropriate size of a corpus needed for DDL practices will be dependent on the specific needs of the users and what the corpus will be used for.

Macro-corpora

Macro-corpora aim for general, genre-balanced representativeness of a particular linguistic domain. Given the amount of work and resources involved in the process, large and very large corpora are usually built, compiled and annotated within the framework of research projects in linguistics and applied linguistics, often in collaboration with commercial entities and public institutions. The BNC itself is an example of this, as the British National Corpus project was led by the Oxford University Press, alongside publishers such as Longman and Chambers (Leech, 1992), while COCA is regularly updated and as of the writing of this entry it contains more than one billion words (https://www.english-corpora.org/coca/). Further—and larger—examples can be found in the use of *Google* as a corpus search engine (Pérez-Paredes et al., 2012), the iWeb Corpus or the TenTen corpus family (Jakubíček et al., 2013), amounting to several billion words each, although the notion of using the 'Web as a corpus' has been questioned in terms of actual representativeness, precisely because of the prioritisation of size over principled design (Egbert et al., 2022).

Using large corpora for language teaching practices through hands-on exploration of concordance lines has been found to positively impact language learning outcomes (Boulton & Cobb, 2017). Römer (2010) outlined how general corpora can be implemented in DDL practices to serve as the basis for teachers to design tasks that highlight specific language features tailored to their learners' specific needs (i.e. future time expressions, modal verbs or if-clauses). Paquot (2022) used macro-corpora as a reference for learners to explore and identify, among others, collocations and patterns, as well as a means to

compare learners' linguistic production to that of native users of the target language.

Large corpora offer language teachers opportunities to explore genre variation and the local grammar of target lexicogrammatical items. They are convenient, often available online and, sometimes, annotated, which facilitates the selection of highly contextualised concordance lines for vocabulary acquisition, writing skills development or grammar analysis (Boulton & Vyatkina, 2021).

Micro-corpora

The appropriate size of a corpus depends on the purpose it is designed for. Micro-corpora refer to collections of smaller-in-scope texts that are often the result of a needs analysis and which are adapted to the specificities of the classrooms where the corpus is used. This category may include usual practices in DDL-based instruction such as ad hoc corpora (Aston, 1999), that is, a corpus compiled to address a target language point; pedagogic and pedagogically-oriented corpora (Braun, 2005), which are collected on the basis of a needs analysis to aid language instruction (Szudarski, 2022), often created and exploited with pedagogic purposes (Farr & Hagen Karlsen, 2022); or personal or student-led, do-it-yourself (DIY) corpora, which can help with discourse analysis-based language instruction (Fahey Palma, 2022).

Micro-corpora are needs-driven and, especially in the cases of ad hoc and DIY corpora, tend to be usable only in the contexts they were designed for. There are, however, examples of pedagogic corpora such as ELISA (Braun, 2007) or SACODEYL (Widmann et al., 2011) which are publicly available. Micro-corpora have been featured repeatedly in the DDL literature. For instance, ad hoc corpora have been used in DDL-based translation instruction (Corpas Pastor, 2001) and to foster reading comprehension in EFL contexts (Curado Fuentes, 2023); pedagogically oriented corpora have been studied with regard to vocabulary acquisition (Szudarski, 2022), while pedagogic corpus implementation has been observed in secondary education lesson plans (Farr & Hagen Karlsen, 2022; Pérez-Paredes, 2019). Moreover, DIY corpora have been found to increase learners' motivation (Lee & Swales, 2006) while helping them with hypothesis verification and intuition checking on language and disciplinary usage as well as providing the opportunity to compare their linguistic production to that of experts (Charles, 2018).

Corpus by Nature

As noted before, a corpus may cover different aspects of linguistic production. As such, it becomes necessary to establish a difference between corpora comprising written, spoken or multimodal language.

Written Corpora

Written corpora cover different genres and styles, allowing users to focus on a wide range of nuances that govern the stylistic and discursive features of language use across written genres. As such, written corpora are by a wide margin the most commonly available type of corpora due to how easy they are to find, compile and annotate for further analyses (Boulton, 2017). Most of the available corpus-related software is designed to interact with language mainly in written form only. In fact, Boulton and Vyatkina's (2021) longitudinal examination of DDL research identified English-Corpora.org, Sketch Engine and BNCWeb as the most commonly used online corpus management software, while the most widely popular corpora are the BNC and COCA. None of these resources allow the consultation of multimedia, multimodal texts and rely solely on the written form of language as the way to approach the study and analysis of the texts.

Written corpora are commonly used in language teaching to analyse grammar, vocabulary and writing styles (Boulton, 2017). The popularity of written corpora is reflected in the amount of DDL research related to the exploration of language in written form (Boulton & Cobb, 2017; Pérez-Paredes, 2022). Apart from the main general corpora, the British Academic Written English corpus (BAWE) is prominently used in several DDL-based contexts across the literature for teacher training (Crosthwaite et al., 2023) or for ESP instruction (Curado Fuentes, 2018).

Szudarski (2020), for example, implemented DDL practices in the form of printed concordance handouts extracted from the BNC to enhance phraseological knowledge of L2 learners.

Spoken Corpora

Spoken corpora, or speech corpora (Tognini Bonelli, 2010), allow us to observe and explore naturally occurring spoken language. It has been suggested that spoken corpora can potentially cover linguistic factors such as intonation, gestures and discursive structure (Adolphs & Knight, 2010). Indeed, pragmatics-based annotation is frequent in the design of spoken corpora (Bardovi-Harlig & Mossman, 2022). Such particularities can also be useful for DDL practices as a means to provide learner error analysis or to raise linguistic awareness (De Cock, 2010). However, the reality is that spoken corpora containing actual sound files and phonological transcriptions are scarce (Crawford, 2022). There is a massive potential for the use of spoken corpora in DDL, but this potential remains to some extent unfulfilled.

Most of the spoken language DDL found in the literature has used conversational corpora, academic spoken corpora, monologic corpora (e.g. presentations) or media-driven corpora (e.g. radio or TV shows). Spoken corpora are somewhat trickier to compile, as they need to be either manually transcribed or supervised if automatic transcription is used (Farr & Hagen Karlsen, 2022). When annotated, the process becomes long and arduous for the corpus designer. Thus, spoken corpora tend to be smaller than written corpora, although examples of relatively large spoken corpora can still be found in, for instance, the spoken BNC, the Trinity Lancaster Corpus (TLC) (Gablasova & Bottini, 2022), or in the Michigan Corpus of Academic Spoken English (MICASE). Spoken corpora have been used in phraseology instruction (Aston, 2015) and the TLC has been found to aid in pedagogical decisions regarding content selection for different proficiency levels, as well as to provide empirical evidence to learners of 'successful' uses of language (Gablasova & Bottini, 2022). Pérez-Paredes (2019) has suggested the use of pedagogic corpora of teenage interviews to teach conversational features and coconstruction of meaning in dialogues and conversations. Language teachers can develop materials based on the findings from spoken corpora, such as worksheets, listening comprehension activities or conversation prompts that reflect real spoken language use, making learning more relevant and engaging for students.

Multimodal Corpora

Multimodal corpora combine audio, visual and language data, allowing users to explore the interaction of lexicogrammatical instances with other semiotic elements in the discourse such as the visual, audio and behavioural aspects of the linguistic interaction (Coccetta, 2022). Designing and preparing multimodal corpora usually entails an increased workload due to the need to synchronise and sometimes render text, audio and video. Available multimodal resources provide means to help annotate audio, video and three-dimensional motion capture data—for example, Anvil (Kipp, 2014) or Elan (Sloetjes, 2014)—while other online resources like YouGlish use the YouTube video database *as a corpus* to provide multimodal input intended to aid in developing phonetic accuracy in learners (https://www.youglish.com).

Even though the design, compilation and annotation of multimodal corpora have become easier due to different technological advances, the implementation of multimodal corpora into DDL practices is still in its very early stages, leaving its potential largely unexplored (Braun, 2005; Coccetta, 2022; Meunier, 2022; Pérez-Paredes, 2019; Pérez-Paredes & Alcaraz-Calero, 2009). There are, however scarce, some examples in the literature, such as Hirata (2019), who used a multimodal corpus tool (Multi-modal Corpus Tool or *MmCT 1.1*) alongside COCA with Japanese student teachers, so they could compare and analyse their queries in both corpora. The results pointed at multimodal corpora being perceived as more suitable for implementing DDL practices for younger learners. Further versions of MmCT (2.0) were presented in Hirata (2022), focusing on making the interface more attractive and user-friendly for children to promote its use in earlier stages of language instruction.

Parallel Corpora

Parallel corpora include texts in at least two different languages, synchronised to phrase or sentence level to allow for direct comparison between the syntactic and morphological constructions of the different languages. Although it may seem that most uses of parallel corpora belong in the domain of translation studies, their applicability in teaching environments is widely attested, offering numerous affordances in terms of the exploration of the cultural interplay between cultures while contributing to areas such as lexicology, pragmatics and contrastive analysis (Hartwell & Kraif, 2022).

Lee (2010) provides examples of parallel corpora, such as the *English-Norwegian Parallel Corpus (ENPC)*, the *English-Swedish Parallel Corpus (ESPC)* or the *Oslo Multilingual Corpus (OMC)*, among others. Other examples of available parallel corpora can be found in *CLUVI (Linguistic Corpus of the University of Vigo)*, which contains 'original texts and their translations from different domains in various language combinations, with the Galician language as the common denominator' (Sotelo, 2015, p. 252), or in the *EPTIC (European Parliament Translation and Interpretation Corpus)*, a compilation of parliamentary speeches in several languages including translations, interpretations and transcriptions (Ferraresi & Bernardini, 2019).

Parallel corpora have many uses in foreign language and DDL instruction, especially when lessons are focused on translation and phraseology. Nishina (2009) implemented DDL practices with Japanese learners of L2 English to address error correction through the use of parallel corpora in a Moodle platform. Alotaibi (2017) described the Arabic-English parallel corpus (AEPC) as a means to implement DDL-based practices, while Gong (2016) explored as well the application of DDL in translation instruction through the use of the Babel English-Chinese Parallel Corpus to develop translation strategies. Spivey (2023) has pioneered the use of multimodal DDL with young Japanese learners of English using the https://e.ddl-study.org website. The resource facilitates aligning parallel corpora and listening options.

Corpus by Thematic Cohesion

A common feature in the design process of any corpus lies in the fact that a corpus is a principled sample that shows a deliberate design (Egbert et al., 2022). Thus, the collection of texts in the field of corpus linguistics must be guided by a set of predetermined design criteria aimed at representing one particular domain of language use, which in turn will influence the quality and quantity of the linguistic texts needed to foster the representativeness of the sample. The dimension of *thematic cohesion* would thus refer to the guiding principles behind design of a given corpus.

According to Boulton and Vyatkina (2021), DDL-based research tends to favour 'general' English over more specific domains such as Language for Specific Purposes. Yet by no means does this entail that there is a lack attention to more 'focused' domains of language in DDL. While using 'general' corpora such as BNC or COCA can help lower-proficiency learners improve their vocabulary skills (Charles, 2015), other areas of linguistic instruction may require a focused approach to material selection to address the participants' needs.

The corpora previously discussed each adhere to distinct guiding principles that set them apart. The BNC concentrates on British English, while the Corpus of Contemporary American English looks at American English. The British Academic Written English (BAWE) corpus specialises in academic texts, and the British Law Report Corpus (BLaRC) (Marín Pérez & Rea Rizzo, 2012) is composed of texts from UK judicial decisions. Charles (2015) has suggested putting together genre- and discipline-specific corpora (e.g. for science and engineering students) to target specific vocabulary which would otherwise get diluted within a larger 'general' corpus. Coxhead (2022) has discussed the potential of specialised corpora in teaching technical vocabulary in areas such as nursing or pharmacology, among many others, while Csomay and Petrović (2012) described the implementation of DDL practices to acquire legal vocabulary through the exploration and exposure to repeated instances of cryptotechnical legal terms within a small corpus composed of TV programmes and movies.

Bringing All Together

Corpora vary according to their size, nature and thematic cohesion. Table 1 offers a practical way for language teachers to understand the type of corpus resource that could be used for DDL. The column on the left shows some of the categories discussed in this entry. The column on the right offers a way in which teachers could conceptualise the SACODEYL English corpus (Widmann et al., 2011).

Conclusion

This entry has provided a rationale to the categorisation of the different types of corpora available for DDL practices. Encased in a three-dimensional framework, corpora can be classified according to their size, nature and thematic cohesion. Thus, the BAWE corpus can be classified as a macro-written-academic corpus, while the Louvain International Database of Spoken English Interlanguage (LINDSEI) can be categorised as a micro-spoken-learner corpus.

The combination of corpora and DDL creates a dynamic and learner-focused learning environment that encourages students to take an active role in their language acquisition. DDL uses real language data from corpora, helping learners understand language in context. It promotes inductive learning and deeper understanding of usage. DDL is flexible and adaptable to various learner needs and teaching goals, but it is essential that language teachers understand how and why a particular corpus was put together. Avoiding overlaps in how different corpora are conceptualised might help DDL educators in the selection and design of appropriate corpora for the specific needs of their particular learning contexts. A more accurate description of the nature of the corpus utilised may help researchers as well in streamlining study design for DDL contexts and conceptualise the range of DDL corpora available.

Types of Corpora in Data-Driven Learning, Table 1 Conceptualising a corpus for DDL

Corpus name and info	SACODEYL English 25 interviews with L1 speakers of English https://www.perezparedes.es/sacodeyl-xml-corpora/
Corpus size	51,500 words Microcorpus
Corpus nature	Spoken corpus Interviews with teenagers (L1 English)
Corpus thematic cohesion	Pedagogic corpus The interviews are topic driven (e.g. holidays, hobbies, school, etc.) The language reflects the domains which are typical of L2 English secondary school curricula
Multimedia available	No
Annotation available and format	Yes. Pedagogic annotation. Corpus annotated using XML The full text can be downloaded and manipulated
Literature on the use of this corpus for DDL	Pérez-Paredes (2019)

Cross-References

- ▶ Corpora as a Reference Resource in Data-Driven Learning
- ▶ Corpus Representativeness in Data-Driven Learning
- ▶ Corpus-Informed Materials in Language Teaching
- ▶ DIY Corpora and Data-Driven Learning
- ▶ Genre Awareness and Data-Driven Learning
- ▶ Learner Corpora in Data-Driven Learning
- ▶ Multimodal Tools and Corpora for Data-Driven Learning
- ▶ Pedagogical–Corpora and Data-Driven Learning
- ▶ Tools for Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Adolphs, S., & Knight, D. (2010). Building a spoken corpus. What are the basics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 38–52). Routledge.

Alotaibi, H. M. (2017). Arabic-English parallel corpus: A new resource for translation training and language teaching. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3053572

Aston, G. (1999). Corpus use and learning to translate. *Textus, 12*, 289–314. https://godzilla.sslmit.unibo.it/~guy/textus.htm

Aston, G. (2015). Learning phraseology from speech corpora. In A. Lenko-Szymanska & Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 65–84). John Benjamins.

Ballance, O. J. (2017). Pedagogical models of concordance use: Correlations between concordance user preferences. *Computer Assisted Language Learning, 30*(3–4), 259–283. https://doi.org/10.1080/09588221.2017.1307228

Bardovi-Harlig, K., & Mossman, S. (2022). Corpora in instructed second language pragmatics. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed., pp. 71–88). Routledge.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–38). J. Benjamins.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching, 50*(4), 483–506. https://doi.org/10.1017/S0261444817000167

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89.

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL, 17*(1), 47–64. https://doi.org/10.1017/S0958344005000510

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*(3), 307–328. https://doi.org/10.1017/S0958344007000535

Charles, M. (2015). Same task, different corpus. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 131–154). John Benjamins.

Charles, M. (2018). Corpus-assisted editing for doctoral students: More than just concordancing. *Journal of English for Academic Purposes, 36*, 15–25. https://doi.org/10.1016/j.jeap.2018.08.003

Coccetta, F. (2022). Multimodal corpora and concordancing in DDL. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed., pp. 361–376). Routledge.

Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology, 26*(6), 10.14742/ajet.1047.

Corpas Pastor, G. (2001). Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada. *TRANS: Revista de Traductología, 5*, 155–184.

Corpas Pastor, G., & Seghiri Domínguez, M. (2010). Size matters: A quantitative approach to corpus representativeness. In *Lengua, traducción, recepción: en honor de Julio César Santoyo, Vol. I = Language, translation, reception: to honor Julio César Santoyo* (pp. 111–145). Publicaciones Universidad de León.

Coxhead, A. (2022). Corpora for teaching and learning vocabulary in ESP. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed., pp. 193–205). Routledge.

Crawford, W. J. (2022). Corpora and speaking skills. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed., pp. 89–101). Routledge.

Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392–1420. https://doi.org/10.1080/09588221.2021.1995001

Csomay, E., & Petrović, M. (2012). "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System, 40*(2), 305–315. https://doi.org/10.1016/j.system.2012.05.004

Curado Fuentes, A. (2018). DDL INTEGRATION IN DIFFERENT EAP SCENARIOS OF ACADEMIC WRITING. *Encuentro Journal, 27*, 157–172. https://doi.org/10.37536/ej.2018.27.1901

Curado Fuentes, A. (2023). Corpus affordances in foreign language reading comprehension. In K. Harrington & P. Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 99–118). Springer International Publishing. https://doi.org/10.1007/978-3-031-11220-1_6

Davies, M. (2015). Corpora: An introduction. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 11–31). Cambridge University Press.

De Cock, S. (2010). Spoken learner corpora and EFL teaching. In M. C. Campoy, B. Bellés-Fortuno, & M. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 123–137). Continuum.

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge University Press.

Fahey Palma, T. (2022). Corpora for teaching culture and intercultural communication. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed.). Routledge.

Farr, F., & Hagen Karlsen, P. (2022). DDL pedagogy, participants, and perspectives. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 329–343). Routledge.

Ferraresi, A., & Bernardini, S. (2019). Building EPTIC: A many-sided, multi-purpose corpus of EU parliament proceedings. In I. D. M. T. S. Nieto (Ed.), *Parallel corpora for contrastive and translation studies: New resources and applications* (pp. 123–139). John Benjamins.

Gablasova, D., & Bottini, R. (2022). Spoken learner corpora for language teaching. In R. R. Jablonkai & E. Csomay (Eds.), *Routledge handbook of corpora in English language teaching and learning* (1st ed., pp. 296–310). Routledge.

Geluso, J. (2013). Phraseology and frequency of occurrence on the web: Native speakers' perceptions of Google-informed second language writing. *Computer Assisted Language Learning, 26*(2), 144–157. https://doi.org/10.1080/09588221.2011.639786

Gong, X.-x. (2016). DDL translation teaching based on an online English-Chinese parallel corpus – Exemplified by teaching translation of English passives. *Journal of Xihua University (Philosophy & Social Sciences), 35*(5), 107–112.

Hartwell, L. M., & Kraif, O. (2022). Parallel corpora in ELT. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed., pp. 478–494). Routledge.

Hirata, E. (2019). The development of a multimodal corpus tool for young EFL learners: A case study on the integration of DDL in teacher education. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation* (pp. 88–105). Routledge.

Hirata, E. (2022). Multi-modal Corpus Tool 2.0 for young EFL learners. *Fukuoka Jo Gakuin University Bulletin. Faculty of International Career Development, 8*, 89–101.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). *The TenTen Corpus Family 7th International Corpus Linguistics Conference CL*. Lancaster.

Kipp, M. (2014). 420ANVIL: The video annotation research tool. In *The Oxford handbook of corpus phonology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199571932.013.024

Lee, D. Y. W. (2010). What corpora are available? In A. O. Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107–121). Routledge.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75. https://doi.org/10.1016/j.esp.2005.02.010

Leech, G. N. (1992). 100 million words of English: The British National Corpus (BNC). In: 서울대학교 언어교육원.

Marín Pérez, M. J., & Rea Rizzo, C. (2012). Structure and design of the British Law Report Corpus (BLRC): A legal corpus of judicial decisions from the UK. *Journal of English Studies, 10*, 131–145. https://doi.org/10.18172/jes.184

Meunier, F. (2022). Revamping DDL: Affordances of digital technology. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 344–360). Routledge.

Nishina, Y. (2009). Lexical study based on parallel corpora, DDL and Moodle. In R. de Marriott & P. L. Torres (Eds.), *Handbook of research on e-Learning methodologies for language acquisition* (pp. 204–218). Information Science Reference.

Paquot, M. (2022). Corpora and second language acquisition. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed.). Routledge.

Pérez-Paredes, P. (2019). The pedagogic advantage of teenage corpora for secondary school learners. In P. Crosthwaite (Ed.), *Data driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 67–87). Routledge.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pérez-Paredes, P., & Alcaraz-Calero, J. M. (2009). Developing annotation solutions for online data driven learning. *ReCALL, 21*(1), 55–75. https://doi.org/10.1017/S0958344009000093

Pérez-Paredes, P., & Chambers, A. (In press). Corpora and the learning and teaching of romance languages. In H. Tyne & S. Spina (Eds.), *Applying corpora in teaching and learning romance languages*. John Benjamins.

Pérez-Paredes, P., Sánchez-Tornel, M., & Calero, J. M. A. (2012). Learners' search patterns during corpus-based focus-on-form activities: A study on hands-on concordancing. *International Journal of Corpus Linguistics, 17*(4), 482–515. https://doi.org/10.1075/ijcl.17.4.02par

Raineri, S., & Debras, C. (2019). Corpora and representativeness: Where to go from now? *CogniTextes, 19*. https://doi.org/10.4000/cognitextes.1311

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.

Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. C. Campoy Cubillo, B. Bellés-Fortuño, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18–38). Continuum.

Sloetjes, H. (2014). 305ELAN: Multimedia annotation application. In *The Oxford handbook of corpus phonology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199571932.013.019

Sotelo, P. (2015). Using a multimedia corpus of subtitles in translation training. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 245–266).

Spivey, M. (2023). Data-driven learning and young learners: Perceptions and attitudes of Japanese elementary school students toward EFL corpus-based activities. *JALT CALL Journal, 19*(3), 394–416. https://doi.org/10.29140/jaltcall.v19n3.1014

Szudarski, P. (2020). Effects of data-driven learning on enhancing the phraseological knowledge of secondary-school learners of L2 English. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 133–149). Routledge.

Szudarski, P. (2022). Corpora and teaching vocabulary and phraseology. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed.). Routledge.

Tognini Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 14–27). Routledge.

Widmann, J., Kohn, K., & Ziai, R. (2011). The SACODEYL search tool: Exploiting corpora for language learning purposes. In A. Frankenberg-García, L. Flowerdew, & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 167–178). Continuum.

L

Learner Corpora in Data-Driven Learning

Sandra Götz
Philipps University Marburg, Marburg, Germany

Abstract

Learner corpora are electronic collections of authentic texts compiled from speech or writing by language learners that can be analysed automatically with the help of special software. While learner corpora, so far, have impacted the field of language pedagogy mainly in the fields of materials design (e.g. through informing the compilers of learner dictionaries or learner grammars) and in language testing and assessment (e.g. through informing current descriptor scales, such as the *Common European Framework of Reference*), the uptake of learner corpora in classroom practice, including for data-driven learning (DDL), is marginal compared to the use of native corpora. Therefore, in this entry, I will give an overview of how and why it can be beneficial to use learner corpora in DDL scenarios in the classroom.

Keywords

Learner corpora · Interlanguage · Error analysis · Peer feedback · Classroom practice · DDL

Introduction

Learner corpora are defined as 'electronic collections of authentic FL [Foreign Language] /SL [Second Language] textual data assembled according to explicit design criteria for a particular SLA [Second Language Acquisition]/FLT [Foreign Language Teaching] purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance' (Granger, 2002, p. 7). Since the first generation of learner corpora was compiled in the 1980s, the field of Learner Corpus Research (LCR) has emerged. Many LCR studies utilise the 'Contrastive Interlanguage Analysis' framework (Granger, 1996, 2015), which involves comparing learner language with native language or different interlanguages with each other. These LCR-based studies have significantly advanced our understanding of interlanguage and have had important implications for SLA and FLT (Granger, 2009). In practical applications, learner corpora have predominantly been used in materials design and language testing and assessment (Götz & Granger, 2024a). However, in the context of classroom practice, particularly in data-driven learning (DDL) scenarios, learner corpora are much less frequently used compared to native corpora. A meta-analysis by Boulton and Vyatkina (2021) on DDL study characteristics revealed that only 4% (21 out of 489) of the studies in their database involved learner corpus-based DDL. Three years later, a literature review

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_69-1

by Götz and Granger (2024a) showed a slight increase to 33 studies. Additionally, Lusta et al.'s (2023) review of classroom-focused DDL studies does not at all mention learner corpora. This lack of uptake is surprising, considering the valuable real-world data learner corpora offer for DDL approaches. The following section will discuss some of these fields of application.

Benefits of Using Different Types of Learner Corpora in DDL

The original concept behind the development of DDL was to 'confront the learners as directly as possible with the data, and to make the learner a linguistic researcher' (Johns, 2002, p. 108). A substantial body of empirical studies has demonstrated that this research-based approach to discovery learning can be highly beneficial and motivating for learners. It has also been observed that 'DDL works pretty well in almost any context where it has been extensively tried' (Boulton & Cobb, 2017, p. 386). Encouraging this type of curiosity in students can also extend to working with learner corpora, providing an authentic database of their own or their peers' language use. Employing learner corpora in DDL offers several advantages for both language learners and educators, which will be explored in this section.

One recurring theme in almost all evaluative studies on DDL is the significant motivational impact of working with corpora. This effect may be even more pronounced with LC-based DDL. For language learners, engaging with authentic data from their peers or even their own work can be especially compelling and motivating, as it provides concrete evidence of language use and progress. This enhances their engagement with DDL activities, as learner corpora offer real examples of language use by individuals with similar backgrounds. This allows learners to see authentic mistakes and successes, making their learning experience more relatable than working with the output of an 'abstract' native speaker to whom they might not easily relate. For instance, DDL exercises that explore successful patterns and structures in the target language produced by slightly more advanced peers can foster a deeper understanding and retention (e.g. André et al. (2024)). Simultaneously, examining common errors in learner corpora from learners with the same L1 background and proficiency level helps students identify frequent and typical mistakes, becoming aware of them and thereby avoiding similar pitfalls (Mendikoetxea et al., 2010). LC-based DDL can also be highly beneficial for personalised learning, as these corpora can be tailored to address the specific challenges and needs of a particular learner group. Using a local learner corpus from one's own classroom, for instance, provides relevant examples for DDL exercises, further boosting learners' motivation (Cotos, 2014; Fuster-Márquez & Gregory-Signes, 2018). Moreover, if learners themselves are trained to compile and analyse corpus data, they can benefit from creating a learner corpus of their own work. This inspires independent learning and fosters autonomy, as students can be encouraged to learn to identify their own strengths and weaknesses and witness their own progression accordingly (e.g. Charles (2018), Hartle (2023)). Such approaches can significantly personalise learning, allowing students to address their specific needs more effectively than working with patterns coming from a native corpus that might not be relevant to their specific needs.

Educators might also benefit from using LC-based DDL, especially when their DDL activities utilise data gathered from their own classrooms. This approach helps teachers to identify common strengths and persistent difficulties within their learner group, enabling them to address these issues more effectively in their teaching and thereby leading to more targeted remediation. These activities of preselecting features from learners' data are the most frequently applied DDL tasks and are often referred to as 'mediated' uses of learner corpora (Götz & Granger, 2024a). Specifically, teachers can use examples from local learner corpora to create DDL exercises and activities that address the specific linguistic trouble spots of their learning cohort, thus making their materials more relevant and effective (Belz & Vyatkina, 2008). Another significant benefit of using learner corpora for

DDL is their ability to be customised to fit various learning contexts, whether it is for specific languages, proficiency levels, or linguistic features. Teachers can always select data with an appropriate proficiency level, as native corpora might sometimes be too difficult for learners to understand (André et al., 2024).

In sum, LC-based DDL not only enhances student engagement and motivation but also provides educators with valuable insights and tools for more targeted and effective teaching. By incorporating authentic learner data, teachers can create more meaningful and impactful learning experiences, which can ultimately lead to improved language acquisition outcomes.

Fields of Application of LC-Based DDL

The possible pedagogical applications of LC-based DDL are extensive. Previous research has primarily concentrated on teaching written features, though applications of LC-based DDL to teaching spoken features are gradually increasing. Learners work with learner corpora (either their own or preexisting ones) either independently or in comparison with native speaker corpora.

In the realm of written language, numerous studies have focused on enhancing learners' academic writing skills. This is often achieved by comparing learner data with native data for the learners to discover which features of learner writing are over-, under- or misused compared to native speakers or by applying LC-based DDL for the learners to discover specific lexical phrases, collocations and formulaic expressions that are commonly used by advanced L2 speakers (e.g. Charles (2018), Cotos (2014), Eriksson (2012), Forti (2017), Fuster-Márquez and Gregori-Signes (2018), Hartle (2023), Staples et al. (2024)). Another common application of LC-based DDL focuses on improving the accuracy of learners' output through error-correction tasks (e.g. Gaskell and Cobb (2004), Mendikoetxea et al. (2010), Papp (2007), Sarré et al. (2024), Todd (2001), Yoon and Jo (2014), Zaki (2021)). There are also some studies reporting on LC-based DDL in the area of grammar (e.g. Pérez Cañado and Díez Bedmar (2006)) and, more recently, on building language and genre awareness (e.g. Staples et al. (2024)).

The application of LC-based DDL to teaching features of spoken language has been much more rarely done and in this context mainly in the form of awareness-raising activities. For instance, Gut (2006) demonstrated how a phonologically annotated learner corpus can be used to teach various pronunciation features, including phonetic and phonological aspects. Peltonen and Lintunen (2024) focus on the area of fluency by employing LC-based awareness-raising activities to enhance L2 fluency development in formal instruction. André et al. (2024) show how video material featuring proficient L2 speakers of French can significantly improve learners' awareness of complex conversational narratives by engaging their metacognitive strategies during their time abroad.

Summary and Outlook

LC-based DDL has great potential to have positive learning outcomes. So far, however, it is only the motivational aspect that seems to be undoubtedly proven. What has become clear is that 'the body of empirical DDL research has been growing rapidly and has by now developed into a fully-fledged CALL subfield' (Boulton & Vyatkina, 2021, p. 81). However, LC-based DDL seems to fall behind in this area, so that Mukherjee's (2009) call for more empirical research still seems to be true, even almost two decades later: 'as in many other areas of applied corpus linguistics, what is needed is much more classroom action research into whether and to what extent learners really profit from the use of learner corpora' (p. 214). We have seen a considerable step forward in this direction in the collection of experimental studies collected in Götz and Granger (2024b), which were all able to report small and/or qualitative increases in learners' performances after LC-based DDL activities. These findings are very promising and have laid a solid foundation for future, more large-scale follow-up studies conducted in a similar vein. In order to document

the true learning benefits of LC-based DDL quantitatively, future empirical research studies would benefit tremendously from including pre- and posttests (as well as delayed posttests), should be conducted with both an experimental and a control group and should be analysed with advanced statistical procedures, complemented with observations from the classroom and further qualitative and/or questionnaire data.

Cross-References

- ▶ Concordancers for Data-Driven Learning
- ▶ Corpus Literacy and Data-Driven Learning
- ▶ DIY Corpora and Data-Driven Learning
- ▶ Evolving Paradigms in Data-Driven Learning
- ▶ Future Challenges and Opportunities for Data-Driven Learning
- ▶ Hands On vs Hands-Off Methods in Data-Driven Learning
- ▶ Inductive and Deductive Approaches to Data-Driven Learning
- ▶ Learner Autonomy and Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Teacher Education for Pedagogical Uses of Corpora
- ▶ Teacher's Digital Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

André, V., Boulton, A., Ciekanski, M., & Cousinard, C. (2024). Learning to interact from conversational narratives: New perspectives for a data-driven approach integrating L2 speaker data. *International Journal of Learner Corpus Research, 10*(1), 67–106. https://doi.org/10.1075/ijlcr.00041

Belz, J., & Vyatkina, N. (2008). The pedagogical mediation of a developmental learner corpus for classroom-based language instruction. *Language Learning & Technology, 12*(3), 33–52. https://hdl.handle.net/1805/2652

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Charles, M. (2018). Corpus-assisted editing for doctoral students: More than just concordancing. *Journal of English for Academic Purposes, 36*, 15–25. https://doi.org/10.1016/j.jeap.2018.08.003

Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL, 26*(2), 202–224. https://doi.org/10.1017/S0958344014000019

Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas & A. Boulton (Eds.), *Input, process and products. Developments in teaching and language corpora* (pp. 195–211). Masaryk University Press.

Forti, L. (2017). Data-driven learning and the acquisition of Italian collocations: From design to student evaluation. In K. Borthwick, L. Bradley, & S. Thouësny (Eds.), *CALL in a climate of change: Adapting to turbulent global conditions—Short papers from EUROCALL 2017* (pp. 110–115). Research-Publishing.Net. https://files.eric.ed.gov/fulltext/ED578177.pdf

Fuster-Márquez, M., & Gregori-Signes, C. (2018). Learning from learners: A non-standard direct approach to the teaching of writing skills in EFL in a university context. *Innovation in Language Learning and Teaching, 12*(2), 164–176. https://doi.org/10.1080/17501229.2016.1142549

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://doi.org/10.1016/j.system.2004.04.001

Götz, S., & Granger, S. (2024a). Learner corpus research for pedagogical purposes: An overview and some research perspectives. *International Journal of Learner Corpus Research, 10*(1), 1–38. https://doi.org/10.1075/ijlcr.00039.got

Götz, S., & Granger, S. (Eds.) (2024b). Learner corpus research for pedagogical purposes [Special Issue]. International Journal of Learner Corpus Research, 10(1). 10.1075/ijlcr.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast: Text-based cross-linguistic studies* (pp. 37–51). Lund University Press.

Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). https://doi.org/10.1075/lllt.6.04gra

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13–33). https://doi.org/10.1075/scl.33.04gra

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7–24. https://doi.org/10.1075/ijlcr.1.1.01gra

Gut, U. (2006). Learner speech corpora in language teaching. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 69–86). Peter Lang.

Hartle, S. (2023). From learner corpus to data-driven learning (DDL): Improving lexical usage in academic writing. *EuroAmerican Journal of Applied Linguistics and Languages, 10*(2), 9–31. https://doi.org/10.21283/2376905X.1.10.2.2748

Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107–117). Rodopi.

Lusta, A., Demirel, O., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*, e22731. https://www.sciencedirect.com/science/article/pii/S2405844023099395/pdfft?md5=8c1dadf8f06fa5391052abcf8b6774aff&pid=1-s2.0-S2405844023099395-main.pdf

Mendikoetxea, A., Murcia-Biels, S., & Rollinson, P. (2010). Focus on errors: Learner corpora and pedagogical tools. In M. Campoy-Cubillo, B. Bellés-Fortuño, & L. Gea-Valor (Eds.), *Corpus-based approaches to English Language Teaching* (pp. 180–194). Continuum.

Mukherjee, J. (2009). The grammar of conversation in advanced spoken learner English: Learner corpus data and language-pedagogical implications. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 203–230). https://doi.org/10.1075/scl.33.17muk

Papp, S. (2007). Inductive learning and self-correction with the use of learner and reference corpora. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 207–220). Rodopi.

Peltonen, P., & Lintunen, P. (2024). Teaching L2 speech fluency with learner-corpus-based awareness-raising activities: Insights from a short-term intervention study. *International Journal of Learner Corpus Research, 10*(1), 39–66. https://doi.org/10.1075/ijlcr.00040.pel

Pérez Cañado, M. L., & Díez Bedmar, M. B. (2006). Data-driven learning and awareness raising: An effective tandem to improve grammar in written composition? iJET. *International Journal of Emerging Technologies in Learning, 1*(3). https://doi.org/10.3991/ijet.v1i3.55

Sarré, C., Brudermann, C., & Grosbois, M. (2024). Using learner corpus data for grammatical accuracy development in written productions: The role of corrective feedback. *International Journal of Learner Corpus Research, 10*(1), 107–145. https://doi.org/10.1075/ijlcr.00042.sar

Staples, S., Conrad, N., Dang, A., & Wang, H. (2024). Building language and genre awareness through learner corpus data in a second language writing course. *International Journal of Learner Corpus Research, 10*(1), 146–181. https://doi.org/10.1075/ijlcr.00043.sta

Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System, 29*(1), 91–102. https://doi.org/10.1016/S0346-251X(00)00047-6

Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology, 18*(1), 96–117. http://dx.doi.org/10125/44356

Zaki, M. (2021). Self-correction through corpus-based tasks: Improving writing skills of Arabic learners. *International Journal of Applied Linguistics, 31*(2), 193–210. https://doi.org/10.1111/ijal.12312

D

DIY Corpora and Data-Driven Learning

Simon Smith
School of Social Sciences and Humanities, Coventry University, Coventry, UK

Abstract

This entry shows how do-it-yourself (DIY) corpora, created by students, have been used in data-driven language learning (DDL) over the last few decades. The entry charts the use of DIYC in the teaching of translation, English for Academic Purposes (EAP), both specific and general, and other languages. It is shown that the approach enables the learning of discipline-specific vocabulary and collocations, grammatical usage, rhetorical functions and reading skills. A variety of task types associated with or following the corpus construction may stimulate and motivate learners. Furthermore, learners often benefit from the engagement with and ownership of a personal corpus resource; these affordances may not follow from the usual corpus consultation to which DDL is typically confined.

Keywords

DIY corpus · Do-it-yourself corpus · Home-made corpus · Virtual corpus · Personal corpus · Disposable corpus · Corpus construction · DDL · Translation · EAP

Introduction

Practitioners of data-driven learning have exploited corpora of all types and sizes, including balanced general corpora such as the BNC (British National Corpus), web corpora, genre-specific corpora such as BAWE (British Academic Written English), and learner corpora. On the DDL model, learners consult or otherwise make use of the corpus, or materials based on it, in order to build their knowledge of the target language.

This entry deals with a further corpus type, the DIY corpus, and its application to DDL. A DIY corpus (DIYC), uniquely, is created or constructed by the learners themselves. Unlike a learner corpus, which of course consists of learner language, a DIYC contains texts usually created by expert language users, often including native speakers, and which typically have previously been published or are otherwise in the public domain. Thus, a learner corpus contains texts created by learners, but is assembled by others; a DIYC is assembled by learners, but its texts are created by others.

DIYCs are also variously known as home-made corpora (Aston, 2002), virtual corpora (Davies, 2023, but this seems to be used by Davies to refer to subcorpora, rather than corpora created by learners from scratch), disposable corpora (Varantola, 2003), and personal corpora (Chen & Flowerdew, 2018). However, "DIYC" now has

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_70-1

general currency in the field and is therefore the term adopted here.

There are a few examples in the literature of corpora whose texts are both created and assembled or partially assembled by the learners (Lee & Swales, 2006). Charles (2024), in her commendable survey of do-it-yourself corpora, would classify these as DIYCs, but this encyclopaedia treats them as a special type of learner corpus.

Affordances of DIYCs

The approaches to DDL described in other contributions to this DDL section of the encyclopaedia involve the *consultation* of corpus resources. It has been claimed that corpus *construction* by learners, followed by consultation, may afford better learning opportunities. According to Tyne (2009), the corpus creation process can motivate learners by conferring on them a sense of "ownership" over the resource. Zanettin (2002) required learners to compile a corpus from the web and analyse it with Wordsmith Tools, noting (p. 7) that "constructing the corpus was as useful as generating concordances from it". Similarly, Charles (2012, p. 101) describes the "truly revelatory moment when they see the patterns appear before their eyes *in their own data*" [emphasis in original].

The use of DIYCs may promote learner autonomy, according to Gavioli (2009), with the teacher taking a less central role, because the knowledge source is the corpus rather than the teacher. Gavioli shows how students analysing corpora may become "active interpreters of the data" while also adding something of value for their research community. Gavioli's students worked with DIYC of film reviews and lonely hearts ads.

Corpus compilation can help learners acquire transferable skills as well as language. For example, Boulton (2010) observed that his students developed problem-solving and ICT competencies. Charles (2014) found that some students were motivated enough to consult and even expand their corpus after the course ended. Additionally, Lee and Swales (2006) noted that some of their students even purchased their own copy of Wordsmith Tools, showing a commitment to continuing with corpus construction and analysis in the future.

The earliest paper setting out the affordances of learner-compiled corpora seems to be Aston (2002), which is evidently inspired by Tribble's (1997) "quick and dirty" corpora—relatively small, ad hoc corpora created by the **practitioner** for a specific teaching purpose. Aston (2002) uses an extended metaphor to equate the task of corpus construction by **learners** to making a fruit salad. When serving a fruit salad, one has the choice of purchasing a factory-prepared product, or buying fresh ingredients and making the salad oneself, and this is analogous to the choice between using an off-the-shelf corpus, and creating a corpus of one's own. In both the fruit salad and corpus cases, the decision to custom-build (according to Aston) means that by forgoing the convenience, reliability and standard format of the ready-made product, and creating one's own, one has (a) control over and certainty about what went into it, and (b) the opportunity to learn through the creation process and through discussions with peers.

DIYCs for Translation Students

Aston was a scholar of translation studies, and teachers and researchers in that field were amongst the earliest adopters of the DIYC approach. Those enrolled on translation courses are typically very advanced learners of the target language(s) who have already mastered most of its grammatical and structural challenges; their needs are likely to be in the area of vocabulary and collocation, particularly the specialized terminology of a domain with which they may not be familiar. Translators have made use of glossaries, term banks and translation memory systems for many decades. Intuitively, a specialized corpus that includes domain-specific vocabulary present in a source text will be of assistance in translating that text.

Maia (1997) conducted an early venture into the compilation of technical glossaries, by students, based on texts they had collected. The

students initially used printed texts, typing them in themselves; later, texts from CD-ROMs were included, and finally Internet access was made available so that the students could include web resources.

Zanettin (2002) encouraged students to attempt a translation task (set by another teacher, or a commercial assignment) using a DIYC they had compiled from the web to help them. They found new texts to add to the corpus as they proceeded, and reported that the process of building the corpus was as useful as analysing it. The ability to link back to the web page (where images and multimedia were often available to help clarify the topic) was especially valued by students.

The advent of the BootCaT software (Baroni & Bernardini, 2004) had the potential to relieve students of some of the heavy lifting here: users engage fully with the corpus construction by supplying seed words (keywords salient to a particular domain with which to "seed" the corpus), and checking the right type of texts are being selected, but much of the process is automated. BootCaT uses n-tuples (e.g. triples, quadruples) of seed words to search the web for relevant web pages which can be included in the DIYC. Castagnoli's (2006) students used BootCaT to create corpora and generate term lists for translation assignments; the study was successful, but Castagnoli concluded that the usefulness of the resulting lists depended on the familiarity of the domain to the person compiling it.

The BootCaT suite used by Castagnoli consisted of a set of Perl programs, but a useful graphic interface https://bootcat.dipintra.it/ is now available. In another development, BootCaT was integrated into the Sketch Engine (SkE, www.sketchengine.eu) software as WebBootCaT (WBC; Baroni et al., 2006). This allows concordances, word sketches, term lists and other affordances of the SkE platform to be directly generated from the bootstrapped web corpora. Frankenberg-Garcia (2015) asked translation students to create corpora using the WBC tools; in the evaluations, 12 out of 13 students agreed or strongly agreed that they were able to build a simple corpus on their own, with 8 stating that they were "likely to build a corpus of my own to help me with my research or with my work in the future".

The research so far described has focused on the needs of translation students. Smith (2011) extended the approach to his General English classes in Taiwan, offering students the option to create corpora based on their academic major by specifying appropriate keywords to WBC. The students were then encouraged to use a recursive function of WBC which expands the corpus based on salient terms found in the first version. The students undertook a mini research project based on the corpus, using SkE, and most students who expressed an opinion reported that the corpus construction was useful, that it had helped to improve their English, that they derived a sense of ownership from creating the corpus, and that they had gained transferable skills. Some students stated that they would continue to use the corpus after the course, but an equal number said that they would not.

DIYCs in EAP

EAP students require support not only in terms of general academic vocabulary, but also in the specialist vocabulary of their area of study. Smith (2020), extending his earlier work, encouraged international students of Finance and Accounting to create focused domain-specific corpora, again using WBC, as part of their EAP module. Students then used SkE to help generate a vocabulary portfolio, which they annotated with definitions, translations, example sentences and other key information about domain-specific terms. Compared to a control group who were taught in a more traditional way, students who had created vocabulary portfolios performed better on a post-test on financial vocabulary. Some of the students of Jablonkai and Čebron (2021) also used WBC to expand DIY corpora they had created.

The potential benefits of DIYC for academic English learners are not confined to the study of vocabulary and collocation. Zhang et al. (2017) and Dong and Lu (2020) had graduate students in medicine and engineering respectively compile their own corpora of published research articles.

Zhang et al. asked their participants to create individual corpora on a medical subdomain of their choice, and use it to check "lexical collocation or grammatical usage while writing and revising their papers". Students reported that the experience was helpful during the course, but when asked whether they had continued to use their corpora 6 months after the intervention, the findings were less encouraging. Dong and Lu's students collaborated with the authors in not only compiling the corpus, but also annotating it by rhetorical function, subsequently searching the corpus to learn about linguistic patterns associated with particular moves.

Wang et al. (2015) reported success with English major students in the use of corpus construction on a reading skills course. They found relevant texts (mainly news) in COCA, then located the original sources online from news websites. Students were also required to annotate the corpora created, for example to identify the topic sentence. Although the full list of tags used is not provided, the authors state that the construction and annotation process motivated students to closely read and analyse the texts.

Other researchers worked with academic English students from mixed disciplinary backgrounds. Charles (2012, 2014) enjoyed considerable success with research students at Oxford University, where students were first given practice with hand-on corpus use, including concordancing, and then asked to build their own disciplinary corpus of research articles, using it to investigate stance (beginning with the use of personal pronouns). Charles (2014) asked students whether they had revisited their corpora over the 12 months following the course, and her findings were generally encouraging.

Lee and Swales (2006) also worked with EAP students from different academic backgrounds. They completed end-of-course projects in which they compared lexical and grammatical patterns between their DIYC and another corpus (in one case, this "other corpus" was a learner corpus of the student's own writing, also compiled by her). This was a small-scale project, with only three projects being submitted. More recently, Chen and Flowerdew (2018) ran workshops across six universities in Hong Kong, involving over 6% of the research students in the territory. Questions such as the choice of texts to be selected for inclusion, and the size of corpus to be constructed, were discussed in the workshops, but disappointingly only 11.73% of participants felt they had benefited from corpus construction. Therova and McKay (2023) was another small-scale study, with six UK-based pre-sessional students creating their own corpora and using SkE and #LancsBox to analyse them. Uniquely, the authors sought the views of PhD supervisors on the value of the intervention, as well as those of students. Most of the supervisor comments related to the subject-specific vocabulary students had acquired, and were largely positive.

Jeaco (2020) had EAP students at a Sino-British university construct DIYC and analyse them using the Prime Machine, a tool developed by the author. Students were able to build their specialist vocabulary by investigating keywords from their DIYC and publicly available corpora, presented side-by-side in the Prime Machine interface. Another tool that is of great potential in generating domain-specific corpora for EAP is AntCorGen (Anthony, 2024). This tool allows the user to selectively download RAs on the multidisciplinary open-access journal platform PLOS (https://journals.plos.org/), and to carry out a variety of searches from within the tool. Different sections (abstract, introduction, etc.) can be saved as separate subcorpora, giving additional flexibility. Song (2020, p. 120) gives a practical example of the use of AntCorGen in a corpus creation task designed to teach the use of hedging devices in RA discussion sections.

Other Languages/Transcription

DIYC work by Tyne (2009, 2012) targets languages other than English—other than the translation studies work described above (where learners would often compile corpora in both their source and target languages), most other studies have been on English. Tyne (2009) called on British university students of sociolinguistics, with a French concentration, to collect spoken

corpora from their own groups of participants, with both the collection procedure and the spoken modality also representing unique aspects of this work. The students were then required to write a transcription of the spoken French, a process which they found beneficial in a number of respects. Tyne (2012) reported on a similar project, in which high-school students in France collected spoken corpora of Spanish by interviewing a Spanish language assistant and transcribed the speech. Feedback from students and teachers (who were ordinary teachers rather than researchers or corpus linguists) indicated that the students' Spanish proficiency had benefited from the transcription, even though the task was challenging.

Conclusion

DIY corpus approaches—the construction and analysis of corpora by learners—have been shown to benefit students of translation, general English, and academic English for general and specific purposes. Learners of other languages have benefited as well. Some studies, naturally, have reported greater success with the approach than others.

DDL tasks and learning outcomes exploiting DIYCs have included translation, the compilation of subject-specific glossaries and vocabulary profiles, rhetorical move/step analysis, a variety of mini research projects, stylistic and structural analysis leading to revision/improvement of the student's own writing, developing reading skills, annotation and transcription. Key themes arising from this conspectus are that:

- The learner often has a sense of ownership of the corpus resource, and may be inclined to use it again after end of a course or module
- The process of construction of a corpus can be as motivating as the use and analysis of the final product
- A specialized corpus created by a learner is more likely to incorporate forms and lexis related to their discipline, topic or task then an off-the-shelf corpus
- The process of corpus construction can help to develop transferable skills as well as language knowledge

Over the course of the last 25 years, advances in technology have enabled learners to create DIY corpora with increasing ease, while generally allowing for learner engagement with the creation process. Early interventions often required learners to seek out and collate texts manually; the wider availability of the web, followed by the arrival or platforms such as WebBootCaT and AntCorGen meant that representative, discipline-specific corpora could be compiled in short order. It remains to be seen whether sophisticated large language models or other developments in artificial intelligence will have a part to play in corpus construction by learners, or whether DIYCs will in coming decades retain their relevance, as new ways of learning and writing emerge.

Cross-References

▶ Learner Autonomy and DDL
▶ Learner Corpora in DDL
▶ Translation and DDL
▶ Vocabulary and DDL
▶ Web as Corpus for DDL
▶ Wordlists and DDL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anthony, L. (2024). *AntCorGen (Version 1.3.0) [Computer Software]*. Waseda University. https://www.laurenceanthony.net/software

Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 9–25). Rodopi.

Baroni, M., & Bernardini, S. (2004). BootCaT: Bootstrapping corpora and terms from the Web. In *Proceedings of 4th international conference on language resources and evaluation, Lisbon, Portugal* (pp. 1313–1316).

Baroni, M., Kilgarriff, A., Pomikálek, J., & Rychlý, P. (2006). WebBootCaT: Instant domain-specific corpora to support human translators. In *Proceedings, 11th annual conference of the European Association for Machine Translation Conference, Oslo, Norway* (pp. 247–252).

Boulton, A. (2010). Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 69–96). Peter Lang.

Castagnoli, S. (2006). Using the web as a source of LSP corpora in the terminology classroom. In M. Baroni & S. Bernardini (Eds.), *Wacky! Working papers on the web as corpus* (pp. 159–172). Gedit. http://wackybook.sslmit.unibo.it/pdfs/castagnoli.pdf

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes, 31*(2), 93–102.

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30–40.

Charles, M. (2024). Benefits and challenges of using do-it-yourself corpora for academic writing development. *TESOL Quarterly 58*(3), 1205–1214.

Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes, 50*, 97–112.

Davies, M. (2023). Creating and using "virtual corpora" to extract and analyse domain-specific vocabulary at English-Corpora.org. In Corpora and Translation Education: Advances and Challenges (pp. 89–108). Springer Nature Singapore.

Dong, J., & Lu, X. (2020). Promoting discipline-specific genre competence with corpus-based genre analysis activities. *English for Specific Purposes, 58*, 138–154.

Frankenberg-Garcia, A. (2015). Training translators to use corpora hands-on: Challenges and reactions by a group of thirteen students at a UK university. *Corpora, 10*(3), 351–380.

Gavioli, L. (2009). Corpus analysis and the achievement of learner autonomy in interaction. In *Using corpora to learn about language and discourse* (Vol. 66, pp. 39–71). Peter Lang.

Jablonkai, R., & Čebron, N. (2021). Undergraduate students' responses to a corpus-based ESP course with DIY corpora. In M. Charles & A. Frankenberg-Garcia (Eds.), *Corpora in ESP/EAP writing instruction: Preparation, exploitation, analysis* (pp. 100–120). Routledge.

Jeaco, S. (2020). DIY needs analysis and specific text types: Using the Prime Machine to explore vocabulary in readymade and homemade English corpora. In *Vocabulary in curriculum planning: Needs, strategies and tools* (pp. 199–223). Palgrave Macmillan.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75.

Maia, B. (1997). Making corpora: A learning process. In S. Bernardini & F. Zanettin (Eds.), *I corpora nella didattica della traduzione* (pp. 47–60). CLUEB. http://web.archive.org/web/20031004185049/http://www.sslmit.unibo.it/cultpaps/paps.htm

Smith, S. (2011). Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning, 24*(4), 291–316.

Smith, S. (2020). DIY corpora for Accounting & Finance vocabulary learning. *English for Specific Purposes, 57*, 1–12.

Song, P. (2020). ESP curriculum case study and analysis. *Professional and Academic English, 55*, 109–123. https://ssrn.com/abstract=3834114

Therova, D., & McKay, A. (2023). Enhancing student writing with do-it-yourself corpora on a PhD pre-sessional programme. *Language Teaching Research*. https://doi.org/10.1177/13621688231221350

Tribble, C. (1997). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In J. Melia & B. Lewandowska-Tomaszczyk (Eds.), *PALC '97 proceedings* (pp. 132–147). Lodz University Press.

Tyne, H. (2009). Corpus oraux par et pour l'apprenant [Spoken corpora by and for the learner]. In A. Boulton (Ed.), *Des documents authentiques oraux aux corpus: Questions d'apprentissage en didactique des langues* (pp. 91–111). Mélanges CRAPEL.

Tyne, H. (2012). Corpus work with ordinary teachers: Data-driven learning activities. In *Input, process and product: Developments in teaching and language corpora* (pp. 136–151). Masarykova univerzita nakladatelství.

Varantola, K. (2003). Translators and disposable corpora. In F. Zanettin, S. Bernardini, & D. Stewart (Eds.), *Corpora in translator education* (pp. 55–68). Taylor & Francis.

Wang, H., Zheng, Y., & Cai, Y. (2015). Application of corpus analysis methods to the teaching of advanced English reading and students' textual analysis skills. In B. Zou, S. Smith, & M. Hoey (Eds.), *Corpus linguistics in Chinese contexts* (pp. 158–177). Palgrave Macmillan UK.

Zanettin, F. (2002). DIY corpora. The WWW and the translator. In B. Maia, J. Haller, & M. Urlrych (Eds.), *Proceedings of the "training the language services provider for the new millennium" international conference, Porto, Portugal, 25–26 May 2001* (pp. 239–248).

Zhang, F., Zheng, Y., & Li, L. (2017). Using medical academic English corpus for graduates students academic writing training. *Theory and Practice in Language Studies, 7*(10), 868–873.

L

(Lexico)grammar and Data-Driven Learning

Gaëtanelle Gilquin
UCLouvain, Louvain-la-Neuve, Belgium

Abstract

Among the different aspects of language that data-driven learning (DDL) can be used for, this entry focuses on lexicogrammar and grammar. It starts by showing that DDL has been applied more often to lexicogrammar than to grammar. It then presents the ways in which each of them can be approached through DDL, with some illustrations taken from the literature. It appears that DDL approaches to lexicogrammar and grammar bear some similarities, including exploratory and corrective functions, and that they each have unique contributions to make to language pedagogy.

Keywords

Collocates · Concordances · Constructions · Corrective data-driven learning · Learner corpus · Patterns · POS tagging · Word sketches

Introduction

Data-driven learning (DDL) is a pedagogical approach that involves the use of corpus data in the language classroom. The term also refers to the research field that explores this pedagogical approach and seeks to assess its effectiveness. In this entry, DDL is discussed, both as a pedagogical approach and a research field, in relation to two aspects of language that DDL can be applied to, namely, lexicogrammar (which straddles lexis and grammar) and grammar.

Research has shown that DDL is especially useful to cover lexicogrammatical aspects of language. According to Boulton and Cobb's (2017) meta-analysis, lexicogrammar is the language aspect that is most often dealt with in DDL publications. Their meta-analysis also reveals that, together with vocabulary, lexicogrammar is the aspect that yields the best results, as measured by effect size. Mizumoto and Chujo (2015) reach similar conclusions in the context of English language teaching in Japan.

Grammar, on the other hand, appears less often in DDL studies. In Boulton and Cobb's (2017) meta-analysis, DDL publications with grammar as their object of study are less frequent than those dealing with lexicogrammar and vocabulary—but more frequent than those dealing with discourse. The effect sizes are also generally smaller than for lexicogrammar and vocabulary, which suggests a lower degree of effectiveness.

In what follows, the focus is first on DDL approaches to lexicogrammar and then on DDL approaches to grammar. Illustrations from the literature are provided and, when evidence is available, the effectiveness of DDL in the reviewed studies is

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_77-1

commented on. Although lexicogrammar and grammar are discussed in different sections, it should be emphasized that they do not represent clearly distinct categories. As its name indicates, lexicogrammar includes some grammatical aspects, which establishes a continuum between the two (as well as between lexicogrammar and vocabulary).

DDL Approaches to Lexicogrammar

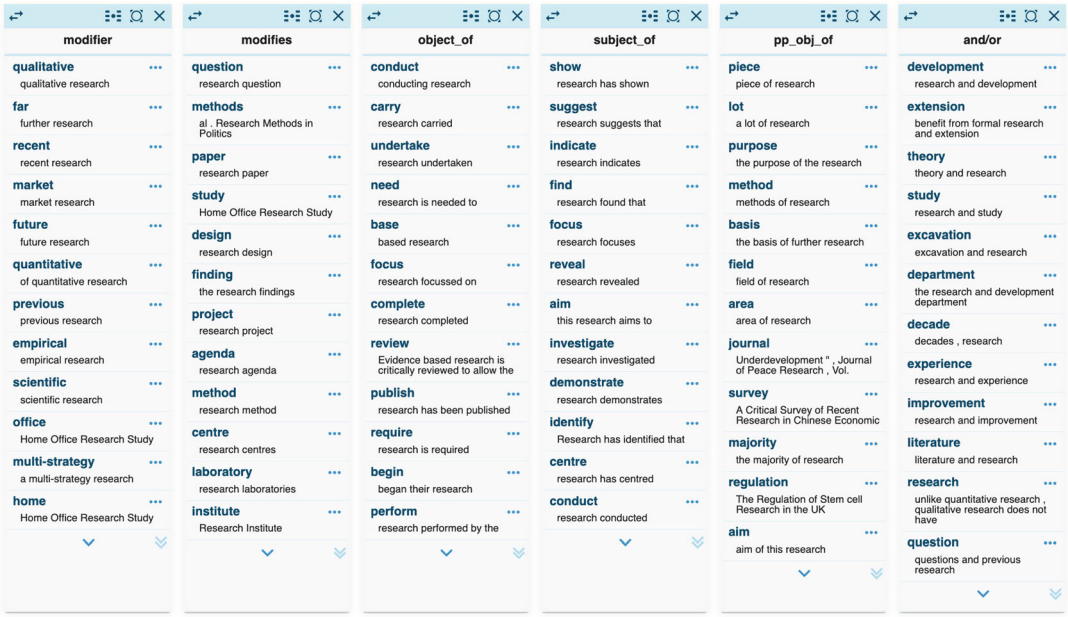
Pearce (2007, p. 109) points out that "[t]he advent of corpus linguistics has made the identification of lexicogrammatical patterns much easier than it once was". It is therefore not surprising that DDL should be particularly suitable to deal with such aspects. DDL has a unique contribution to make in this respect, since speakers' intuitions tend to be rather poor when it comes to lexicogrammatical patterns (Hunston, 2002, pp. 20–21) and traditional pedagogical resources like dictionaries, grammar books or textbooks may not be readily amenable to the teaching of such patterns (O'Sullivan & Chambers, 2006, p. 50).

In DDL approaches to lexicogrammar, the starting point is usually a single word, but unlike DDL approaches to vocabulary (see ▶ "Vocabulary and DDL"), the focus is "less on meaning than on how the target word fits into its surroundings" (Boulton & Cobb, 2017, p. 380). Charles (2015), for example, describes how students were invited to use the Collocate tool of AntConc (Anthony, 2011) in order to examine the collocates of the verbs *show* and *suggest*, that is, the words occurring regularly in their close textual environment. The students first looked for the presence of *that* to the right of the verbs, which helped them identify the VERB + *that* pattern. They then considered noun and pronoun collocates to the left of the verbs, in an attempt to describe the nature of their subjects—either personal (e.g. *I*, *we*) or impersonal (e.g. *evidence*, *results*).

A DDL activity can also start from a pattern and encourage students to learn more about its use in context. In Vyatkina (2016), the DDL intervention centred around certain verb-preposition combinations in German (e.g. *warten auf*, "wait for"). Students were given worksheets with concordances that had been extracted from the DWDS corpus (http://www.dwds.de) and that showed occurrences of the combinations in context. On that basis, the students were able to observe how these combinations were used and, more particularly, what grammatical cases they took (e.g. accusative in the case of *warten auf*) and then to produce their own example sentences.

Interestingly, certain corpus tools give direct access to lexicogrammatical patterns. This is the case of Sketch Engine and its "word sketches" (Kilgarriff et al., 2010), which show lists of collocates organized according to their grammatical relationship with the search item. Figure 1 reproduces part of the word sketch of the noun *research* in the British Academic Written English (BAWE) corpus. It includes modifiers of *research* (e.g. *qualitative*, *recent*), nouns modified by *research* (e.g. *question*, *paper*), verbs used with *research* as a direct object (e.g. *conduct*, *undertake*) or as a subject (e.g. *show*, *indicate*), nouns followed by the preposition *of* and *research* (e.g. *piece*, *field*) and nouns coordinated with *research* (e.g. *development*, *theory*). Thomas (2015) refers to such information as "extended collocations". He explains how students can group them into semantic groups so as to derive "word templates", such as the following for the noun *allegation*: "[an injured party] makes an allegation about [a negative abstraction (potentially criminal, e.g. corruption) or a public figure]" (Thomas, 2015, p. 97). Thanks to these templates and through a process of "guided discovery learning" (Thomas, 2015, p. 102), students are expected to acquire productive knowledge of the lexicogrammatical uses of the target words.

The above examples are instances of what Kennedy and Miceli (2017) call "pattern hunting", which involves letting students explore a corpus to find language patterns and information about them. DDL can also serve the purpose of "pattern refining", which provides students with "ways of editing their text for lexico-grammatical accuracy" (Kennedy & Miceli, 2017, p. 94). Li (2023) illustrates pattern refining, with students using the Corpus of Contemporary American English (COCA) to correct the collocation errors underlined in their essays. The study demonstrates



(Lexico)grammar and Data-Driven Learning, Fig. 1 Word sketch of the noun *research* extracted by means of Sketch Engine from the BAWE corpus

the superiority of the DDL approach over the use of a dictionary and indicates a sustained learning effect.

DDL Approaches to Grammar

The relatively low number of DDL studies that have dealt with grammar may be due to the impression that DDL is "difficult to reconcile with the 'big themes' of language teaching, such as 'tenses' or 'articles'" (Hunston, 2002, p. 184). Misguided perceptions may also explain some of the less positive evaluations of DDL approaches to grammar. In Yoon and Hirvela (2004), students having experienced DDL evaluated it as somewhat less helpful for grammar than for most other aspects (usage of vocabulary/phrases, writing skills, academic writing ability, etc.). This, however, is arguably linked to the fact that "grammar as presented through a corpus orientation did not conform to their notion of learning about grammar" (Yoon & Hirvela, 2004, p. 269).

DDL may also seem less crucial for grammar than for lexicogrammar because many pedagogical resources include materials for grammar teaching. However, it should be emphasized that DDL approaches to grammar are quite different from traditional approaches. Traditionally, grammar is taught by means of abstract rules that learners have to study. The problem is that learners may master the rules without being able to apply them (Gaskell & Cobb, 2004). DDL, by contrast, encourages learners to observe the grammatical regularities found in corpus data, which makes them more active in the process and hence more likely to remember and reproduce the discovered patterns. In Boulton (2007), students proved able to detect patterns on the basis of concordances from the British National Corpus (BNC) illustrating the future uses of *will* and *going to* and to apply these patterns to new contexts. While the results were similar to those of students using traditional grammatical rules, Boulton (2007, p. 10) argues that they would have been better with "more appropriate corpora, more carefully selected concordances, and especially learner training in interpreting the concordances".

As is the case with lexicogrammar, DDL can also start from a pattern, for example, a grammatical

construction whose features are examined in the context of corpus data. In Gilquin (2021), three constructions were the focus of a DDL intervention: the *MAKE* causative construction (e.g. *She tried to make him look at her*), the *way* construction (e.g. *She elbowed her way through the crowd*) and the *into* causative construction (e.g. *He had blackmailed her into marrying him*). Students received a printed handout with concordance lines from the BNC in which the constructions had been highlighted. They were also given a worksheet with some tasks to carry out with the help of the concordance, for example, determining the syntactic category of the infinitive in instances of the *MAKE* causative construction (transitive, intransitive or copular) and identifying the prepositional phrase in instances of the *way* construction. As a reflection of the overlap between grammar and lexicogrammar, the students also had to discover the words typically occurring in some slots of the constructions (e.g. the main verb in the *into* causative construction).

Certain types of corpora are particularly useful for DDL activities centring around grammatical aspects. This includes part-of-speech (POS) tagged corpora, in which each word has been assigned a label indicating its word class. Friginal (2018, p. 32), for instance, explains how students can be asked to define a noun on the basis of the POS tags of the words found before and after nouns in a POS tagged corpus (e.g. preceded by an article, followed by another noun). In Hegelheimer (2007), it is a different type of corpus that is used to help students improve their grammatical accuracy, namely, an error-tagged learner corpus, that is, a corpus of texts produced by language learners in which errors have been identified and corrected. Hegelheimer (2007) describes how, through the iWRITE program, students can select an error type, such as tense inconsistency, and see examples of this error type from the corpus as well as corrected versions.

It has been suggested that DDL approaches to grammar may be less suited to "present sequences of new structures that learners are hypothesized to be ready for" than to "shed light on structures they have already shown a need for" (Gaskell & Cobb, 2004, p. 303). This points to the relevance of the corrective function of DDL for grammar, that is, the use of corpus data to make one's texts grammatically more accurate. In O'Sullivan and Chambers (2006), learners of French were asked to correct and improve their essays with the help of a French corpus and WordSmith Tools (Scott, 1996). The analysis of the changes made after consulting the corpus revealed that, of all errors found in the essays, it was the category of grammatical errors that led to the highest number of revisions.

Conclusion

DDL has gained respectability as a pedagogical approach to lexicogrammar, which it is ideally suited to deal with, since corpus techniques help make lexicogrammatical patterns more noticeable. While DDL has been less often used to cover grammatical aspects of language, it has been shown to be suitable for this purpose too.

DDL approaches to grammar and lexicogrammar actually bear some similarities. They can both involve exploring the textual environment of a linguistic item or taking a whole pattern as a starting point. They can also fulfil a corrective function, when learners use a corpus as a source of information to correct and improve their texts. In addition, both have unique contributions to make to language pedagogy: DDL approaches to lexicogrammar provide insights into an aspect of language whose treatment in traditional pedagogical materials tends to leave much to be desired, while DDL approaches to grammar give learners an active role of pattern discoverers rather than rule memorizers.

Although both types of approaches inherit the challenges associated with DDL, they also offer considerable opportunities (see ▶ "Future Challenges and Opportunities for the Field of DDL"), which teachers and students alike should seize to help with the development of (lexico)grammatical competence.

Cross-References

▶ Future Challenges and Opportunities for the Field of DDL
▶ Vocabulary and DDL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anthony, L. (2011). *AntConc 3.2.4 [Computer Software]*. Waseda University. https://www.laurenceanthony.net/software

Boulton, A. (2007). DDL is in the details… and in the big themes. In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.), *Proceedings of the corpus linguistics conference 2007*. University of Birmingham. https://ucrel.lancs.ac.uk/publications/cl2007/paper/126_Paper.pdf

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Charles, M. (2015). Same task, different corpus: The role of personal corpora in EAP classes. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 131–154). John Benjamins. https://doi.org/10.1075/scl.69.07cha

Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. Routledge.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://doi.org/10.1016/j.system.2004.04.001

Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics, 31*(2), 229–247. https://doi.org/10.1111/ijal.12317

Hegelheimer, V. (2007). Helping ESL writers through a multimodal, corpus-based, online grammar resource. *CALICO Journal, 24*(1), 5–32. https://www.jstor.org/stable/24156292

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning, 30*(1–2), 91–114. https://doi.org/10.1080/09588221.2016.1264427

Kilgarriff, A., Kovář, V., Krek, S., Srdanović, I., & Tiberius, C. (2010). A quantitative evaluation of word sketches. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the 14th EURALEX international congress* (pp. 372–379). Fryske Akademy.

Li, L. X. (2023). Promoting accuracy of collocation use in L2 writing: The role of data-driven learning in indirect corrective feedback. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2023.2292554

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies, 22*, 1–18. https://jaecs.com/jnl/jnl_22.pdf

O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing, 15*(1), 49–68. https://doi.org/10.1016/j.jslw.2006.01.002

Pearce, M. (2007). *The Routledge dictionary of English language studies*. Routledge.

Scott, M. (1996). *WordSmith Tools [Computer Software]*. Oxford University Press. https://lexically.net/wordsmith

Thomas, J. (2015). Stealing a march on collocation: Deriving extended collocations from full text for student analysis and synthesis. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 85–108). John Benjamins. https://doi.org/10.1075/scl.69.05tho

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL, 28*(2), 207–226. https://doi.org/10.1017/S0958344015000269

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

T

Translation and Data-Driven Learning

Silvia Bernardini
University of Bologna, Forlì, Italy

Abstract

Data-driven learning (DDL) is particularly well suited to translator education, providing an adequate pedagogic approach for two related yet distinct learning objectives: the development of translation competence and the learning of foreign languages. Interest in data-driven learning in the translation classroom started in the 1990s, as an obvious development of corpus-based translation studies and corpus-assisted language learning. Two main types of translation-driven corpora have been used so far in this setting: bilingual comparable corpora (sets of nontranslated comparable texts in two languages) and parallel corpora (sets of original texts and their translations, aligned at sentence level). While the former raise challenges at the level of analysis, due to the difficulties involved in establishing cross-linguistic equivalences, the latter are easier to consult, but the evidence they offer should be carefully evaluated, as it may be influenced by the translation process. Attempts at assessing the effectiveness of DDL in translator education are still limited. Evidence obtained by comparing translations carried out with and without corpora remains inconclusive, even though a positive learning effect seems to exist, particularly for translation into the foreign language. The perceptions of students are also mixed, due to the cognitive burden of corpus analysis and its time-consuming nature. Yet a pedagogy that favours the acquisition of research skills and flexibility, like DDL, would seem to be particularly well suited at a time when the translation profession is faced with important societal and technological challenges. To address them, educators should favour a change of attitude in students, so that they embrace complexity.

Keywords

Data-driven learning · Language learning · Monolingual comparable corpora · Parallel corpora · Translation competence · Translation studies · Translator education

Introduction

This contribution describes the specificities of data-driven learning (DDL) when employed as a didactic approach in translator education contexts. The term DDL itself can be understood both in a more general sense, to refer to the use of 'the tools and techniques of corpus linguistics for pedagogical purposes' (Gilquin & Granger, 2021, p. 430), or in a narrower sense, as '[t]he hands-on use of authentic corpus data (concordances) by

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_82-1

advanced, sophisticated foreign or second language learners in higher education for inductive, self-directed language learning of advanced usage' (Boulton, 2011, p. 572). It is the latter definition that is adopted here, taking into consideration the profiles and needs of translation students, which differ in several ways from those of most other learners.

The most obvious difference is that translators, as future language professionals, are also professional language learners. Throughout their careers, they will need to extend and fine-tune their knowledge of different domains and text types, depending on their areas of specialisation. Learning to quickly and efficiently discern the typical textual and contextual features of texts they are not familiar with, and to adapt their translation strategies accordingly, are skills that need to be learnt at university and then practised throughout one's career. Furthermore, as mediators between two languages and cultures, translation students need to develop a keen sensitivity for contrastive differences between their working languages. A focus on form and access to multilingual textual resources are thus essential to their education. While form-focused instruction is no longer considered with suspicion in language teaching in general (Ellis, 2016), its role in translator education has always been undisputed. Finally, translators often resort to textual materials for reference purposes in their work, to investigate domain-specific phraseology, terminology and concepts. Turning this material into a corpus and turning instances of reference use into a learning experience are obvious extensions of these traditional, well-established practices.

Throughout this contribution, I will focus specifically on direct uses of corpora by students of translation. This includes both learning contexts geared more towards the learning of *translation* skills and others where *language* skills are the main focus, since drawing a clear line between the two is not always possible nor advantageous. After a brief survey of the origins of work on corpora in the translation classroom (section "Origins"), section "Translation-Driven Corpora Used for DDL" describes the most relevant corpus typologies, section "DDL and Translation: Issues and Findings" reports on the main research topics and findings and section "Summing Up and Looking Ahead" summarises the contribution and offers some thoughts on future trends.

Origins

Interest in DDL with translation students can be traced back to two main research strands in the 1990s. On the one hand, following Baker's (1993) proposals for the extension of corpus linguistics assumptions and methods to translation studies, translation scholars started to explore pedagogic applications within the applied branch of corpus-based translation studies. In a seminal issue of the translation studies journal *Meta* dedicated to the corpus-based approach (Laviosa, 1998), Zanettin (1998) describes activities relying on small bilingual comparable corpora through which translation students seek confirmation for their hypotheses about cross-linguistic correspondences and at the same time learn about domain terminology and content and explore texts contrastively, to establish correspondences and differences in an open-ended manner.

The other fundamental input came from work on DDL and pedagogic applications within corpus linguistics. Along with the large corpora being developed in the 1990s for research purposes (such as the British National Corpus, BNC Consortium, 2007), some scholars advocated the construction of small, specialised corpora for use in the language teaching classroom (Ghadessy et al., 2001), by and for learners. Gavioli (2005) showed how learners can use such corpora to observe and thus learn to produce English for Special Purposes (ESP) texts. She suggested that they act as 'spies' or 'double agents' (drawing on Swales, 1990), that is, outsiders trying to infer the communicative practices of a discourse community observing its texts, with the aim to produce (or translate) texts of their own conforming to such practices.

Substantial relevant work from these two strands converged in the CULT (Corpus Use and Learning to Translate) and TaLC (Teaching and Language Corpora) conference series and is

available from the corresponding edited volume proceedings.

Translation-Driven Corpora Used for DDL

Focusing specifically on translation-driven corpora, that is, corpora created and/or used for some translation-related purpose (Zanettin, 2014), two types are especially relevant to DDL applications: bilingual comparable corpora and parallel corpora. Bilingual comparable corpora are collections of texts in two languages, ideally belonging to the same text type and dealing with the same subject matter, but 'genetically' unrelated to each other. Parallel corpora are also collections of texts in two languages, but in this case, the texts are genetically related, in the sense that one subcorpus contains the *source*, or original texts, and the other the *target*, or translated texts.

Building ad hoc bilingual comparable corpora, usually for specific translation assignments, is one of the learning tasks translation students may be asked to carry out, to develop their text-analytic and documentation competences (Herget & Alegre, 2023). The process does not require technical expertise but may prove cognitively complex due to the effort involved in evaluating the comparability and reliability of texts (usually found on the Web).

For instance, in a translation course focusing on a single, relatively specialised topic such as alternative proteins, students may be asked to collect comparable texts in their source and target languages. The process would involve first of all making decisions as to why the corpus is being built: for instance, to clarify ambiguous expressions found in the source text, to extract terminology and phraseology and to observe register and stylistic preferences contrastively. Next, decisions concerning text types to be included/excluded should be made accordingly: is one going to collect scientific and/or popularising articles, institutional and/or company reports, etc.? Once the target text population is delimited in this way, several other inclusion/exclusion criteria should be agreed upon collectively, for example, which time span to use, how to assess reliability, whether to exclude (machine) translated texts, etc. After a first round of text collection, the results are jointly appraised, and the selection criteria further refined. As learners evaluate whole texts for inclusion in the corpus, they also develop a level of familiarity with them, which comes in handy once they access them through corpus analysis tools, to extract word lists for a glossary or to check the attestedness of specific terms and collocates. Thanks to this familiarity, and to the textual cues offered by the texts in their original form (with images, formatting etc.), it should be easier for learners to 'authenticate' corpus findings with respect to the translation task they are performing.

Parallel (translated) texts are easier to spot than comparable ones, but their availability is more limited in general, and their inclusion in a parallel corpus requires that a series of preliminary steps are carried out to align the texts at subtextual, usually sentence, level. The consultation of an aligned parallel corpus also requires specific software (parallel concordancers), making these resources overall less easily accessible. When a parallel corpus is available, its analysis is more straightforward than is the case with bilingual comparable corpora, since parallel concordancers display matching segments of texts, and users can easily spot cross-linguistic correspondences, without the need to *guess* at such correspondences based on cotextual cues, as would be the case in comparable corpus analysis. However, evaluating translation equivalence based on parallel corpus analysis also requires a level of sophistication in the interpretation of results and consequently teacher guidance.

For instance, one term that students are likely to come across when working on alternative proteins is the modifier *plant-based*, as in *plant-based meat* or *plant-based products*. ChatGPT offers *carne vegetale* and *carne a base di piante* as Italian equivalents (OpenAI, 2024). A search for *plant-based* in the English-Italian subset of the InterCorp parallel corpus (https://www.korpus.cz/, Čermák & Rosen, 2012) returns a single occurrence of *a base di piante* and 34 occurrences of *vegetale* (out of 69 total occurrences of *plant-based*). Other translations also exist however,

such as *vegano* (*vegan*), or *a base di verdura* (*made of vegetables*). Evaluating the appropriacy of such translation equivalents would involve not only taking into account the frequency of the equivalents but also the cotext: *vegano* generally appears as a collocate of *dieta* (*diet*) not of *meat* or *products*. It would also, crucially, involve critical evaluation of the actual meaning correspondence between the two terms. Translating *plant-based* as *a base di verdura* might indeed lead an Italian reader to think that plant-based products only include vegetables, to the exclusion of grains, beans, nuts etc. Summing up, the parallel concordance from InterCorp does suggest a correct equivalent term (*vegetale*) but also potentially creates ambiguity due to the presence of other expressions that are arguably less cotextually (*vegano*) or semantically (*a base di verdure*) appropriate. Learning to choose the most appropriate option among a set of possible alternatives, and to justify one's decision with reference to observable textual features, is an essential component of translator competence (see also Pym, 2003), which can be practised and honed through DDL.

A further issue to consider when selecting a corpus typology for DDL activities is related to the impact of the translation process on parallel corpora. This is known to make translated texts differ subtly from comparable nontranslated texts (Kotze, 2020). Therefore, the direction of translation (which is the source and which is the target language) should always be known to the corpus user and held constant across all the texts in the corpus. Furthermore, evidence from the target side of the corpus should be carefully evaluated when one's aim is to observe communicative practices in a given discourse community in general. This is because the translation process may leave traces on texts, resulting from the effort of interpreting the source text and adapting it to the target language. This is not to say that parallel corpora should not be used: they offer valuable, easy-to-access, contextual evidence about decision-making in translation and have been used extensively in DDL activities in the translation classroom (see section "DDL and Translation: Issues and Findings").

Beyond bilingual comparable and parallel corpora, pedagogic tasks can also be designed that take advantage of other translation-driven corpus types, such as translation learner corpora (collections of multiple student translations of the same source text), as in Castagnoli et al. (2011) and Munday (2012). This is in line with the very nature of DDL, that is, an active learning approach based on the exploration of authentic textual data sources.

DDL and Translation: Issues and Findings

Starting with the seminal work by Zanettin (1998), several studies have described the use of corpora in translation courses, where the main objective is the acquisition of translation competences and skills (see, for example, Kruger, 2012).

Aston and Kübler (2009) see the translation process as consisting of three phases, namely, documentation, drafting and revision. They illustrate the ways in which different types of corpora can assist in all three phases, providing practical advice for effective didactic use and exemplification. Focusing on the first phase, namely, documentation, Symseridou (2018) guided translation students to build a specialised bilingual comparable corpus using WebBootCat (Baroni et al., 2006) through SketchEngine (https://www.sketchengine.eu) and then use the corpus to translate a specialised text from French into Greek. When interviewed about their learning experience, students voiced their concerns about the time and effort required to arrive at solutions based on corpora. Such concerns, consistently observed in studies on DDL for translation, constitute one of the major limiting factors for its adoption. As claimed by Aston (2009), a change in attitude is needed since '[n]ot all translators, be they learners or professionals, appreciate that corpus use may have a medium- and long-term payoff which can override what they often perceive as short-term disadvantages' (p. x).

A few studies have attempted to evaluate the effects of DDL methods on the learning of translation skills. Results remain largely inconclusive

due to the difficulty of singling out the variable of corpus use when measuring improvements in translation quality in an ecologically valid way. Yet some findings appear promising. Liu et al. (2023) observed a significant difference in translation performance by a group of Chinese students translating extracts of company profiles from English into Chinese with the help of a parallel corpus when compared to a control group using online reference resources. No difference was instead found in translation into English, their foreign language. However, in a similar experiment, Liu (2020) had reached the opposite conclusion, finding clearer evidence of improvement in translation into the foreign language than into the first language. Frérot (2016) and Giampieri (2020) also focused on L2 translation. They observed overall positive effects of using a specialised corpus of the target language for translation and revision tasks into English as a foreign language by French and Italian students. This may be related to a more general positive effect of corpus use on L2 writing, to which L2 translation is clearly germane. However, they also reported that some students relied uncritically on corpus evidence or made hasty decisions, which led to mistakes. In general, the cognitive effort of drawing conclusions based on corpus data, particularly if nonparallel, remains a hurdle which needs to be addressed more systematically in DDL pedagogy.

Studies addressing foreign language learning as a main rather than side effect specifically for translators are more limited, possibly since they overlap with other areas of DDL (ESP learning, advanced general language learning). An exception is Cresswell (2007), who experimented with DDL for the teaching of English logical connectors to advanced translation and interpreting students. He found that the method was moderately effective but pointed to the strong effect of individual variables, particularly learning styles, affecting the success of DDL in this setting. More methodologically oriented reflections were offered by Contreras (2016), who pointed out the specificity of this language learning setting, where 'languages are both the means by which translation theory and practice are taught, and the object of the learning process itself' (p. 165).

Summing Up and Looking Ahead

As suggested in this contribution, DDL seems especially suitable to translator education, contributing to the acquisition of both language- and translation-related competences and skills. Corpus types employed to these effects have been mainly of the bilingual comparable and parallel types. Apart from providing methodological suggestions about tasks and observational data, some studies have attempted to empirically measure the effects of DDL on learning, with mixed results due to the hurdles involved in designing ecologically valid experiments in the translator education setting.

At a time when the translation profession faces major disruptions due to societal and technological changes (widespread multilingualism, non-professional translation, neural and large language model-driven machine translation), translators should try to become more flexible and adaptive, as well as capable of interpreting (and questioning) the output of ever more sophisticated technological aids (Bernardini et al., 2020). DDL, not just based on corpora but also on other data-driven technologies (Enríquez Raído et al., 2019), would seem to provide the perfect learning approach for the development of such essential research-oriented and transversal skills.

Cross-References

- ▶ Constructivist Approaches to Data-Driven Learning
- ▶ Data-Driven Learning and Languages Other than English
- ▶ Machine Translation Literacy
- ▶ Tools for Data-Driven Learning
- ▶ Types of Corpora in Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Aston, G. (2009). Foreword. In A. Beeby, P. Rodríguez Inés, & P. Sánchez-Gijón (Eds.), *Corpus use and translating* (pp. ix–x). Benjamins.

Aston, G., & Kübler, N. (2009). Using corpora in translation. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 501–515). Routledge.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini Bonelli (Eds.), *Text and technology* (pp. 223–250). Benjamins.

Baroni, M., Kilgarriff, A., Pomikalek, J., & Rychly, P. (2006). WebBootCaT: Instant domain-specific corpora to support human translators. In *Proceedings of the 11th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation.

Bernardini, S., Bouillon, P., Ciobanu, D., van Genabith, J., Hansen-Schirra, S., O'Brien, S., Steiner, E., & Teich, E. (2020). Language service provision in the 21st century: Challenges, opportunities and educational perspectives for translation studies. In S. Noorda, P. Scott, & M. Vukasovic (Eds.), *Bologna process beyond 2020* (pp. 297–303). Bononia University Press.

BNC Consortium. (2007). *The British national corpus* (XML ed.). Oxford Text Archive. http://hdl.handle.net/20.500.14106/2554

Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563–580). Peter Lang.

Castagnoli, S., Ciobanu, D., Kunz, K., Kübler, N., & Volanschi, A. (2011). Designing a learner translation corpus for training purposes. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 221–247). Peter Lang.

Čermák, F., & Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics, 17*(3), 411–427. https://doi.org/10.1075/ijcl.17.3.05cer

Contreras, N. (2016). A proposal for language teaching in translator training programmes using data-driven learning in a task-based approach. *International Journal of English Language & Translation Studies, 4*(2), 155–167.

Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267–287). Rodopi.

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research, 20*(3), 405–428.

Enríquez Raído, V., Austermühl, F., & Sánchez Torrón, M. (2019). Computer-assisted L2 learning and translation (CAL2T). In S. Laviosa & M. González-Davies (Eds.), *The Routledge handbook of translation and education* (pp. 278–299). Routledge.

Frérot, C. (2016). Corpora and corpus technology for translation purposes in professional and academic environments: Major achievements and new perspectives. *Cadernos De Tradução, 36*(1), 36–61. https://doi.org/10.5007/2175-7968.2016v36nesp1p36

Gavioli, L. (2005). *Exploring corpora for ESP learning*. Benjamins.

Ghadessy, M., Henry, A., & Roseberry, R. L. (Eds.). (2001). *Small corpus studies and ELT: Theory and practice*. Benjamins.

Giampieri, P. (2020). Volcanic experiences: Comparing non-corpus-based translations with corpus-based translations in translation training. *Perspectives, 29*(1), 46–63. https://doi.org/10.1080/0907676X.2019.1705361

Gilquin, G., & Granger, S. (2021). Using data-driven learning in language teaching. In A. O'Keeffe & M. J. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 430–442). Routledge.

Herget, K., & Alegre, T. (2023). Data-driven project-based learning in specialized translation classes: The case of comparable corpora. In *Proceedings of the 9th International Conference on Higher Education Advances* (pp. 1097–1104). Universitat Politècnica de València.

Kotze, H. (2020). Translation as constrained communication: Principles, concepts and methods. In S. Granger & M.-A. Lefer (Eds.), *Extending the scope of corpus-based translation studies* (pp. 67–98). Bloomsbury.

Kruger, R. (2012). Working with corpora in the translation classroom. *Studies in Second Language Learning and Teaching, 2*(4), 505–525. https://doi.org/10.14746/ssllt.2012.2.4.4

Laviosa, S. (Ed.). (1998). The corpus-based approach. *Meta, 43*(4), 474. https://doi.org/10.7202/003424ar

Liu, K. (2020). *Corpus-assisted translation teaching: Issues and challenges*. Springer Nature.

Liu, K., Su, Y., & Li, D. (2023). How do students perform and perceive parallel corpus use in translation tasks? Evidence from an experimental study. In J. Pan & S. Laviosa (Eds.), *Corpora and translation education: Advanced and challenges* (pp. 135–157). Springer Nature.

Munday, J. (2012). *Evaluation in translation: Critical points of translator decision-making*. Routledge.

OpenAI. (2024). *ChatGPT* (July 25 version) [Large language model]. Retrieved from https://www.openai.com

Pym, A. (2003). Redefining translation competence in an electronic age: In defence of a minimalist approach. *Meta, 48*(4), 481–497. https://doi.org/10.7202/008533ar

Swales, J. (1990). *Genre analysis*. Cambridge University Press.

Symseridou, E. (2018). The web as a corpus and for building corpora in the teaching of specialised translation: The example of texts in healthcare. *FITISPos International Journal, 5*(1), 60–82. https://doi.org/10.37536/FITISPos-IJ.2018.5.1.160

Zanettin, F. (1998). Bilingual comparable corpora and the training of translators. *Meta, 43*(4), 616–630. https://doi.org/10.7202/004638ar

Zanettin, F. (2014). *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. Routledge.

C

(Critical) Discourse Analysis and Data-Driven Learning

Kevin Frank Gerigk
Aston University, Birmingham, UK
Lancaster University, Lancaster, UK
Universidad de Santiago de Chile, Santiago, Chile

Abstract

Multiculturalism and diversity are crucial societal aspects of life in the twenty-first century. In this entry, I discuss how corpus linguistics in the shape of data-driven learning (DDL) for critical discourse analysis (CDA) should, could, and has been used in (mainstream) L2 pedagogy for its potential to develop and foster criticality amongst aspiring global citizens. First, I provide definitions of the key concepts *DDL* and *CDA*, before I address the considerations around cultural diversity and critical awareness in the L2 classroom. I then discuss the status quo of criticality in programmes of study in higher education in the United Kingdom and the United States. Then, I criticise the representation of culture and diversity in EFL, which is still often based on 'othering'. I continue that section by addressing how DDL can be put to action, conducting CDA in the classroom to promote criticality through authentic, published input. This discussion is supported by the summary of a case study on how DDL has been used for a CDA lesson on the representation of immigration and gender through a mini corpus project. Thereafter, I address longstanding practical concerns and offer a discussion of partially available recent answers to these concerns, drawing upon analysis tools that can be used in the DDL classroom. In the last section of this entry, I address avenues of future research and practical needs in terms of teacher education and the issue of Eurocentrism in DDL research.

Keywords

Data-driven learning · (Critical) Discourse analysis · Corpus linguistics · Criticality · Cultural awareness

Introduction

Data-driven learning (DDL) enables learners to engage directly with linguistic data by using a linguistic corpus for language input. I argue in this entry that, due to the access to large amounts of linguistic data, corpus-based DDL is a suitable vehicle to conduct critical discourse analysis (CDA) in the diverse and multicultural English as a foreign language (EFL) classroom of the twenty-first century. Today, English is used increasingly between non-native speakers with different L1 backgrounds. This has led to two phenomena: (1) the distinction between the learners' L1 culture and the target culture (e.g. Britain or the United

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_83-1

States) has become blurry; (2) the cultural influences and identities of L2 English users seem as diverse as never before. This produces difficulties and potential issues regarding attempts to emphasise a model L2 target culture in the EFL classroom. Furthermore, learners and teachers alike may find themselves exposed to transcultural biases detached from the L1 English target culture. In order to expose L2 learners to various cultural aspects of their L2 and their peers' L1s, and to foster cultural awareness, criticality and empathy, corpus-based CDA has emerged as a promising approach in the diverse L2 classroom. As proposed by Fahey Palma (2022), corpus-based CDA allows L2 learners to experience and analyse intercultural contexts and to critically engage with the representations of minority groups. This may facilitate discussions around diversity and inclusivity through the exposure to authentic data and the promotion of critical questions regarding specific discourse contents and cultural and societal representations. Ultimately, this may help combat biases and prejudices.

In this entry, I first provide definitions of the key concepts *corpus linguistics and DDL* as well as *(critical) discourse analysis*. Thereafter, I discuss the affordances of CDA for critical awareness in the L2 classroom and how this approach can be realised through DDL. In this, I focus particular attention on the aspects of cultural diversity and critical awareness in L2 pedagogy, before I outline how DDL and CDA in action can promote criticality through authentic input. I conclude this entry by discussing future directions for research and teaching.

Data-Driven Learning and Critical Discourse Analysis

Data-Driven Learning

DDL was first introduced by Johns (1990) and refers to the use of linguistic corpora by the language learners (see discussion by Scott, this volume). Using corpus data in the classroom allows learners to engage directly with linguistic data and turns them into language detectives. A linguistic corpus (plural: corpora) is a large, principled collection of written or spoken text which can be analysed for linguistic patterns within the speech context using computer tools (McEnery & Hardie, 2012). DDL can be distinguished into two strands: indirect and direct (Farr & O'Keeffe, 2019). In indirect DDL, corpora are used to create teaching materials and reference materials, e.g. dictionaries. Moreover, the teacher can create task sheets in the form of curated concordance lines and devise tasks from them, e.g. gap fill exercises. On the other hand, direct DDL refers to the learners' direct use of a corpus in the classroom to research and identify patterns in the selected genre. In direct DDL, the teacher assumes a rather passive role, whilst the learners explore linguistic data on their own. Depending on the learners' L2 proficiency and corpus literacy, the teacher may have to curate and scaffold the learners' exploratory journey. Nonetheless, direct DDL is a teaching approach that offers a large degree of learner autonomy and allows for independent learning (see Charles, this volume).

Critical Discourse Analysis

CDA focuses on the macrostructure of language use and the social context in which the language is produced by a member of a specific speech community (Schiffrin et al., 2001). As such, discourse in social interaction can be defined as a social practice and is highly context-dependent in terms of appropriacy (McCarthy & McCarten, 2022). This means that discourse is a form of social practice where language influences societal norms and where societal norms influence the use of language (Fairclough, 1989). CDA helps to reveal these intricate relationships and the connections between language, power, and ideology (Cots, 2006) and allows us to uncover patterns of the representation of (marginalised) social, societal, geographical, or ethical communities, belief systems, and religions. In this context, 'critical' denotes an enquiry into 'making transparent taken-for-granted assumptions' (Koller, 2014, p. 149) about how discourse participants reinforce and/or subvert power relations. By critically engaging with (potentially controversial) texts and genres, learners are enabled and empowered to critically use the L2 to examine and judge the world and challenge their belief systems.

Data-Driven Learning and (Critical) Discourse Analysis for Critical Awareness in the L2 Classroom

Corpus-Driven CDA

Corpus-driven CDA offers access to large numbers of words and linguistic patterns from various text genres, exposing the user/learner to ideological trends within the discourse. This means that learners can consult corpora to quantitatively analyse patterns of language use in the context in which they were produced by or in relation to a speech or cultural community. These patterns can be further analysed qualitatively, allowing the L2 learner to critically assess social phenomena and the linguistic representation of (minority) communities and resulting societal issues, in practice (Baker, 2006). This makes DDL and CDA an invaluably rich source for the development of the L2, criticality, and cultural understanding. Using this approach has a particular value in terms of meaningful language use as the learners use the L2 to engage with real-world issues and not just for the sake of language learning. I discuss this argument from the perspective of CDA and criticality in L2 pedagogy and from the perspective of applied DDL/CDA in the following subsections.

Cultural Diversity and Critical Awareness in L2 Pedagogy

Culture and identity are crucial concepts in L2 classrooms. This is particularly prominent in EFL contexts, where multiple cultures use a single language for cross-cultural communication. Due to the growing level of diversity amongst the users of English, intercultural awareness in conjunction with language proficiency should lie at the heart of instructed EFL. Yet, it appears that cultural awareness is not effectively covered in Anglophone higher education contexts, which tend to neglect the complexities of cultural and societal identities (Hua et al., 2017), leading to potentially prejudiced assumptions about the L2 target culture (Sayer & Meadows, 2012). Here, I should question the use of the term *target culture* since it constitutes the main criticism of intercultural competence teaching: *othering*. As Coupland (2010) states: *Othering* refers to the negative or simplified representation of a foreign culture. This culturally loaded distinction may imply superiority of one culture over another. Sayer and Meadows (2012) state that in L2 teaching, the cultural component is taken for granted in the teaching content. As such, the target culture is presented as unique and alien compared to the learners' L1 cultural background, which further intensifies potential cultural biases, creating a boundary or separation between 'them' and 'us'. In order to avoid this potentially hostile practice, L2 pedagogy should address cultural differences where they exist and promote approaches that give students the tools to analyse cultural representation. I argue that engaging with different cultural aspects of the target language and beyond may foster open-mindedness and reduce bias in the L2 learner, helping them to join a global community of L2 English users. This is particularly relevant as the global spread of English as a vehicle for communication between speakers from different cultural backgrounds necessitates it to address pre-existing socio-cultural biases to encourage the learners (and teachers alike) to exercise criticality when engaging with the outside world. From a critical language pedagogy perspective, learners need to be aware of cultural nuances (Byram, 1997). Considering the complexities of culture and the challenges presented through incorrect and negative representations of minority groups in public discourse and ELT materials, Fahey Palma (2022) emphasises that it is crucial to provide students with a robust methodology to enable the development of intercultural and intersocietal critical thinking skills that can transcend the L2 classroom and provide opportunities for lifelong critical thinking.

Corpus-Driven DDL and CDA for Criticality in the L2 English Classroom: Fostering Criticality Through Authentic Input

Due to a lack of inclusion of societally and culturally critical content in ELF coursebooks (see, for example, Bagheri and Erfan (2024), for intercultural awareness), I argue that a suitable, if not the most suitable, approach to analysing mainstream representation of societal and cultural communities is teacher-curated CDA based on appropriate corpus linguistics data. As Fahey Palma (2022) states, 'it is [. . .] intuitive that

[corpus linguistic] analysis of the representations of social groups in the mainstream media and beyond can yield fruitful results for understanding intercultural issues and through these discussions enhance [. . .] critical awareness' (p. 119). In other words, through a DDL task design, learners are given the tools to find, extract, and critically examine how a foreign culture or indeed a domestic minority community is pitched against a seemingly favourable norm. Attempts have been made to include CDA into the L2 classroom; however, the use of corpora in these contexts seems underresearched.

Corpus-driven CDA would kill the metaphorical two birds with one stone: A DDL design offers insights into linguistic patters, whilst the learners are enabled to critically analyse how these linguistic patterns are used to represent sociocultural groups (Mautner, 2022). Hence, the use of corpus linguistics allows for a principled analysis of the linguistic devices employed in defining and positioning (outsider) communities through the identification of specific identity markers such as lexical choice, naming and referencing, and pronominal features. This has been showcased convincingly by Fahey Palma's (2022) mini corpus projects to increase cultural awareness. In these projects, the students work with small, specialised corpora to investigate the representation of 'others' and minority groups through textual evidence by close reading of relevant text excerpts and frequency data of the co-occurrence of words and phrases. She presents case studies on the refugee crisis and gender minorities and summarises that students had the opportunity to critically engage with the discourse and its macrofunctions, whilst they also built the skills to conduct corpus linguistics research and engage with rich L2 input. The case studies have shown that DDL and CDA can be an important vehicle for the development of criticality and L2 proficiency by enabling the learner to use the L2 to analyse and discuss sociocultural issues, through the combination of qualitative and quantitative measures. Hence, DDL and CDA can be used qualitatively and quantitatively to promote progressive views on interculturality through the analysis of authentic texts, where students can critically discuss the representation of a variety of social groups and movements (Fahey Palma, 2022).

At this point, it must also be critically considered that apparently convincing arguments have been raised against the use of corpus linguistics and DDL in mainstream educational contexts. Widdowson (2000) was concerned that the learners may not be able to process and work with authentic materials due to their lack of cultural exposure and language proficiency, which is necessary to contextualise the text. This claim fails to account for the learners' actual capability to recontextualise the input that they receive and use it for their own learning experience. Another, and more tangible, concern amongst scholars is the issue of access to (affordable) concordancers (Templeton & Timmis, 2023). It must be acknowledged that access to computer and a stable Internet connection cannot be taken for granted in every part of the world. This may pose a particular challenge to emerging or rural regions in the Global South (Wimpenny et al., 2022). However, recent global events, such as the COVID-19 pandemic, have led to progress in the digitisation of educational contexts (Suárez-Guerrero et al., 2023). Today, English teachers globally have access to freely available concordance tools online (for a comprehensive review, see Gerigk (2023)). For instance, CorpusMate (https://corpusmate.com/; Crosthwaite & Baisa, 2024) offers access to various genres, ranging from political discourse to academic writing. The easy-to-use web interface allows learners to critically analyse linguistic devices and patterns in a large variety of published texts as I demonstrate in the following task proposal on the representation of *transgender* people (for a more detailed discussion on pedagogical guides for DDL in the L2 classroom, see LeFoll, this volume; for a detailed discussion and guide to DDL lesson planning, see Goulart, this volume).

The task presented in Fig. 1 is a proposal for reading comprehension and critical engagement with the lived experiences of the *transgender* community. Due to the wide range of specific lexicogrammatical requirements of the sources, I suggest this task be used with learners who have a proficiency of B2 or higher.

Task. In this exercise, we will use CorpusMate to learn about the representation of members of *transgender* communities.

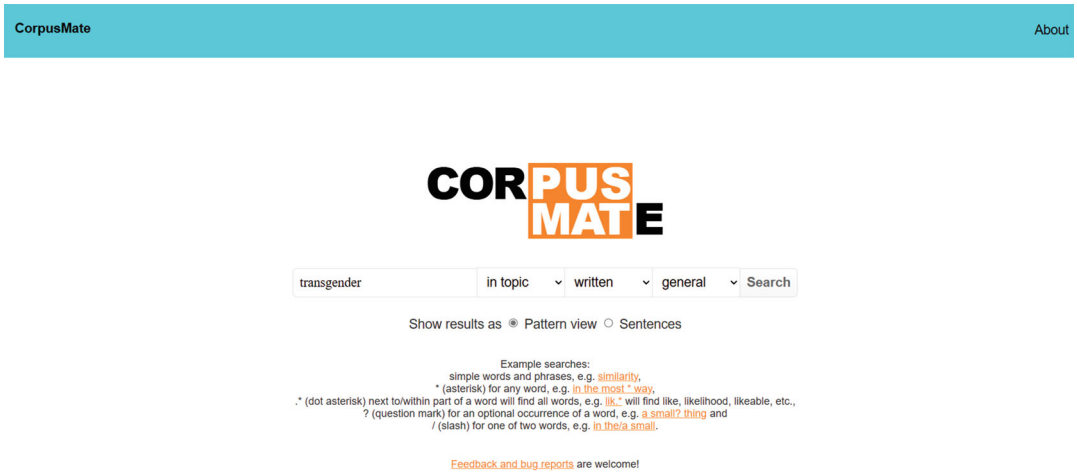
1) Read the following informative article about *transgender* communities critically and identify the main themes. Access to the article: https://www.hrc.org/resources/understanding-the-transgender-community
2) Summarise this article briefly and discuss with a partner the main claims made for each theme. Focus in particular on the following aspects:
   a. Mental health
   b. The law and prosecution
   c. The workplace
3) Now navigate to CorpusMate (https://corpusmate.com/) and enter the search term *transgender* into the query field. Manipulate the parameters as follows:
   a. Topic: default
   b. Mode: written
   c. Genre: general
4) With a partner, expand the first 10 concordance lines and identify the main themes regarding the transgender community in the corpus. Compare the results to the themes in 2) above. Do they agree or disagree with your previous findings?
5) Now click on a few examples of the concordance lines and navigate to the source text. Read a minimum of 5 texts carefully and analyse the reasons for discrimination of and the effects on mental health for the *transgender* community.
6) Now reflect with your partner: Have you observed similar problems in your own community? How have these texts impacted your perception of the problems and struggles of this particular community? What action can you take to support this community? Present your arguments to the whole class.

(Critical) Discourse Analysis and Data-Driven Learning, Fig. 1 Corpus-driven CDA task for CorpusMate

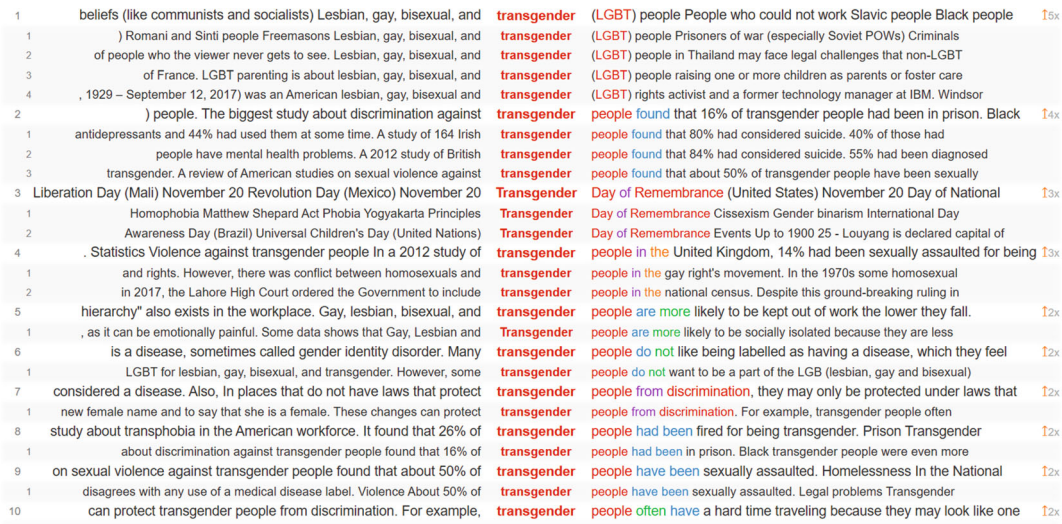
Using CorpusMate, the task allows learners to analyse quantitative, frequency-driven linguistic data and identify linguistic patterns to show how language is used to represent members of the *transgender* community. After introducing the topic through a reading task, students navigate to CorpusMate, enter the query term (*transgender*), and set the appropriate parameters (written mode, nonacademic publications) as shown in Fig. 2.

Students then click on **Search** to run the query and are presented with a number of concordance lines that have the node word *transgender* sorted to the right. This shows the students the most frequent co-occurrences with words in the right-side context of the node word. As can be seen in Fig. 3, these combinations include topics regarding mental health (lines 2.1–2.3), emancipation (lines 3.1–4.2), discrimination (5.1–8.1), and sexual assault (9). Learners are then asked to link these topics back to the initial reading text. Whilst it must be acknowledged here that the concordance lines regard harsh topics which may not be suitable for every individual, I argue that they show valuable insights into the representation and struggles of communities. Hence, the teacher may curate these findings before offering this task to the students according to the learners' needs.

Subsequently, selected concordance lines or larger excerpts from the corpus can be accessed and close read for qualitative analysis (e.g. the reasons for the reported hardship of members of the *transgender* community). This offers a comprehensive picture of the representation of groups in a specific speech community and text genre, whilst



(Critical) Discourse Analysis and Data-Driven Learning, Fig. 2 CorpusMate start page



1	beliefs (like communists and socialists) Lesbian, gay, bisexual, and	transgender	(LGBT) people People who could not work Slavic people Black people	↑5x
1	) Romani and Sinti people Freemasons Lesbian, gay, bisexual, and	transgender	(LGBT) people Prisoners of war (especially Soviet POWs) Criminals	
2	of people who the viewer never gets to see. Lesbian, gay, bisexual, and	transgender	(LGBT) people in Thailand may face legal challenges that non-LGBT	
3	of France. LGBT parenting is about lesbian, gay, bisexual, and	transgender	(LGBT) people raising one or more children as parents or foster care	
4	, 1929 – September 12, 2017) was an American lesbian, gay, bisexual and	transgender	(LGBT) rights activist and a former technology manager at IBM. Windsor	
2	) people. The biggest study about discrimination against	transgender	people found that 16% of transgender people had been in prison. Black	↑4x
1	antidepressants and 44% had used them at some time. A study of 164 Irish	transgender	people found that 80% had considered suicide. 40% of those had	
2	people have mental health problems. A 2012 study of British	transgender	people found that 84% had considered suicide. 55% had been diagnosed	
3	transgender. A review of American studies on sexual violence against	transgender	people found that about 50% of transgender people have been sexually	
3	Liberation Day (Mali) November 20 Revolution Day (Mexico) November 20	Transgender	Day of Remembrance (United States) November 20 Day of National	↑3x
1	Homophobia Matthew Shepard Act Phobia Yogyakarta Principles	Transgender	Day of Remembrance Cissexism Gender binarism International Day	
2	Awareness Day (Brazil) Universal Children's Day (United Nations)	Transgender	Day of Remembrance Events Up to 1900 25 - Louyang is declared capital of	
4	. Statistics Violence against transgender people In a 2012 study of	transgender	people in the United Kingdom, 14% had been sexually assaulted for being	↑3x
1	and rights. However, there was conflict between homosexuals and	transgender	people in the gay right's movement. In the 1970s some homosexual	
2	in 2017, the Lahore High Court ordered the Government to include	transgender	people in the national census. Despite this ground-breaking ruling in	
5	hierarchy" also exists in the workplace. Gay, lesbian, bisexual, and	transgender	people are more likely to be kept out of work the lower they fall.	↑2x
1	, as it can be emotionally painful. Some data shows that Gay, Lesbian and	Transgender	people are more likely to be socially isolated because they are less	
6	is a disease, sometimes called gender identity disorder. Many	transgender	people do not like being labelled as having a disease, which they feel	↑2x
1	LGBT for lesbian, gay, bisexual, and transgender. However, some	transgender	people do not want to be a part of the LGB (lesbian, gay and bisexual)	
7	considered a disease. Also, In places that do not have laws that protect	transgender	people from discrimination, they may only be protected under laws that	↑2x
1	new female name and to say that she is a female. These changes can protect	transgender	people from discrimination. For example, transgender people often	
8	study about transphobia in the American workforce. It found that 26% of	transgender	people had been fired for being transgender. Prison Transgender	↑2x
1	about discrimination against transgender people found that 16% of	transgender	people had been in prison. Black transgender people were even more	
9	on sexual violence against transgender people found that about 50% of	transgender	people have been sexually assaulted. Homelessness In the National	↑2x
1	disagrees with any use of a medical disease label. Violence About 50% of	transgender	people have been sexually assaulted. Legal problems Transgender	
10	can protect transgender people from discrimination. For example,	transgender	people often have a hard time traveling because they may look like one	↑2x

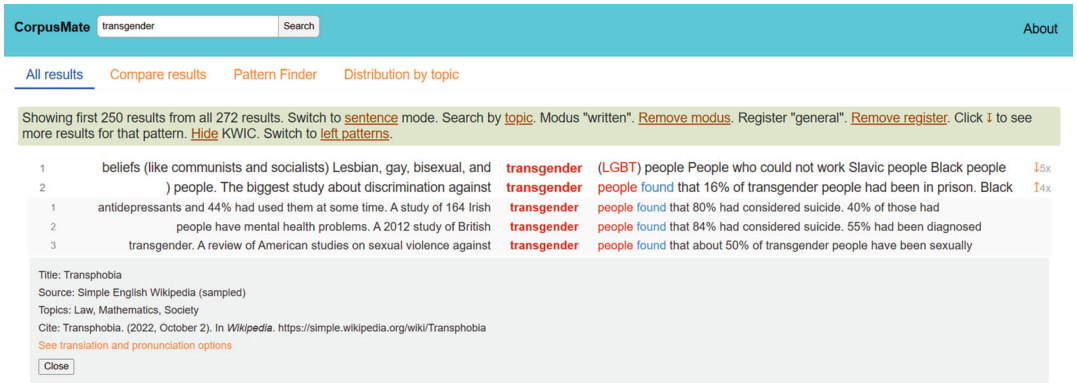
(Critical) Discourse Analysis and Data-Driven Learning, Fig. 3 Concordance lines of transgender in CorpusMate

it simultaneously increases learner autonomy. To do so, learners click on a selected concordance line and enter the source domain into their browsers to access the source text in its original version through their browsers as shown in Fig. 4.

The above tasks deal with a rather sensitive topic. Hence, it is important to debrief the learners and facilitate their contextualising the information and language used. Task 6 is designed to allow the learners to critically reflect on the information and to recontextualise their findings to their own context. This may sensitise the learners for certain socially and politically charged topics and facilitate their development of critical thinking skills, which may lead to becoming more socially just and culturally aware learners and members of society.

Future Directions of Research and Practice

In this entry, I have discussed current research and trends in the use of corpus-based CDA through DDL. I have discussed that biases and prejudices



(Critical) Discourse Analysis and Data-Driven Learning, Fig. 4 Accessing the source domain for close reading of the concordance line in their original text

in the field of ELT can be combatted by promoting critical thinking skills into the L2 classroom. Furthermore, I have provided a rationale for using corpus-driven DDL as the vehicle for such culturally sensitive analysis as it provides the L2 learners with a rich source of linguistic data, which can be analysed qualitatively and quantitatively for patterns in their linguistic context of discourse about potentially marginalised groups. These factors may eventually lead to more open-mindedness in the learners' lives, even beyond the L2 classroom, by fostering criticality in the context of lifelong learning.

What remains is the question: *What is next?* I would like to re-emphasise two demands recently made by researchers in the field. First of all, the practical use of CDA needs to be further emphasised and included into teacher education programmes (see Leńko-Szymańska, this volume). Arellano (2019) finds that whilst CDA is part of teacher education, the modules are likely to be theory-heavy and offer only little value for practical application to the future teaching context. This is also reflected in the underrepresentation of corpus linguistics modules in teacher education programmes (Farr & O'Keeffe, 2019). Hence, it is vital for the success of DDL-driven CDA and the development of criticality amongst L2 learners that teacher education programmes reconsider their syllabi on critical discourse analysis and corpus linguistics, introducing or adapting courses which offer authentic and practical skills for preservice and in-service teachers. Secondly, Boulton (2024) argues that DDL-related output and applications are still rather Eurocentric. I support this call and highlight that research and reports from the Global South are likely to offer novel perspectives on criticality and cultural awareness, as well as cultural representation, which are necessary in the diverse L2 classrooms of the twenty-first century. DDL is an excellent approach to critically analyse discourse and promote cultural awareness and sensitivity amongst the growing population of L2 users of English in an increasingly diverse and multicultural world.

Cross-References

- ▶ Critical Approaches to CALL
- ▶ Critical Digital Literacy
- ▶ Integrating Corpora and Data-Driven Learning into Lesson Planning
- ▶ Language Teacher Education for CALL
- ▶ Learner Autonomy and Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Teacher Education for Pedagogical Uses of Corpora
- ▶ Tim Johns and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Arellano, R. (2019). Learning discourse analysis from a TESOL perspective. *English Australia Journal, 35*(2), 35–44. https://search.informit.org/doi/10.3316/informit.700624057096164

Bagheri, R., & Erfan, R. (2024). Representation of culture and intercultural interactions in Iranian and Afghan secondary English textbooks. *Journal of Foreign Language Teaching and Translation Studies, 8*(2), 109–138. https://doi.org/10.22034/efl.2024.425553.1274

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Boulton, A. (2024). Data-driven learning: In conversation with Alex Boulton. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 45–57). Routledge.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Cots, J. M. (2006). Teaching 'with an attitude': Critical discourse analysis in EFL teaching. *ELT Journal, 60*(4), 336–345. https://doi.org/10.1093/elt/ccl024

Coupland, N. (2010). 'Other' representation. In J. Jaspers, J. Ostman, & J. Verschueren (Eds.), *Society and language use, handbook of pragmatics highlights* (pp. 241–260). Benjamins.

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics, 29*, 595. https://doi.org/10.1075/ijcl.23056.cro

Fahey Palma, T. (2022). Corpora for teaching culture and intercultural communication. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 116–130). Routledge.

Fairclough, N. (1989). *Language and power*. Longman.

Farr, F., & O'Keeffe, A. (2019). Using corpus approaches in English language teacher education. In S. Walsh & S. Mann (Eds.), *The Routledge handbook of English language teacher education* (pp. 268–282). Routledge.

Gerigk, K. F. (2023). Review. CQPweb, BNClab, and BNClab and their applicability to the DDL classroom. *Arboles y Rizomas, 5*(2), 144–150. https://doi.org/10.35588/ayr.v5i2.6408

Hua, Z., Handford, M., & Young, T. (2017). Framing interculturality: A corpus-based analysis of on-line promotional discourse of higher education intercultural communication courses. *Journal of Multilingual and Multicultural Development, 38*(3), 283–300. https://doi.org/10.1080/01434632.2015.1134555

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria, 10*, 14–34.

Koller, V. (2014). Applying social cognition research to critical discourse studies: The case of collective identities. In C. Hart & P. Cap (Eds.), *Contemporary critical discourse studies*. Bloomsbury.

Mautner, G. (2022). What can corpus tell us about discourse? In A. O'Keeffe & M. McCarthy (Eds.), *Routledge handbook of corpus linguistics* (pp. 250–262). Routledge.

McCarthy, M., & McCarten, J. (2022). Corpora for teaching social conversation. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 102–115). Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics – Method, theory and practice*. Cambridge University Press.

Sayer, P., & Meadows, B. (2012). Teaching culture beyond nationalist boundaries: National identities, stereotyping, and culture in language education. *Intercultural Education, 23*(3), 265–279. https://doi.org/10.1080/14675986.2012.702446

Schiffrin, D., Tannen, D., & Hamilton, H. (2001). *The handbook of discourse analysis*. Blackwell.

Suárez-Guerrero, C., Lloret-Catalá, C., & Mateu-Luján, B. (2023). Didactics and technology. Teaching lessons from the long-term emergency remote school. *RIED-Revista Iberoamericana de Educación a Distancia, 27*(1), 397–415. https://doi.org/10.5944/ried.27.1.37686

Templeton, J., & Timmis, I. (2023). A flexible framework for integrating data-driven learning. In K. Harrington & P. Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 39–58). Palgrave Macmillan.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 3–25. https://doi.org/10.1093/applin/21.1.3

Wimpenny, K., Finardi, K. R., Orsini-Jones, M., & Jacobs, L. (2022). Knowing, being, relating and expressing through third space global south-north coil: Digital inclusion and equity in international higher education. *Journal of Studies in International Education, 26*(2), 279–296. https://doi.org/10.1177/10283153221094085

G

Genre Awareness and Data-Driven Learning

J. Elliott Casal[1] and Matt Kessler[2]
[1]University of Memphis, Memphis, TN, USA
[2]University of South Florida, Tampa, FL, USA

Abstract

This encyclopaedia entry introduces readers to data-driven learning (DDL) research that has investigated issues involving the development of second language (L2) learners' genre awareness. The piece begins by defining key terms, which include both DDL and genre awareness. Next, readers are introduced to some of the most common foci involved in DDL studies of genre awareness. Such studies have typically focused on L2 English learner populations, and the extent to which DDL influences these learners' awareness of rhetorical moves and formulaic features of written academic genres. Lastly, the piece concludes with a discussion of common methods involved in DDL research on genre awareness. This discussion covers common design trends in such research, including longitudinal data collection and genre analysis components, as well as common sources of data, including surveys and student writing.

Keywords

Academic English · Applied linguistics · Corpus · Data-driven learning · DDL · ESP · Genre awareness · L2 writing · Research methods

Introduction to DDL and Genre Awareness

Data-driven learning (DDL) refers to a variety of pedagogical implementations that involve language learners interacting with linguistic corpora as a core pedagogical activity. Associated with the pioneering work of Tim Johns (1991), DDL was originally conceived as an inductive approach to language learning that leveraged learner autonomy and agency—as well as corpora of naturally occurring language data—in a student-centred language pedagogy. Corpus linguists and Computer-Assisted Language Learning (CALL) researchers and educators have since implemented DDL in a diverse range of manners, including those that involve direct interaction with corpora or corpus tools and those that involve learner interaction with corpus data that has been reproduced and/or selected by educators. Considerable attention has been paid to lexical and lexicogrammatical learning targets, second language (L2) writing contexts, and intermediate or advanced adult learners as study participants. However, it should be noted that applications have been diverse and are diversifying, as can be seen

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_84-1

by the other entries on DDL in this encyclopaedia. DDL has been examined in the teaching of oral communication (e.g. Hirschi & Kang, 2024), some scholars are dedicating attention to the affordances of DDL for young learners (e.g. Crosthwaite, 2020a), and research on how to develop and support teachers for DDL pedagogies represents a vibrant and expanding line of inquiry (e.g. Farr & Leńko-Szymańska, 2024; Ma et al., 2023).

Within the commonly studied L2 writing context, one area that has captured numerous researchers' attention has been DDL's impact on L2 learners' *genre awareness*. It is worth noting that L2 writing scholars have sometimes made distinctions between the terms *genre awareness* and *genre knowledge* (for more, see Tardy et al., 2020); however, such distinctions historically have not been made by DDL researchers. DDL researchers have also tended to use the term *awareness* interchangeably with, or at least in tandem with, related terms such as *noticing* (see Schmidt, 1990) or *metacognition* (see the encyclopaedia entry on 'Metacognition and DDL' for more). At the same time, genre awareness is not always specified as such, but is sometimes labelled as awareness of genre-specific rhetorical functions (e.g. Charles, 2007). Thus, in the DDL literature, genre awareness has been conceptualized as a term that refers to a learner's conscious recognition or noticing of different features of a focal genre; this includes (but is not limited to) the extent to which DDL can foster noticing of different aspects of commonly used rhetoric, organizational structures, lexicogrammar, and other linguistic features that may be frequently used in a written or spoken genre (see Hyland, 2023, for a related discussion).

Foci of Genre Awareness Studies

When broadly examining the existing body of DDL literature, studies involving genre awareness are less common compared to other targets such as error correction (Crosthwaite, 2020b). Notably, most DDL studies targeting genre awareness have examined written academic genres, such as conference abstracts or research articles (RAs), particularly with English learners at high-intermediate or advanced levels of L2 proficiency. In such studies, the targets of inquiry have also remained relatively consistent. Studies typically focus on raising learners' awareness of formulaic features such as lexical bundles or phrase frames (for more, see Cortes, 2024) or rhetorical moves used in academic genres (see Casal & Kessler, 2024). Importantly, such studies often make use of custom-made and manually annotated corpora.

For example, Cotos et al. (2017) and Dong and Lu (2020) have investigated the impact of DDL on aspects of L2 learners' awareness of written academic genres with a primary focus on rhetorical moves (e.g. Swales, 2004) in research articles, such as building a gap in an introduction or justifying methodological decisions. Both studies examined how DDL activities can promote the development of genre awareness and genre knowledge by examining the linguistic features in their rhetorical contexts. These studies make use of custom-built research article corpora that are annotated for rhetorical moves locally. In the former study, Cotos and colleagues examined whether DDL engagement through the web-based learning and automated feedback package *Research Writing Tutor* helped writers deepen their genre awareness and improve their research writing. Analysis targeted learners' developing genre awareness in terms of discourse patterns (distribution, order, occurrence, amount, and overlap of rhetorical moves) and more local linguistic cues of rhetorical aims that reflected the corpus content (e.g. common lexical item-rhetorical function associations, such as the word 'issue' with gap building). Their findings suggested that the DDL intervention was impactful, particularly on novice writers' awareness of the move structure and conventions of RAs. Similarly, Dong and Lu investigated the influence of DDL on L2 English learners' awareness of rhetorical moves (based on Swales, 2004) in the introduction sections of engineering RAs. Like Cotos et al., the researchers found that engaging with the annotated corpus over several weeks helped students notice the rhetorical structures in RA introductions, along with becoming more aware of specific linguistic features.

Similarly, Casal and Yoon (2023) explored how the implementation of three DDL pedagogical sessions raised L2 English learners' awareness of formulaic language used in conference abstracts. Prior to the instruction, the authors first examined the frequency and functions of phrase frames in conference abstracts (e.g. '*the * of this study'*, [*aim, purpose, goal*]; '*it is * to note that*', [*interesting, important, surprising*]), which were then used as the source material for the pedagogical intervention. The findings of their study highlighted how DDL sessions subsequently impacted students' awareness (i.e. their decision-making processes) of both rhetorical and linguistic choices when composing conference abstracts, as well as their ability to notice and interpret phraseological features and rhetorical moves when reading. Other DDL studies have addressed the teaching of linguistic features to signal academic writers' rhetorical goals (e.g. Charles, 2007; Chen & Flowerdew, 2018; Lee & Swales, 2006), although not always with the explicit aim to target genre awareness.

Although most studies involving DDL and genre awareness have examined written academic genres, some researchers have targeted spoken academic genres. For instance, Zare et al. (2023) examined the impact of DDL activities in helping L2 English learners become aware of the discourse organizers used in oral academic lectures (e.g. *the second point is, to sum up*). In the study, students participated in nine sessions in which they worked with a concordancer, exploring how L1 English speakers conveyed topics of importance in authentic academic lectures. Results from a post-test showed that the DDL group outperformed a comparison group (who received explicit, lecture-style instruction) in comprehending academic lectures. Students in the DDL group also self-reported that they felt DDL was effective for helping them notice key points in oral lectures.

Methodological Approaches

Studies on DDL and L2 genre awareness have many methodological similarities in design, but there have been important innovations in this area with regard to the broader field of DDL research. Firstly, as many studies analysing DDL and L2 genre awareness adopt an English for Specific/Academic Purposes orientation, such projects tend to include or be based on a genre analysis of a representative target corpus that is used in the pedagogical intervention. All studies discussed in the previous section either included a corpus analysis and annotation stage or were otherwise based on a corpus that had been annotated previously for rhetorical and/or linguistic features. Many DDL studies make use of specialized or annotated corpora for DDL activities, and corpus selection is an essential consideration in all DDL endeavours. Considering what is known about genre- and discipline-based language specialization (see Hyland, 2004), corpus construction and/or annotation also play a key role in DDL studies of genre awareness. As a second major trend, studies in this area tend to adopt a longitudinal within-subjects perspective, as genre-awareness and knowledge is theorized and documented to develop over longer periods of time (for a discussion, see Kessler, 2021).

Studies on DDL and L2 genre awareness draw from data sources that allow for learner introspection (perception data), or evaluation of changes in writing processes, products, or changes in learner decision-making (outcomes data). Regarding perception data, researchers of genre awareness employ methodological tools similar to those used by CALL researchers who are interested in investigating learners' perceptions. Learner introspection is often targeted through Likert-based surveys, open-ended questionnaires, interviews, and written reflections of various forms, all of which ask learners to consider and share their own perceptions of learning, confidence, or growth. With regards to outcomes data, most studies assessing the development of genre awareness through DDL use learner writing as a primary source of data, assessing draft revisions for changes or evidence of changes in genre knowledge across writing tasks. Much less common in DDL research on genre awareness (and DDL research more generally) is the use of stimulated recalls or text-based interviews that ask learners to articulate the principles and intentions behind their linguistic and rhetorical choices. In this sense, qualitative data is not used as a means of

evaluating learner perceptions, but rather provide insights into learner decision-making and genre awareness. Some rare examples include Yoon (2008), who implemented this method in case study of six learners' linguistic awareness, and Casal (2020), who used this method as a means of examining changes in genre awareness in L2 English graduate student RA writing. This approach resonates with recent arguments for moving beyond changes in learner writing when assessing genre awareness and developing genre knowledge (see Negretti & McGrath, 2020). Therefore, it represents an important direction for future research on the impacts of DDL on L2 learner genre awareness.

Conclusion

There is robust evidence to support the use of DDL for language teaching in general (see Boulton & Vyatkina, 2021), and many studies provide support for DDL as an effective means of raising linguistic awareness (e.g. Cotos et al. 2017; Yoon, 2008). In this encyclopaedia entry, studies which examine how DDL influences L2 genre awareness were considered. Overall, there are particular benefits in this area regarding learner perceptions of pedagogical experiences and changes in learner writing and decision-making, although research explicitly targeting genre knowledge and awareness through DDL is still somewhat limited. The research that does exist often integrates large-scale corpus annotation projects or makes use of existing annotated corpora, and the increasing availability of automated analytical tools and annotated corpora may facilitate future research in this area. For example, if rhetorically annotated corpora are made publicly available, DDL pedagogies for enhancing genre awareness can be more broadly implemented. Additionally, as researchers of genre knowledge continue to develop new ways of examining development beyond changes in texts, research into how DDL may bring about such development can be expected to expand as well.

Cross-References

- ▶ DIY Corpora and Data-driven Learning
- ▶ Learner Autonomy and Data-driven Learning
- ▶ Metacognition and Data-driven Learning
- ▶ Qualitative Designs in Data-driven Learning Research
- ▶ Quantitative Designs in Data-driven Learning Research

References

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Casal, J. E. (2020). *An integrated corpus and genre analysis approach to writing research and pedagogy: Development of graduate student genre knowledge* [Unpublished doctoral dissertation, The Pennsylvania State University].

Casal, J. E., & Kessler, M. (2024). Rhetorical move-step analysis. In M. Kessler & C. Polio (Eds.), *Conducting genre-based research in applied linguistics: A methodological guide* (pp. 82–104). Routledge.

Casal, J. E., & Yoon, J. (2023). Frame-based formulaic features in L2 writing pedagogy: Variants, functions, and student writer perceptions in academic writing. *English for Specific Purposes, 71*, 102–114. https://doi.org/10.1016/j.esp.2023.03.004

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes, 6*(4), 289–302. https://doi.org/10.1016/j.jeap.2007.09.009

Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes, 50*, 97–112. https://doi.org/10.1016/j.esp.2017.11.004

Cortes, V. (2024). Lexical bundles and phrase frames. In M. Kessler & C. Polio (Eds.), *Conducting genre-based research in applied linguistics: A methodological guide* (pp. 105–126). Routledge.

Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology, 21*(3), 104–130. http://hdl.handle.net/10125/44623

Crosthwaite, P. (Ed.). (2020a). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge.

Crosthwaite, P. (2020b). Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of*

Applied Linguistics, 43(2), 169–195. https://doi.org/10.1075/aral.00031.cro

Dong, J., & Lu, X. (2020). Promoting discipline-specific genre competence with corpus-based genre analysis activities. *English for Specific Purposes, 58*, 138–154. https://doi.org/10.1016/j.esp.2020.01.005

Farr, F., & Leńko-Szymańska, A. (2024). Corpora in English language teacher education: Research, integration, and resources. *TESOL Quarterly, 58*. https://doi.org/10.1002/tesq.3281

Hirschi, K., & Kang, O. (2024). Data-driven learning for pronunciation: Perception and production of lexical stress and prominence in academic English. *TESOL Quarterly, 58*. https://doi.org/10.1002/tesq.3302

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.

Hyland, K. (2023). Genre-based instruction and corpora. *TESOL Quarterly, 58*. https://doi.org/10.1002/tesq.3273

Johns, T. (1991). Should you be persuaded – Two samples of data-driven learning materials. *ELR Journal, 4*, 1–16.

Kessler, M. (2021). The longitudinal development of second language writers' metacognitive genre awareness. *Journal of Second Language Writing, 53*, 100832. https://doi.org/10.1016/j.jslw.2021.100832

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*, 56–75. https://doi.org/10.1016/j.esp.2005.02.010

Ma, Q., Chiu, M. M., Lin, S., & Mendoza, N. B. (2023). Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL, 35*(1), 19–39. https://doi.org/10.1017/S0958344022000180

Negretti, R., & McGrath, L. (2020). English for specific playfulness? How doctoral students in Science, Technology, Engineering and Mathematics manipulate genre. *English for Specific Purposes*, *60*, 26–39. https://doi.org/10.1016/k.esp.2020.04.004.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Tardy, C. M., Sommer-Farias, B., & Gevers, J. (2020). Teaching and researching genre knowledge: Toward an enhanced theoretical framework. *Written Communication, 37*(3), 287–321. https://doi.org/10.1177/0741088320916554

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12*(2), 41–48. http://llt.msu.edu/vol12num2/yoon/

Zare, J., Karimpour, S., & Delavar, K. A. (2023). Classroom concordancing and English academic lecture comprehension: An implication of data-driven learning. *Computer Assisted Language Learning, 36*(5), 885–905. https://doi.org/10.1080/09588221.2021.1953081

M

Metacognition and Data-Driven Learning

Amelie Xiaohan Sun and Atsushi Mizumoto
Kansai University, Osaka, Japan

Abstract

This entry explores the intersection of metacognition and data-driven learning (DDL) in language education. It illustrates how the integration of metacognitive strategies within DDL frameworks significantly enhances the efficacy of learning processes by enabling learners to better understand, regulate and direct their own learning experiences. The core argument of the chapter is that metacognitive awareness—comprising both knowledge of one's cognitive processes and the ability to regulate them—empowers learners to utilize DDL more effectively. This integration fosters a personalized, reflective and ultimately more effective learning environment. This chapter proposes the benefits of this synergistic approach include improved autonomy, critical thinking and deeper understanding of language use. Additionally, practical strategies for teachers to implement this integrated approach in classroom settings are discussed, highlighting the role of technology and collaborative learning in enhancing metacognition and DDL practices.

Keywords

Metacognition · Data-driven learning (DDL) · Resource use · Individual differences · Learner factor

Introduction

Data-driven learning (DDL) is a learner-centred learning approach. It optimizes the language learning process by using the massive amounts of data collected in real-world language environments (Poole, 2020). Meanwhile, through the examination of patterns and trends within the data, teachers can adapt instructional strategies to align with the specific needs of individual learners, thereby enhancing the efficacy of teaching methodologies. This approach not only personalizes learning experiences but also provides actionable insights that facilitate continuous improvement.

In conjunction with DDL, the concept of metacognition, which can be described as "knowledge about knowledge," "thinking about thinking" or "reflections about actions" (Flavell, 1971), plays a crucial supportive role. Metacognition refers to the awareness and regulation of one's cognitive processes, encompassing the ability to plan, monitor and evaluate one's understanding and performance. By empowering learners to take charge of their own educational directions, this self-regulation system encourages a more thoughtful approach to learning.

When combining metacognition with DDL, learners may better comprehend and apply data-

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_85-1

driven feedback, which allows them to modify their strategy for learning. This integration fosters learners' ability to understand and retain knowledge, while also improving a culture of self-improvement and lifelong learning. Besides, teachers can create more dynamic and responsive learning environments that meet different needs of their students.

Metacognition

According to Flavell (1971), metacognition is a critical cognitive process that involves the awareness and regulation of one's own thought processes. It helps individuals comprehend and regulate their cognitive processes, which is key to gaining knowledge and solving problems. Typically, metacognition is divided into two main components (Sato, 2022): metacognitive knowledge and metacognitive regulation.

Metacognitive knowledge refers to what individuals know about their own cognitive processes and learning strategies. It includes knowledge of self, knowledge of general strategies that might be used for different tasks and knowledge of the conditions under which these strategies might be used (Pintrich, 2002). Accordingly, metacognitive knowledge can be grouped into three domains: person, task and strategy. "Person" symbolizes knowledge about oneself as a learner. It includes awareness of one's own cognitive abilities, recognizing one's strengths and weaknesses in different areas and understanding personal learning preferences. "Task" refers to knowledge about the nature of the tasks being undertaken. It involves understanding the specific requirements of different tasks and recognizing what strategies are best suited to accomplish them. "Strategy" includes knowledge about various cognitive and metacognitive strategies and when to use them. It involves being aware of different techniques for learning, remembering and problem-solving and understanding how to apply these strategies effectively. For example, learners might know that they are good at memorizing facts (which they may perceive as a strength) but struggle with applying concepts in practical situations. This realization could lead them to change their approach to the task, such as adjusting their study methods to focus more on practical application exercises and less on rote memorization. By doing so, they can develop a more balanced understanding of the material, which not only leverages their strengths but also addresses their weaknesses. Finally, learners can activate the relevant situational or conditional knowledge for solving a problem in a certain context, such as recognizing that multiple-choice tests require recognition of the correct answers, while essay tests demand actual recall of information.

In contrast to metacognitive knowledge, which serves as a foundational knowledge base, metacognitive regulation involves the processes used to control and oversee one's cognitive activities during learning and problem-solving (Marantika, 2021). It includes planning, monitoring and evaluating. "Planning" involves setting goals, selecting appropriate strategies and allocating resources before beginning a learning task. Effective planning can help learners approach tasks in an organized manner, increasing the likelihood of success. During the learning process, "monitoring" involves tracking one's own understanding and performance. This can include self-questioning to ensure comprehension, checking progress towards goals and recognizing when adjustments are needed. After completing a task, "evaluating" involves assessing the outcomes and the effectiveness of the strategies used. This can help learners understand what worked well and what needs improvement, thereby informing future learning activities. For example, learners might evaluate their language acquisition by reviewing their written assignments to see if they have correctly applied the language patterns studied. They might also engage in peer review sessions to gain feedback on their usage of new vocabulary and grammar structures. By reflecting on these evaluations, learners can identify specific strategies that were successful and those that require modification, leading to a more targeted and effective learning approach in subsequent activities.

By developing strong metacognitive skills, learners can enhance their ability to manage their own learning, leading to improved understanding, retention and application of knowledge. This self-

regulatory mechanism is crucial for effective learning and problem-solving across various educational contexts.

Integrating Metacognition and DDL

Data-driven learning involves using large collections of authentic language data, known as corpora, to inform and guide language learning. By analysing these extensive databases of real-world text and spoken language, learners can observe how language is used in various contexts and identify common collocations, grammatical structures and discourse patterns. This method encourages learners to become "language detectives" (Johns, 1997, p. 101), exploring language use empirically and independently.

At its core, DDL utilizes authentic language data to facilitate language learning and proficiency. DDL empowers learners to explore real-world language usage, uncovering patterns and gaining insights that are often not visible in traditional language learning materials (Lusta et al., 2023). On the other hand, metacognition focuses on self-regulation and reflection, allowing learners to monitor and control their cognitive processes. Metacognition plays a pivotal role in mediating language learning when employing DDL tools in educational activities (Sato, 2024).

Integrating metacognitive strategies into DDL may enhance learners' ability to comprehend, monitor and evaluate their linguistic development. The process can be loosely conceptualized through three interrelated stages: planning, monitoring and evaluating. These stages, informed by principles of metacognitive theory, could help guide learners to engage more effectively with DDL tools and resources.

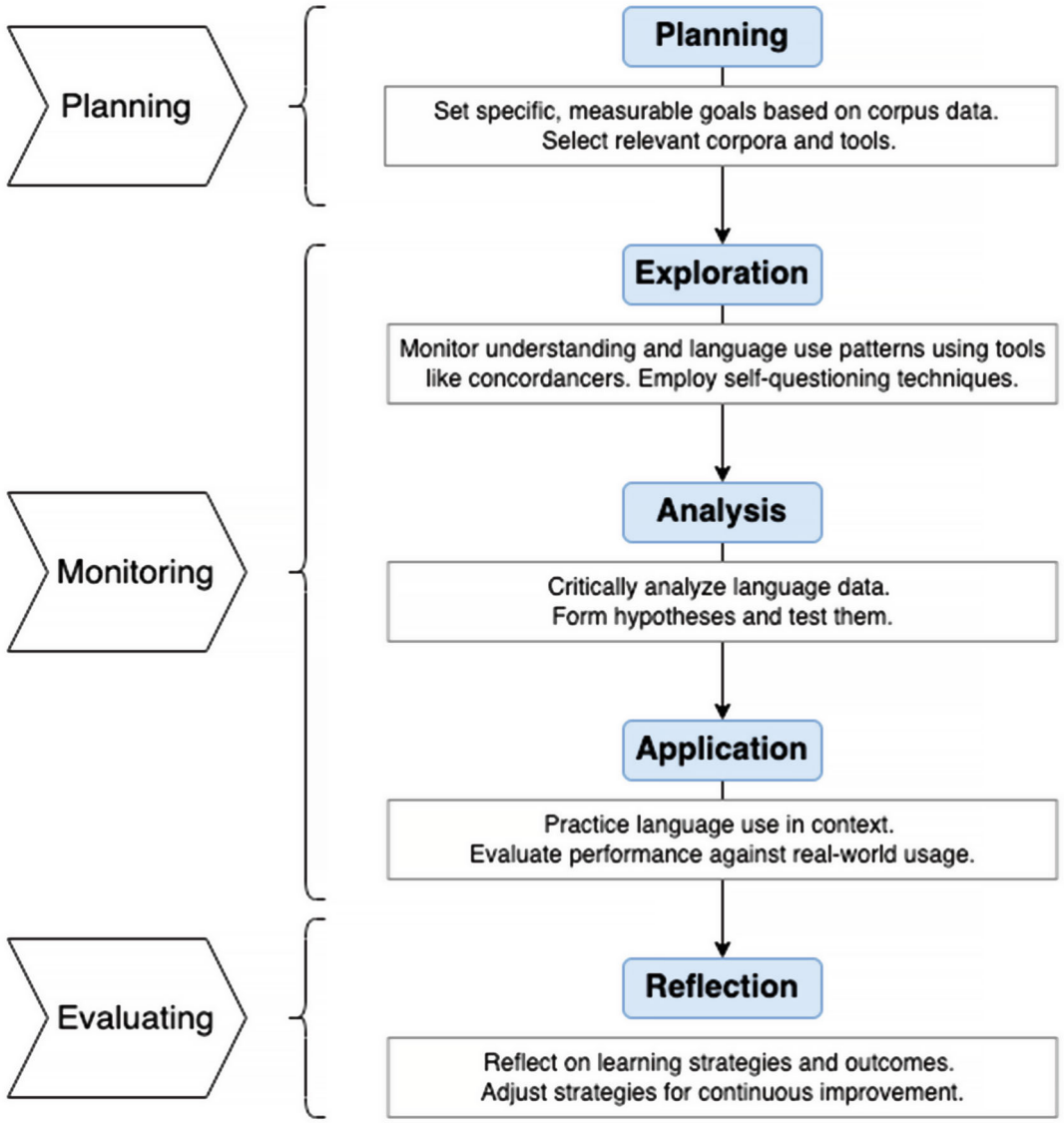
In the planning stage, learners can set specific, measurable goals based on their analysis of corpus data. For example, they might aim to master the use of particular collocations or grammatical structures frequently encountered in authentic texts. Additionally, learners can choose the most relevant corpora and tools that align with their learning objectives. This might involve selecting specialized corpora that focus on particular genres or registers of language, ensuring that the resources they use are tailored to their specific learning needs.

During their exploration of corpus data which is the monitoring stage, learners can regularly check their understanding and use of language patterns. Tools like concordancers can help them see how their hypotheses about language use hold up against actual data. Additionally, learners can employ self-questioning techniques to ensure they understand the data they are analysing. Questions such as "How is this word typically used?" or "What patterns do I notice in this structure?" can guide their investigation and help them make sense of the language data they encounter.

After working with corpus data, learners can evaluate their understanding and application of new language patterns during the evaluating stage. This might involve practical exercises where they use the language in context, followed by a review of their performance. Additionally, learners can reflect on the effectiveness of their strategies and make adjustments as needed. This reflection helps them understand what works best for them and how they can improve their use of DDL tools, fostering continuous improvement in their learning process (Fig. 1).

Benefits of Integrating Metacognition and DDL for Language Learners

Integrating metacognitive strategies with DDL might offer numerous educational benefits, enhancing the overall effectiveness of language acquisition through more sophisticated and autonomous learning processes (Sato, 2024). Learners develop the capability to self-direct their learning process by making informed decisions about what linguistic elements to focus on and how to approach their studies. Consequently, this autonomous learning empowers students to take control of their learning paths, fostering a sense of ownership and responsibility. Furthermore, the investigative nature of DDL, when combined with metacognitive reflection, significantly enhances critical thinking skills. Learners engage in the critical analysis of language data, formulating hypotheses about language use and rigorously



Metacognition and Data-Driven Learning, Fig. 1 Framework of applying metacognition to DDL

testing these hypotheses against authentic language data. This process promotes deeper analytical skills, enabling learners to question, evaluate and refine their understanding of language patterns and usage. By exploring language use in authentic contexts, learners achieve a more detailed and in-depth comprehension of linguistic structures and functions. Thus, this approach transcends rote memorization, encouraging learners to understand the "why" and "how" behind language use. Additionally, metacognitive strategies further deepen this understanding by prompting learners to continually evaluate and refine their knowledge. The integration of metacognitive strategies allows learners to tailor their DDL experiences to their specific needs and goals. This personalization ensures that learning activities are relevant and engaging, enhancing motivation and sustained interest. By focusing on areas of personal relevance and importance, learners are more likely to remain engaged and achieve higher levels of proficiency. Ultimately, the skills cultivated through the integration of metacognition and DDL extend beyond language learning,

fostering a lifelong learning mindset. Learners develop the tools and strategies necessary for continuous self-improvement and independent learning throughout their lives. Therefore, this lifelong learning approach equips learners with the flexibility and adaptability to handle various learning contexts and challenges.

Practical Implementation for Teachers

To effectively integrate metacognitive strategies with DDL, teachers could employ several approaches (Mizumoto, 2023). Initially, teachers should offer comprehensive training on the use of corpus tools and data interpretation. This training should include instruction on metacognitive strategies such as planning, monitoring and evaluating learning activities. By equipping students with these foundational skills, teachers empower them to effectively utilize DDL resources. Secondly, creating learning activities that incorporate both DDL and metacognitive strategies is crucial. For instance, tasks that involve analysing corpus data to identify linguistic patterns can be followed by reflective exercises that prompt students to consider their learning strategies and outcomes. These activities should encourage active engagement with authentic language data and promote critical reflection on learning processes. Furthermore, regularly prompting students to reflect on their learning experiences and strategies is necessary for the development of metacognitive skills. This can be facilitated through various methods such as reflective journals, group discussions and self-assessment questionnaires. Reflection helps students to internalize their learning experiences, understand their cognitive processes and make informed adjustments to their learning strategies. Additionally, incorporating technology can enhance the integration of metacognition and DDL. Tools such as language learning apps, online corpora and concordancers can provide learners with immediate feedback and data-driven insights. Teachers should guide students in the effective use of these technologies, ensuring they are utilized to support metacognitive and data-driven learning strategies. Lastly, providing continuous feedback and support is important. Teachers should offer constructive feedback on both the content of learners' language use and their metacognitive strategies. This feedback should be specific, actionable and aimed at helping learners refine their approach to DDL and metacognitive regulation.

Conclusion

The integration of metacognitive strategies with DDL creates a comprehensive framework for effective and autonomous language learning in modern educational practices. By utilizing authentic language resources and applying self-regulatory strategies, teachers and learners can achieve a more dynamic, personalized and effective learning environment. This approach not only enhances immediate learning outcomes but also equips learners with the skills necessary for lifelong, independent learning.

While the integration of metacognition and DDL holds great promise, to fully realize the potential of this integration, challenges such as data literacy and the effective training of teachers must be addressed. Teachers need robust professional development programmes that enhance their ability to use data-driven strategies and apply metacognitive principles in their teaching. By addressing these challenges and making use of the combined power of metacognition and DDL, educational practices can be significantly enhanced, leading to improved learning outcomes and fostering a more adaptive, personalized and effective educational experience for all learners.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Boulton, A. (2017). Data-driven learning and language pedagogy. In S. L. Thorne & M. Stephen (Eds.), *Language education and technology* (pp. 181–192). Springer.

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? *Human Development, 14*(4), 272–278.

Johns, T. (1997). Contexts: The background, development, and trialing of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100–115). Longman.

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*(12), e22731. https://doi.org/10.1016/j.heliyon.2023.e22731

Marantika, J. E. R. (2021). Metacognitive ability and autonomous learning strategy in improving learning outcomes. *Journal of Education and Learning (EduLearn), 15*(1), 88–96. https://doi.org/10.11591/edulearn.v15i1.17392

Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics, 3*(3), 100074. https://doi.org/10.1016/j.acorp.2023.100074

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice, 41*(4), 219–225. https://doi.org/10.1207/s15430421tip4104_3

Poole, R. (2020). "Corpus can be tricky": Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning, 35*(7), 1620–1641. https://doi.org/10.1080/09588221.2020.1825095

Sato, M. (2022). Metacognition. In I. S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 95–110). Routledge. https://doi.org/10.4324/9781003270546-8

Sato, M. (2024). Metacognition and data-driven learning. *TESOL Quarterly, 58*(3), 1246–1255. https://doi.org/10.1002/tesq.3271

T

Teaching Sociolinguistics with Data-Driven Learning

Dana Gablasova and Emil Tangham Hazelhurst
Lancaster University, Lancaster, UK

Abstract

The field of sociolinguistics focuses on different aspects of the relationship between language and society. Language use typically varies systematically according to social characteristics of the users, social context and aims of the communication. Languages offer a variety of ways for conveying a single message, allowing speakers to communicate a range of social meanings through their choice of linguistic means. Understanding the role of social factors in language use is therefore a key aspect in successful communication for both first language (L1) and second language (L2) speakers. The importance of sociolinguistic topics in language learning has been long recognised, and a variety of methods have been used to integrate social aspects of language communication into language classrooms. Corpus-based approaches to language learning, such as data-driven leaning (DDL), represent a very effective method of introducing learners to sociolinguistic concepts. In particular, they help learners develop their awareness of language variation by drawing on empirical data representing real-life language use by different social actors and in different social contexts. Such use of DDL serves to broaden learners' understanding of the relationship between language and society, providing opportunities to explore the values and belief systems of a given language community.

Keywords

Sociolinguistics · Data-driven learning · Social variation · Social characteristics · Language change · L1 production · L2 production · L2 pedagogy · Language teaching

Introduction

Languages offer a variety of ways for conveying a single message, allowing speakers to communicate a range of social meanings through their choice of linguistic means. Language use typically varies systematically according to social characteristics of the users, social context and aims of the communication (Holmes & Wilson, 2022). Understanding the role of social factors in language use is therefore a key aspect in successful communication for both first language (L1) and second language (L2) speakers of the target language. This understanding plays a role in language production as well as in the interpretation of meaning conveyed by interlocutors. Sociolinguistic aspects of language communication therefore deserve particular attention in language learning and teaching.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_86-1

The importance of sociolinguistic topics in language learning has been long recognised, and a variety of methods have been used to integrate social aspects of language communication into language classrooms (e.g., Bucholtz et al. (2014), Farrell (2017), McKey and Hornberger (1995)). Corpus-based approaches to language learning, such as data-driven leaning (DDL), represent a highly effective method of introducing learners to sociolinguistic concepts. In particular, they help learners develop their awareness of language variation by drawing on empirical data representing real-life language use by different social actors and in different social contexts. Such use of DDL serves to broaden learners' understanding of the relationship between language and society, providing opportunities to explore the values and belief systems of a given language community.

Integrating Sociolinguistics into Language Teaching

Key Topics in Sociolinguistics

Sociolinguistics as a field is concerned with understanding how language communication varies according to different social and context-related factors (Holmes & Wilson, 2022; Baker, 2010). First, it investigates the variation in language use according to the *social characteristics of language users*, such as their age, gender, socioeconomic status, level of education and their occupational background. Social groups defined by one or more of these characteristics tend to share certain linguistic preferences, related to the shared aspects of their identity and communicative patterns. For example, language use by speakers of a certain age or gender typically shows some shared attributes, which distinguish these speakers from other social groups. The second major topic in sociolinguistics concerns the *social context* in which the communication takes place and how it affects language use in different situations. For instance, informal conversation between friends differs from that of professional workplace interaction, and this difference is reflected in the linguistic choices of the speakers. Other major areas of interest in the field of sociolinguistics include the investigation of the ways in which speakers convey, negotiate and interpret meaning in *social interaction* and how *language use changes over a period of time* (i.e., diachronic variation). Diachronic variation can refer to changes in language more broadly—in the whole of the language or in the language use of major social groups—or describe changes in language use of individual speakers. Finally, sociolinguistics also involves the study of *attitudes towards language use* and why some varieties of language are considered more prestigious, 'better' or 'more correct' than others.

There are two major sociolinguistic constructs that underly variation in language use according to the factors described above (Holmes & Wilson, 2022): (i) Language can serve a number of different functions and it can express information and social meaning simultaneously. For example, in one utterance, we are able to convey information but, at the same time, we can also express emotions (e.g., respect, annoyance). (ii) This is possible because at each level of language (e.g., vocabulary, grammar, discourse, pronunciation and gestures), there are a large number of linguistic forms and structures that can express the same information but are encoded with different social meanings. As a result, language offers speakers a choice of how to express meaning, using different linguistic styles in different social contexts. For example, speakers have a choice of how to address others, express criticism, accept offers or give information. These choices reflect speakers' awareness of and response to social factors such as interpersonal relationships, speaker roles, aims of communications and the context in which the interaction is taking place.

Importance of Sociolinguistics in Language Pedagogy

The importance of incorporating sociolinguistic topics into language teaching has been repeatedly acknowledged in pedagogical literature (Farrell, 2017; Folse & Vitanova, 2006; Geeslin & Long, 2014; McKay & Hornberger, 2005). The relevance of sociolinguistics has been highlighted for different educational contexts, including

second language learning, learning about speakers' first or heritage languages (e.g., in secondary schools or heritage classes), as well as in university-level linguistic courses.

When discussing the benefits of incorporating sociolinguistics into language learning, different types of linguistic knowledge have been highlighted: first, at the most concrete level, learners may become aware of specific linguistic patterns (at the level of lexicon, grammar, sound and pragmatics) typical of different social groups or contexts of use. Second, at a more general level, this approach will allow learners to understand that language varies systematically according to social variables, such as social groups, contexts, interlocutors and aims of the interaction. This will help them recognise that variation and linguistic choice are inherent characteristics of language use, highlighting 'the role of speaker agency in choosing among different linguistic forms and styles' (Leeman & Serafini, 2016, p. 68). Further, the understanding that language varies naturally also plays an important role in recognising that usually there is not one 'correct' form but rather that 'good language users are able to respond appropriately to a broad range of speech contexts' (Geeslin & Long, 2014, p. 259). As Geeslin and Long (2014) further argue, the ability to vary one's linguistic choices to achieve different aims is part of communicative competence and 'a complex skill that takes time to develop' (p. 259); therefore, learners should be given support to develop these skills as part of their language learning (Gutiérrez & Fairclough, 2006). This teaching approach thus differs significantly from teaching language in isolation, focusing on developing vocabulary, grammatical knowledge and skills such as writing and speaking without situating them in terms of social characteristic of the interaction (Farrell, 2017).

The third reason for bringing sociolinguistics into language classrooms relates to raising awareness about the types/purpose of language variation and developing learners' ability to critically evaluate attitudes towards language use. Demonstrating that language variation is a natural and common phenomenon in language communication is 'useful for promoting positive attitudes towards linguistic diversity and multiple varieties' (Leeman, 2018, p. 350) and helps to address prescriptivist and negative attitudes to specific types of language (e.g., as used by some social groups). As Shin and Hudgens Henderson (2017, p. 205) point out, 'whereas most people agree that it is not acceptable to judge someone based on the color of their skin, judging other peoples' language patterns is widespread and rarely questioned'. This is especially important given the major role played by language 'in the reproduction of a wide range of social constructs, including but not limited to gender, race, nation and culture' (Leeman, 2018, p. 355). Raising awareness of social variation can thus contribute not only to better linguistic knowledge and more effective communicative skills but also, more broadly, to the issues of social justice (Bucholtz et al., 2014).

Teaching about Sociolinguistic Topics with Corpora

Affordances of Corpora and Corpus Tools for Teaching Sociolinguistics

To achieve the aims discussed in section "Importance of Sociolinguistics in Language Pedagogy," teachers have been encouraged to present learners with language use that reflects a variety of speakers and language situations and to 'diversify the input to which students have access' (Geeslin & Long, 2014, p. 265). Corpus-based language teaching is particularly suitable for teaching about key sociolinguistic topics based on real-life data representing a variety of social actors and settings (Leeman & Serafini, 2016). Corpora are an important source of information about the effect of social variables on communication as they provide empirical evidence about how language is used by different groups of users and in different contexts, to achieve different aims and about how language varies over time (Andersen, 2010; Baker, 2010). Corpus data and methods can be integrated into language teaching of sociolinguistics following two main approaches: First, corpus-based research findings can be used in developing teaching resources, including textbooks and exercises, that include examples from corpora.

Second, corpus data can be used directly by learners to explore social variation in language use.

Two types of corpora are especially effective in DDL approaches to sociolinguistic topics in language classrooms: First, these are general language corpora, typically large in size, which include balanced data on language use by speakers from different social groups, contexts and communicative purposes. An example of such corpora is the British National Corpus 2014 (BNC2014; Brezina et al., 2021; Love et al., 2017) which contains 100M words of spoken and written data. In particular, the spoken component contains transcripts of 1,251 recordings that involve 668 speakers in informal conversations (Love et al., 2017), with demographic information (gender, age and social class) available for the speakers. The language in the corpus can be thus investigated according to social variables at different levels of language use such as lexicon, grammar or pragmatics (e.g., Brezina et al. (2018)). Such datasets can provide teachers and learners with rich language evidence to explore according to their needs. What is especially valuable is that general corpora are available for a number of languages other than English. One such example is the Croatian Adult Spoken Language Corpus (Kraljević & Hržica, 2016), which contains 250,000 words from 617 speakers collected in the period of 2010–2016, along with information on speakers' age, gender and the location of the conversation. The second group of corpora with particular relevance for corpus-based teaching of sociolinguistic topics are corpora that represent language from a specific demographic group of language users and/or a specific domain of use. An example of such corpus is the Bergen Corpus of London Teenage Language (COLT; Stenström et al., 2002) which represents language production from 13- to 17-year-old speakers of English and which can thus be explored for linguistic features characteristics of teenage communication. Another example can be seen in a corpus of academic speech such as the Michigan Corpus of Academic Spoken English (MICASE; Simpson et al., 2002) which, in addition to representing speech from a particular language domain, also provides information on the age, gender and academic role of the speakers.

In order to let students use corpus data to learn about sociolinguistic topics, the corpora can be accessed in different ways: first, they can be searched according to social information via existing general corpus tools and interfaces. For example, the BNC2014 can be explored using free access through #LancsBox X (Brezina & Platt, 2024) or CQPweb (Hardie, 2012; CQPweb currently includes only the spoken component of the BNC2014). Both of these corpus tools can be used in the classroom, with teachers and students using different functionalities to search the language of different social groups. These tools also provide access to the original British National Corpus (BNC; Burnard, 2007), which represents British English from the 1990s, allowing students to compare language change over time in British English. Using existing general corpus tools in classrooms may require more time for training the students in the use of the tools and is thus typically more suitable for longer-term pedagogical purposes (e.g., a series of lessons that would use the corpus tool).

The second option for bringing corpus data into language classes with a sociolinguistic focus is through dedicated tools and resources. Some available resources are corpus-based and include access to a limited amount of data but do not let students explore the corpus evidence more fully. This can be illustrated by the teaching resources based on the corpora representing London English (Kerswill et al., 2004–2007, 2007–2010) and the materials based on them (Cheshire & Fox, 2016). These materials, designed primarily for secondary-school students in the UK, focus on a number of sociolinguistic topics (e.g. language change, language and gender, language and media). While several of the teaching resources are based on corpus data and students are also encouraged to gather further language samples, these materials only include specific examples (e.g., a transcript) from the larger collection and do not employ corpus methods in the language learning process.

Other tools allow students to explore corpus data directly. An example of such resources is BNClab (Brezina et al., 2018) which offers a user-friendly pedagogical platform for accessing demographically balanced five-million-word samples from the spoken components of both the BNC and the BNC2014. The platform allows searching language use according to social variation (e.g., gender, age, social class and regional affiliation of the speakers). Due to the inclusion of both the BNC and the BNC2014, it also offers the option of directly comparing British English in 1990s and 2010s and study language change in this variety over time. BNClab also provides an accompanying set of teaching resources that follow principles of data-driven learning and explore different types of social factors in language use. Each of these worksheets is guided by a question which is then explored through a series of exercises that involve students working with the data in the corpus and identifying and discussing patterns of social variation in language use. The worksheets cover the key sociolinguistic topics described in section "Key Topics in Sociolinguistics," such as language variation according to social groups (e.g., 'Does age affect how people talk about their emotions?'), attitudes towards language varieties (e.g., 'Why are people judged by the way they speak?'), changes in language over time (e.g., 'Why does English keep on changing?') and how language fulfils different linguistic and social functions simultaneously (e.g., 'Many faces of swearing: Are swearwords always bad?'). Due to the rich social metadata available, BNClab can be used flexibly by learners and teachers to explore a variety of sociolinguistic topics.

Empirical Studies on Using DDL to Teach Sociolinguistics

DDL represents a particularly effective way of engaging students with real-life data and teaching about sociolinguistic variation in language use. However, while there are corpus data and resources available, as discussed above, there have been relatively few studies so far that report on the use of corpora for studying sociolinguistic topics (although the actual practice may be much more widespread). For example, Miguel (2022) investigated the effectiveness of DDL for teaching about language variation to university students learning Spanish. In the study, a particular focus was placed on the geographical variation in the use of grammatical structures, and the Corpus del Español (Davies, 2002) was employed as the main resource. Findings from the study showed that the use of corpus data contributed to raising students' awareness of the role of linguistic variation in language. However, the study also revealed that some students found using corpus tools difficult, raising issues that have often been encountered across DDL classes (e.g., the unfamiliarity with the corpus tools and the frustration with learning through discovery learning, see Geluso and Yamaguchi (2014)). The study thus argues that in DDL approaches to teaching sociolinguistics, it is crucial to focus on both raising awareness of language variation as well as helping students understand the benefits of corpus-based learning (Miguel, 2022).

In another study that focused on a university-level Spanish course, Shin and Hudgens Henderson (2017) investigated the effectiveness of using corpus data and methods for teaching grammatical structures that are subject to social variation in spoken language, challenging the prescriptivist view of language and raising questions of linguistic discrimination. The study found that the course and the use of corpus data had a beneficial effect on both learners' grammatical knowledge and sociolinguistic awareness, resulting in an increased 'understanding of and appreciation for linguistic variation and diversity' (Shin & Hudgens Henderson, 2017, p. 205). In addition, learners reported an improvement in self-esteem when speaking their own dialect after engaging with more language varieties in the classroom.

Practical Recommendations and Open Access Materials for DDL Approaches to Teaching Sociolinguistics

The potential of corpus data for teaching about sociolinguistic aspects of language use is considerable, with corpus-based language teaching providing an opportunity to work with real-word data and to analyse the key sociolinguistic topics described in section "Key Topics in Sociolinguistics." In the current section, we will offer practical

recommendations for integrating corpus evidence into language classrooms using examples from the Corpus for Schools project (Gablasova, 2024; https://www.lancaster.ac.uk/corpusforschools//) which focuses on creating DDL materials for L1 and L2 users of English in secondary and university-level educational contexts. The resources produced as part of the project contribute to the growing body of open-access DDL materials; in particular, they address the sociolinguistic and sociocultural aspects of language communication—a topic not yet commonly covered in other available DDL materials. When integrating sociolinguistic topics into language classes, two major approaches to using corpora and data-driven learning are available. First, we can use *information about frequency* of occurrence of the target features (e.g., grammatical structures or lexical items) to explore the social factors in language use, such as belonging to a particular social group or community. Such use is illustrated by an exercise shown in Fig. 1 in which the learners are guided to use the data from the BNC2014 (accessed via BNClab) to investigate the use of swearwords according to three social variables—gender, age and social class of the speakers.

Depending on the previous sociolinguistic knowledge and language experience of language learners, the exercises can offer different degrees of scaffolding, which is an effective way of controlling the complexity of DDL tasks. Thus, while learners can be given the search terms and the hypothesis to test directly (as illustrated by the task in Fig. 1), more advanced learners can be asked to formulate the hypothesis about social variation themselves in order to evaluate beliefs about language with empirical data. Such use of corpora is demonstrated in the task shown in Fig. 2. The exercise is taken from a worksheet which explores the effect of speakers' age on their linguistic choices when expressing emotions. In this exercise, participants are asked to first formulate a hypothesis and then check it with the aid of the BNC2014 data.

In addition to using corpora to explore the frequency of linguistic features that vary according to social variables, the information about the *language use in context* can be very valuable when helping students learn more about sociolinguistic and pragmatic aspects of communication. This approach is especially useful when demonstrating two major sociolinguistic topics discussed in section "Key Topics in Sociolinguistics": (i) the fact that language can realise different functions (and it can do so simultaneously) and (ii) that language offers a number of ways to deliver the same message, with different social

Task 2

Who uses swearwords? It is sometimes believed that swearwords are used predominately by certain groups of people, for example, by men rather than women, and by people from the working class rather than students and people from the middle class. In this activity, we are going to check these hypotheses. Look up the following swearwords in BNClab and note down which groups in terms of age, gender and social class use them most.

	GENDER	AGE	SOCIAL CLASS
bullshit	*e.g. Women slightly more than men*	*Mostly younger people*	*Mostly students & middle class*
bloody			
dickhead			
fucking			
shit			

Teaching Sociolinguistics with Data-Driven Learning, Fig. 1 A task on the use of swearwords by different social groups (Corpus for Schools, 2024)

Task 3

Interjections and exclamations serve to convey a broad range of emotions, such as joy, surprise, pain and excitement. Look at the set of exclamations below – do you think some of them are more typical for younger or for older people? First note down your hypothesis and then check your intuition in BNClab by typing the expressions in and using the Age button. You can also try searching for exclamations of your own choice. Before checking the corpus, briefly discuss the factors that you have considered when forming your hypothesis.

Bloody hell	*Hypothesis:*	*Findings:*
Oh dear		
Oh man		
Oh my god		
Oh shit		

Looking at the patterns, can you think of the reasons why some of the expressions may be more or less common for speakers from different age groups?

Teaching Sociolinguistics with Data-Driven Learning, Fig. 2 A task on the linguistic choices according to speakers' age (Corpus for Schools, 2024)

meaning. The effectiveness of looking at linguistic patterns and language use in context is illustrated in Fig. 3, which shows an exercise on the expression of disagreement and politeness strategies. In this task, learners are asked to work with the concordance lines from the Trinity Lancaster Corpus of L2 spoken English (Gablasova et al., 2019) to explore the notion of politeness and different ways in which it can be conveyed in spoken English.

In another example of the usefulness of contextual information, the task in Fig. 4 demonstrates how learners can use corpus data to study semantic change in the meaning of lexical items (in this case, *sexist, sexism, racist* and *racism*) over the period of 20 years.

In general, DDL activities that focus on sociolinguistic topics can start with students using the corpus data in order to answer a set of specific questions and to learn about language variation of specific linguistic forms. However, as explained in section "Importance of Sociolinguistics in Language Pedagogy," as a next step, this foundational knowledge can then be expanded to raise students' awareness of broader sociolinguistic topics such as why languages vary, why there are different choices available in language systems and where negative attitudes to some forms of language use stem from. Students can be further guided to consider possible reasons for the presence of these sociolinguistic phenomena and to discuss variation as a central phenomenon in human communication.

Conclusion

Teaching about sociolinguistic topics in language classrooms is important for developing skills to facilitate successful communication as well as for exploring broader questions of social justice. Corpus data and data-driven learning represent valuable resources in this pedagogical area, with great potential to achieve the aims identified by educators to date. While there are currently only a limited number of studies reporting on the use of corpus-based approaches to teaching about sociolinguistic topics, the practice is most likely more widespread with a number of corpora (e.g., the BNC2014) and corpus tools (e.g., BNClab) available to support sociolinguistic analysis of language data in the classrooms.

Task 3

Look at the following examples of disagreement from conversations of English learners in the Trinity Lancaster Corpus. Although they all express disagreement with the previous speaker, they differ in the strategies used to express the opposing view.

Do you think that some of the examples can be considered more polite than others? Rank the examples using the following scale: (1) not polite – (2) polite – (3) very polite

	Example	Your score
A	I completely disagree with this because er I I repeat as I said …	
B	I agree with this point but don't you think maybe the fact that times are changing is a good thing?	
C	but I personally would disagree that that money would necessarily be spent on that	
D	erm no no it's not so	
E	well I 'm not totally convinced but er you know I live in a really traditional family	
F	mm I can understand your opinion erm but I was still wondering…	
G	I can't agree with you	
H	er er I I think erm I I think they I I think they are wrong	
I	I think they're completely wrong	
J	no way	
K	I I I can understand what you 're saying but I'm not I don't agree with that	

When you are done, compare your rating with a partner and discuss the following questions:

- Did you have similar ratings?
- What expressions made the disagreement more or less polite?
- What strategies of polite communication have been demonstrated in the sentences ranked as polite and very polite?

Teaching Sociolinguistics with Data-Driven Learning, Fig. 3 A task on different ways of conveying disagreement in English and the social implications of these choices (Corpus for Schools, 2024)

Task 4

Over the last twenty years, our perception of the world has changed and people have become more aware of various social issues. Search for *sexis** and *racis** in BNClab and use the Change button to see how the use of these words changed overtime (Note: the asterisk acts as a wildcard allowing us to search for different forms of a word at the same time. For example, *sexis** will search for *sexism* as well as *sexist*).

- First, discuss what these words (*sexist* and *sexism*, *racist* and *racism*) mean
- What trends have you observed in the use of these words?
- Can you explain the changes in how people used these words in the 1990s and today? [Tip: To help you understand the changes, have a look at the context in which these words appeared in conversations in 1990s and today. To see the context, click in turn on each bar under the Usage button].

Teaching Sociolinguistics with Data-Driven Learning, Fig. 4 A task on semantic change in the use of the same forms at two points in time (Corpus for Schools, 2024)

Cross-References

- ▶ Communicative Data-Driven Learning
- ▶ Corpus-Informed Materials in Language Teaching
- ▶ First Language Acquisition and Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Pragmatics and Data-Driven Learning

Competing Interest Declaration The authors have no competing interests to declare that are relevant to the content of this manuscript.

References

Andersen, G. (2010). How to use corpus linguistics in sociolinguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 547–562). Routledge.

Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh University Press.

Brezina, V., & Platt, W. (2024). *#LancsBox X* [Software]. Lancaster University. http://lancsbox.lancs.ac.uk

Brezina, V., Love, R., & Aijmer, K. (2018). Corpus approaches to contemporary British speech. In *Sociolinguistic studies of the Spoken BNC2014* (pp. 654–657). Routledge.

Brezina, V., Hawtin, A., & McEnery, T. (2021). The written British National Corpus 2014 – Design and comparability. *Text & Talk, 41*(5–6), 595–615.

Bucholtz, M., Lopez, A., Mojarro, A., Skapoulli, E., VanderStouwe, C., & Warner-Garcia, S. (2014). Sociolinguistic justice in the schools: Student researchers as linguistic experts. *Language and Linguistics Compass, 8*(4), 144–157. https://doi.org/10.1111/lnc3.12070

Burnard, L. (2007). *BNC user reference guide*. http://www.natcorp.ox.ac.uk/docs/URG/BNCdes.html

Cheshire, J., & Fox, S. (2016). From sociolinguistic research to English language teaching. In P. Corrigan & A. Mearns (Eds.), *Creating and digitizing language corpora* (Vol. 3, pp. 265–290). Palgrave Macmillan.

Farrell, T. S. C. (2017). *Sociolinguistics and language teaching*. TESOL Press.

Folse, K., & Vitanova, G. (2006). Sociolinguistic factors in TESOL: The least teachers and teacher educators should know. *TESL Reporter, 39*(1), 48–58.

Gablasova, D. (2024). *Corpus for Schools*. Lancaster University. https://wp.lancs.ac.uk/corpusforschools/

Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research, 5*(2), 126–158.

Geeslin, K. L., & Long, A. Y. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context*. Routledge.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*, 225–242.

Gutiérrez, M., & Fairclough, M. (2006). Incorporating linguistic variation into the classroom. In R. Salaberry & B. Lafford (Eds.), *The art of teaching Spanish* (pp. 173–191). Georgetown University Press.

Hardie, A. (2012). CQPweb – Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics, 17*(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Holmes, J., & Wilson, N. (2022). *An introduction to sociolinguistics*. Routledge.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2004–2007). *Linguistic innovators: The English of adolescents in London*. UK Economic and Social Research Council project RES 000 23 0680.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2007–2010). *Multicultural London English*. UK Economic and Social Research Council project RES 062 23 0814.

Kraljević, J. K., & Hržica, G. (2016). Croatian adult spoken language corpus. *Fluminensia: Časopis Za Filološka Istraživanja, 28*(2), 87–102. https://hrcak.srce.hr/174013

Leeman, J. (2018). Critical language awareness and Spanish as a heritage language: Challenging the linguistic subordination of US Latinxs. In K. Potowski (Ed.), *Handbook of Spanish as a minority/heritage language* (pp. 345–358). Routledge.

Leeman, J., & Serafini, E. (2016). Sociolinguistics for heritage language educators and students: A model for critical translingual competence. In M. Fairclough & S. Beaudrie (Eds.), *Innovative approaches in heritage language pedagogy: From research to practice* (pp. 56–79). Georgetown University Press.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics, 22*(3), 319–344.

McKay, S. L., & Hornberger, N. H. (Eds.). (2005). *Sociolinguistics and language teaching*. Cambridge University Press.

Miguel, N. M. (2022). Exploring the use of corpus tools for teaching language variation to L2 Spanish majors. *Language, 98*(2), e80–e107. https://doi.org/10.1353/lan.2022.0001

Shin, N. L., & Hudgens Henderson, M. (2017). A sociolinguistic approach to teaching Spanish grammatical structures. *Foreign Language Annals, 50*(1), 195–213. https://doi.org/10.1111/flan.12249

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. The Regents of the University of Michigan.

Stenström, A., Andersen, G., & Hasund, I. K. (2002). *Trends in teenage talk: Corpus compilation, analysis and findings*. John Benjamins.

C

Complexity, Accuracy, and Fluency and Data-Driven Learning

Aysel Şahin Kızıl
İzmir Bakırçay University, İzmir, Türkiye

Abstract

Data-driven learning (DDL) is considered a promising approach for effectively enhancing the writing performance of second language (L2) learners. Although extensive research has explored the use of DDL in developing various aspects of writing skills, few studies have focused on its potential for revising writing. Moreover, few existing studies have examined DDL at the revision stage using complexity, accuracy, and fluency (CAF) measures as objective and reliable indices of writing quality. This entry explores the intersection of DDL and the CAF framework in L2 writing development. Following a brief discussion of DDL and CAF measures within L2 writing, this entry presents the empirical studies applying these measures within DDL contexts, discussing the methodologies and findings in detail.

Keywords

CAF measures · Complexity, accuracy, and fluency · Data-driven learning · L2 writing · Second language writing · Writing quality indices

Introduction

In the ever-evolving landscape of computer-assisted language learning (CALL), language corpora—large, systematically organized, and computer-readable collections of texts that enable the analysis of language use patterns across various contexts (Szudarski, 2018)—have become a prominent area of research, generating a wealth of studies in both linguistic analysis and language pedagogy. One particularly impactful method combining insights from CALL and corpus linguistics is what is dubbed data-driven learning (DDL). Pioneered by the work of Johns (1991), positioning learners as "research workers whose learning needs to be driven by access to linguistic data" (p. 2), DDL emphasizes the active role of learners in their own language acquisition process. In DDL as a pedagogical approach, learners interact with language data either indirectly through paper-based materials derived from the corpus (i.e. hands-off DDL) or directly by accessing online corpus data to generate concordance lines themselves (i.e. hands-on DDL).

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_87-1

Whether learners interact directly with corpora or use teacher-curated materials derived from them, the core principle of DDL lies in its potential to provide learners with intensive exposure to frequent language patterns. This aligns perfectly with usage-based models of acquisition, which posit that language learning happens through encountering and internalizing these recurring patterns in real-world use (Ellis, 2012). DDL can be viewed as a powerful tool to accelerate the process of language learning through "input flooding" and "condensed exposure" to lexical use and patterns in the target language (O'Keeffe & Mark, 2022, p. 1).

Together with these affordances, DDL has firmly established itself as "both a distinct teaching method and a distinct subfield of language teaching research" (Vyatkina & Boulton, 2017, p. 1). This understanding has spurred a surge of research interest in DDL, with scholars investigating its effectiveness in various learning contexts and exploring its potential to enhance language learning outcomes. One area in which the impact of DDL has been rigorously investigated is L2 English writing development (Boulton & Vyatkina, 2021). While a significant portion of research investigating the efficacy of DDL in L2 writing development reports positive findings (e.g. Chang, 2014), a closer examination reveals limitations in how these outcomes are measured. Many studies rely on measures such as student self-perceptions, gap-filling exercises, multiple-choice questions, or error frequency counts, which may not fully capture the complexities of writing development. This is verified in recent systematic reviews as well. Boulton's (2021) analysis of 351 DDL studies identified a relative scarcity of research employing detailed analysis of written production. Similarly, Pérez-Paredes (2022) noted that only 23% of studies analysed actual writing output. Among the studies gauging the impact of DDL on writing performance of learners, only a few (e.g. Huang, 2014; Luo, 2016; Samoudi & Modirkhamene, 2020) examined the impact of DDL at revision stage of L2 writing using complexity, accuracy, and fluency (CAF) measures.

CAF measures, being the three significant dimensions of writing, are considered as objective and reliable indices of writing quality and are reported to be effectively used in measuring the impact of pedagogical interventions (Barrot & Agdeppa, 2021; Skehan & Foster, 2012). Developmental measures allow teachers and researchers to evaluate and describe the learner's developmental level in a more precise way. In addition, they can be used to examine the effect of a particular pedagogical procedure on language use.

This entry explores the intersection of DDL and the CAF framework in L2 writing development. Given the substantial body of research supporting the efficacy of DDL in enhancing language learning, this entry aims to present the findings related to how DDL influences the development of CAF in L2 writing. The following sections delve into the theoretical underpinnings of DDL followed by a brief account of CAF measures and their relevance in L2 writing. It also reviews empirical studies that have applied these measures within DDL contexts. Additionally, this entry proposes directions for future research to further explore the potential of DDL in fostering complex, accurate, and fluent language use among learners.

DDL and L2 Writing

Crafting written communication requires the capacity to convey ideas effectively through text, drawing upon various cognitive and linguistic resources. Therefore, writers must possess not only a solid grasp of vocabulary and grammar but also a nuanced understanding of discourse conventions and rhetorical strategies. However, developing writing skills presents significant challenges for L2 learners and users especially in English as a Foreign Language (EFL) contexts, which are characterized by various constraints, including limited access to naturally occurring language data (Barrot & Gabinete, 2019). This restricted exposure is identified as one of the factors impeding L2 growth and language competence, ultimately resulting in lower writing

performance (Wu & Zhang, 2017). Despite prolonged effort and practice, many EFL learners, compared to skilled native writers, are reported to struggle with various aspects of writing, reaching only a modest level of proficiency (Samoudi & Modirkhamene, 2020). Their writings are typically characterized by issues in syntactic complexity, lexical diversity and accuracy (Barrot & Gabinete, 2019).

DDL is a promising pedagogical approach, drawing on constructivist language learning theory, which entails concepts such as induction, inference, hypothesizing, learner-centredness, and discovery learning. The DDL literature is also enriched with links to several SLA concepts including the noticing hypothesis, the usage-based model, input enhancement, and the involvement load hypothesis (for a fuller account of DDL and language learning theories, see O'Keeffe, 2021). Building on these theoretical foundations, DDL demonstrates considerable potential for enhancing writing skills in L2 learners.

DDL activities provide a unique foundation for stimulating inductive learning, particularly in the formulation and testing of hypotheses (O'Keeffe, 2021). Engaging in DDL activities, learners are exposed to extensive language input, enabling them to compare their usage of specific language structures with authentic language examples. Under the guidance of teachers, learners can scrutinize discrepancies in their language usage and analyse language units within the corpus data. This process allows them to test and formulate their own generalizations based on observed linguistic features, which is particularly beneficial for refining writing skills, as it requires learners to consciously compare their output with target language input (Schmidt, 2010). Moreover, increased and intensive exposure to language input through DDL activities may facilitate noticing of new structures and features that learners were previously unaware of (Leńko-Szymańska & Boulton, 2015).

These advantages have prompted extensive scholarly investigation into DDL within the domain of language acquisition, with a particular focus on writing instruction. A thorough examination of the literature concerning DDL in writing reveals two primary research strands based on the treatment of corpora: (1) Corpora as a research tool, which involves the systematic use of language corpora by L2 learners to deduce underlying rules pertaining to linguistic elements (e.g. Sun & Hu, 2023; (2) Corpora as a reference tool, where language corpora are consulted to address linguistic challenges encountered during writing (e.g. Huang, 2014).

Studies examining the efficacy of DDL for writing revision, albeit relatively limited in quantity (Liou, 2019), demonstrate variability in their approaches to assessing writing proficiency. Most of the studies in this direction have predominantly relied on error frequency as a metric to evaluate the writing performance of participants using DDL in their learning endeavours. A noteworthy finding from this line of research indicates that, when employed for revision purposes, DDL yields positive outcomes in correcting specific types of errors (Crosthwaite, 2017; Crosthwaite et al., 2020, among others). Additionally, other investigations on DDL for revision assess learner writing through the use of CAF measures (Huang, 2014; Luo, 2016; Samoudi & Modirkhamene, 2020; Şahin Kızıl, 2023). Further elaboration on CAF measures in L2 writing in conjunction with DDL will be provided in the subsequent discussion.

CAF Measures and L2 Writing

Given the multifaceted nature of L2 writing, CAF measures are considered the most comprehensive approach for its evaluation. Since their introduction by Skehan (1989) as the three key dimensions of proficiency, CAF measures have served as an established metric of writing quality. From a theoretical perspective, CAF measures are regarded as a fundamental construct in L2 writing research that are used to identify valid and reliable indices of language development and overall proficiency in L2 learners. From a practical perspective, they provide a valuable framework for understanding L2 performance in writing classrooms and enable researchers to objectively and holistically assess learners' progress in L2 writing (Barrot & Agdeppa, 2021).

The application of CAF measures has been the subject of rigorous investigation across four key areas: language development, language performance, writing quality, and proficiency (Ortega, 2012). When used as an index of language performance, some studies have investigated the variations in CAF measures and their connection to cognitive task conditions imposed on L2 learners (e.g. Johnson, 2017). As an index of writing quality, research has explored how CAF measures can differentiate high-performing from low-performing L2 writers (e.g. Casal & Lee, 2019). In terms of language development, studies have set out to determine if certain CAF measures serve as valid and reliable indicators of L2 writing development over time (e.g. Yoon & Polio, 2017). For proficiency, research suggests that some complexity measures are valid indicators, with indices of CAF increasing as L2 users become more proficient (e.g. Martínez, 2018).

The concept of complexity broadly refers to a learner's ability to use a diverse range of grammatical structures and forms, thereby demonstrating linguistic sophistication. Bulté and Housen (2014) define it as "an absolute, objective, and inherently quantitative property of language units, features, and (sub)systems" determined by the number and nature of distinct elements and their interrelationships (pp. 45–46). Syntactic complexity in L2 writing has been operationalized through various metrics, often focusing on specific syntactic units like T-units, sentences, and clauses. For a comprehensive list of complexity metrics, see Lu (2010). Although research findings regarding the most effective measures are not entirely conclusive, a consensus among L2 writing scholars is that that higher proficiency in L2 learners manifests in the production of more intricate syntactic structures.

Accuracy is defined as L2 writers' ability to employ language that adheres to established norms of grammatical form and structure by producing error-free language (Skehan & Foster, 2012). As for the indices to measure accuracy, researchers suggest various metrics including holistic scores, error-free units (EFT), weighted error-free T-unit, and error-free clauses (EFC). Polio and Shea (2014), in their investigation of accuracy measure validity, found that none of the metrics emerged as definitively superior.

Finally, fluency within L2 writing contexts is defined as a writer's ability to generate syntactic units and written words within a specific timeframe (Barrot & Agdeppa, 2021). Various indices have been identified to measure fluency, with some of the earlier ones being error-free T-unit length (EFT length), error-free clause length (EFC length), and T-unit length (Wolfe-Quintero et al., 1998). However, relevant literature approaches this cautiously, noting that a strict focus on error-free T-units and clauses could potentially conflate with accuracy measures. Therefore, some studies opt to use alternative metrics such as the number of words per text, number of T-units per text, number of clauses per text and mean length of error-free T-units (MLEFT) (Barrot & Agdeppa, 2021; Samoudi & Modirkhamene, 2020; Şahin Kızıl, 2023). The effectiveness of these measures has been explored through the use of DDL as a source of intervention. The following section provides a concise overview of empirical findings at the intersection of DDL, CAF, and L2 writing.

L2 Writing at the Intersection of DDL and CAF

Studies examining the writing performance of the learners at the intersection of DDL and CAF measures in second language writing are notably sparse, with limited empirical investigations reporting inconclusive results.

Huang (2014) investigated the potential impact of DDL activities on enhancing the lexico-grammatical use of five abstract nouns in second language (L2) writing with 40 Chinese learners of English. Paper-based concordance lines led to a significant improvement in both accuracy and complexity over dictionary use. The author notes that DDL helps in recognizing, internalizing, and preserving lexico-grammatical patterns and "enables learners to generate more accurate and complex syntactic structures" (p. 175).

Luo (2016) conducted a comparative study on the effectiveness of direct DDL vs. search engines

as revision tools on writing CAF. The study revealed significant improvements in fluency and accuracy for the DDL group compared to the control group; however, no significant difference was observed in writing complexity. This suggests that DDL might be helpful in developing awareness of lexicogrammatical patterns, and facilitating fluency and accuracy in writing. Enhanced accuracy through the implementation of DDL for revision is also reported by Liou (2019) who observed a decrease in student errors as their use of corpus tools for revision increased. The results supporting DDL's effectiveness in enhancing accuracy are contradicted by Samoudi and Modirkhamene's (2020) study that investigated the impact of direct DDL vs. indirect DDL activities on the writing development. The study found no significant difference in accuracy improvement between the direct DDL and control group. More recently, Şahin Kızıl (2023) investigated the impact of integrating DDL into the writing process of EFL learners. The study's findings revealed a significant positive impact of DDL on writing fluency and lexical complexity. The DDL group outperformed the non-DDL group in producing writing that was both lexically diverse and fluent. However, no significant difference was observed between the groups regarding accuracy or grammatical complexity.

The empirical findings sketched out above showcase mixed findings regarding the impact of DDL on various aspects of writing. While some research highlights positive effects on fluency, lexical complexity, and even accuracy, others report no significant improvement in these areas.

Conclusion

This entry examines the intersection of DDL and CAF measures in the context of L2 writing development. While DDL holds promise as a pedagogical approach, the research on its effectiveness in enhancing complexity, accuracy, and fluency in L2 writing yields mixed results. The relevant literature shows that DDL activities potentially improve L2 writing fluency and lexical complexity: learners may develop a broader vocabulary and awareness of lexicogrammatical patterns, leading to more diverse and fluent writing. The impact of DDL on accuracy, on the other hand, is inconclusive. Some studies report positive effects, while others show no significant improvement. Additionally, limited research has investigated the influence of DDL on syntactic complexity, with no clear consensus on its effectiveness. Also limited is the number of studies handling the CAF measures in connection with DDL jointly as an index of writing quality. Existing research investigating DDL for revision in relation with CAF mostly focused on accuracy and neglected the other dimensions of writing—fluency, in particular, has been underexplored.

Given the potential of DDL for enhancing L2 writing skills, particularly fluency and lexical complexity, further research exploring its role in L2 writing using CAF measures is highly recommended. Future studies should incorporate robust pre-tests, post-tests, and control groups to establish a causal relationship between DDL and CAF improvement. Additionally, investigating the effectiveness of different DDL implementation methods (direct vs. indirect) in relation to CAF measures is warranted.

Overall, DDL presents a promising avenue for enhancing L2 writing skills, particularly fluency and lexical complexity. However, more research is required to fully understand its influence on various aspects of CAF and identify the most effective implementation strategies.

Cross-References

► Hands-On Versus Hands-Off Methods in DDL
► Input Enhancement and Input Flood in DDL

References

Barrot, J., & Gabinete, M. K. (2019). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. *International Review of Applied Linguistics in Language Teaching, 59*(2), 209–232. https://doi.org/10.1515/iral-2017-0012

Barrot, J. S., & Agdeppa, J. Y. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers'

proficiency. *Assessing Writing, 47*(2021), 1–11. https://doi.org/10.1016/j.asw.2020.100510

Boulton, A. (2021). Research in data-driven learning. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond the concordance: Corpora in language education* (pp. 9–34). John Benjamins. https://doi.org/10.1075/scl.102.01bou

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*(2014), 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing, 44*(June), 51–62. https://doi.org/10.1016/J.JSLW.2019.03.005

Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL, 26*(2), 243–259. https://doi.org/10.1017/S0958344014000056

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning, 30*(6), 447–473. https://doi.org/10.1080/09588221.2017.1312462

Crosthwaite, P., Storch, N., & Schweinberger, M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing, 49*, 100729. https://doi.org/10.1016/j.jslw.2020.100729

Ellis, N. (2012). Frequency effects. In P. Robinson (Ed.), *The Routledge encyclopedia of SLA* (pp. 260–265). Routledge.

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL, 26*(2), 163–183. https://doi.org/10.1017/S095834401400002

Johns, T. (1991). Should you be persuaded: Two examples of data driven learning. *ELR Journal, 4*(1), 1–16.

Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing, 37*(June), 13–38. https://doi.org/10.1016/j.jslw.2017.06.001

Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 1–14). https://doi.org/10.1075/scl.69.01int

Liou, H. C. (2019). Learner concordancing for EFL college writing accuracy. *English Teaching and Learning, 43*(2), 165–188. https://doi.org/10.1007/s42321-019-00022-5

Lu, X. (2010). What can corpus software reveal about language development? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 184–193). Routledge.

Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *Springerplus, 5*(1), 1–13. https://doi.org/10.1186/s40064-016-2935-5

Martínez, A. C. L. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing, 35*, 1–11. https://doi.org/10.1016/j.asw.2017.11.002

O'Keeffe, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching, 54*(2), 259–272. https://doi.org/10.1017/S0261444820000245

O'Keeffe, A., & Mark, G. (2022). Principled pattern curation to guide data-driven learning design. *Applied Corpus Linguistics, 2*(3), 100028. https://doi.org/10.1016/j.acorp.2022.100028

Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). Mouton de Gruyter. https://doi.org/10.1515/9783110229226.127

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10–27. https://doi.org/10.1016/J.JSLW.2014.09.003

Şahin Kızıl, A. (2023). Data-driven learning: English as a foreign language writing and complexity, accuracy and fluency measures. *Journal of Computer Assisted Learning, 39*(4), 1382–1395. https://doi.org/10.1111/jcal.12807

Samoudi, N., & Modirkhamene, S. (2020). Concordancing in writing pedagogy and CAF measures of writing. *International Review of Applied Linguistics in Language Teaching, 60*(3), 699–722. https://doi.org/10.1515/iral-2020-2014

Schmidt, R. (2010). Attention, awareness and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). National University of Singapore, Centre for Language Studies.

Skehan, P. (1989). *Individual differences in second language learning*. Edward Arnold.

Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 199–220). John Benjamins). https://doi.org/10.1075/lllt.32.09fos

Sun, X., & Hu, G. (2023). Direct and indirect data-driven learning: An experimental study of hedging in an EFL writing class. *Language Teaching Research, 27*(3), 660–688. https://doi.org/10.1177/1362168820954459

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge. https://doi.org/10.4324/9781315107769

Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology, 21*(3), 1–8. https://doi.org/10.10125/44621

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawaii Press.

Wu, H., & Zhang, L. J. (2017). Effects of different language environments on Chinese graduate students' perceptions of English writing and their writing performance. *System, 65*, 164–173. https://doi.org/10.1016/j.system.2017.02.001

Yoon, H.-J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly, 51*(2), 275–301. https://doi.org/10.1002/tesq.296

C

Corpus Literacy and Data-Driven Learning

Qing Ma
Department of Linguistics and Modern Language Studies, The Education University of Hong Kong, Hong Kong, China

Abstract

This entry examines the need for basic corpus literacy among language teachers and learners to expand the benefits of data-driven learning (DDL) to a wider audience beyond corpus linguists and researchers. It explores the nature and typical contexts of DDL, defines corpus literacy, and reviews existing measures of corpus literacy. The entry synthesizes previous research on DDL in language education and examines the limitations in the emphasis on advanced corpus knowledge and skills for corpus educators at the university settings, neglecting basic corpus literacy needed for language teachers and learners without corpus expertise. The entry highlights the need to prioritize basic corpus literacy for language teachers and learners. In addition, it offers practical recommendations for effective corpus literacy training, including prioritizing essential corpus skills, scaffolding corpus website introduction, designing user-friendly corpus tools, emphasizing hands-on corpus analysis practice, and addressing corpus limitations to foster users' metacognitive awareness. By advocating a basic and attainable corpus literacy for the vast number of language teachers and learners, this entry aims to democratize access to DDL resources, empowering language educators to make informed pedagogical decisions and enhance learners' language learning experiences.

Keywords

Corpus literacy · Data-driven learning · Language education · Language teachers · Language learners

Introduction

The field of language education has witnessed a growing emphasis on the integration of corpus linguistics in research and teaching methodologies. Corpus-based language learning, commonly known as data-driven learning (DDL), utilizes authentic language data from corpora to actively engage learners in the analysis and exploration of language patterns and usage. While DDL has primarily been explored in higher education settings, it has yet to gain widespread familiarity among ordinary school teachers and their students. This discrepancy highlights the need for basic corpus literacy among language teachers and learners, ensuring that the benefits of DDL research can be extended to a much wider

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_88-1

audience beyond tertiary settings. Corpus literacy refers to the skill of employing corpus linguistics technology for language investigation and development (Mukherjee, 2006).

In this entry, we first examine the nature and typical contexts of DDL, recognize its achievements and identify the direction regarding how to make DDL truly benefit the vast language teacher and learner communities who do not possess corpus expertise. We then elaborate the concept of corpus literacy (CL) which can be divided into basic and advanced CL. The former is particularly suitable for language teachers and learners new to corpus technology (e.g., what knowledge to acquire and what online resources to refer to), whereas the latter is typically needed by corpus linguists or researchers for analysing and researching languages. We also review how teachers' and learners' basic CL can be measured. Based on which, we propose some pedagogical implications regarding how to provide effective training to help teachers and learners gain a sufficient CL to harness corpus technology to benefit their language teaching and learning.

DDL: Definition and Overview

Johns (1990) defines DDL as an approach where learners are actively engaged in analysing authentic linguistic data, typically from corpora. Learners are encouraged to discover patterns, test hypotheses, and draw conclusions about language use based on the evidence present in the data. The teacher acts as a facilitator, guiding learners through the process of linguistic investigation. Later on, Boulton defines prototypical DDL as "the hands-on use of authentic corpus data (concordances) by advanced, sophisticated foreign or second language learners in higher education for inductive, self-directed language learning of advanced usage" (2011, p. 572). In this context, DDL primarily emphasizes learners' independent exploration of corpora, rather than the pedagogical use of corpora by teachers, although the latter is not entirely precluded. For this reason, the majority of DDL literature has investigated advanced learners' use and learning with corpus technology at the university level or beyond (Boulton & Vyatkina, 2021, Pérez-Paredes, 2020). Therefore, DDL has mainly benefited corpus linguists/educators and language learners in university settings, but it largely remains unfamiliar to ordinary school teachers and their students (Callies, 2019; Ma et al., 2022).

DDL in language education, as reflected in various key texts published in the past two decades (e.g., Sinclair, 2004; McEnery et al., 2006; Pérez-Paredes, 2020), reveals an evolving landscape where corpus linguistics is increasingly central to both research and teaching methodologies. The research, ranging from the work by McEnery et al. (2006) on comprehensive guide to corpus-based language studies to the work by Taylor et al. (2021) focusing on university-level language teaching, encompasses a wide array of approaches and applications. However, a critical examination of these papers uncovers a prevalent issue: the emphasis is placed on advanced corpus knowledge and skills for researchers and practitioners, often at the expense of basic corpus literacy for the broader audience of language teachers and learners.

To truly integrate DDL into language learning and teaching, the field must place greater emphasis on developing and promoting basic corpus literacy among the key stakeholders: language teachers and learners. This includes simplifying corpus tools, developing user-friendly interfaces, and creating more intuitive teaching materials that can be used by educators without specialized training in corpus linguistics. By doing so, the benefits of DDL research can permeate more deeply into everyday language learning and teaching, making corpora accessible to a wider audience and ultimately enhancing the learning experience for students across diverse linguistic backgrounds.

Basic Corpus Literacy for Language Teachers and Students

Mukherjee (2006) raised the concept of "corpus literacy" (CL) to help DDL reach the school teacher community. Originally, the term "literacy"

was used to describe the fundamental skills and knowledge related to books and printing, as noted by McMillan in 1996. However, due to advances in digital technologies, efforts have expanded the concept of literacy beyond just reading and writing (Buckingham & Burn, 2007). Later on, Heather and Helt (2012) formally defined corpus literacy as the skills to employ corpus linguistics technology for language investigation and to improve students' language development.

Mukherjee (2006) emphasizes the importance of equipping ordinary school language teachers, who may not have prior expertise in corpus linguistics, with foundational CL. This involves four key dimensions. First, teachers need to understand what a corpus is—essentially, a large, structured collection of texts used for linguistic analysis. Second, they should be aware of what can and cannot be achieved with a corpus. For instance, while a corpus can reveal frequency and usage patterns, it might not explain *why* certain linguistic forms are used. Third, teachers must develop skills in analysing corpus data, such as using software to identify common phrases or grammatical structures in large text samples. Fourth, they should be capable of drawing meaningful conclusions about language use from their analyses, such as understanding trends in language evolution or variations in usage between different demographic groups. Lastly, Callies (2016) emphasizes the importance of knowing how to search corpora effectively using available resources and tools as a key component of CL. Developing corpus literacy involves hands-on experience with corpora, particularly using concordance software. This practical experience of searching concordance lines or other corpus information (e.g., collocates, key words in context—KWIC) is essential for developing a deeper understanding of language use. Similarly, language learners are expected to develop these five dimensions of basic CL to truly benefit from a DDL approach.

In recent years, a number of free resources have been developed to facilitate developing teachers' and students' corpus literacy, including the website Corpus-aided Platform for Language Teachers (CAP; https://corpus.eduhk.hk/cap/) and a streamlined MOOC course based on the CAP (https://pressbooks.pub/cmtry1/). Other online resources include the pressbooks created by Le Foll (2021) (https://pressbooks.pub/elenlefoll/) and Pinto et al. (2023) (https://uq.pressbooks.pub/using-language-data/).

Advanced Corpus Literacy for Language Teachers in the AI Era

As language education evolves alongside progress in AI, teachers face new challenges in integrating DDL with AI-driven instruction, requiring them to cultivate advanced corpus skills that extend beyond basic CL. First, teachers may learn how to effectively present concordance lines in a clear and accessible manner, reducing data density and complexity for student comprehension (Lusta et al., 2023; Bennett & Dhonnchadha, 2023). Furthermore, they should expand their use of corpora beyond language structures to explore cultural insights and generate engaging discussion topics, as well as analyse stylistic devices like juxtaposition and metaphors to enrich understanding of language use (Poole, 2018). Creating one's own corpus, tailored to specific learning objectives, allows teachers to provide highly relevant and targeted language data (Friginal, 2018; Poole, 2018), from which they can extract targeted wordlists and collocation lists to enable focused learning of specific language patterns within a given topic. By mastering these advanced corpus skills, language teachers can effectively leverage the power of DDL in the age of AI, fostering a deeper and more nuanced understanding of language for their students.

How to Measure Teachers' and Learners' Basic Corpus Literacy?

We need practical and easy-to-use measurement of CL since our goal is to help the vast number of language teachers and learners gain a basic corpus literacy. Research on measuring teachers' and learners' corpus literacy (CL) predominantly employs qualitative methods like interviews, reflections, and lesson analysis (Latif, 2021; Leńko-Szymańska, 2017; Heather & Helt, 2012; Zareva, 2017). Some studies use surveys with

closed and open-ended questions (Leńko-Szymańska, 2014; Callies, 2019). Most explore teacher and learner attitudes and perceptions towards using corpora, focusing on their training experiences (Leńko-Szymańska, 2014; Zareva, 2017) and attitudes towards corpus use in learning and teaching (Mukherjee, 2006; Latif, 2021).

The five components of basic CL (understand ing, advantages and disadvantages, analysis, and search skills) suggested by Mukherjee (2006) and Callies (2016) can be used as a framework to help us understand how teachers' and students' CL has been measured. Among the five components, understanding is relatively straightforward to assess through open-ended or Likert-scale que stions (Leńko-Szymańska, 2014; Özbay & Kayaoğlu, 2015). Perceived advantages and d isadvantages, confidence in corpus analysis, and interpretation skills have also been investigated (Callies, 2019; Leńko-Szymańska, 2014; Mu kherjee, 2006).

The above studies, often limited by small samples and qualitative methods, have explored specific CL subskills (under the five-component framework) but lack a comprehensive quantitative study investigating all five critical CL subskills. The study by Ma et al. (2023) filled this gap by investigating teachers' perceived corpus literacy using a survey developed based on the five-component framework with reference to Mukherjee (2006) and Callies (2019). The study also hypothesized a relationship between corpus literacy and teachers' intention to use corpora in their teaching. 183 pre-service and in-service teachers participated, receiving corpus training before completing a survey assessing their corpus literacy and intentions. Confirmatory factor analysis revealed a hierarchical factor structure for corpus literacy, encompassing the five subskills. Structural equation modelling indicated a positive correlation between corpus literacy and the intention to integrate corpora into teaching. The findings emphasize the importance of all five subskills, particularly search and analytical skills, considered crucial for effective corpus literacy training. Additionally, recognizing the limitations of corpora is vital for informing tool development, instructional design, and teachers' metacognitive awareness in DDL classroom applications.

Although the research by Ma et al. (2023) has provided a useful and practical measure for measuring teachers' and learners' basic corpus literacy, self-reported data might not accurately reflect actual CL performance. Future research should develop and validate objective CL performance tests for a more accurate and objective assessment of teachers' and learners' CL skills.

How to Provide Initial Corpus Literacy Training for Teachers and Learners?

The aforementioned findings regarding how to measure teachers' and learners' basic corpus literacy can provide us with insights into enhancing initial corpus literacy effectively for teachers and students who are new to corpus linguistics and technology. The literature reveals that the following strategies can be adopted to help develop teachers' and students' initial corpus literacy.

Prioritizing Essential Corpus Skills

While establishing a foundational understanding of corpora and their benefits is essential (, training should prioritize developing practical skills. Search skills, particularly using concordances to find words, collocates and phrases, are fundamental for navigating and utilizing corpus data (Leńko-Szymańska, 2017; Zareva, 2017). Equally important is developing analytical skills to interpret corpus data and understand its inherent limitations, moving beyond surface-level observations.

Scaffolding Corpus Website Introduction

Introducing multiple corpus websites simultaneously can overwhelm learners, hindering their ability to grasp the nuances of each platform. Begin with one website, focusing on its unique interface and search functions. Once learners are comfortable, gradually introduce additional websites, highlighting similarities and differences. This scaffolded approach facilitates a smoother learning curve, preventing cognitive overload.

Designing User-Friendly Corpus Tools

Corpus website developers should prioritize user-friendly interfaces and simplified search syntax to make corpora accessible to a wider audience. The revamped COCA website (https://www.english-corpora.org/coca/) (since 2020), simplifying complex search functions previously understood primarily by corpus linguists, may serve as a model for other platforms. Other user-friendly corpus tools suitable for teacher and student training include LexTutor (https://www.lextutor.ca/conc/), SKELL (https://skell.sketchengine.eu/#home?lang=en), Netspeak (https://netspeak.org/), and CorpusMate (https://corpusmate.com/). Accessible corpus tools encourage greater uptake among teachers and learners, promoting wider adoption of corpus-based methods among the teacher and learner communities.

Emphasizing Hands-on Corpus Analysis Practice

Corpus analysis can be challenging, as noted by Zareva (2017), who found that teacher trainees struggled with the time-consuming nature of analysis, summarizing language patterns, and interpreting data presented in tables and charts. Training should provide ample hands-on practice, allowing learners to interact with corpora, analyse data, and draw meaningful conclusions. Activities should focus on interpreting tables and charts, identifying language patterns, and presenting findings in clear and accessible ways, as advocated by Farr (2008). From the learner's perspective, Heather and Helt (2012) suggest that teacher trainees should learn how to "organize and present concordance data in ways that lead more clearly to autonomous learning for their students" (p. 436).

Addressing Corpus Limitations and Fostering Metacognitive Awareness

Understanding the limitations of corpora is crucial for effective CL, enabling users to critically evaluate findings and avoid misinterpretations. Training should explicitly address these limitations and encourage users to develop strategies to compensate for them. This fosters metacognitive awareness, enabling teachers to plan and manage corpus-based teaching effectively. As Heather and Helt (2012) suggest, understanding corpus limitations may even be indicative of a higher level of CL.

There are several benefits of understanding corpus limitations. First, such understanding can inform corpus linguists and developers about user challenges in interacting with corpus technology and propel them to create more user-friendly and effective tools, better aligned with the needs of teachers and students. Second, teachers can design targeted training activities to address specific corpus limitations, facilitating better learning outcomes and equipping teachers with practical instructional strategies. Third, understanding limitations empowers teachers to critically evaluate corpus data and make informed pedagogical decisions, ensuring responsible and effective corpus use in the classroom.

By focusing on the above key areas, CL training can empower language teachers and learners to effectively utilize corpora and corpus technology, enriching language learning and teaching experiences. After they gain a basic corpus literacy, tailored training targeting those more advanced CL skills as shown above can be provided to further improve their corpus skills and empower their teaching and learning.

Conclusion and Future Directions

DDL has significant potential for enhancing language learning and teaching methodologies. However, to fully realize these benefits, there is a need to bridge the gap between advanced corpus knowledge and skills for researchers and basic corpus literacy for language teachers and learners. This requires simplifying corpus tools, developing user-friendly interfaces, and creating more intuitive teaching materials that can be used by teachers without specialized training in corpus linguistics.

A basic corpus literacy for language teachers and students involves understanding what a corpus is, being aware of its advantages and limitations, developing skills in analysing corpus data, and being capable of drawing meaningful conclusions about language use. Teachers should also be competent in effective corpus search techniques. Similarly, language learners need to develop these

basic corpus literacy skills to effectively benefit from a DDL approach.

Measuring teachers' and learners' basic corpus literacy can be done through qualitative methods such as interviews, reflections, lesson analysis, and surveys with closed and open-ended questions. However, there is a need for more comprehensive quantitative studies that investigate all the critical subskills of basic corpus literacy, and compare self-assessment against more objective criteria of actual ability.

To provide effective initial corpus literacy training, it is essential to prioritize practical skills such as search skills and analytical skills. This can be achieved through scaffolding the introduction of corpus websites, designing user-friendly corpus tools, emphasizing hands-on corpus analysis practice, and addressing corpus limitations to foster metacognitive awareness. By focusing on these areas, language teachers and learners can gain the necessary corpus literacy skills to harness the full potential of corpus linguistics in language learning and teaching.

Incorporating corpus literacy into language education has the potential to enhance language learning and teaching practices for a wider audience. By simplifying corpus tools, developing user-friendly interfaces, and providing effective training, language teachers and learners can gain the necessary skills to effectively utilize corpus linguistics technology and enhance their language development. Further research is needed to develop and validate objective measurements of corpus literacy skills and to explore more advanced corpus literacy skills for researchers and practitioners. Overall, corpus literacy holds great promise for advancing language education and empowering language teachers and learners.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Bennett, C., & Dhonnchadha, E. U. (2023). Becoming corpus literate: An in-service EFL teacher education framework for integrating corpora into EFL teaching. *Applied Corpus Linguistics, 3*(1), 100048. https://doi.org/10.1016/j.acorp.2023.100048

Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563–580). Peter Lang.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. 10125/73450

Buckingham, D., & Burn, A. (2007). Game literacy in theory and practice. *Journal of Educational Multimedia and Hypermedia, 16*(3), 323–349.

Callies, M. (2016) Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. In Kreyer, R., Schaub, S. & Güldenring, B. A. (eds.), angewandte linguistik in schule und hochschule [applied linguistics in secondary school and at university]. Frankfurt am Main: Peter Lang, 391–415. Google Scholar

Callies, M. (2019). Integrating corpus literacy into language teacher education: The case of learner corpora. In S. Götz & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245–263). John Benjamins. https://doi.org/10.1075/scl.92.12cal

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness, 17*(1), 25–43. https://doi.org/10.2167/la414.0

Friginal, E. (2018). Developing corpus-based lessons and activities: An introduction. In E. Friginal (Ed.), *Corpus linguistics for English teachers* (pp. 187–212). Routledge. https://doi.org/10.4324/9781315649054-7

Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for student language teachers: Six case studies. *Journal of Technology and Teacher Education, 20*(4), 415–440.

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria, 10*, 14–34.

Latif, M. M. M. A. (2021). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL, 33*(1), 34–48. https://doi.org/10.1017/S0958344020000129

Le Foll, E. (Ed.). (2021). *Creating corpus-informed materials for the English as a foreign language classroom*. Pressbooks. Retrieved from https://pressbooks.pub/elenlefoll

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL, 26*(2), 260–278. https://doi.org/10.1017/S095834401400010X

Leńko-Szymańska, A. (2017). Training teachers in data driven learning: Tackling the challenge. *Language Learning & Technology, 21*(3), 217–241. 10125/44628.

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data-driven learning (DDL) in language classrooms: A systematic review. *Heliyon,*

9(12), e22731. https://doi.org/10.1016/j.heliyon.2023.e22731

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning, 35*(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

Ma, Q., Chiu, M. M., Lin, S., & Mendoza, N. B. (2023). Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL, 35*(1), 19–39. https://doi.org/10.1017/S0958344022000180

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

McMillan, S. (1996). Literacy and computer literacy: Definitions and comparisons. *Computers & Education, 27*(3–4), 161–170. https://doi.org/10.1016/S0360-1315(96)00026-7

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art–and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpora and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Peter Lang.

Özbay, A., & Kayaoğlu, M. N. (2015). EFL teacher's reflections towards the use of computerized corpora as a teaching tool in their classrooms. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 44*(1), 85–104. https://doi.org/10.14812/cuefd.54367

Pérez-Paredes, P. (2020). *Corpus linguistics for education: A guide for research*. Routledge. https://doi.org/10.4324/9780429243615

Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. (Eds.). (2023). *Using language data to learn about language: A teachers' guide to classroom corpus use*. University of Queensland.

Poole, R. (2018). *A guide to using corpora for English language learners*. Edinburgh University Press. https://doi.org/10.1515/9781474427180

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins. https://doi.org/10.1075/scl.12

Taylor, C. T., Ackerley, K., & Castello, E. (2021). *Corpora for university language teachers*. Peter Lang.

Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal, 71*(1), 70–79. https://doi.org/10.1093/elt/ccw045

T

Teacher Education for Pedagogical Uses of Corpora

Agnieszka Leńko-Szymańska
University of Warsaw, Warsaw, Poland

Abstract

Corpora use among language teachers, especially in primary and secondary schools, remains limited. Addressing this reluctance requires specific teacher education which involves enhancing teachers' technical skills, corpus linguistics expertise, as well as pedagogical abilities to create corpus-based materials and activities and integrate them into classroom procedures. Various approaches to teacher education have been reported, which include intensive workshops and courses on corpus linguistics, corpus-assisted language learning, as well as training in corpus-based language pedagogy. In addition, complementary methods like corpus-based reflective practice have been proposed. Research on the effectiveness of these different approaches uses various methods but faces generalisation challenges. However, even well-designed training programmes may not suffice. Accessible resources and tools are needed, including simplified text collections, user-friendly corpus-querying tools, and resource banks with ready-to-use corpus-based materials and activities. These resources can ease preparation time, inspire creative use, and foster gradual confidence in incorporating corpora into teaching.

Keywords

Corpus-based language pedagogy · Corpus-based reflective practice · Corpus linguistics skills · Corpus literacy · Pedagogical corpora · Pedagogical corpus tools · Pedagogical skills · Teacher education in corpus-based pedagogy · Technical skills

Introduction

The applications of corpora in language teaching are twofold. First, they can inform the creation of syllabuses and support the development of pedagogical materials. These materials use well-established formats for presenting selected language points (e.g. descriptions and exemplifications), and for practice (e.g. gap-filling and matching tasks), but include more precise and accurate accounts of language use, such as information on frequency, patterns, and variation. Second, corpora also enable a new approach to language teaching—data-driven learning, where learners themselves query large collections of language samples for relevant linguistic information and learn directly from language data, with no or little mediation of the teacher and curated language materials.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_89-1

These opportunities offered by language corpora, especially the former one, have been rapidly taken up by professional materials writers and publishers. Over the last 40 years, corpora have had an enormous impact on the production of language teaching materials such as dictionaries, reference grammars, activity books, and course books. Unfortunately, despite technological advances and improved access to technology, as well as the increasing availability of corpus resources and tools, corpora have not been embraced by language teachers, particularly those working in primary and secondary schools. Teachers are reluctant to use corpora, and when they do, it is mainly for their own reference (Callies, 2019).

The antidote for this lack of enthusiasm among language instructors has been seen in teacher education. As early as in the first decade of the twenty-first century, corpus linguists identified the need for teacher education in the context of pedagogical uses of corpora (Mukherjee, 2004), and the first reports of such training for language teachers began to appear in the literature.

Teachers' Competences and Skills

Researchers and educators promoting the use of corpora for language instruction have identified at least three key sets of skills that teachers must possess in order to take full advantage of corpus resources (Leńko-Szymańska, 2022). The most evident of these are technical skills related to setting up access to corpora, operating software and troubleshooting. Corpus linguistics is developing rapidly, with new resources and tools being created, and older ones being updated and redesigned on a regular basis. Consequently, teachers must be able to continually adapt to these changes and develop a dexterity in dealing with ever-evolving computer programmes, functionalities, and interfaces. A separate, though related set of skills, which goes beyond the mere operation of software, pertains to expertise in corpus linguistics. Teachers need to understand what kind of linguistic information can be retrieved from corpora, know how to search for this information (i.e. how to write effective queries), and be able to interpret the results of their searches (i.e. observe patterns and generalise from messy data). Such procedures pose high cognitive demands, which teachers must be prepared to meet. Finally, it is not sufficient to be capable of using corpus resources and tools and finding information that is relevant to language learners. In addition, teachers also must be able to transform this information into meaningful corpus-informed or DDL materials and activities for their students, and to integrate them with their everyday pedagogical practices. Corpus-based tasks cannot be an end in themselves; rather, they must be well integrated into larger teaching sequences in order to have a lasting impact on language acquisition.

It is evident that mastering these three sets of skills by language teachers requires dedicated and well-designed training. Consequently, a debate about an appropriate model for such education has opened up, coupled with research on the effectiveness of various approaches.

Approaches to Teacher Education

Teacher education in the application of corpora in language instruction has been approached from different angles. One approach emphasises the acquisition of knowledge and skills related to corpus-based research. This training may take the form of short intensive workshops, particularly for in-service teachers, or comprehensive courses within the academic component of pre-service programmes (e.g. Abdel-Latif, 2021). This approach involves the introduction of essential concepts in the field of corpus linguistics and the presentation of various corpus resources and methods for querying text collections using a range of corpus tools. Its primary objective is to help trainees understand the types of information that can be extracted from a corpus and the types of analyses that can be performed with corpus tools. The syllabus is typically designed around available resources and software, with a focus on developing technical and corpus-analytical skills in language teachers.

An alternative approach involves introducing trainees—typically student teachers—to corpora as part of language development or language awareness courses, which are often included in teacher education programmes. In these courses, future teachers interact with corpora as language learners, performing tasks that focus on analysing specific language features (e.g. Callies, 2016). The trainees are presented with data or information derived from corpora or are guided to manipulate corpus tools in order to complete language activities. This approach prioritises demonstrating pedagogical applications of corpora to the detriment of a solid foundation in corpus linguistics.

However, neither of these two approaches warrants that teachers will gain competence in the appropriate and effective application of corpus-based resources into classroom pedagogical practices and their integration with larger teaching sequences such as the popular Present-Practice-Produce model. Consequently, within some teacher education programmes trainees are instructed explicitly in planning, designing, and creating corpus-informed and DDL materials and tasks for learners of different ages and at different proficiency levels, including learners of specialised languages (e.g. Leńko-Szymańska, 2014, Le Foll, 2023). Ma et al. (2022) have proposed that such training should be a two-step programme, comprising the corpus literacy (CL) and corpus-based language pedagogy (CBLP) components. The latter component may take the form of learning by design in which student teachers collaborate in groups to create their own corpus-based lessons (e.g. Le Foll, 2023; Ma et al., 2023). The process of creation, revision, and reflection on the final product enables future teachers to gain knowledge and skills related to the effective application of corpus-derived information and corpus-based activities in their teaching practice.

Nevertheless, research has demonstrated that even such dedicated courses may not be sufficient (Leńko-Szymańska, 2014, 2017). Therefore, Farr and Leńko-Szymańska (2023a) proposed that the two-step programme can be supplemented with another method of corpus integration into language teacher education which they termed corpus-based reflective practice (CBRP). This practice involves the use of corpora of classroom discourse as data for student teachers' reflection on teaching and learning procedures and practices. By applying corpus methodology, in-service teachers can develop a deeper insight into the patterns and features of classroom interaction and observe specific aspects of language instruction and acquisition. Thus, within this extended framework, corpora are not only utilised as essential tools to improve language teaching and learning but also provide valuable insights into the pedagogical and acquisitional processes.

Research on the Effectiveness of Teacher Education

The effectiveness of different approaches to teacher training in the pedagogical applications of corpora has been subjected to empirical investigation using a variety of research methods (see Farr & Leńko-Szymańska, 2023a, b for a comprehensive account of such studies). Student teachers' confidence in newly acquired corpus-related knowledge and skills, as well as their opinions on the usefulness of corpus-informed and DDL materials and activities for language instruction, have been explored using questionnaires (e.g. Farr & Riordan, 2023). Qualitative methods such as interviews, focus groups, reflective essays, and diaries have also been used for this purpose (e.g. Abdel-Latif, 2021). In order to collect tangible evidence of the technical, corpus linguistic, and pedagogical skills acquired by student teachers, researchers have examined corpus-based tasks or lesson plans developed by participants in corpus-based language pedagogy courses (e.g. Crosthwaite et al., 2023) or observed real lessons (e.g. Xodabande & Nazari, 2022). It should be noted, however, that these studies inevitably involve limited numbers of participants and are heavily conditioned by specific institutional and geographical settings. Consequently, it is difficult to generalise their findings in order to establish a standardised model for incorporating corpora into teacher education that would fit all contexts.

Resources for Teachers

The results of the research on the effectiveness of different training programmes—in particular the opinions and attitudes expressed by pre-service and in-service practitioners—have demonstrated that teacher education, although effective, is unlikely to solve the problem of slow uptake of corpora by language teachers on its own. Among the obstacles frequently mentioned in the questionnaires and interviews is the lack of dedicated resources and tools that would be suitable for both language teachers and learners. Another frequently cited barrier is the time required to create tailor-made corpus-based and DDL materials and activities (Farr & Leńko-Szymańska, 2023a).

Teachers often express the need for easily accessible, stable, and free text collections that are appropriate to the cognitive and linguistic levels of language learners. There are only a handful of pedagogical corpora that contain a selection of less complex samples of spoken and written language on topics relevant to language learners, or simplified speech and texts (e.g. BackBone; Kohn et al., 2009). In addition, teachers require dedicated pedagogical tools for querying corpora that prioritise a clear and intuitive interface over a large number of settings and options (e.g. SKELL; Baisa & Suchomel, 2014). Finally, language teachers and learners would profit from materials and tools that can act as an intermediary between the corpus and the user, providing access to rich information derived from corpora, but in a streamlined and curated way (e.g. ColloCaid; Frankenberg-Garcia et al., 2019; Academic Phrasebank; Morely, 2023). Such solutions can offer the benefits of observing and analysing linguistic data without the impediments of messy and unrelatable real language use, an overabundance of data, and complex and unfriendly tools, all of which can be overwhelming for language teachers and inappropriate for their learners.

Another solution to the problems reported by language teachers can be provided by easily accessible, free, and well-stocked resource banks, which would offer ready-to-use corpus-informed and DDL materials and activities suitable for learners of different ages and proficiency levels. Ideally, the items in such a bank should include detailed instructions on which corpus to use and how to build a relevant query, as well as a step-by-step description of classroom procedures, a student answer sheet to copy, and the key. The resources of this kind recently started to be published online (e.g. Le Foll, 2021; Pinto et al., 2023; The EdUHK Corpus Team, 2024). It is notable that they are frequently the outcome of specific corpus-based language pedagogy courses and contain modules developed by student teachers.

These resources can save teachers the preparation time required to create a corpus activity from scratch, thus encouraging them to use corpus-based resources in their teaching. However, these collections can also serve another purpose. They can showcase a range of ideas for using corpora in the classroom and inspire teachers to adapt them to their own needs and settings. In this way, language instructors can gradually gain the confidence and experience to create their own original corpus-informed and DDL materials and activities. It can be argued that such resources could therefore play a complementary role in educating teachers in pedagogical uses of corpora.

Cross-References

- ▶ Corpus Literacy and Data-Driven Learning
- ▶ Corpus-Informed Materials in Language Teaching
- ▶ Integrating Corpora and Data-Driven Learning into Lesson Planning
- ▶ Pattern-Hunting and Pattern-Defining in Data-Driven Learning
- ▶ Pedagogical Corpora and Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Tools for Data-Driven Learning

References

Abdel-Latif, M. M. M. (2021). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL, 33*(1), 34–48. https://doi.org/10.1017/s0958344020000129

Baisa, V., & Suchomel, V. (2014). Sketch engine for English language learning. In A. Horák & P. Rychlý (Eds.), *Proceedings of recent advances in Slavonic*

natural language processing (pp. 63–70). Lexical Computing Ltd.

Callies, M. (2016). Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. In R. Kreyer, S. Schaub, & A. Gulderning (Eds.), *Angewandte Linguistik in Schule und Hochschule* (pp. 391–414). Peter Lang.

Callies, M. (2019). Integrating corpus literacy into language teacher education. The case of learner corpora. In S. Götz & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245–263). Benjamins. https://benjamins.com/catalog/scl.92.12cal

Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392–1420. https://doi.org/10.1080/09588221.2021.1995001

Farr, F., & Leńko-Szymańska, A. (2023a). Corpora in English language teacher education: Research, integration, and resources. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3281. Online Early View.

Farr, F., & Leńko-Szymańska, A. (2023b). Corpus linguistics in second language teacher education. *Second Language Teacher Education, 2*(2), 117–132. https://doi.org/10.1558/slte.28536

Farr, F., & Riordan, E. (2023). *Re*-evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Second Language Teacher Esucation, 2*(2), 133–160. https://doi.org/10.1558/slte.26146

Frankenberg-Garcia, A., Rees, G., Lew, R., Roberts, J., Sharma, N., & Butcher, P. (2019). ColloCaid: A tool to help academic English writers find the words they need. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouësny (Eds.), *CALL and complexity – Short papers from EUROCALL 2019* (pp. 144–150). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.1000.

Kohn, K., Hoffstaedter, P., & Widmann, J. (2009). *BACKBONE – Pedagogic corpora for content & language integrated learning*. EuroCALL, Universidad Politecnica de Valencia, Spain.

Le Foll, E. (Ed.). (2021). *Creating corpus-informed materials for the English as a foreign language classroom*. Pressbooks. https://pressbooks.pub/elenlefoll/

Le Foll, E. (2023). 'Opening up' Corpus Linguistics: An open education approach to developing corpus literacy among pre-service language teachers. *Second Language Teacher Esucation, 2*(2), 161–186. https://doi.org/10.1558/slte.25371

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL, 26*(2), 260–278. https://doi.org/10.1017/s095834401400010x

Leńko-Szymańska, A. (2017). Training teachers in data driven learning: Tackling the challenge. *Language Learning and Technology, 21*(3), 217–241.

Leńko-Szymańska, A. (2022). Training teachers and learners to use corpora. In E. Csomay & R. Jablonkai (Eds.), *The Routledge handbook of corpora in English language teaching and learning* (pp. 509–524). Routledge.

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning, 35*(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

Ma, Q., Lee, H. T. H., Gao, X. (Andy), & Chai, C. (2023). Learning by design: Enhancing online collaboration in developing pre-service TESOL teachers' TPACK for teaching with corpus technology. *British Journal of Educational Technology.* https://doi.org/10.1111/bjet.13458. Online Early View.

Morely, J. (2023). *Academic Phrasebank*. The University of Manchester. https://www.phrasebank.manchester.ac.uk/

Mukherjee, J. (2004). Bridging the gap between applied Corpus linguistics and the reality of English language teaching in Germany. *Applied Corpus Linguistics*, 239–250. https://doi.org/10.1163/9789004333772_014

Pinto, P. T., Crosthwaite, P., de Carvalho, C. T., Spinelli, F., Serpa, T., Garcia, W., & Ottaiano, A. O. (Eds.). (2023). *Using language data to learn about language: A teachers' guide to classroom corpus use*. The University of Queensland. https://doi.org/10.14264/3bbe92d.

The EdUHK Corpus Team. (2024). *Enhancing corpus literacy and competence for corpus-based language pedagogy for in- and pre-service English teachers*. Pressbook. https://pressbooks.pub/cmtry1/

Xodabande, I., & Nazari, M. (2022). Impacts of a corpus linguistics course on in-service EFL teachers' corpus literacy. *Computer Assisted Language Learning, 23*(1), 318–346.

Q

Qualitative Designs in Data-Driven Learning Research

Seiko Hayashi and J. Elliott Casal
University of Memphis, Memphis, TN, USA

Abstract

This encyclopaedia entry discusses qualitative designs in data-driven learning (DDL) research. After a brief introduction to DDL and qualitative research methods, the authors consider what qualitative designs offer researchers in this area. More specifically, they discuss research designs, methods and topics associated with DDL research on three primary areas: learning attitudes and learning experiences, instructor attitudes and corpus literacy. Recent studies which demonstrate the potential of qualitative perspectives in DDL research are highlighted and discussed as illustrative examples. Overall, qualitative methods and designs have allowed for more longitudinal and introspective insights into learning and teaching language with DDL.

Keywords

Data-driven learning · Corpus linguistics · Research methods · Qualitative research methods · Case studies · Interviews

Introduction

Building on Johns' (1991) pioneering work, language researchers and educators have dedicated considerable attention to developing, implementing and evaluating corpus-based language teaching methodologies, referred to broadly as data-driven learning (DDL). DDL is a language teaching approach that places learners' interaction with language corpora and/or corpus data (directly or indirectly) at the centre, with a range of theoretical motivations (O'Keeffe, 2021) and pedagogical implementations (Cotos et al., 2017).

Importantly, numerous recent meta-analyses and research syntheses have demonstrated that DDL is not just a motivating and novel pedagogy but also that it is effective (e.g. Boulton & Cobb, 2017; Boulton & Vyatkina, 2021; Lee et al., 2019; Perez-Paredes, 2022). With various foci, these studies highlight that DDL is an increasingly popular and increasingly global pedagogy, that empirical research provides strong evidence that DDL leads to enhanced learning outcomes and that the preponderance of such scholarship targets the teaching and learning of L2 writing, vocabulary or lexicogrammatical features. Authors of meta-studies have been careful to note "the many excellent qualitative DDL studies" (Boulton & Cobb, 2017, p. 381) not included by their methods, while other scholars and authors of other review types (e.g. scoping reviews, such as Boulton & Vyatkina, 2021) have noted the power of DDL in

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_90-1

less empirically observable phenomena, such as learner autonomy (Charles, 2023; Flowerdew, 2012; see also chapter ▶ "Learner Autonomy and Data-Driven Learning" by Charles).

This encyclopaedia entry provides a brief outline of the valuable perspectives on DDL that result from qualitative approaches to research. Noting the diversity of qualitative research methods in applied linguistics, Heigham and Croker (2009) broadly define qualitative research as "collecting primarily textual data and examining it using interpretive analysis" (p. 5). Further, Lew et al. (2018) frame qualitative views as concerned not with "regularities" and associations between variables—as in quantitative research perspectives—but rather with events and "processes of changes across time" (p. 80). These perspectives afford important points of insight into developmental processes, experiences and attitudes of both learners and educators involved in DDL activity. It is important to note the differences between the use of qualitative data collection (often used alongside quantitative data collection procedures) and qualitative coding or analysis of data.

Qualitative Research in DDL

Student's Attitudes and Experiences

Much of the qualitative research on DDL has examined students' attitudes towards and evaluation of corpus-based writing instruction, often through longitudinal case studies (e.g. Lee & Swales, 2006) and/or employing qualitative data collection instruments such as questionnaires (oftentimes alongside Likert-scale questions; for example, Casal, 2020), interviews (e.g. Chen & Flowerdew, 2018), written reflections (e.g. Jones & Oakey, 2024) and observation (e.g. Yoon, 2008). Particularly in attitude and perception studies, qualitative data collection instruments (such as open-ended survey items, interviews and reflections) are typically employed alongside quantitative surveys that capture trends and changes in student perspectives. The aim has often been to adopt qualitative methods as a means of capturing rich, personal perspectives and experiences of individual learners in relation to DDL pedagogies and activities over time. Many of these studies present thematic analysis of emerging trends, but there is not consistent reporting of coding procedures (e.g., intracoding/intercoding reliability). Studies have also adopted qualitative designs for analysing learner's development with more varied theoretical and methodological orientations and procedures, as is discussed below.

Charles (2014) and Chen and Flowerdew (2018) exemplify the important role that qualitative data collection instruments play alongside quantitative approaches in learner perception studies on DDL. In terms of DDL implementation, learners in Charles' (2014) project were EAP students who compiled their own discipline-specific corpora for an L2 English research writing course that implemented DDL. While numerical survey results in a previous study (Charles, 2012) indicated that most students provided positive evaluations of DDL for L2 academic writing learning, Charles (2014) distributed a follow-up questionnaire and adopted a case study approach to investigate long-term learner use of their personal corpora. Some trends and generalizations were presented based on survey results, but the close focus on learner responses to open-ended survey items and case studies of students "whose attitude to their personal corpus developed along opposing trajectories" (p. 37) allowed for rich thematic insights into what learners value (or do not value) regarding DDL, how they make use of DDL affordances, the dynamic nature of their relationships with and use of DDL over time and the individuality of such trajectories. Similar approaches have been adopted by Chen and Flowerdew (2018), who tracked their PhD students' corpus use after a workshop through interviews. These two studies demonstrate the value of qualitative methods for collecting and analysing students' perspectives (and stories) regarding DDL, and they represent a broader trend of using qualitative instruments—especially learner surveys—in what individual challenges, needs and preferences shape students' attitudes towards and evaluation of corpus-based instruction.

Qualitative research methods have also been invaluable for analysing DDL and language learning from a developmental perspective, particularly to investigate changes in students' writing experiences, processes and behaviours, through interactions with a corpus. With a focus on lexicogrammatical awareness and collocations, Yoon (2008) utilized a case study design of EAP students, drawing from observational notes, interviews, stimulated recalls, corpus search records, class DDL tasks and written reflections to investigate learner's writing development. Studies such as this employ qualitative designs to examine how DDL can foster student awareness and autonomy. Relatedly, Park (2012) adopted a Vygotskian sociocultural theory perspective in an innovative study to examine student writing behaviour and development in terms of learner's interaction with corpora. Park analysed screen recordings, corpus search queries, oral/written reflections and written essays from three students via case studies. This allowed for a close analysis of learners' interaction with the corpus over time, rather than relying indirectly on reconstruction and retrospection or cross-sectional snapshots.

Qualitative instruments and designs are often used in research on DDL and L2 disciplinary academic writing development. For example, Cotos et al. (2017) analysed research article revisions after DDL activities for evidence of changes in awareness of linguistic conventions and development of rhetorical genre knowledge. Casal (2020) also implemented a qualitative stimulated recall writing conference, prompting students to articulate the motivations behind their revisions and the role that DDL activities may have played in such changes. In this study, Casal also analysed the data through a Vygotskian microgenetic lens, leveraging changes in understandings of target course concepts, decreased reliance on mediational support and innovation in application of course concepts as evidence of development. Taken together, these studies demonstrate that qualitative approaches to assessing development are valuable to DDL scholarship as a means of foregrounding processes in change, the means by which development occurs in DDL activity and the diversity of learner trajectories and learning experiences (see chapter ▶ "Genre Awareness and Data-Driven Learning" by Casal and Kessler).

Qualitative designs and instruments are also employed in DDL studies beyond writing and lexicogrammatical contexts. For example, Jones and Oakey (2024) examine DDL in spoken English grammar instruction. Using a case study approach, the researchers analysed personal diaries that learners kept to reflect on their use of spoken grammar features from a DDL-based pedagogical intervention in conversations outside of the classroom. Such contextualized and reflective data provided insights into what learners notice and how they are (or are not) able to transfer concepts from DDL pedagogy into their own language use, allowing Jones and Oakey to gain a rich "ecological snapshot of the contexts and types of conversations" (p. 10) of the students as a site of learning and consciousness-raising itself. Importantly, it also provided a lens to understand how corpus-based instruction may impact and enhance students' spoken language learning beyond the classroom.

Teacher's Attitudes and Corpus Literacy

In addition to leveraging qualitative methods to investigate learner's attitudes and development in DDL pedagogies, researchers have implemented qualitative designs to investigate teachers' perceptions of DDL activity and their own relevant expertise (see chapter ▶ "CALL Teacher Communities of Practice" by Ma). Research on DDL-based teacher training, including teachers' experience of and knowledge development through the training, is an increasingly important line of inquiry. Many of these studies employ interviews, open-ended survey questions and lesson plan analysis to examine teacher's (or teacher trainee) experiences and perceptions of DDL pedagogies and training programmes (e.g. Leńko-Szymańska, 2014; Zareva, 2017). These studies highlight particular challenges and concerns that educators face when learning about or implementing DDL, but more broadly they show that qualitative data allows us to examine the experiences, emotions and identities that training can shape and provoke in teachers.

Other studies provide insights into the skills and knowledge teachers may need and what support is most suitable to enhance teachers' corpus use in classroom instruction. Corpus literacy (Mukherjee, 2006) and corpus-based language pedagogy (Ma et al., 2022) have surfaced as two important constructs targeted in corpus-based language teacher education programmes, with qualitative designs and instruments playing prominently in the analysis of their effectiveness. Common methods include, once more, introspective data sources including open-ended survey questions and interviews (e.g. Ma et al., 2022), instructor lesson plans themselves (e.g. Crosthwaite et al., 2023) and the combination of these sources (e.g. Chung et al., 2024). In the latter of these studies, Chung et al. are unique for targeting the preparation of teachers for DDL in primary educational contexts. Qualitative approaches were essential for that most participant teachers struggled with aligning DDL with curriculum goals, expanding corpus-based activities beyond microtasks and selecting appropriate corpus data and tools and that there existed a gap between teachers' (self-reported) high corpus literacy and their ability to implement DDL in their language classrooms, in spite of the workshops.

Thus, qualitative approaches to DDL and teacher education have allowed researchers to outline and investigate complex constructs such as teachers' corpus-based language pedagogy skills, develop and assess teacher training programmes and uncover difficulties of successfully integrating appropriate data-driven instruction and activities into classroom language pedagogy. Research on DDL training for language educators is flourishing and diversifying with emerging scholarship considering issues such as reflective practices in DDL (see Farr & Leńko-Szymańska, 2023).

Conclusion

This encyclopaedia entry was concerned with the affordances of qualitative designs for DDL research and the methodological considerations of such scholarship. In general, much research in this area leverages qualitative methods' exploratory power and potential for close examination of processes and complex constructs over time to provide rich, longitudinal perspectives on what learners experience during and think of DDL, what learning looks like with DDL and how language teachers may develop helpful competencies for implementing DDL into their own language teaching. While key studies exist in this area, powerful technologies, such as screen capture, allow for researchers to pursue exciting new investigations into what development looks like during DDL.

Cross-References

- ▶ Constructivist Approaches to Data-Driven Learning
- ▶ Genre Awareness and Data-Driven Learning
- ▶ Learner Autonomy and Data-Driven Learning
- ▶ Metacognition and Data-Driven Learning
- ▶ Narrative Inquiry and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Casal, J. E. (2020). *An integrated corpus and genre analysis approach to writing research and pedagogy: Development of graduate student genre knowledge* [Unpublished doctoral dissertation]. The Pennsylvania State University.

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes, 31*(2), 93–102. https://doi.org/10.1016/j.esp.2011.12.003

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30–40. https://doi.org/10.1016/j.esp.2013.11.004

Charles, M. (2023). Corpora and autonomous language learning. In R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 406–419). Routledge.

Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes, 50*, 97–112. https://doi.org/10.1016/j.esp.2017.11.004

Chung, T., Crosthwaite, P., Cao, H., & deCarvalho, T. (2024). Walking the walk? (Mis)alignment of EFL teachers' self-reported corpus literacy skills and their competence in planning and implementing corpus-based language pedagogy. *TESOL Quarterly, 58*, 1046. https://doi.org/10.1002/tesq.3299

Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology, 21*(3), 104–130. http://hdl.handle.net/10125/44623

Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392–1420. https://doi.org/10.1080/09588221.2021.1995001

Farr, F., & Leńko-Szymańska, A. (2023). Corpora in English language teacher education: Research, integration, and resources. *TESOL Quarterly, 58*, 1181. https://doi.org/10.1002/tesq.3281

Flowerdew, L. (2012). Corpora in the classroom: An applied linguistic perspective. In K. Hyland, C. M. Huat, & M. Handford (Eds.), *Corpus applications in applied linguistics* (pp. 208–224). Continuum.

Heigham, J., & Croker, R. (2009). *Qualitative research in applied linguistics: A practical introduction*. Palgrave Macmillan.

Johns, T. (1991). Should you be persuaded – Two samples of data-driven learning materials. *ELR Journal, 4*, 1–16.

Jones, C., & Oakey, D. (2024). Learners' perceived development of spoken grammar awareness after corpus-informed instruction: An exploration of learner diaries. *TESOL Quarterly, 58*, 1138. https://doi.org/10.1002/tesq.3305

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75. https://doi.org/10.1016/j.esp.2005.02.010

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL, 26*(2), 260–278. https://doi.org/10.1017/S095834401400010X

Lew, S., Yang, A., & Harklau, L. (2018). Qualitative methodology. In A. Phakiti, P. I. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 79–101). Palgrave Macmillan.

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning, 35*(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Peter Lang.

O'Keeffe, A. (2021). Data-driven learning – A call for a broader research gaze. *Language Teaching, 54*(2), 259–272. https://doi.org/10.1017/S0261444820000245

Park, K. (2012). Learner–corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics, 33*(4), 361–385. https://doi.org/10.1093/applin/ams012

Perez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12*(2), 31–48. http://llt.msu.edu/vol12num2/yoon

Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal, 71*(1), 69–79. https://doi.org/10.1093/elt/ccw045

Q

Quantitative Designs in Data-Driven Learning Research

Matt Kessler and Kübra Çekmegeli
University of South Florida, Tampa, FL, USA

Abstract

This encyclopaedia entry introduces readers to quantitative research methods involved in investigations of data-driven learning (DDL). The piece opens by briefly defining two key terms, DDL and quantitative research methods, in addition to the underlying assumptions of quantitative approaches more broadly. Then, readers are introduced to common quantitative designs and samples that have been used in prior DDL research. Next, the piece transitions to highlighting some of the most common targets of inquiry in quantitative DDL designs. Typically, this has consisted of L2 linguistic foci involving writing skills, vocabulary, and lexico-grammar, among others. Finally, the piece closes with a discussion of frequently used instruments such as surveys, tests, student-produced essays, and tracking data, which have been analyzed using both descriptive and inferential statistics.

Keywords

Applied linguistics · Corpus · Data-driven learning · DDL · L2 writing · Research methods · Questionnaires · Surveys · Statistics · Vocabulary

Introduction to DDL and Quantitative Methods

Approximately three decades ago, Tim Johns coined the term data-driven learning (DDL), sparking research interest in its scope and significance. Boulton and Vyatkina (2021) provide a comprehensive definition, emphasizing that DDL consists of the integration of corpus data and corpus linguistics tools into language education, particularly in the teaching of foreign or second languages (L2). This definition highlights DDL's use of digital tools and its evolution into the applied linguistics subfield of Computer-Assisted Language Learning (CALL). Crosthwaite and Boulton (2024) further elaborate on this understanding, advocating for innovation and methodological improvements in DDL research, including examining how DDL can effectively be integrated with pedagogy, and exploring its long-term effects on L2 learning outcomes.

Notably, DDL is a dynamic field that is continuously refining its methods to better serve learners and educators, with distinctions commonly made

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_91-1

between hands-on and hands-off procedures. In this sense, hands-on DDL involves students directly engaging with electronic corpora and tools, while hands-off DDL entails students using printouts or other pre-processed data derived from corpora, typically provided by the teacher, without direct manipulation of electronic tools.

When it comes to researching various facets of DDL, one of the primary approaches taken by researchers involves using quantitative methods. In quantitative designs, statistics and large sample sizes are often used, in which the goal is to employ objective analyses to yield generalizable findings (Fryer et al., 2018). Notably, a quantitative research paradigm operates under the assumption that by dissecting constituent parts, the broader world and its mechanisms can be understood (Loerts et al., 2020). In the context of DDL, quantitative methods enable researchers to assess the effectiveness of DDL interventions and to understand DDL's impact on other areas related to L2 acquisition.

Designs and Samples

When it comes to investigating DDL, multiple systematic reviews of DDL research such as Lusta et al. (2023) have estimated that since the late 1990s, mixed methods that combine quantitative and qualitative designs have been adopted most frequently (used in approximately 41% of studies), followed by quantitative-only designs (approximately 39%), with qualitative-only designs being less common (approximately 20%). Meanwhile, more recent snapshots focusing on the past 5 years have suggested that mixed methods and quantitative-only designs account for nearly 90% of all recently published research on DDL (see Boulton & Vyatkina, 2023). With such studies that adopt quantitative methods, many have tended to employ quasi-experimental designs (i.e. typically classroom-based studies involving students who are enrolled in a language course) in addition to experimental designs (i.e. lab-based studies). The majority of these quantitative studies focus on either a single classroom or one focal group of learners who are engaging in DDL (e.g. Crosthwaite, 2017); however, other studies have used both experimental and control groups as a means of comparing the relative impact of DDL on learners' L2 skills or content knowledge (e.g. Huang, 2014).

In terms of determining the average study duration, scholars have noted this is relatively difficult. This is because some researchers have reported their treatments in terms of days or weeks (of a course), yet others have reported treatments in terms of hours/minutes. For nearly 100 recent studies, however, Boulton and Vyatkina (2023) noted that interventions had a mean duration of 11 h and 9 min. Relatedly, in terms of participant samples, Boulton and Vyatkina have remarked that the large majority of quantitative DDL studies focus on intermediate or advanced L2 English learners within university contexts. These participants are typically recruited from language courses such as English for General/Academic Purposes, with a median sample size of 55 (although it should be noted that there is considerable variation among studies).

Targets of Inquiry

In terms of *what* DDL researchers investigate via the use of quantitative methods, thus far, the primary focus has been on exploring language gains. Typically, explorations of language gains have consisted of examining areas such as L2 learners' development of vocabulary knowledge, lexico-grammar, and academic writing skills (Crosthwaite & Boulton, 2024). Because of the large number of studies involving specific targets such as DDL's influence on vocabulary gains, there have also been meta-analyses conducted which investigate its effectiveness. For example, in Lee et al.'s (2019) meta-analysis, the authors aggregated 29 quantitative studies and found a positive, medium-sized effect of using DDL for vocabulary learning purposes. Despite the overwhelming focus on topics such as vocabulary, lexico-grammar, and writing skills, a number of quantitative studies have also focused on other targets. These include areas such as error correction, pragmatics, and speaking skills (e.g. Geluso & Yamaguchi, 2014).

Another major target of interest for researchers who adopt quantitative methods (and also mixed methods and qualitative methods) has involved investigating students' attitudes and perceptions towards DDL and corpus use. As Lusta et al. (2023) have noted, when adopting quantitative, (quasi)-experimental designs, the aim has often been to understand learners' perceptions of DDL's utility, along with what challenges they face when using corpora for CALL purposes. Some studies have also targeted learners' actual use of corpora, including how they search and navigate within a given corpus. Relatedly, some quantitative studies have targeted teachers' attitudes and perceptions as well. Such studies have often attempted to understand the intersections of teacher training and DDL, including focal areas such as teachers' corpus literacy and the extent to which they plan to integrate DDL into their future classrooms (e.g. Ma et al., 2022).

Instruments and Data Analysis

When it comes to *how* DDL researchers have investigated the aforementioned topics, most studies have used more than one instrument, with a mean of 1.9 instruments (as reported in Boulton & Vyatkina, 2021). Learning gains (e.g. in vocabulary, lexico-grammar, or writing skills) are often measured via the use of tests (e.g. using pretest/posttest designs) or learner productions (e.g. student-produced essays). For other targets of inquiry such as students' and teachers' attitudes and perceptions of DDL, apart from qualitative data, questionnaires/surveys have been the most popular quantitative instrument. In fact, estimates of the use of questionnaires/surveys range from 53% in some syntheses to as high as 67% in other studies (see Boulton & Vyatkina, 2021; Lusta et al., 2023). In many cases, questionnaires/surveys are the only quantitative instrument used within a study. Finally, other instruments have also been used as a means of tracking learners' behaviours when engaging in DDL. Various software and tools have been used for eye-tracking, keystroke-logging, and for collecting corpus query syntax data (e.g. Pérez-Paredes et al., 2011); however, the use of such instruments is relatively less common.

Finally, in terms of data analysis, this also varies widely depending on the aims of the study and the instruments employed by the researchers. However, broadly speaking, approximately one-third of all DDL studies (regardless of design) present basic descriptive statistics in the form of mean percentages or raw numbers (Boulton & Vyatkina, 2023); meanwhile, approximately 58% perform some sort of inferential statistical test(s). Such tests have typically included various types of correlations, *t*-tests, and ANOVAs. A small but growing number of DDL studies have responded to calls for SLA researchers to adopt more advanced analyses used in sister disciplines such as psychology and economics (for more, see Gass et al., 2021). For instance, some scholars have leveraged techniques such as path analysis (e.g. Lee et al., 2020) as well as other types of regression models (e.g. Crosthwaite et al., 2020) for understanding the relationships between different variables and L2 learning outcomes. However, the use of such advanced statistics still tends to be quite rare in quantitative DDL designs.

Conclusion

In summary, quantitative research involving DDL has often involved quasi-experimental or experimental designs, with university-level students who are at intermediate or advanced levels of L2 English proficiency. Researchers have typically targeted the learning of vocabulary, grammar, and other areas such as learners' perceptions of DDL activities. Thus, there is considerable room for innovation in terms of targets of inquiry (see Casal & Kessler, 2024). Although the targets of quantitative inquires have remained relatively consistent through the years, more researchers are now adopting a greater diversity of methods, instruments, and statistical analyses for making sense of their data.

References

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Boulton, A., & Vyatkina, N. (2023). Expanding methodological approaches in DDL research. *TESOL Quarterly.* Advance online publication. https://doi.org/10.1002/tesq.3269

Casal, J. E., & Kessler, M. (2024). New theoretical and practical horizons in data-driven learning: Introduction to the special issue. *TESOL Quarterly.* Advance online publication. https://doi.org/10.1002/tesq.3331

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning, 30*(6), 447–473. https://doi.org/10.1080/09588221.2017.1312462

Crosthwaite, P., & Boulton, A. (2024). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. *HAL Open Science*, 1–17. hal-03506624. https://hal.science/hal-03506624

Crosthwaite, P., Storch, N., & Schweinberger, M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing, 49*, 100729. https://doi.org/10.1016/j.jslw.2020.100729

Fryer, L. K., Larson-Hall, J., & Stewart, J. (2018). Quantitative methodology. In A. Phakiti, P. I. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 55–78). Palgrave Macmillan.

Gass, S., Loewen, S., & Plonsky, L. (2021). Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching, 54*(2), 245–258. https://doi.org/10.1017/S0261444819000430

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL, 26*(2), 225–242. https://doi.org/10.1017/S0958344014000044

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL, 26*(2), 163–183. https://doi.org/10.1017/S0958344014000020

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal, 104*(2), 345–362. https://doi.org/10.1111/modl.12634

Loerts, H., Lowie, W., & Seton, B. (2020). *Essential statistics for applied linguistics: Using R or JASP.* Bloomsbury.

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*, 1–19. https://doi.org/10.1016/j.heliyon.2023.e22731

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning, 35*(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J. M., & Jiménez, P. A. (2011). Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning, 24*(3), 233–253. https://doi.org/10.1080/09588221.2010.539978

C

Corpora as a Reference Resource in Data-Driven Learning

Adrian Jan Zasina
Institute of Czech and Deaf Studies,
Charles University, Prague, Czech Republic

Abstract

A corpus is an electronic database of texts gathered from a variety of disciplines and genres in written and spoken registers for a common purpose to analyse a language. Corpora provide much more information than traditional resources for language teachers and learners. They may be used as a reference resource in data-driven learning (DDL) in a direct way by learners themselves or in an indirect way with a help of a teacher or by using ready-to-use corpus-based materials. Two unquestionably beneficial aspects of corpus consultation are the actuality and authenticity of language data. Contemporary studies continue to explore corpus consultation, with a growing emphasis on user-friendly tools and ready-to-use materials. They have shown that corpus consultation positively influences non-native language production. Recent trends also indicate an increasing interest in corpus consultation for languages other than English. However, a noticeable obstacle is a corpus consultation underuse by language teachers. Future advancements should prioritize the normalization of corpus consultation as a reference resource, enhancing accessibility and efficacy for a broader spectrum of users.

Keywords

Concordance · Corpora · Corpus consultation · Corpus resource · DDL

Introduction

The rapid development of language corpora and corpus linguistics in the 1970s is related to the progress in information technology, specifically computers, which made it possible to collect and process increasingly larger amounts of corpus data. A language corpus is an electronic database of texts gathered for a common purpose to analyse a language. In addition to texts, it typically contains linguistic information on the grammatical categories of lexical units and metainformation (author's name, year of publication, publisher, etc.) about each text included. This enables users to search through thousands of texts using concordancers and other functions such as frequency and collocation lists. As a result, a corpus allows for both qualitative and quantitative analyses of language data (McEnery & Hardie, 2012, pp. 1–2).

The usefulness of corpora in language learning was recognized by linguists in the late 1980s (Leech & Candlin, 1986). However, the very first application of corpus methods appeared already in

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_92-1

the 1970s in chemistry vocabulary acquisition (Butler, 1974). At the beginning of the 1990s, the corpus approach was popularized by Johns (1991) who coined the term data-driven learning (DDL) and defined it as an inductive approach to foreign language learning that uses corpus concordance to provide learners with evidence of authentic language. Learners, based on observation, classify language features and generalize rules about how particular language phenomena work.

Since then, numerous studies focusing on DDL have been published (cf. Boulton & Vyatkina, 2021). Recent papers can be divided into four categories as proposed by Boulton (2017, pp. 484–485): (a) theoretical underpinnings, (b) descriptive papers, (c) surveys and syntheses, (d) empirical evaluations. Empirical evaluations can be further categorized into studies focusing on (d1) learners' attitudes and perceptions of corpus use, (d2) learners' behaviour and processes in corpus use, (d3) effects of corpus use as a learning aid, and (d4) effects of corpus use as a reference resource. The latter includes information on the results of corpus consultation and/or learner reactions and evaluations of corpus data use. Corpus consultation may be conducted directly by learners themselves or indirectly with the help of a teacher or by using ready-to-use corpus-based materials (Chambers, 2007, p. 4).

Supplementary material from Boulton and Vyatkina (2021) attests that the early evidence of corpus use as a reference resource in DDL was captured in Ma's study (1994). The study presents experiences with concordance use in writing by second language learners of English. It highlights the possible obstacles for corpus consultation and comments on learning strategies. As one of the first studies in this domain, it provides ideas for further development of classroom concordancing. Since then, many studies describing corpus consultation have appeared, predominantly focusing on writing skills, particularly academic writing, accuracy in writing, creative writing, and occasionally on reading, interpreting skills, and translation, whereas listening and speaking are often neglected. Other foci have included grammar, lexis, lexicogrammar, vocabulary, spelling, error correction, as well as stylistics, phraseology, pragmatics, and many other areas.

Corpus Consultation in DDL

Despite the broad availability of language corpora in the early 2000s, there had been little evidence of regular consultation of corpora instead of traditional dictionaries or textbooks (Chambers, 2005, p. 121). Therefore, the very first research synthesis (Chambers, 2007) of specific outcomes of corpus use attempted to identify which aspects facilitated the popularization of corpus use as a reference resource since the early 1990s. It showed that the interest in corpus consultation had been growing and that overcoming obstacles related to working with corpora could positively contribute to the language-learning environment in the future. However, none of the studies under examination answered the questions that could facilitate corpus popularization. The main obstacle here was the fact that the only facilitators were teachers/researchers conducting the analysed studies. Chambers (2007, p. 13) also underlined the need to get corpora out of the classroom and provide research directly focusing on independent corpus consultation by learners themselves. Another aspect is that analysed studies presented individual corpus consultation rather than work in pairs or groups. A follow-up discussion paper by Chambers (2019) underlined that using corpora by learners and teachers had not been normalized yet, and the emphasis should lie in giving them appropriate skills for corpus consultation and analysis rather than trying to turn them into corpus linguists.

Specific outcomes of corpus use were later summarized in recent research syntheses (e.g., Boulton, 2017, 2021; Boulton & Cobb, 2017; Boulton & Vyatkina, 2021). These studies have shown that more research has evaluated corpus use for learning rather than reference purposes (Boulton & Cobb, 2017, p. 378). The most comprehensive research synthesis (Boulton & Vyatkina, 2021, pp. 78–79) confirms this statement but also concludes that there is a growing interest in consulting corpus data as a reference tool, accounting for over a quarter of the studies under examination. There is still more focus on English than other languages, and DDL methods are predominantly used in university environments by teachers who are also researchers (Boulton, 2021, pp. 21–24).

Corpora can be used either as a reference resource or for language learning, or for both purposes simultaneously with the boundaries being sometimes unclear. While corpus use as a learning aid involves using corpus data to develop language skills through corpus-based exercises or as authentic material in place of coursebooks and grammars, corpus use as a reference resource is a means of acquiring information about language usage. For instance, corpora can be consulted to: check the context of a word, verify sentence structure, identify the exact difference in meaning between lexemes, check idiomatic expressions, find more examples of lexeme usage, check for grammar and spelling mistakes, examine word valency (e.g. combinations of verbs and prepositions), and search for collocations (O'Sullivan & Chambers, 2006, p. 62). Additionally, corpora are also used to: find an L2 equivalent for a single word, find an L2 equivalent for a string of keywords, verify and confirm the part of speech, find an alternative usage, or find the correct inflected form (Lai & Chen, 2015, p. 363).

Benefits of Corpus Consultation

Two unquestionably beneficial aspects of corpus consultation are the actuality and authenticity of data. Corpora in general provide real-life texts covering a variety of disciplines and genres in written and spoken registers (Römer, 2011). In comparison with dictionaries or grammars, they provide corpus users with instant information on grammatical patterns. They show the natural context of a search word that helps foreign language users better understand a target language, which may differ from their first language or languages they know. Thanks to this, users have the opportunity to follow actual trends and work with language that is up to date. Dealing with corpus data is an inductive approach that develops the individual work of the learner and supports their independent thinking (Zasina, 2022, p. 127). It also increases learners' autonomy, language awareness, and skills that may be used in new, self-initiated areas (Boulton, 2021, pp. 30–31).

However, a noticeable obstacle is the underuse of corpus consultation by language teachers (Vyatkina, 2020, pp. 362–363), which prevents the spread of awareness of the corpus as a reference source in DDL. Vyatkina lists three reasons for this: (1) DDL studies predominantly focus on university students of English as a second or foreign language with intermediate to advanced language proficiency levels; (2) corpora and corpus tools are not always freely available; (3) corpora and their interfaces are not user-friendly for non-specialist users without expertise in corpus linguistics, such as regular language teachers and learners. Fortunately, steps have been taken to reduce these obstacles by accommodating corpus tools for regular users and delivering ready-to-use corpus materials.

Effectiveness of Corpora as a Reference Resource

The effectiveness of corpora as a reference resource has been predominantly explored in the context of written production. Research primarily focuses on comparing learners' outcomes with and without access to corpora for reference, often through pre-test and post-test comparisons. Occasionally, cross-sectional studies compare the use of dictionaries and corpora. Other less frequently employed methods include questionnaires, interviews, and observations.

Earlier empirical studies, such as Boulton (2008), using the pre-test/post-test method, highlighted that corpus consultation was beneficial for learners at any language level, with pattern detection proving more useful than simple rules or instructions. These studies also argued for the need for more empirical research to convince a wider audience of the usefulness of DDL. More recent research (Muftah, 2023) emphasizes the importance of factors such as task type, methodology, language proficiency, training, and the selection of corpora, all of which may influence learners' results when consulting a corpus. It appears that some corpus tools may be more beneficial than others, making it crucial to compare different corpus resources. Additionally, corpus guidance and training are still necessary to optimize corpus exploitation (Wu, 2021).

Direct comparisons between dictionaries and corpora (Lai & Chen, 2015; Yoon, 2016) have shown that learners use these resources for different purposes. While dictionaries are primarily used for searching meanings, corpora are consulted for collocations and grammatical patterns. However, corpus use often involves prior training, leading learners to revert to familiar solutions (Yoon, 2016). Therefore, future developments in corpus consultation require user-friendly tools that facilitate corpus searches.

In general, studies have shown that corpus consultation positively influences non-native language production. Corpora provide much more information than traditional resources such as coursebooks, grammars, or dictionaries (O'Sullivan & Chambers, 2006, p. 65), and they develop language skills more effectively, which would be difficult to achieve through classical methods such as searching for collocations, stylistic differences between single registers or finding the most frequent expressions.

Further Possibilities

Despite the increasing availability of materials and publications that facilitate the use of corpora as a reference resource in DDL, specific guidance for languages other than English remains limited. A promising solution would be the creation of an extensive open-access database of corpus consultation examples that systematically compiles materials for various languages, proficiency levels, and educational stages. Additionally, it is imperative to disseminate information about the relevance and utility of corpora as a reference resource among educators, learners, and educational institutions.

Another significant advancement could involve integrating corpora with artificial intelligence (AI) to automatically generate instructional materials for teachers and learners, thereby enhancing the ease of corpus consultation. AI technology holds the potential to supplant traditional manuals and workbooks, offering customized, instant solutions tailored to individual user needs. The rapid progress of modern technology could fundamentally transform our approach to corpus usage, making it more accessible and efficient through the application of large language models to corpus searches.

Cross-References

▶ AI Chatbots and DDL
▶ Concordancers for DDL
▶ Data-driven and corpus-based CALL
▶ Tools for DDL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Boulton, A. (2008). Looking (for) empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk (Ed.), *Corpus linguistics, computer tools, and applications: State of the art* (pp. 581–598). Peter Lang. https://hal.science/hal-00326983/

Boulton, A. (2017). Research timeline: Corpora in language teaching and learning. *Language Teaching, 50*(4), 483–506.

Boulton, A. (2021). Research in data-driven learning. In P. Pérez-Paredez & M. Geraldine (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 9–34). John Benjamins.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89.

Butler, C. (1974). Examples of course and materials development: German for chemists. *Teaching languages to adults for special purposes. CILT reports and papers*, *11*, 50–53.

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology, 9*(2), 111–125.

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In *Corpora in the foreign language classroom* (pp. 1–16). Brill. https://doi.org/10.1163/9789401203906_002

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching, 52*(4), 460–475. https://doi.org/10.1017/S0261444819000089

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *Classroom Concordancing: ELR Journal, 4*, 1–16.

Lai, S.-L., & Chen, H.-J. H. (2015). Dictionaries vs concordancers: Actual practice of the two different tools in EFL writing. *Computer Assisted Language Learning, 28*(4), 341–363. https://doi.org/10.1080/09588221.2013.839567

Leech, G., & Candlin, C. (1986). *Computers in English language teaching and research: Selected papers from the 1984 Lancaster symposium 'computers in English language education and research'*. Longman.

Ma, B. K. C. (1994). Learning strategies in ESP classroom concordancing: An initial investigation into data-driven learning. In J. Flowerdew & A. Tong (Eds.), *Entering texts* (pp. 197–214). Hong Kong University of Science and Technology Language Centre. https://www.brucema.com/pdfs/USTPAPA3.pdf

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Muftah, M. (2023). Data-driven learning (DDL) activities: Do they truly promote EFL students' writing skills development? *Education and Information Technologies, 28*, 13179–13205. https://doi.org/10.1007/s10639-023-11620-z

O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing, 15*(1), 49–68. https://doi.org/10.1016/j.jslw.2006.01.002

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225. https://doi.org/10.1017/S0267190511000055

Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals, 53*(2), 359–370. https://doi.org/10.1111/flan.12464

Wu, Y. A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology, 25*(2), 192–214.

Yoon, C. (2016). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology, 20*(1), 209–229.

Zasina, A. J. (2022). Designing a corpus workbook for students of Czech as a foreign language. *Studie z Aplikované Lingvistiky – Studies in Applied Linguistics, 13*(2), 125–132.

A

AI Safety in Computer-Assisted Language Learning

Lucas Kohnke
Department of English Language Education, The Education University of Hong Kong, Tai Po, Hong Kong

Abstract

Integrating artificial intelligence (AI) in computer-assisted language learning (CALL) has revolutionized language education. However, the rapid advancement of AI also raises significant safety concerns that must be addressed to ensure responsible and ethical use. This entry explores key AI safety challenges in CALL, including robustness, assurance, and specification. It discusses strategies for addressing these challenges, such as rigorous testing, continuous improvement, responsible deployment practices, content moderation, data privacy protection, and fact-checking. It also emphasizes the importance of ongoing research, stakeholder collaboration, and standardized guidelines to harness AI's potential while mitigating risks and promoting learners' well-being in CALL.

Keywords

AI safety · Computer-assisted language learning (CALL) · Responsible AI · Language education · Ethical AI

Introduction

The integration of large language models (LLMs) has significantly transformed the field of computer-assisted language learning (CALL). These artificial intelligence (AI) systems offer personalised, adaptive learning experiences that accelerate language acquisition and support learners in their language learning journeys (Kohnke et al., 2023). However, the widespread adoption of these powerful tools has also raised concerns about their safety and potential risks (Humphreys et al., 2024). As AI becomes deeply embedded in CALL applications, addressing these safety concerns is crucial to ensure it is used responsibly and ethically.

AI Safety Challenges in CALL

AI safety comprises several critical areas, including robustness, assurance and specification. The subsequent sections will provide more details about each of them.

Robustness

Robustness refers to the ability of an AI system to function safely in unfamiliar situations. For instance, an AI language tutor may encounter colloquialisms or regional dialects not included in its training data. A robust system would recognise these as unfamiliar inputs and adapt its responses accordingly rather than providing

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_95-1

incorrect or irrelevant feedback. In these cases, techniques such as predictive uncertainty estimation enable the AI program to assess its confidence and recognise when it is venturing into unknown territory (Rudner & Toner, 2021). This ability is crucial in CALL, as language learning involves a wide range of contexts, topics and user inputs that may not have been encountered during training.

Assurance

Assurance involves making the operations of the AI system transparent and understandable to human operators. This is crucial in educational settings, where both teachers and students rely on the system's output (Memarian & Doleck, 2023). For example, if an AI-generated language lesson provides unexpected results, the system should offer explanations that educators can easily understand and trust. Approaches to enhance assurance include developing interpretable AI models and providing clear, concise explanations of their decision-making processes. For instance, developers can generate visualisations of the mathematical operations inside a machine learning system and identify the input characteristics that influence its outputs most significantly.

Specification

Proper specification involves defining goals for an AI system that align precisely with the outcomes desired by human operators. Misalignment can lead to behaviours that—while technically correct—are undesirable or harmful. For example, in CALL, an AI system may generate inappropriate or offensive content, promote ineffective learning strategies or fail to adapt to individual learners' needs.

Strategies for Addressing AI Safety

Rigorous Testing and Evaluation

Before AI models are released publicly, rigorous testing and evaluation must be conducted (Miao & Holmes, 2023) to mitigate potential risks and ensure that they are safe for CALL. This process should involve extensive internal testing and engaging external experts to provide feedback and identify potential safety issues (Memarian & Doleck, 2023). Collaborating with researchers and institutions specialising in AI safety is crucial to ensure models are thoroughly vetted before they are deployed in educational settings. This can also help establish best practices, develop standardised evaluation frameworks and promote knowledge sharing across disciplines.

Continuous Improvement

Refining AI models is necessary to enhance their performance and safety. Techniques such as reinforcement learning with human feedback allow models to modify their behaviour based on input from human users (Ouyang et al., 2022). This iterative fine-tuning helps identify and address potential safety concerns (Dai et al., 2023). However, it is crucial to balance the benefits of improvement with the need for caution. The need for caution arises from the potential risks of unintended biases, privacy infringements, and unpredictable behaviours that can emerge as AI systems are updated. Incremental updates and rigorous testing at each stage can ensure that AI systems in CALL remain safe and effective as they evolve.

Responsible Use

Deploying AI in CALL necessitates a cautious and gradual approach. Implementing age restrictions and verification processes to access AI tools can help protect children from potential harm (UNESCO, 2021). Establishing clear guidelines and policies for appropriate AI use in language learning is also crucial to promoting responsible use and preventing misuse (European Commission, 2019). These guidelines should cover data privacy, content moderation and ethical considerations. Working with educational institutions and developers to create tailored safety measures can further enhance safety and efficacy (OpenAI, 2023).

Content Moderation and Filtering

Various mechanisms are essential to prevent generating harmful, hateful or inappropriate content in AI-powered CALL systems. They include filters and detection systems that can identify and

flag potentially abusive or improper material (Anderljung & Hazell, 2023). Natural language processing techniques such as sentiment analysis and keyword filtering can also detect and remove offensive or harmful content (Shanmugavadivel et al., 2022). Collaborating with organisations like Thorn, which specialises in combating child sexual abuse material, can help ensure that AI systems do not inadvertently contribute to its spread.

Privacy and Security

Because AI systems gather so much personal information for training and personalisation, it is critical to ensure that these data remain private and secure. Some key strategies include removing personal information from datasets when feasible and responding promptly to individual requests for data deletion. Techniques such as data anonymisation, differential privacy and secure multi-party computation can help minimise the risk of AI models generating responses containing private individuals' personal information (Zhu et al., 2022). Adhering to data protection regulations, such as the General Data Protection Regulation (GDPR) and the Children's Online Privacy Protection Act (COPPA), is crucial to maintain user trust and ensure compliance.

Factual Accuracy

Ensuring that AI-generated content is accurate is a significant challenge (Hockly, 2023), as LLMs, which form the basis of many AI tools, cannot engage in fact-checking. This can be particularly problematic in educational contexts, where learners rely on the information AI tools provide. One way to address this is to solicit user feedback regarding factual errors, using it to continuously refine the models (OpenAI, 2023). In addition, incorporating knowledge from trusted sources, such as educational databases and experts, can further enhance accuracy.

Ethical Considerations and Future Research

Ethical Considerations

Using AI for language learning raises important ethical considerations that must be addressed (Holmes et al., 2022). Ensuring these systems are fair, transparent and accountable is essential to prevent bias and discrimination (Baker & Hawn, 2022). AI should provide equal opportunities for learners regardless of their background, language proficiency or learning abilities. It should also be transparent about how it makes decisions (Nguyen et al., 2023). Finally, accountability measures, such as regular audits and impact assessments, can help identify and mitigate potential ethical risks associated with AI in CALL (Kazim et al., 2021).

Promoting Responsible AI Use

Guaranteeing that AI is deployed responsibly in education requires ongoing collaboration among researchers, educators and policymakers (Miao & Holmes, 2023). Engaging stakeholders from diverse backgrounds and expertise can ensure that AI systems align with educational goals and values in CALL. In addition, establishing ethics review boards and involving learners, parents and educators can help identify and address potential issues early on.

As AI advances and becomes more integrated into CALL, prioritising ongoing research and collaboration is paramount. Developing standardised guidelines and best practices for using AI in education can help ensure consistent and reliable implementation across different contexts. These guidelines should cover data privacy, content moderation, bias mitigation and transparency (Nguyen et al., 2023). Interdisciplinary approaches that unite experts in AI, education and psychology can foster a more comprehensive understanding of the challenges and opportunities associated with AI in CALL.

Future Research Directions

Future research should consider developing more robust and interpretable AI models, creating standardised frameworks for evaluating AI in education and exploring various approaches to content moderation and fact-checking. Collaboration with stakeholders in language education, including teachers, learners and policymakers, will be essential to ensure that the measures adopted are practical, effective and aligned with the needs and goals of language learning.

Conclusion

AI has the potential to revolutionise CALL by offering personalised and adaptive learning experiences that can accelerate language acquisition (Kohnke et al., 2023). However, its rapid advancement also raises significant concerns that must be addressed to ensure that it is used responsibly and ethically (Dignum, 2019). Addressing these challenges requires rigorous testing, continuous improvement, responsible deployment practices, content moderation, data privacy protection and fact-checking. Ongoing research, collaboration among stakeholders and standardised guidelines and best practices are crucial to ensure that AI can be integrated into CALL safely and effectively. As AI continues to shape the future of language learning, it is essential to harness its full potential, mitigate potential risks and promote the well-being of learners.

Cross-References

- ▶ AI Ethics
- ▶ ChatGPT
- ▶ Educating Teachers About AI
- ▶ Generative AI
- ▶ Generative AI and Teacher Education
- ▶ Language Teacher Education for CALL
- ▶ LLMs
- ▶ Technology standards for language teachers

References

Anderljung, M., & Hazell, J. (2023). Protecting society from AI misuse: When are restrictions on capabilities warranted. arXiv. https://arxiv.org/pdf/2303.09377

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Dai, J., Pan, X., Sun, R., Ji, J., Xu, X., Liu, M., Wan Y., & Yang, Y. (2023). SAFE RLHF: Safe reinforcement learning from human feedback. arXiv. https://arxiv.org/pdf/2310.12773

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way.* Springer.

European Commission. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hockly, N. (2023). Artificial intelligence in English language teaching: The good, the bad and the ugly. *RELC Journal, 54*(2), 445–451. https://doi.org/10.1177/00336882231168504

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T. B., Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Humphreys, D., Koay, A., Desmond, D., & Mealy, E. (2024). AI hype as a cyber security risk: The moral responsibility of implementing generative AI in business. *AI and Ethics, 4*, 791. https://doi.org/10.1007/s43681-024-00443-4

Kazim, E., Denny, D. M. T., & Koshiyama, A. (2021). AI auditing and impact assessment: According to the UK information commissioner's office. *AI and Ethics, 1*, 301–310. https://doi.org/10.1007/s43681-021-00039-2

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers & Education: Artificial Intelligence, 5*, 100152. https://doi.org/10.1016/j.caeai.2023.100152

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Nguyen, A., Ngo, H. H., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*, 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

OpenAI. (2023). Our approach to AI safety. https://openai.com/index/our-approach-to-ai-safety/

Ouyang, L., Wu, J., Jiang, X., Almedia, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv. https://doi.org/10.48550/arXiv.2203.02155

Rudner, T. G. J., & Toner, H. (2021). *Key concept in AI safety: An overview.* Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/key-concepts-in-ai-safety-an-overview/

Shanmugavadivel, K., Sathishkumar, V. E., Raja, S., Lingaiah, B., Neelakandan, S., & Subramanian, M. (2022). Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. *Scientific Reports, 12*, 21557. https://doi.org/10.1038/s41598-022-26092-3

UNESCO. (2021). Policy guidance on AI for children. https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf

Zhu, T., Ye, D., Wang, W., Zhou, W., & Yu, P. S. (2022). More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Transactions on Knowledge and Data Engineering, 34*(6), 2824–2843. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9158374

A

Artificial Intelligence (AI) and DGBLLT

Daniel H. Dixon and Yiwen Zheng
Department of Applied Linguistics & ESL, Georgia State University, Atlanta, GA, USA

Abstract

Digital games have relied on forms of artificial intelligence (AI) since the industry began in the 1970s. The AI used in early games, like those made for the Atari 2600, primarily drew on rule-based scripts to control the behaviour of its nonplayer characters (NPCs) (Yannakakis and Togelius, Artificial intelligence and games. Springer, 2018). The AI technology of today, however, can beat all 57 original Atari games 'with performance above the human benchmark' (Badia, Agent57: outperforming the Atari human benchmark. In: Proceedings of the 37th International Conference on Machine Learning, vol 119, pp 507–517. Available from https://proceedings.mlr.press/v119/badia20a.html, 2020, p. 8). While AI has a long history with digital games designed for entertainment, the use of AI in educational apps or games designed for second-language (L2) learning is garnering great interest from educators, researchers and language learners. This interest has been driven by the now widely available large language models (LLMs) that power generative AI platforms like ChatGPT (Open AI, *ChatGPT* [Large language model application]. https://chat.openai.com/, 2022) and Claude (Anthropic, *Claude* [Large language model]. https://claude.ai/, 2023), among others. Generative AI can process and generate language that is remarkably similar to human language, leading to exciting new areas of investigation in the domain of Digital Game Based Language Learning and Teaching (DGBLLT). In this entry, the evolution of AI in digital games is summarized as well as the technology driving generative AI. Findings and implications from this new area of research that investigates its use in the DGBLLT domain are also reviewed. This review is followed by a discussion of the ethics and risks related to generative AI that stem from the inherent bias in the language data with which the LLMs are trained. We conclude with direction for future research on incorporating generative AI into L2 classrooms and DGBLLT contexts.

Keywords

AI · Generative AI · LLMs · Digital games · DGBLLT · AI for education · L2 education

Introduction

Artificial intelligence (AI) is a term that has become ubiquitous in the modern digital era, especially since the public release of *ChatGPT*

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_106-1

(OpenAI, 2022) in late 2022. Since then, the term AI seems to get attached to anything related to technology; from vacuum robots and washing machines to investment platforms and language learning apps, AI seems to be omnipresent. Given this ubiquity, Musiol (2024) argues that the term AI is consistently misused and oversimplified in the mainstream media, which makes gaining a clear understanding of the technology difficult. Musiol suggests that misunderstandings can be ameliorated through 'consistent and clear language' (p. 1) and offers two broad classes to describe the types of AI: *generative AI* and *discriminative AI*. One of the main tasks of discriminative AI centres around classification using algorithms to discriminate among predefined groups. A few examples include email spam filtering, facial recognition and natural language processing (NLP) applications that classify words into part-of-speech categories (e.g. *spaCy*; Honnibal & Johnson, 2015) or rate the sentiment of texts. Besides classification, other tasks used in discriminative AI include regression, clustering, dimensionality reduction and reinforcement learning (see Musiol, 2024). While discriminative AI is used to classify and identify patterns, generative AI learns 'patterns and characteristics of given datasets [to] create new data samples that are similar to the original data' (Musiol, 2024, p. 14).

Generative AI

Generative AI, in contrast to discriminative AI, can process and generate language that is remarkably similar to human language, which has garnered both excitement and concern in higher education and industry alike. While generative AI can produce nonlanguage-related content like videos, images and code, large language models (LLMs) have been developed to understand and generate language that 'captures the complexities and nuances of human language' (Ozdemir, 2023, p. 4). As Ozdemir (2023) explains, 'language modeling is a subfield of NLP that involves the creation of statistical/deep learning models for predicting the likelihood of a sequence of tokens in a specified vocabulary' (p. 4). In general, language modelling tasks fall into two categories: autoencoding and autoregressive. Autoencoding models are used to fill in missing words from any phrase while autoregressive models, like GPT, are used to generate the next most likely words in any phrase (Ozdemir, 2023). It is important to note that the type of AI that has traditionally been used in digital games is characterized as rule-based AI and not generative AI; however, the evolution of AI does have roots in digital games.

The Evolution of AI in Digital Games

Although sharing few characteristics with the modern generative AI of the mid-2020s, rule-based and other forms of AI have been used widely in digital games since the industry began in the 1970s (Jusoh & Al Fawareh, 2023; Yannakakis & Togelius, 2018). In these early days of the industry, Yannakakis and Togelius (2018) report that the AI in digital games primarily relied on simple rules to script the behaviour of nonplayer characters (NPCs). Creating autonomous NPCs draws on 'agent-based AI' in which characters 'take in information from the game data [to] determine what actions to take' (Millington & Funge, 2009, p. 11). Millington and Funge (2009) point to the ghosts in *Pac-Man* (Midway Games West, Inc., 1979) as an early example of 'a state machine' that allowed the ghosts to seemingly conspire against the player (p. 7). Depending on their state, the ghosts would chase or runaway from Pac-Man or randomly select paths to travel.

Decades later, games like *Metal Gear Solid* (Konami Corporation, 1998) incorporated more advanced AI techniques by adopting 'sense simulation'. This allowed NPCs to 'see their colleagues and notice if they were killed' (Millington & Funge, 2009, p. 8), creating more realistic and responsive game environments. Later, an AI technique known as 'reinforcement learning' was first used in a game titled *Black and White* (EA, 2000) where a creature 'learns through positive rewards and penalties' (Yannakakis & Togelius, 2018, p. 12). Other AI techniques relevant to gaming include ad hoc authoring, tree

search, evolutionary computation, supervised learning and unsupervised learning (see Yannakakis & Togelius, 2018).

Although generative AI in commercial digital games is still quite rare as of this writing, the development of the hardware and software that makes generative AI possible has been attributed to the success and innovation from the digital game industry. Nvidia is one of the main producers of the hardware and chips on which generative AI relies. As reported by Lieberman and McDonald (2024), since being founded in 1993, Nvidia's CEO, Jensen Huang, originally developed the architecture of its chips to improve the graphics of modern digital games. Although initially failing at designing a new faster chip and almost bankrupting the company in 1996, Nvidia was eventually successful in developing a new GPU which vastly improved the graphics quality of games. In 2016, Nvidia sold one of its new supercomputers with its improved chip architecture to OpenAI, the developers of ChatGPT. The architecture of Nvidia's new chips, although original motivated by Jensen's desire to improve game graphics, was found to be 'perfect for AI' by deep learning researchers (Lieberman & McDonald, 2024).

Generative AI in Digital Games

As discussed above, AI in games is not a new concept, although the application of generative AI in games is currently very limited. With the advancement of generative AI, some commercial entertainment games have tried to incorporate generative AI into the development and improvement of games, such as developing game mechanics or rules (Werning, 2024), improving sensor-based interfaces for music games (Diaz et al., 2019), creating audio effects (Marrinan et al., 2024) and predicting player's identification and their next move, which would facilitate player analysis and improve player engagement and game security (Fields & Licato, 2023), among others. More specifically, it has been found that the Five-Dollar Model, which is a lightweight text-to-image generative AI model, is a powerful tool for generating maps in retro role-playing games (RPGs) like *Pokémon* (Merino et al., 2023). The AI model was able to process text input and generate appropriate maps based on a set of criteria and requirements. Nevertheless, Merino et al. found that the model behaved poorly when it came to generating emojis compared to generating maps. Apart from generating maps, Buongiorno et al. (2024) found that Procedural Artificial Narrative using Generative AI (PANGeA) can generate searchable data from the actions of players and NPCs. PANGeA can use the data to predict player actions and generate general descriptions of the game environment. Merino et al. report that PANGeA can remain consistent when interacting with free-form input from players, demonstrating a new approach in using large language models (LLM) for generating content in turn-based RPGs and more generally digital games. Additionally, Sun et al. (2023) demonstrated how the GPT-4 LLM model (Open AI, 2023) can be used as a cocreator in a storytelling game.

Besides the aforementioned research and games, generative AI has also demonstrated a great deal of potential in educational games designed for L2 learning in recent years. Several commercial educational games have incorporated AI to enhance their functionality and improve players' engagement and experiences, such as *Duolingo Max* (Duolingo, 2023), *Newcomer* (Kappes, 2023) and *the AI Dungeon* (Latitude, Inc., n.d.). Handini et al. (2022) investigated how the Duolingo application has utilized AI to improve English learning, especially listening and speaking. They found that learning is more effective when using the Duolingo application than without it. *Duolingo MAX*, which requires a paid subscription, incorporates the GPT-4 model to provide two new functions to learners: Explain My Answer and Roleplay (Duolingo Team, 2023). By subscribing to this plan, learners can get more information on their responses to Duolingo and interact with the app in multiple scenarios. For instance, when learners made a mistake in learning, they can use the 'Explain My Answer' function to ask for a detailed explanation on their mistakes and more examples to learn from. The Roleplay function allows learners

to practise having real-world conversations with Duolingo's NPCs.

DGBLLT Research into Generative AI

The area of generative AI and DGBLLT is nascent, and published research is currently limited. Research that has been published tends to focus on using generative AI to create interactive, adaptive and personalized experiences for L2 learners. For instance, Chien et al. (2024) designed an online role-playing game using *ChatPDF* (ChatPDF, n.d.), a chatbot platform that utilizes the GPT API to generate answers based on a PDF file provided by the user. In this game, players can practise their basic scientific knowledge by interacting with *ChatPDF*, designed into the game's story as an inserted memory chip of a character. During this process, players can also develop AI prompting skills to extract key clues from the chip. The game employs two scaffolding mechanics: instant feedback, cognitive scaffolding and strategic scaffolding. The former is a two-dimensional mechanic that allows players to 'obtain scientific knowledge and task clues through questions and answers with AI' (p. 6). The latter provides hints and suggestions for developing AI prompting strategies and spatial mission exploration tactics. Chien et al. report that learners reached a state of high flow and low task anxiety while playing the RPG powered by *ChatPDF*. Their findings suggest that generative AI in edutainment games has the potential to be a helpful study agent in and out of the classroom to provide instant feedback and interactive scaffolding experiences to learners.

Similar to Chien et al., Tang et al. (2022) report on the use of generative AI to create an intelligent tutoring system for L2 learners. The researchers combined generative AI with an existing narrative game *Gridlock* (Tang et al., 2020). In this study, learners have to repair a traffic light, and AI is responsible for providing feedback and hints on their operation based on their performance. Tang et al. provide insight on how a personalized instruction and need-aware gaming (PING) system can be integrated into existing games with AI to facilitate learning. More recently, Verhelst et al. (2024) designed a chatbot game, with content completely generated by AI, to interact with students in an L2. In this game, students listen to a sentence that contains the target word, and then they choose the picture that correctly describes the sentence from a list of five pictures. Both the sentence and the images are generated by AI. The researchers found a significant learning effect of the game despite a potential ceiling effect in the posttest. However, there were no significant differences between the control and experimental group in this study, which might indicate that the AI-based game may require further development to increase its L2 learning efficacy.

Despite the limited number of studies on digital games using generative AI for L2 learning, recently published qualitative research has explored the potential of this novel technology. Seow (2023) demonstrates how generative AI can create immersive language environments that allow learners to interact freely with characters in realistic scenarios. This approach contrasts sharply with traditional games, where learners are often confined to predesigned environments and missions. The key advantage of generative AI lies in its ability to create highly individualized experiences for learners, setting it apart from more traditional forms of AI. This level of adaptability and creativity enables generative AI to be integrated into various aspects of game design, potentially revolutionizing language learning experiences. For instance, generative AI could dynamically adjust dialogue complexity, create personalized quests based on a learner's proficiency level or generate culturally relevant content on the fly. As research in this field progresses, a significant shift in how language learning games are designed and implemented may be seen, offering more engaging, effective and tailored experiences for language learners (Godwin-Jones, 2024).

Bias in Generative AI and Potential Consequences in DGBLLT

While generative AI offers promising applications in DGBLLT, it is crucial to consider the potential

biases inherent in these systems and their implications for DGBLLT contexts. Generative AI, including LLMs, often carries biases that emerge from its training data (Baker & Hawn, 2022; Pack & Maloney, 2023). For example, Caliskan et al. (2017) demonstrated that LLMs can replicate gender and racial biases that exist in their training data, capturing these biases through word embeddings without explicit cultural context. As a result, AI-generated game characters might reflect gender, racial or cultural stereotypes, potentially reinforcing biased perceptions among language learners. In fact, Ferrara (2023) pointed out that the predominance of Western materials in current LLM training data suggests that DGBLLT applications using generative AI are likely to reflect inherited Western biases related to gender, occupation, race and other factors. Additionally, the language produced by AI characters could contain biases in vocabulary choice, sentence structure or idiomatic expressions, inadvertently exposing learners to biased language patterns. Creely (2024) emphasizes that while AI can provide adaptive and personalized learning experiences, it is critical to be aware of the constraints it imposes on language education. These findings necessitate a more cautious approach towards incorporating generative AI in language education games, and practitioners should be careful with the engineering of prompts given to AI platforms (see also Critical DGBLLT and Inclusive DGBLLT).

Direction for Future Research in DGBLLT with Generative AI

Research and applications of generative AI in DGBLLT contexts remain largely unexplored, providing exciting opportunities for innovation in this domain. Future research can benefit from collaboration across disciplines, particularly between applied linguistics and computer science. Applied linguists can draw on expertise in SLA theory, pedagogy, task-based language teaching and of course DGBLLT frameworks to work with computer scientists to develop generative AI games, platforms and applications tailored for L2 learners. Although future research is needed to verify, the nascent research in this domain suggests that generative AI can reliably produce language that uses the linguistic features that are characteristic of a targeted register (i.e. a situation of language use). That is, through careful prompt engineering, LLMs like *GPT-4o* (OpenAI, 2024) can be instructed to generate language using linguistic features frequent in registers like everyday conversations, university applications or academic writing, to name a few (see also Game Mechanics and Design Features for DGBLLT). With the AI acting as an NPC, prompts can be further refined to simulate specific scenarios, tasks or 'quests' like requesting an extension from a professor, ordering food or coffee or even discussing artwork as GPT-4o is capable of generating language based on an image or any other type of digital file. In the language learning RPG *Newcomer* (Kappes, 2023), some of the dialogue of NPCs is generated by GPT, allowing the player to have open-ended conversations with the game's characters. Although the L2 learning efficacy of *Newcomer* has yet to be empirically investigated, the game provides an innovative example of generative AI applied to DGBLLT contexts. Given the popularity of generative AI across industry, disciplines and research domains, future innovation is certainly imminent, and DGBLLT researchers and practitioners are well positioned to positively contribute to this innovation and help determine the role of generative AI in higher education.

Cross-References

- ▶ Critical DGBLLT
- ▶ Game Mechanics and Desing Features for DGBLLT
- ▶ Inclusive DGBLLT
- ▶ Interaction and DGBLLT
- ▶ Research Practices in DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anthropic. (2023). *Claude* [Large language model]. https://claude.ai/

Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, Z. D., & Blundell, C. (2020). Agent57: Outperforming the Atari human benchmark. In *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 507–517). Available from https://proceedings.mlr.press/v119/badia20a.html

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*, 1–41.

Buongiorno, S., Klinkert, L. J., Chawla, T., Zhuang, Z., & Clark, C. (2024). PANGeA: Procedural artificial narrative using generative AI for turn-based video games. *arXiv preprint arXiv:2404.19721*.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186.

ChatPDF. (n.d.). *ChatPDF: Transform your PDFs into interactive chatbots* [Software]. https://www.chatpdf.com

Chien, C. C., Chan, H. Y., & Hou, H. T. (2024). Learning by playing with generative AI: Design and evaluation of a role-playing educational game with generative AI as scaffolding for instant feedback interaction. *Journal of Research on Technology in Education*, 1–20. Published online: 08 Apr 2024

Creely, E. (2024). Exploring the role of generative AI in enhancing language learning: Opportunities and challenges. *International Journal of Changes in Education, 1*, 158.

Diaz, C. G., Perry, P., & Fiebrink, R. (2019, August). Interactive machine learning for more expressive game interactions. In *2019 IEEE Conference on Games (CoG)* (pp. 1–2). IEEE.

Duolingo Team. (2023, March 14). *Introducing Duolingo Max, a learning experience powered by GPT-4*. Duolingo. https://blog.duolingo.com/duolingo-max/

Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. SSRN Preprint. https://doi.org/10.2139/ssrn.4615421.

Fields, L., & Licato, J. (2023). Player identification and next-move prediction for collectible card games with imperfect information. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 19, No. 1, pp. 43–52). https://doi.org/10.1609/aiide.v19i1.27500.

Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *arXiv preprint arXiv:2403.20216*.

Handini, B. S., Nurhasanah, N., & Panly, F. I. (2022). The effect of artificial intelligent technology used (Duolingo application) to enhance English learning. *ELLITE: Journal of English Language, Literature, and Teaching, 7*(2), 86–94.

Honnibal, M., & Johnson, M. (2015, September 19–21) An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1373–1378). Association for Computational Linguistics

Jusoh, S., & Al Fawareh, H. (2023). AI games and algorithms: An overview of categories. In *15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, 2023*. https://doi.org/10.1109/ECAI58194.2023.10194176

Kappes, J. (2023). *Newcomer: A language learning RPG*. [digital game]. Jason Kappes.

Latitude, Inc. (n.d.). *AI Dungeon*. https://aidungeon.com

Lieberman, M. & McDonald, C. (Producers). (2024, April 28). Children of War; Nvidia; Crisis at Pearl Harbor (Season 56, Episode 30) [TV series episode]. In B. Owens (Executive Producer), *60 minutes.* CBS.

Marrinan, T., Akram, P., Gurmessa, O., & Shishkin, A. (2024). Leveraging AI to generate audio for user-generated content in video games. *arXiv preprint arXiv:2404.17018*.

Merino, T., Negri, R., Rajesh, D., Charity, M., & Togelius, J. (2023). The five-dollar model: Generating game maps and sprites from sentence embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 19, No. 1, pp. 107–115).

Millington, I., & Funge, J. (2009). *Artificial intelligence for games* (2nd ed.). Elsevier.

Musiol, M. (2024). *Generative AI: Navigating the course to the artificial general intelligence future*. Wiley.

OpenAI. (2022). *ChatGPT* [Large language model application]. https://chat.openai.com/

OpenAI. (2023). *GPT-4* [Large language model]. https://www.openai.com

OpenAI. (2024). *GPT-4o* [Large language model]. https://www.openai.com

Ozdemir, S. (2023). *Quick start guide to large language models: Strategies and best practices for using ChatGPT and other LLMs*. Pearson.

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly, 57*(4), 1571–1582.

Seow, O. (2023, October). LingoLand: An AI-assisted immersive game for language learning. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–3).

Sun, Y., Li, Z., Fang, K., Lee, C. H., & Asadipour, A. (2023, October). Language as reality: A co-creative storytelling game experience in 1001 nights using generative AI. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 19, No. 1, pp. 425–434).

Tang, Y., Liang, J., Hare, R., & Wang, F.-Y. (2020). A personalized learning system for parallel intelligent

education. *IEEE Transactions on Computational Social Systems, 7*(2), 352–361.

Tang, Y., Hare, R., & Ferguson, S. (2022, October). Classroom evaluation of a gamified adaptive tutoring system. In *2022 IEEE Frontiers in Education Conference (FIE)* (pp. 1–5). IEEE.

Verhelst, E., Janssens, R., Demeester, T., & Belpaeme, T. (2024, March). Adaptive second language tutoring using generative AI and a social robot. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 1080–1084).

Werning, S. (2024). Generative AI and the technological imaginary of game design. In *Creative tools and the softwarization of cultural production* (pp. 67–90). Springer Nature Switzerland.

Yannakakis, G. N., & Togelius, J. (2018). *Artificial intelligence and games*. Springer.

G

Game-Related Social Practices and LLT

Boris Vazquez-Calvo
Department of Language Education, University of Málaga, Málaga, Andalucía, Spain

Abstract

Game-related social practices in language learning involve activities, interactions and language use surrounding digital games, extending beyond direct gameplay into social contexts. This entry explores how these practices enhance linguistic and other skills essential for twenty-first-century language learning and teaching. It examines activities such as participation in social media, content creation like fan translation or modding, community engagement and the development of digital literacies. Pedagogical implications are discussed, suggesting the integration of these practices into formal education to create engaging and effective language learning environments. This includes connecting with relevant pedagogical frameworks and addressing challenges such as curriculum alignment and technological access. Future research directions propose using a wider range of qualitative methods to further understand the impact of game-related social practices on language learning and teaching from an emic perspective. Overall, the entry underscores the value of leveraging game-related social practices in language learning and teaching to capitalize on learners' interests, motivations and identities.

Keywords

Beyond gaming · Community engagement · Content creation · Digital gaming · Digital literacies · Fan translation · Modding · Language learning · Social media

Introduction

Learning through play and affinity is essential in contemporary language teaching. Informally, playing and connecting with affinities keeps learners engaged and motivated. In classrooms, these activities help teachers meet learners' needs and interests. Both in classroom contexts and outside, games provide immersive environments for practising language skills, enhancing vocabulary proficiency (Soyoof et al., 2024; Sundqvist, 2019; see also ▶ "Vocabulary and DGBLLT"), pragmatic skills (Taguchi, 2023; see also ▶ "Pragmatics and DGBLLT") and overall communicative competence (Godwin-Jones, 2014). Games foster language learning not only through human-machine interaction but also through the social participation they encourage (Reinhardt, 2019; Sykes, 2019).

Overall, digital games represent language-intensive opportunities, offering meaningful,

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_124-1

contextualized input. The social dimensions of gaming help form additional language learner and gamer identities (see also ▶ "Identity and DGBLLT"), build communities and create affinity spaces, afforded by deeper exploration of texts, discourses and practices beyond actual gameplay. Providing rich, informal language learning opportunities and serving as potential inspiration for language teaching (see also ▶ "Pedagogical Approaches to DGBLLT"), an operational denomination encompassing these activities is 'game-related social practices'.

Definition

According to Reinhardt (2019), specific subfields of digital gaming and their impact on language learning and teaching include game-enhanced learning, where learning is supported and enhanced by digital games; game-based learning, with games intentionally designed for learning; game-informed learning, where learning and teaching are guided by game design principles; and game-attendant discourses and practices, which mostly refer to game-related social practices.

Game-related social practices thus encompass the activities, interactions and language use that occur around digital games and beyond direct gameplay on a social dimension. Sykes (2019) explains that digital games foster diverse discourse patterns in forums and online communities, where players discuss strategies, resolve disagreements and offer help, thereby expanding their language skills and understanding of cultural discourse norms. As an example of how digital games influence broader societal contexts, Sykes notes how gaming culture may lead to the creation of novel words like 'fortnitis' for school absences due to Fortnite. Reinhardt (2019) emphasizes that game-attendant discourses and practices include interactions with game texts, emergent gameplay discourses and *broader sociocultural practices*. These paratextual, sociocultural practices, which arise from player interactions and the wider gaming culture, encompass social media engagement, gaming forums, fan content creation and community interactions and are the focus of this entry. Both authors agree on the value of these game-related social practices for informal and incidental language learning and the development of digital literacies (see also ▶ "Gaming Literacies and LLT").

Game-Related Social Practices, Case Studies and Examples

Game-related social practices transcend the gameplay experience and yet are central to the learning games can provide. They offer a dynamic environment where language is learned and practised in socially and culturally rich contexts. For instance, Thorne et al. (2012) highlight the complexity of both in-game and out-game discourses, concluding that games like *World of Warcraft* and related gamer-created texts about the game, such as forums about the game, represent 'a diverse and linguistically complex social-semiotic environment for L2 learners of English' and presumably other languages (Thorne et al., 2012, p. 298). This complexity makes game-related social practices valuable for informal and formal language learning and teaching. These practices include participation in social networking sites, content creation, community engagement and the development of digital literacies.

Participation in Social Networking Sites

Social networking sites dedicated to gaming, such as forums, Discord servers or Twitch channels and other social media interest groups, serve as pulsating spaces for linguistic exchange. These platforms allow players worldwide to engage in discussions, share experiences and negotiate meaning, offering opportunities for cultural understanding and language practice in authentic contexts. For instance, Brevik's study (2019) found that among Norwegian teenagers, gamers were the only group actively engaging in offline novels compared to less engaged social media users and online surfers. This suggests that digital gaming uniquely promotes sophisticated reading, understanding of narratives and extended social interaction.

Reinhardt (2019) notes that social media engagement facilitates meaningful participation,

enabling players to develop professional skills, communicative competence or intercultural awareness. In this light, some studies emphasize the role of active social media engagement in enhancing socialization and creativity among teenagers. For example, Vazquez-Calvo et al. (2020) explored how Jova, Sema and Marse, three plurilingual teenagers from Catalonia, extended their socialization to YouTube for fun and social acceptance. They produced, edited, commented on and posted their own Minecraft gameplays and composed short movies, acting as their favourite gaming characters, which later shared online. By skilfully building their online personas, they gained social capital within their local contexts, including among their families and school peers, who viewed them as creative individuals. Unfortunately, they also remarked that no recognition was given by teachers or the school context, generally, for their developed digital and gaming literacies (see also ► "Gaming Literacies and LLT").

Content Creation

Creation of fan content is a key aspect of game-related practices, reflecting fans' deep connections with popular figures or media. Fans use digital tools and communication technologies to 'discuss, share, create, or otherwise respond to a public performance or figure including, for example, music and musicians, literature, sports and athletes, theatre, television, and film and performers' (Sauro, 2017, p. 132). Reinhardt (2019) explains that content creation promotes creative and purposeful language use, enhancing reading, writing and collaborative discussion skills. These activities provide meaningful opportunities for language development.

In gaming, fan literacy practices include writing game-inspired fanfiction, translating video games, producing game guides or tutorials, creating video-based gameplays (video gameplays, often commented orally by the player[s] and/or spectators) and walkthroughs (video guides, often commented orally by the player[s]) and developing mods, which involve bug fixing, altering the game or adding content (Sotamaa, 2022). All these activities require extensive language skills. To illustrate, fan translation of games, typically from L2 to L1, involves advanced L2 reading comprehension, sociocultural and interlinguistic mediation and proficient L1 writing skills.

Fan translators must accurately capture linguistic nuances to ensure the translated game retains the original meaning and intent, which is particularly challenging with humour and phraseology, requiring high levels of interlinguistic creativity (Chik, 2014; Vazquez-Calvo et al., 2019; Vazquez-Calvo, 2021). For instance, Chik (2014) describes the case of Jing, a Beijing gamer who joined a team of amateur translators to create Chinese versions of popular games. Jing's task was to translate in-game dialogues from English to Chinese, which were then proofread by more experienced fan translators acting as chief editors. Despite facing challenges and occasional criticism regarding the quality of translations, this activity validated Jing's English proficiency and provided valuable resources to the gaming community.

Fan translation extends identities beyond those of 'player' or 'gamer.' Fan translators often work to translate games into minoritized languages, such as Catalan, for those languages to gain social recognition and address a lack of commercial interest (Vazquez-Calvo & Thorne, 2022). This practice, along with other game-inspired content creations, reactivates intrinsic motivation and self-directed learning among adult language learners. For example, Shintaku (2023) describes how Calvin, who modded his PlayStation as a child to play Japan-distributed games in English, revisited these games in their original Japanese to learn the language. Similarly, Selo is an adult gamer who translates games from English to Spanish out of nostalgia, reliving his affinity identity as a retro gamer (Vazquez-Calvo, 2018). Such activities allow learners to explore various genres, fostering ownership and agency in their informal and rather incidental language learning experiences. With rekindled motivation for language learning, benefiting both younger and older learners, the primary intent of gamers who engage in content creation is to connect with broader gaming communities, which potentially leads to increased interaction and language learning opportunities.

Community Engagement

Highly engaged gamers create content not just for their own use but to share with others, reflecting their membership and appreciation for online affinity spaces (Gee, 2005). These spaces, characterized by shared interests and goals, become platforms for further opportunities in group discussions and community engagement. By generating content, gamers also facilitate deeper connections and interactions within these communities and adopt specific roles and identities, often calling forth diverse degrees of expertise in knowledge within the communities.

By adopting specific roles and personas, players practise language in various social contexts. Reinhardt (2019) notes that such engagement offers authentic language learning environments, mirroring real-life interactions. Through community participation, players navigate social norms, develop negotiation skills and build lasting relationships, enhancing their language development. Gamer identity is fluid, ranging from casual players to highly engaged 'gamers,' influenced by factors like play frequency and online social gaming involvement (Stone, 2019; see also ▶ "Identity and DGBLLT"). Vazquez-Calvo et al. (2020) classify gamer identities based on their engagement, such as sporadic watchers who occasionally observe gameplays, active followers who regularly subscribe to content, commenters who engage in online discussions and beta testers who participate in testing and publicizing games before commercialization. These varied, often complementary roles enrich the gaming experience and provide valuable language practice and skill development.

Opportunities to adopt roles and identities in gaming communities often stem from specific game-related practices. For example, when a group of gamers undertakes fan translations, they typically form a hierarchical organization with specialized roles, such as group coordinator, senior or novice translator, proofreader or group applicant. Vazquez-Calvo (2021) observed that, in order to join a group of Catalan fan translators, applicants must pass rigorous translation tests reviewed by current members. These revisions occur in group chats with complex metalinguistic discourse, focusing on L2-L1 interlinguistic mediation against the group's linguistic and stylistic standards which attempt to revitalize Catalan as a language for cultural production. This structured, collaborative approach not only improves the translation process but also provides a source for language practice, especially metalinguistic reflection.

Contrary to the traditional perception of gamers as socially inept, professionally unsuccessful or idle and unmotivated (Kowert et al., 2014), the studies above highlight that community engagement and socialization within interest-driven affinity spaces are frequent, discursively sophisticated and conducive to complex language exposure (Thorne et al., 2012). These interactions facilitate situated language learning, underscoring the potential of gaming communities to support and scaffold meaningful social as well as language and digital literacy development.

Digital Literacy Development

Engaging in game-related social practices cultivates digital literacies, such as successfully navigating digital platforms, producing complex multimodal texts and participating in online communities with audience awareness. For example, gamers who mod or translate games face challenges like maintaining format consistency, addressing graphic design issues and adding characters for new languages, such as graphical accents or letters not contained in the original language alphabet, often English (e.g. *á*, *à*, *ñ*) (Vazquez-Calvo, 2018). Concomitantly, due to the language-intensive nature of these activities, they also develop high digital literacy levels using tools like automated translation software or specialized terminological resources, such as the *Urban Dictionary* (https://www.urbandictionary.com), *Emojipedia* (https://emojipedia.org), or *Know Your Meme* (https://knowyourmeme.com), to ensure linguistic comprehension or adequate production and interaction with other gamers (Vazquez-Calvo et al., 2019).

Reinhardt (2019) and Darvin and Hafner (2022) highlight the importance of digital literacies for academic and professional success. They note that players, as language learners and users, learn to critically evaluate information, collaborate

effectively and manage digital resources and identities. This often involves curating their online personas and developing context-specific communicative abilities to enhance in-group social reputation, similar to the digital literacies required by Jova, Marse and Sema to produce their gameplays (Vazquez-Calvo et al., 2020). Producing gameplays involves digital literacies such as advanced video production and editing skills, online content curation, digital identity curation and social networking abilities via comments and replies.

Based on a corpus of 73 interviews with teenagers focusing on emerging professionalization practices, Establés, Guerrero-Pico and Contreras-Espinosa (2019) concluded that while game-related social practices and other forms of social media engagement, such as fanfiction or TikTok content creation, enabled teenagers to acquire skills often neglected in academic contexts, only a minority were aware of the sociopolitical economy and dynamics of media. The authors call for tailored support and education to foster digital literacies effectively, including awareness of the socioeconomic and sociopolitical dimensions of online media, the gaming industry and not-for-profit fandom labour.

Beyond digital literacies, digital game cultures foster workplace attributes such as continual feedback for improvement, appreciating diverse teams, coping with rapid change and viewing learning as overcoming obstacles (Thorne et al., 2012). These qualities, sometimes called 'the gamer disposition' (Brown & Thomas, 2008), align with twenty-first-century soft skills and have relevant pedagogical implications today (Dooly & Thorne, 2018).

Pedagogical Implications, Challenges, and Future Directions

The exploration of game-related social practices reveals a dynamic interplay of language, culture and digital gaming. These practices provide insights into how language is used, developed and learned within and around gaming communities through activities beyond gameplay. Participating in social media, creating content for gaming and fan communities and engaging in these communities while developing digital literacies offer a fresh approach to language learning in informal contexts. Driven by their interests and proactivity, gamers, players and fans of games generally voluntarily engage in various activities that, even without crystal-clear language learning intent, result in meaningful and rich language learning experiences.

Gamers embody proactive language learning behaviours (Papi & Hiver, 2024) by actively seeking linguistic input through in-game text, dialogues in and out of game interfaces, forums or tutorials (for links with other SLA theories, see also ▶ "SLA Theories and DGBLLT"). They engage in interactions in multiplayer games and online discussions, enhancing their language skills. Gamers also search for explicit language information via game-related content like walkthroughs and fan translations. Moreover, they seek feedback from peers and community members, refining their language use. These proactive behaviours showcase gamers' and players' agency, strategic engagement and persistent motivation in gaming. Additionally, these behavioural traits triggered by an affinity to gaming transfer, to some degree, to language learning.

Integrating game-related social practices into formal language teaching implies capitalizing on these naturally occurring, interest-driven discourses and practices to create engaging and effective learning environments conducive to proactive language learning behaviours (see also ▶ "Pedagogical Approaches to DGBLLT" and ▶ "Task-Based DGBLLT"). York's study (2023) demonstrates the potential of such an approach by drawing inspiration from connected learning theory (Ito et al., 2020) and the 'Bridging Activities' pedagogical framework (Thorne & Reinhardt, 2013) in a Japanese university EFL classroom. Students participated in online Reddit communities related to their interests, such as gaming and anime, and engaged in multimodal communication by creating and analysing posts. This learning experience allowed students to connect their classroom learning with real-world language use, enhancing their vocabulary, pragmatic skills and

intercultural awareness. The study highlights how integrating game-related social practices can bridge the gap between formal education, highly engaging vernacular literacy practices connected to learners' affinities and digital literacies essential for academic and professional success.

However, incorporating these practices into formal education is not without challenges. Their informal and dynamic nature makes systematic study and scaffolding difficult. Aligning them with curriculum objectives and standards poses pedagogical challenges, particularly in contexts with strict curricular expectations. Additionally, the digital divide and access to technology, including school- or system-wide policies banning or limiting educational technology use, can impact the integration of game-related social practices in language teaching.

Looking ahead, research could zoom in on what and how gamers learn by using various data collection tools for triangulation (see also ▶ "Research Practices in DGBLLT"). Qualitative, emic and ethnographic methods such as screencasting, participant-generated photography, photo-elicitation interviews, first-person video diaries and capturing data through wearable devices like smart and point-of-view cameras (Yi et al., 2022) can enhance our understanding of game-related discourses and their impact on language learning. For instance, some of the studies reported in this entry use screencast to inform fan translation processes (Vazquez-Calvo, 2018; Vazquez-Calvo et al., 2019). These participatory methods empower participants to shape the research process by choosing and interpreting what is meaningful to them, thereby co-constructing knowledge. This approach encourages participants to examine their game-related practices, deepening their understanding of themselves as digitally literate and linguistically competent individuals. It also humanizes the research by balancing the roles of researcher and participant and offers potential for pedagogical use in capturing informal learning experiences for reflection and analysis in self-directed learning or classroom contexts.

Game-related social practices advance our understanding of the connections between digital gaming and language learning. Gamers learn not only through gameplay but also through a variety of texts and activities, engaging with broader interest-driven communities. This often leads to incidental language learning and the development of digital literacies. Beyond 'tedious language drills in dopamine-generating gamification loops' provided by some language learning apps (Lotherington & Bradley, 2024, p. 4), these practices foster proactive, autonomous language learning (Papi & Hiver, 2024). With expanding online participation and the ever more seamless integration of AI, AR or VR technologies (see also ▶ "Artificial Intelligence and DGBLLT" and ▶ "Extended Reality XR and DGBLLT"), game-related social practices are expected to thrive, diversify and grow. Language educators and researchers must remain alert and agile, ready to adapt to these dynamic shifts and continue to examine and test the potential of game-related social practices for language learning and teaching.

Cross-References

▶ Artificial Intelligence and DGBLLT
▶ Extended Reality XR and DGBLLT
▶ Gaming Literacies and LLT
▶ Identity and DGBLLT
▶ Pedagogical Approaches to DGBLLT
▶ Pragmatics and DGBLLT
▶ Research Practices in DGBLLT
▶ Task-Based DGBLLT
▶ Vocabulary and DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Brevik, L. M. (2019). Gamers, surfers, social media users: Unpacking the role of interest in English. *Journal of Computer Assisted Learning, 35*(5), 595–606. https://doi.org/10.1111/jcal.12362

Brown, J. S., & Thomas, D. (2008). The gamer disposition. *Harvard Business Review Blogs*. http://blogs.hbr.org/cs/2008/02/the_gamer_disposition.html

Chik, A. (2014). Digital gaming and language learning: Autonomy and community. *Language Learning and Technology, 18*(2), 85–100. https://doi.org/10125/44371

Darvin, R., & Hafner, C. A. (2022). Digital literacies in TESOL: Mapping out the terrain. *TESOL Quarterly, 56*(3), 865–882. https://doi.org/10.1002/tesq.3161

Dooly, M., & Thorne, S. L. (2018). Knowledge for network-based education, cognition & teaching: Key competences for the 21st century. KONECT white paper. https://doi.org/10.6084/m9.figshare.7366982

Establés, M., Guerrero-Pico, M., & Contreras-Espinosa, R. (2019). Gamers, writers and social media influencers: Professionalisation processes among teenagers. *Revista Latina de Comunicación Social, 74*. https://doi.org/10.4185/RLCS-2019-1328en

Gee, J. P. (2005). Semiotic social spaces and affinity spaces: From the age of mythology to today's schools. In D. Barton & K. Tusting (Eds.), *Beyond communities of practice* (pp. 214–232). Cambridge University Press.

Godwin-Jones, R. (2014). Games in language learning: Opportunities and challenges. *Language Learning and Technology, 18*(2), 9–19. https://doi.org/10125/44363

Ito, M., Arum, R., Conley, D., Gutierrez, K., Kirshner, B., Livingston, S., Michalchik, V., Penuel, W., Peppler, K., Pinkard, N., Rhodes, J., Tekinbas, K. S., Schor, J., Sefton-Green, J., & Watkins, S. C. (2020). The Connected Learning Research Network: Reflections on a decade of engaged scholarship. *Connected Learning*. https://clalliance.org/wpcontent/uploads/2020/02/CLRN_Report.pdf

Kowert, R., Festl, R., & Quandt, T. (2014). Unpopular, overweight, and socially inept: Reconsidering the stereotype of online gamers. *Cyberpsychology, Behavior, and Social Networking, 17*(3), 141–146. https://doi.org/10.1089/cyber.2013.0118

Lotherington, H., & Bradley, N. (2024). Hey Siri: Should # language, ☹, and follow me be taught?: A historical review of evolving communication conventions across digital media environments and uncomfortable questions for language teachers. *Language Learning & Technology, 28*(1), 1–19. https://hdl.handle.net/10125/73552

Papi, M., & Hiver, P. (2024). Proactive language learning theory. *Language Learning, 1–35*. https://doi.org/10.1111/lang.12644

Reinhardt, J. (2019). *Gameful second and foreign language teaching and learning*. Springer. https://doi.org/10.1007/978-3-030-04729-0

Sauro, S. (2017). Online fan practices and CALL. *CALICO Journal, 34*(2), 131–146. https://doi.org/10.1558/cj.33077

Shintaku, K. (2023). Learning of L2 Japanese through video games. *Language Learning & Technology, 27*(1), 1–17. https://doi.org/10125/73540

Sotamaa, O. (2022). Modding. In Grabarczyk, P. (Ed.), *Encyclopedia of ludic terms* (Spring 2022 edition). https://eolt.org/articles/modding

Soyoof, A., Reynolds, B. L., Shadiev, R., & Vazquez-Calvo, B. (2024). A mixed-methods study of the incidental acquisition of foreign language vocabulary and healthcare knowledge through serious game play. *Computer Assisted Language Learning, 37*(1–2), 27–60. https://doi.org/10.1080/09588221.2021.2021242

Stone, J. A. (2019). Self-identification as a "gamer" among college students: Influencing factors and perceived characteristics. *New Media and Society, 21*(11–12), 2607–2627. https://doi.org/10.1177/1461444819854733

Sundqvist, P. (2019). COTS games in the digital wild and L2 learner vocabulary. *Language Learning & Technology, 23*(1), 87–113. https://doi.org/10125/44674

Sykes, J. M. (2019). Emergent digital discourses: What can we learn from hashtags and digital games to expand learners' second language repertoire? *Annual Review of Applied Linguistics, 39*, 128–145. https://doi.org/10.1017/S0267190519000138

Taguchi, N. (2023). Game-based instruction of pragmatics: Learning request-making through perlocutionary effects. *Language Learning & Technology, 27*(1), 1–18. http://hdl.handle.net/10125/73519

Thorne, S. L., & Reinhardt, J. (2013). "Bridging activities," new media literacies, and advanced foreign language proficiency. *CALICO Journal, 25*(3), 558–572. https://doi.org/10.1558/cj.v25i3.558-572

Thorne, S. L., Fischer, I., & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL, 24*(3), 279–301. https://doi.org/10.1017/S0958344012000158

Vazquez-Calvo, B. (2018). The online ecology of literacy and language practices of a gamer. *Educational Technology & Society, 21*(3), 199–212. https://doi.org/10.30191/ETS.201807_21(3).0018

Vazquez-Calvo, B. (2021). Guerrilla fan translation, language learning, and metalinguistic discussion in a Catalan-speaking community of gamers. *ReCALL, 33*(3), 296–313. https://doi.org/10.1017/S095834402000021X

Vazquez-Calvo, B., & Thorne, S. L. (2022). Catalan identity and language attitudes through fan translation of video games in the digital wilds. *Journal of Language, Identity & Education*, 1–18. https://doi.org/10.1080/15348458.2022.2137168

Vazquez-Calvo, B., Zhang, L. T., Pascual, M., & Cassany, D. (2019). Fan translation of games, anime, and fanfiction. *Language Learning & Technology, 23*(1), 49–71. https://doi.org/10125/44672

Vazquez-Calvo, B., Elf, N., & Gewerc, A. (2020). Catalan teenagers' identity, literacy and language practices on YouTube. In M. Freiermuth & N. Zarrinabadi (Eds.), *Technology and the psychology of second language learners and users* (pp. 251–278). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-34212-8_10

Yi, Y., Cho, S., & Jang, J. (2022). Methodological innovations in examining digital literacies in applied linguistics research. *TESOL Quarterly, 0*(0), 1–11. https://doi.org/10.1002/tesq.3140

York, J. (2023). Engaging with the world: Applying connected learning in a university language learning context. *Foreign Language Annals*, March, 334–361. https://doi.org/10.1111/flan.12691

G

Gamification and LLT

Caroline Cruaud
University of South-Eastern Norway, Drammen, Norway

Abstract

Gamification of language learning and teaching (LLT) is a pedagogical method to engage students by making the learning activities more like games. The emergence of gamification in the 2010s was strongly connected to the marketing and business sectors. The concept later found its way into educational research and practice. This article will introduce two main approaches to the understanding and implementation of gamification in education: an artefact-centred approach and a situated approach. Central themes in gamification research within LLT will be explored, such as autonomy and engagement.

Keywords

Gamification · Playful learning · Game-based learning · LLT · Engagement · Educational research

Introduction

Gamification is often defined as 'the use of game elements in non-game contexts' (Deterding et al., 2011, p. 1). In other words, an activity usually not linked to game and play (e.g. work, training) is adapted through the use of game mechanics to make it more engaging. If informally the term gamification can be applied to any use of games and play in a given field, gamification in its definition does not refer to the use of full-fledged games but the framing of an activity with playful elements and techniques (see ▶ "Digital Game-Based Language Learning and Teaching").

The term gamification was first coined in the 2010s as a marketing technique and created a new awareness of already existing strategies to increase consumer engagement (e.g. reward programs, badges). Researchers, and maybe especially educational researchers, contested the term gamification and its application. The construction of the word itself—with the suffix –*fication* from latin *ficare* to make/to do—seems to imply an automatic transformation of an activity into a game (Silva, 2013).

In reaction to this, students and learning were brought back in focus in the application of gamification to education. Gamified learning is 'using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems' (Kapp, 2012, p. 10).

In language education, gamification first spread in informal learning settings through the development of language apps, for example, *Duolingo*. The use of games and play being already seen as central in foreign language learning (Council of Europe, 2001), examples of research and practice within the LLT field multiplied rapidly (e.g. Silva,

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_125-1

2013; Perry, 2015; Lombardi, 2015; Berns et al., 2016; Cruaud, 2016).

Definitions of Gamification: The Boundary Challenge

A first wave of research, often within the fields of human computer interaction (HCI) and game studies, focused on defining gamification and game elements and especially on cutting a clear boundary between gamification and similar approaches as serious games or playful interactions (e.g. Deterding et al., 2011, Huotari et al., 2012, Mekler et al., 2013; see also ▶ "Game Mechanics and Design Features for DGBLLT"). This need for a clear-cutting definition, perhaps expected in a new field, was enhanced by the challenge of the separation between games and gamified activities.

Gamification is defined as the use of game elements and not of fully developed computer games. However, the boundary between games and gamified activities can be difficult to trace (Werbach, 2014, p. 267). Is the training game *Zombies Run!* that adds story, points and levels to the sport of running a full-fledged game, or is it an example of gamification? If the game started as the gamification of the running activity, does its existence as a separate game remove it from the gamification field? There have been two main ways of addressing this issue in the field of research that also represent two different understandings of gamification: an artefact-based approach and a situated approach. The former aims at more precise descriptions of elements and their application; the latter moves the focus towards the experience of play. First, let us discuss the artefact-centred approach.

Game Elements: An Artefact-Centred Approach to Gamification

The focus on artefacts and systems is already present in the word gamification itself, a word built from 'game' and not 'play' (Silva, 2013, p. 4). In addition, the concept of 'game' in gamification is influenced by video games and focused on macrogameplay, also called the objective-challenge-reward loop (Philippette, 2014, p. 188). A first approach to gamification accepts these premises and focus on game elements, systems and artefacts of play. Gamification is described as a process, with a list of game elements that one can pick from depending on what one wants to achieve (Kapp, 2012). These may include, for example, the following:

- Reward structures: often taking the form of leaderboards, points or badges, they represent the student's progress in the learning activity
- Levels: either as a way of structuring the learning activity in small chunks or as the choice of difficulty for the tasks
- Storytelling: adding a narrative (e.g. characters, plot) to the activity to contextualise the learning situation

However, this does not equate with a mindless applying of points and leaderboards to a learning activity but rather 'a careful and considered application of game thinking to solving problems and encouraging learning using all the elements of games that are appropriate' (Kapp, 2012, pp. 15–16). Gamification has to be combined with a good instructional design to give positive effect on learning outcomes (Huang et al., 2020, p. 1897).

A study following this approach is interested in effects: more participation, more engagement, more content reviewed, and which strategies or game elements will maximise these effects in the learning situation (e.g. Barata et al., 2013; Huang & Hew, 2015; Huang et al., 2020). Within the educational field, this often translates to increasing student participation in a quantifiable way, through completed tasks, higher ranking in the leaderboard, length of presence online or attendance at lectures.

Playful Frame: A Situated Approach to Gamification

Advocating for a reappropriation of the term gamification from the field of marketing and its focus on consumer engagement, a group of researchers proposes a more holistic understanding

of gamified situations. Where the artefact-centred approach focuses on the game aspect and elements, the situated approach turns towards play and the students' experience. The separation between game and nongame is then addressed by 'considering the game to be an experience rather than a clear system of objects' (Philippette, 2014, p. 189). Gamification is then defined as the opening of a playful frame in an activity, for example, giving space for a playful interpretation of the learning situation (Cruaud, 2018b). Central in this approach is the focus on a play-based approach (Sanchez et al., 2016), on the playful attitude as the fourth dimension of play (Silva, 2013) and on the playful frame (Cruaud, 2018a, b). All build on the premise that 'the essence of the game does not lie [...] in the system of objects but in the relationship that develops between the player and the game' (Philippette, 2014, p. 192).

In this approach, the development of a gamified activity will begin by looking at the context (e.g. learning French as a foreign language in high school), the people within it and their interaction with different artefacts. It will focus 'on the learners and tak[e] into consideration the situation that emerges when they play rather than the artifact dedicated to play' (Sanchez et al., 2016, p. 484).

Another consequence of this approach is highlighting that gamification does not need to be digital but can also be used in analogue processes (see Lombardi, 2015 for an example of paper gamification).

Central Themes in Gamification Research for LLT

Research on gamification has developed rapidly in the last 15 years. Several meta-analyses and review studies have shown that gamified systems can have positive effects on student engagement, motivation, attitude and performance (Subhash & Cudney, 2018; Sailer & Homner, 2020; Huang et al., 2020). A short overview of central themes within the field of LLT research will now be presented.

Gamification and Learner Engagement

Unsurprisingly, a main focus of research studies has been on learner engagement. Findings in this theme are that learners are engaged in gamified learning activities. Engagement, when defined, is understood as participation in activities and satisfaction. Gamified learning activities generally have a positive effect on students' participation: students are active and complete more tasks (e.g. Perry, 2015; Lombardi, 2015; Cruaud, 2016). There also seems to be a positive effect on students' level of enjoyment (e.g. Dehghanzadeh et al., 2021; Lombardi, 2015). Finally, student engagement is often linked to motivation, and most studies find a positive effect of gamification on student motivation (e.g. Berns et al., 2016, Shen et al., 2024; see ▶ "Motivation and Affect in DGBLLT").

Learner Autonomy in Gamified Learning

Gamified learning opens up for the emergence of learner autonomy (Cruaud, 2018a). Students can create their own learning paths in the language classroom, through the use of gamified structures (e.g. levels, tasks) and when they are offered the possibility for choice. Findings also show that students in gamified courses are actively selecting tasks and can mediate their learning activities (e.g. Perry, 2015, Lombardi, 2015, Cruaud, 2018a; see ▶ "Autonomy and DGBLLT").

Gamification and Learner Anxiety

Taking a playful stance towards a situation, for example, when a gamified activity leaves space for a playful interpretation, gives learners the possibility to distance themselves from the fear of performing the language and of participating in the class activities. This playful distance can help reduce anxiety and give learners the confidence to speak in their target language (Reinders & Wattana, 2012). Students within a gamified learning situation feel less anxious about performing in the language (e.g. Perry, 2015; Lombardi, 2015; Berns et al., 2016) and more confidence towards the instructional activities (Subhash & Cudney, 2018). They start playing with the language through jokes, exploration of terms and language play finding their own way of being in the foreign language (Cruaud, 2016; see ▶ "Motivation and Affect in DGBLLT").

Cross-References

- ▶ Autonomy and DGBLLT
- ▶ Game Mechanics and Design Features for DGBLLT
- ▶ Motivation and Affect in DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Improving participation and learning with gamification. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 10–17). https://doi.org/10.1145/2583008.2583010

Berns, A., Isla-Montes, J.-L., Palomo-Duarte, M., & Dodero, J.-M. (2016). Motivation, students' needs and learning outcomes: A hybrid game-based app for enhanced language learning. *Springerplus, 5*(1), 1–23. https://doi.org/10.1186/s40064-016-2971-1

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Cruaud, C. (2016). The playful frame: Gamification in a French-as-a-foreign-language class. *Innovation in Language Learning and Teaching, 12*(4), 330–343. https://doi.org/10.1080/17501229.2016.1213268

Cruaud, C. (2018a). Learner autonomy and playful learning: Students' experience of a gamified application for French as a foreign language. *Alsic, 21*(1). https://doi.org/10.4000/alsic.3166

Cruaud, C. (2018b). *The playful frame: Design and use of a gamified application for foreign language learning* [Dissertation]. University of Oslo.

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 34*(7), 934–957. https://doi.org/10.1080/09588221.2019.1648298

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. E. (2011). Gamification: Toward a definition. In *CHI 2011, Vancouver, Canada*.

Huang, B., & Hew, K. (2015). Do points, badges and leaderboard increase learning and activity: A quasi-experiment on the effects of gamification? In *International Conference on Computers in Education, ICCE 2015 Proceedings* (pp. 275–280).

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Natercia, V., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology, Research and Development, 68*(4), 1875–1901. Social Science Premium Collection.. https://doi.org/10.1007/s11423-020-09807-z

Huotari, K., Hamari, J., & Lugmayr, A. (2012). Defining gamification. In *Proceeding of the 16th International Academic MindTrek Conference* (pp. 17–22). ACM. https://doi.org/10.1145/2393132.2393137

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.

Lombardi, I. (2015). Fukudai hero: A video game-like English class in a Japanese National University. *EL.LE – Educazione Linguistica Language Education, 4*(3), 483–500.

Mekler, E., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Disassembling gamification: The effects of points and meaning on user motivation and performance. In *Chi'13 extended abstracts on human factors in computing systems* (pp. 1137–1142). Association for Computing Machinery.

Perry, B. (2015). Gamifying French language learning: A case study examining a quest-based, augmented reality mobile learning-tool. *Procedia-Social and Behavioral Sciences, 174*, 2310–2317.

Philippette, T. (2014). Gamification: Rethinking 'playing the game' with Jacques Henriot. In M. Fuchs, S. Fizek, P. Ruffino, & N. Schrape (Eds.), *Rethinking gamification* (pp. 187–200). meson Press by Hybrid Publishing Lab.

Reinders, H., & Wattana, S. (2012). Talk to me! Games and students' willingness to communicate. In *Digital games in language learning and teaching* (pp. 156–188). Springer.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sanchez, E., Piau-Toffolon, C., Oubahssi, L., Serna, A., Marfisi-Schottman, I., Loup, G., & George, S. (2016). Toward a play management system for play-based learning. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, Lyon, France, September 13–16, 2016, Proceedings* (pp. 484–489). Springer International Publishing. https://doi.org/10.1007/978-3-319-45153-4_47

Shen, Z., Lai, M., & Wang, F. (2024). Investigating the influence of gamification on motivation and learning outcomes in online language learning. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1295709

Silva, H. (2013). La «gamification» de la vie: Sous couleur de jouer? *Sciences Du Jeu*, 1.

Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behavior, 87*, 192.

Werbach, K. (2014). (Re)Defining gamification: A process approach. In A. Spagnolli, L. Chittaro, & L. Gamberini (Eds.), *Persuasive technology* (Vol. 8462, pp. 266–272). Springer International Publishing. https://doi.org/10.1007/978-3-319-07127-5_23

L

Learner Variables in DGBLLT

Pia Sundqvist
Department of Teacher Education and School Research, University of Oslo, Oslo, Norway

Abstract

Learner variables, also known as individual difference (ID) variables, are individual characteristics of learners of a foreign/second language (L2) that influence L2 learning processes and outcomes. Learner (or ID) variables that are commonly used in L2 research are, for example, age, anxiety, aptitude, attitude, beliefs, cognitive abilities, gender, language background, learning strategies, learning styles, motivation, personality traits (e.g. extroversion/introversion), proficiency, working memory and willingness to communicate. In digital game-based language learning and teaching (DGBLLT), an area of research within the broad field of computer-assisted language learning (CALL) research, it is always relevant to consider learners' informal L2 learning outside the walls of educational institutions, that is, their extramural engagement in activities such as playing digital games and watching streamed media in the target language, as this learner variable complements any formal learning. It can be noted that informal L2 learning often comprises learners' use of digital technology. Access to technology is a typical socioeconomic status (SES) learner variable connected with the so-called digital divide. In DGBLLT as well as in CALL research, writing is clearly most researched when it comes to various L2 proficiency variables. In this entry, the concept of *learner variable* and associated concepts/variables are defined. Research findings are summarized along with an attempt at identifying trends and research gaps related to learner variables in DGBLLT. Finally, suggestions for future research are provided.

Keywords

Age · DGBLLT · Extramural engagement · Gaming · Gender · Individual difference variables · Situational variables · Language background · SES · Informal language learning

Introduction

Learner variables are individual characteristics of learners of a foreign/second language (L2) that influence learning processes and outcomes. Learner variables are also known as individual difference (ID) variables (see, for example, Skehan, 1989). In L2 research, some commonly used and investigated learner (or ID) variables are, for example, age, anxiety, aptitude, attitude, beliefs, cognitive abilities (e.g. working memory), extramural engagement, gender, language background (first language

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_130-1

(s), L1(s)), learning strategies, learning styles, motivation, personality traits (e.g. extroversion/introversion), proficiency and willingness to communicate (see also Attitudes Towards DBGLLT and Motivation and Affect in DGBLLT). More general variables will vary according to the extent to which learners control them; for example, while learners cannot control their age, they might be able to control—and therefore also able to change—their learning style. Many of these learner variables are interrelated, such as learner beliefs and strategies. As mentioned, ID variables influence learning, but it can also be the other way around that learning outcomes affect ID variables. Ellis (1994) points out that it is likely that "the more mutable factors exist in a symbiotic relationship with achievement" (p. 472). As an example, he argues that anxiety and motivation impact learning positively or negatively, but learners' perceived learning success or failure will also influence motivation. There are other types of learner variables too, such as a learner's socioeconomic status (SES). It is generally important to consider learners' SES as a variable, since factors such as parental educational level and financial resources are known to play a role in academic performance. In DGBLLT research, access to technology is clearly an important SES variable. In contrast to other learner/ID variables, SES can be described as a situational variable, as it depends on learners' specific external circumstances rather than internal characteristics. When it comes to examined L2 proficiency variables in DGBLLT as well as CALL research, writing is clearly researched the most (Mohsen et al., 2024). In what follows, learner variables (individual and situational) in DGBLLT are defined, and brief summaries of research findings are discussed. Finally, some trends and directions for future research are provided.

Defining Individual Difference Variables in DGBLLT

Digital games are without a doubt potential resources for L2 teaching and learning (see, for example, Reinders, 2012, for an edited volume), and researchers who take an interest in digital games, whether for teaching or learning, can approach their chosen research problem in different ways. Reinhardt (2017) provides a helpful account of such research orientations, where one orientation is to study a specific problem from the perspective of the game, another to do so from the perspective of the player (a player-learner orientation) and a third from the perspective of pedagogy. The learner variables outlined below—age, gender, language background, extramural engagement, and affective variables—have all been included in different studies that exemplify the three research orientations given in Reinhardt (2017) (see also ▶ "Research Practices in DGBLLT").

Age

The age variable is a perennial issue in discussion of L2 acquisition and consequently a thoroughly researched variable, not least as part of work investigating whether there is a critical (or sensitive) period for L2 learning, as there is for first language learning (the critical period hypothesis, Lenneberg, 1967). The concept of a critical period "implies a declining learning capacity within a specific age range and a maturational, ultimately biological reason for this decline" (DeKeyser, 2012, p. 443). Most scholars agree that there is such a critical/sensitive period for first-language learning, but not all agree about it for L2 learning. Therefore, as explained by DeKeyser (2012), in L2 research, the term "age effects" is instead commonly employed; this term is more neutral and pertains to empirical observations with fewer theoretical implications compared to the term "critical period".

As an ID variable in DGBLLT research, age must be defined. Age can, of course, simply refer to the age of research participants at the time of data collection. However, depending on the research aim, it may also be relevant to consider participants' age of onset, age of arrival and age of acquisition. *Age of onset* refers to the age at which a participant was first exposed to or began learning the L2. *Age of arrival* only applies to potential immigrant participants and denotes the age at which an immigrant arrived in the country (see also Language Background). *Age of acquisition* refers to the age at which a participant became

proficient in the L2 (a certain level of mastery is attained). It should be noted that sometimes age of onset and age of acquisition are used interchangeably or with overlapping meanings. The nuances between these terms can influence how findings are interpreted, especially when examining variables that contribute to L2 learning and the effects of early versus late L2 learning.

Age effects have frequently been examined in naturalistic language learning settings, often including adult immigrant learners, but less so in classrooms (DeKeyser, 2012). However, classroom-based studies are relatively common in DGBLLT. Further, DGBLLT concerns learners of all ages, from children to adults. The fact that technology has advanced tremendously since the turn of the millennium and considering that L2 learners of all ages can access technology, the actual age of research participants has become increasingly relevant to examine, as age impacts how L2 learners can interact with and benefit from digital games. For instance, if the research orientation is from the perspective of pedagogy, age would be a crucial ID variable to consider.

To better understand the impact of age in DGBLLT, examined topics include, for example, the age appropriateness of digital games. Researchers can ask and examine how digital games may vary in their effectiveness for L2 learners of different ages or what game features may be especially engaging for young as opposed to adult learners. Cognitive and developmental factors, instructional design and in-game social interaction are other topics clearly connected with age as an ID variable. Age-related differences in attitudes towards technology may also impact learners' acceptance of and beliefs about the efficiency of DGBLLT (compared with a player-learner research orientation, Reinhardt, 2017) and are, thus, yet another topic of research interest.

Gender

Prensky (2001) noted already in 2001 that gender would be an issue "to keep in mind as we create Digital Game-Based Learning" (p. 160). Early on, the stereotype was that digital games were created for and by men and boys, which might be based on some truth. However, as anticipated by Prensky, the noted gender-related difference diminished over time, and today digital games are played by all genders. Results from L2 gaming studies which have included gender as an ID variable tend to show that men (or boys) report playing digital games most frequently. This is one reason why gender can be relevant to include in DGBLLT when planning game-based teaching/research (an example of pedagogy-oriented research, using Reinhardt's (2017) terminology) or when examining whether L2 learners' voluntary game play in the home using the target language has an effect on their L2 proficiency (an example of player-learner-oriented research). Another reason why gender can be relevant to consider has to do with the fact that males generally tend to show more favourable attitudes and greater self-efficacy towards technology than females (Lee, 2022). It can also be mentioned that when school children have been given the opportunity to design their own games, significant gender differences have been observed as regards the choice of game genre and game worlds (Kafai, 1996).

Language Background

Including learners' language background as an ID variable in DGBLLT research is important for several reasons. For example, in experimental DGBLLT designs, it can be highly relevant to measure participants' baseline (preintervention) L2 proficiency level. In such measurements, considering learners' full language repertoire can prove useful. For immigrant participants, it would in addition be relevant to know of their age of arrival. Learning about language backgrounds provides insight into learners' previous exposure to and experience with the target language and other languages. Moreover, learners' respective individual learning context can influence how they interact with game-based learning environments as well as their engagement and motivation.

Learners' language backgrounds are tied to different cultural contexts, and different cultural contexts influence how learners perceive and respond to game narratives, game characters and

game scenarios. Thus, it is important to ensure cultural sensitivity in game design. In particular, if vernacular/commercial-off-the-shelf games are used, it may be necessary to critically discuss cultural ideologies and biases that come with the game (Reinhardt, 2019). Moreover, to include language background as a learner variable is also important as there might be so-called transfer effects. That is, skills and knowledge from one language may influence the learning and use of the target language in a DGBLLT context. Finally, many L2 learners are multilinguals, which naturally calls for an inclusion of language background (see ▶ "Multilingualism and DGBLLT").

Extramural Engagement

In DGBLLT research that aims to measure proficiency or development in the L2, there are strong arguments for including learner's *extramural* (i.e. "outside the walls" of the educational institution) engagement in the target language as a learner variable (Sundqvist, 2024). There is plenty of evidence from empirical DGBLLT research that extramural engagement has a positive impact on the learning of the target language (for a collection of studies, see Reinders et al., 2022; for a research overview, see Zhang et al., 2021). Learners' engagement in various extramural activities is self-initiated and never connected with school or schooling, so it exemplifies informal L2 learning. Further, learners' engagement is self-regulated and interest-driven, typically leading to incidental learning, even though intentional learning is also part of the extramural concept (for details, see Sundqvist & Sylvén, 2016; see also ▶ "Autonomy and DGBLLT"). More exposure, attention and time will add to learning the abilities and/or content knowledge connected with specific extramural activities, including game-based activities.

The concept *engagement* encompasses at least three distinct, yet interrelated, dimensions: behavioural, cognitive and emotional/affective engagement. First, in DGBLLT, behavioural engagement may be the amount of time spent playing and active participation in game-related activities. Second, cognitive engagement can be the mental effort necessary to play the game using the L2 and the strategies used while playing a game. Third, emotional/affective engagement may correspond to the responses and attachment learners have towards a game, such as excitement and feelings of being immersed. There is some overlap between engagement and affective variables, the focus of the next section.

Affective Variables

Affective variables can briefly be defined as the emotional, attitudinal and motivational factors that influence an individual's L2 learning process and performance (see also Motivation and Affect in DGBLLT). They include a vast range of psychological states (and traits) that can either facilitate or hinder L2 learning, such as anxiety (specifically foreign language anxiety), attitudes, boredom, enjoyment, frustration, grit, self-confidence and self-efficacy. In DGBLLT, affective variables play a critical role in shaping learners' willingness to engage in games and also their persistence when faced with challenges, because research has shown an interrelationship between attitude and action (Lee, 2022). Ultimately, affective variables matter for learners' overall success in attaining L2 proficiency.

Considering affective variables is helpful in designing more effective DGBLLT environments that not only support L2 development but also foster positive emotional and motivational states in learners. Thus, both for researchers and educators, it is necessary to understand that affective variables are integral to improving the L2 learning process.

Defining Situational Variables in DGBLLT

Situational variables can be defined as various external circumstances and conditions that shape and influence individuals' learning, behaviour, decisions and interactions. In classroom-based pedagogically oriented DGBLLT research, situational variables can refer to the classroom environment, teaching methods, peer interactions and teaching materials and resources that can impact L2 learning and learner behaviour, examining, for

example, whether to use vernacular or educational games (see, for example, Reinhardt & Sykes, 2012, for more examples and details). Like the learner variables discussed above, situational variables have also been researched using various research orientations, as proposed in Reinhardt (2017).

Setting and Pedagogy

The effectiveness of DGBLLT depends on a thoughtful integration of games into formal educational settings and the application of sound pedagogical principles tailored to meet the needs of L2 learners. As regards setting and pedagogy, in experimental DGBLLT, it is common to investigate learning outcomes (i.e. the dependent variable(s)) from the employment of situational game-related variables, such as the use of, for example, different types of interaction (e.g. player-to-player or player-to-nonplayer character), a certain task design or feedback (i.e. the independent variables). Moreover, it is not uncommon to include ID variables (such as motivation) as dependent variables. For example, in research comparing results between experimental (i.e. digital game-based teaching) and control groups (i.e. "traditional" teaching), motivation has frequently been investigated in order to find out whether digital game-based teaching leads to increased motivation (see Motivation and Affect in DGBLLT). This type of experimental research, then, typically approaches the problem from the perspective of the game or from the perspective of pedagogy (see Research Approaches to DGBLLT and also Reinhardt, 2017).

Formal and informal L2 learning is another situational variable considered in DGBLLT research. Whereas formal digital game-based L2 learning takes place intramurally in schools and higher education institutions, informal digital game-based L2 learning takes place extramurally (see Extramural Engagement). When involved in informal game play, a learner is both a player and a learner at the same time, and the learner's gaming experience will undoubtedly be "impacted by variables like gender, age, L2 proficiency, and game literacy" (Reinhardt, 2017, p. 211). Research seeking to uncover the influence of such individual and situational variables on L2 proficiency and sociocollaborative learning, within the context of game-mediated interactions, has often been undertaken by adopting a player-learner orientation (Reinhardt, 2017; Sundqvist & Sylvén, 2016). It should be added that setting is not limited to the distinction between the levels of formality; setting is a particularly interesting situational variable when the research focus is on examining cultural differences in DGBLLT research, where more transnational studies are needed.

Socioeconomic Status

As mentioned in the introduction, learners' SES is an important situational variable to consider in DGBLLT, since it impacts academic performance. In a globalized world, access to technology is crucial but accessibility varies. A frequently used concept to describe this problem is *the digital divide*, that is, the gap between individuals (here L2 learners), households and geographical areas at different socioeconomic levels in terms of both opportunities to access information and communication technologies and the ability to use such technologies effectively (see, for example, Ortega, 2017). The digital divide also encompasses disparities regarding access to digital devices and Internet connectivity, digital literacy and learners' skills necessary to utilize digital resources fully.

Summary of Research Findings and Trends

As shown in numerous studies and volumes, both intramural and extramural game-based language learning opportunities form a rich environment for L2 learners (Reinders, 2012; Reinders et al., 2022; Reinhardt, 2019; Reinhardt & Sykes, 2012; Sundqvist & Sylvén, 2016; Zhang et al., 2021), but the DGBLLT environment per se will not guarantee that L2 learning will be successful. To some extent, the degree of success will always depend on different individual and situational learner variables. Thus, although research findings on DGBLLT overall tend to be positive, the role of individual and situational variables in this line of research still remains to be explored in greater depth, not least in

terms of the interrelatedness of learner variables and L2 learning through DGBLLT. This entry has provided some insight into a few important learner variables: age, gender, language background, extramural engagement and affective variables (ID variables), as well as setting and pedagogy and socioeconomic status (situational variables). Some trends in DGBLLT research encompass experimental studies focusing on the interrelatedness of gaming, affective variables and L2 learning, explorative studies on informal extramural engagement in games and L2 learning, game-enhanced L2 learning and pedagogy and an increased interest in DGBLLT studies with young research participants.

Future Directions for Research

There are gaps in current research. In formal settings, to accompany experimental research, there is a need for more classroom-based studies to examine how educational as well as vernacular games are used. There are also few studies on L2 teachers' attitudes towards using games. In informal settings, as suggested by Lee (2022, p. 107), to deepen our understanding of the complexity of L2 learning outside educational institutions and contribute to research-informed pedagogy or curriculum design, "emic investigation into the interactions between each learner factor and other socio-cultural factors of institutional contexts, as well as examination of discrepancies between intention and actual behavior" is necessary. Her suggestion certainly applies to DGBLLT research. For example, current research findings on the role of extramural gaming for L2 learning are inconclusive. Therefore, it would be beneficial with comparative studies focusing on extramural gaming in different cultural and national contexts. Preferably, learners of different ages would be included. Moreover, time is ripe to move on to theory-led studies on informal extramural engagement in gaming, to tease out learners' use of the L2 in that type of natural game settings. Such studies would elucidate the uniqueness of individual learners and demonstrate the dynamic interplay between the more malleable components of ID variables and learners' L2 learning experiences. With regard to methodology in future DGBLLT, quantitative, qualitative and mixed methods are welcomed, to allow for generalizations but also to provide more in-depth knowledge about L2 learning processes.

Cross-References

- ▶ Attitudes Towards DBGLLT
- ▶ Autonomy and DGBLLT
- ▶ Extramural Engagement
- ▶ Language Background
- ▶ Motivation and Affect in DGBLLT
- ▶ Multilingualism and DGBLLT
- ▶ Research Approaches to DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

DeKeyser, R. M. (2012). Age effects in second language learning. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 442–460). Routledge.

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Kafai, Y. B. (1996). Gender differences in children's constructions of video games. In P. M. Greenfield & R. R. Cocking (Eds.), *Interacting with video* (pp. 39–66). Ablex Publishing Corporation.

Lee, C. (2022). English language learning beyond the classroom. Do learner factors matter? In H. Reinders, C. Lai, & P. Sundqvist (Eds.), *The Routledge handbook of language learning and teaching beyond the classroom* (pp. 98–111). Routledge.

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.

Mohsen, M. A., Althebi, S., Alsagour, R., Alsalem, A., Almudawi, A., & Alshahrani, A. (2024). Forty-two years of computer-assisted language learning research: A scientometric study of hotspot research and trending issues. *ReCALL, 36*(2), 230–249. https://doi.org/10.1017/S0958344023000253

Ortega, L. (2017). New CALL-SLA research interfaces for the 21st century: Towards equitable multilingualism. *CALICO Journal, 34*(3), 285–316. https://doi.org/10.1558/cj.33855

Prensky, M. (2001). *Digital game-based learning*. Pragon House.

Reinders, H. (Ed.). (2012). *Digital games in language learning and teaching*. Palgrave Macmillan.

Reinders, H., Lai, C., & Sundqvist, P. (Eds.). (2022). *The Routledge handbook of language learning and teaching beyond the classroom*. Routledge. https://doi.org/10.4324/9781003048169

Reinhardt, J. (2017). Digital gaming in L2 teaching and learning. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology in second language teaching and learning* (pp. 202–216). John Wiley & Sons.

Reinhardt, J. (2019). *Gameful second and foreign language teaching and learning: Theory, research, and practice*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-04729-0

Reinhardt, J., & Sykes, J. (2012). Conceptualizing digital game-mediated L2 learning and pedagogy: Game-enhanced and game-based research and practice. In H. Reinders (Ed.), *Digital games in language learning and teaching* (pp. 32–49). Palgrave Macmillan.

Skehan, P. (1989). *Individual differences in second-language learning*. Edward Arnold.

Sundqvist, P. (2024). Extramural English as an individual difference variable in L2 research: Methodology matters. *Annual Review of Applied Linguistics*, 1–13. https://doi.org/10.1017/S0267190524000072

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-46048-6

Zhang, R., Zou, D., Cheng, G., Xie, H., Wang, F. L., & Au, O. T. S. (2021). Target languages, types of activities, engagement, and effectiveness of extramural language learning. *PLoS One, 16*(6), e0253431. https://doi.org/10.1371/journal.pone.0253431

M

MMOGs and LLT

Nasser Jabbari and Seyyed Pedram Allahveysi
Department of Language and Linguistics,
University of Essex, Colchester, UK

Abstract

The integration of play into educational frameworks has long intrigued educators, especially in language learning. One contemporary approach gaining traction is the use of massively multiplayer online games (MMOGs). This entry discusses the potential of MMOGs as educational tools for language learning and teaching. It begins with an overview of digital games and their growing popularity and then reviews the specific attributes of MMOGs that make them promising media for language education. The entry also highlights practical applications of MMOGs in educational settings, demonstrating how their affordances can be leveraged to foster student engagement and improve language learning outcomes. Additionally, it addresses potential challenges that educators may encounter when integrating MMOGs into their curriculum and provides recommendations for effectively navigating these challenges. By examining the affordances and limitations of MMOGs, this entry aims to offer some insights into their role in modern language education and suggest best practices for their implementation.

Keywords

Massively multiplayer online games · Language learning · Language teaching

Introduction

The gaming market has experienced remarkable growth, with the industry projected to generate nearly $282.30 billion in 2024 (Statista, n.d.). This surge is largely attributed to the increased demand for online games following the COVID-19 pandemic in 2020, which saw a significant rise in the popularity of mobile games as people stayed home. Even in the postpandemic period, interest in online gaming continues to rise, with approximately 1.1 billion online gamers worldwide (Clement, 2024). Forecasts indicate that the gaming industry will sustain a compound annual growth rate of 12% from 2020 to 2025 (Linares et al., 2021). Within this expanding market, massively multiplayer online games (MMOGs) have attracted considerable attention, particularly for their potential applications in language learning and teaching.

MMOGs are networked video games featuring highly graphical three-dimensional fantasy worlds. Accessible on personal computers, gaming consoles and mobile devices, MMOGs allow thousands of players across the globe to explore the game world, compete, collaborate and socialise simultaneously in real time within the game's persistent virtual world. This entry focuses on

© Crown 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_132-1

commercial or 'vernacular' (Reinhardt & Sykes, 2012) MMOGs that are developed solely for entertainment purposes (see ▶ 'DGBLLT and Entertainment Games'), with some adopting a free-to-play version supplemented by paid features. 'Scalability, fast response time, and low cost are of utmost importance in designing a successful massively multiplayer online game' (Yahyavi & Kemme, 2013, p. 1). This business model has resulted in the proliferation of MMOGs in recent years, with numerous games experiencing exponential and widespread growth (Ghobaei-Arani et al., 2019).

MMOGs have transformed the gaming industry by enabling the simultaneous participation of thousands of players from diverse geographical, social, linguistic, national, gender and age backgrounds. This innovation within the entertainment industry, and more broadly within social media platforms (Kaplan & Haenlein, 2010), has facilitated global connectivity and interaction within virtual gaming environments (Tay, 2005). The design, storyline, audiovisual effects, aesthetic features, social interaction opportunities, immersive environments, regular content updates and competitive elements of MMOGs have all contributed to their global popularity. Released with various scenarios, game features and mechanics, these games have attracted players of diverse tastes worldwide, making them commercially successful (Griffiths et al., 2003).

Given their increasing popularity, it is crucial to examine the impact of MMOGs on various aspects of gamers' lives, including language and culture learning (Zhong, 2011). This entry aims to explore the potential applications of MMOGs in language learning and teaching. It begins with an overview of MMOGs, discussing their characteristics and popularity. Next, it examines their practical applications in classroom settings and discuss how these games can facilitate language learning and enhance intercultural understanding. Finally, this entry discusses the challenges and considerations associated with integrating MMOGs into language education. It offers language teachers some insights to comprehend both the potential benefits and inherent limitations of MMOG-enhanced language teaching as an innovative pedagogical approach.

Overview of MMOGs

At the onset of a game, MMOG players select and customise their avatars—fictional characters within the game's virtual world—and subsequently control their actions throughout the gameplay (Zhong, 2011). Players get involved in various quests (or missions), which are assigned by computer-controlled or nonplaying characters (NPCs). Quests involve a range of different activities such as collecting a specific number of items or resources (e.g. herbs and lost artifacts) from the game world, eliminating a specific number of monsters in a region and solving riddles and deciphering clues to unlock a treasure chest. Quests grow more challenging and thus require more collaboration and active communication among gameplayers.

Another key feature of MMOGs is their rewards system. Upon completing quests, players receive rewards such as experience points (XPs), which help gameplayers to level up their characters and unlock new skills; in-game currency (e.g. coins, gems or tokens), by which they can purchase equipment or services within the game world; special titles or badges, which signify their achievements or completion of specific challenges; and mounts and pets, which gameplayers can ride or keep as companions. The pursuit of rewards, quests and storylines, customised characters, persistent and expansive game worlds, progression systems throughout the gaming experience and immersive audiovisual features are among a plethora of different key features that help immerse players in the game's virtual world (see also ▶ 'Game Mechanics and Design Features for DGBLLT').

MMOGs' increasing popularity can also be attributed to their significant social features that can contribute to various aspects of language learning (see also ▶ 'Game-Related Social Practices'). MMOGs foster a sense of community among players participating in guilds and alliances to collaborate towards shared goals, for

example, completing quests and winning battles (Ma et al., 2019). Such behaviours create an 'affiliative bond' (Thorne, 2008, p. 321) among gamers and promote socially supportive gaming environments (Cole et al., 2020). Partaking in social events and in-game celebrations can further enhance MMOG players' communal experience and make the game's virtual world feel lively and socially interconnected. MMOGs facilitate extensive interactions among players, thereby fostering cooperation and social support akin to the dynamics observed in socially and culturally bonded groups in the real world (Cole et al., 2020).

The provision of a persistent world is another crucial factor that distinguishes MMOGs within the gaming industry. In these games, the games' virtual world remains continuously active, regardless of player presence, and the game data is maintained even in the event of server failures (Zhang et al., 2008). Persistence is highly valued, as it mirrors real-world dynamics where objects change over time (Gustafsson et al., 2020). For example, in *Player Unknown Battleground* (PUBG), aspects of the game such as seasons, climate and calendar events get constantly updated, creating a real-world impression.

Another notable aspect of MMOGs is their capacity to enable real-time embodied interactions through avatars, utilising both text and voice chat channels. These interactions closely simulate real-world dynamics, allowing players to effectively communicate within the game environment. In these settings, players are required to collaborate with teammates to complete various in-game tasks. Sustained connectivity and attentiveness during gameplay are crucial for optimal interaction, including responding to immediate changes, updating teammates, adapting tactics and strategies and issuing new instructions (Jabbari & Eslami, 2023).

MMOGs come in different genres and types, each offering unique gameplay experiences, mechanics and player interactions (Reinhardt, 2021), which accommodate diverse preferences within the gaming community. Some of the most popular genres include Massively Multiplayer Online Role-Playing Games (MMORPGs), First-Person Shooters (MMOFPS), Third Person Shooters (MMOTPS), Real-Time Strategy Games (MMORTS), Simulation Games (MMOSG), Sandbox MMORPGs and Battle Royale games.

In MMORPGs (e.g., *World of Warcraft* and *Final Fantasy XIV*), players customise avatars, complete quests and gain experience points to advance. These games feature expansive worlds where players explore, interact and engage in various challenges, developing skills and acquiring equipment along the way. MMOFPS games (e.g. *Call of Duty: Warzone and Planetside 2*) feature large-scale, real-time battles where players compete in expansive environments, using various weapons and tactics to achieve objectives. MMOTPS games (e.g. *Destiny 2*) involve players engaging in competitive and cooperative gameplay from a third-person perspective. These games offer expansive worlds where players navigate missions, battle enemies and participate in large-scale events. The third-person viewpoint allows for a broader field of vision and tactical awareness, enhancing strategic combat and teamwork in both player-versus-environment (PvE) and player-versus-player (PvP) scenarios. MMRTS games (e.g. *StarCraft 2*) require players to build bases, gather resources and command armies in real time to defeat opponents and achieve strategic objectives. MMOSGs (e.g. *Eve Online*) simulate real-world activities and systems, offering an immersive experience by allowing players to engage in complex economic, social and political interactions within a persistent virtual world. Sandbox MMORPGs (e.g. *Black Desert Online*) allow gamers extensive freedom to explore vast game worlds and engage in various activities. Players can pursue diverse interests, from crafting and trading to combat and exploration, without being restricted by predefined paths. In Battle Royale games (e.g. *PUBG* and *Fortnite*), numerous players are dropped into a shrinking play area where they fight until only one player or team survives. This variety in game genres has attracted a diverse range of players with various gaming motivations (e.g. interacting and socialising, competing and achieving and immersing in and discovering the game world) (Yee, 2006) and experiences around the globe.

MMOGs and Language Learning

MMOGs have emerged as a new genre of play culture that evolves through real-time users' interaction, blending communication and entertainment within a computer-mediated setting (Freeman, 2018). MMOG players immerse themselves in linguistically rich and semiotically complex environments (Thorne et al., 2012), where language is used along with other modes of communication (Jabbari & Eslami, 2023) to complete various in-game tasks. In these settings, MMOG players are highly engaged, constantly participating in meaningful collaborations to accomplish collective in-game goals. Considering MMOGs' immersive and interactive nature as well as their popularity, accessibility and affordances for providing opportunities for contextualised language use, many language teachers and second language (L2) acquisition scholars have been curious to know how these games can serve as an effective tool for enhancing language learning experience and teaching effectiveness.

MMOGs have been widely recognised as valuable tools that can facilitate language learning (e.g. Chotipaktanasook and Reinders (2018), Dixon and Christison (2021), Jabbari and Eslami (2023), Jabbari and Peterson (2023), Peterson et al. (2020), Rankin et al. (2009)). For instance, Rankin et al. (2009) showed that ESL students improved their vocabulary knowledge as the result of playing *Ever Quest II* collaboratively with native English speakers. They highlighted the role of language socialisation in L2 learning. Chotipaktanasook and Reinders (2018) also found that playing MMORPGs significantly increased L2 interactions, which encompassed a broader range of discourse functions compared to those typically observed in classroom-based English language interactions. Dixon and Christison (2018) also concluded that MMORPGs benefit L2 acquisition by prompting target language interactions during collaborative problem-solving tasks within the game. Their study highlighted that playing *Guild Wars 2* fosters modified language output and provides opportunities to negotiate language input to complete in-game tasks. In the same vein, Jabbari and Eslami's (2023) microanalyses of L2 gamers' interactions also revealed many instances of successful negotiations for meaning, which are considered beneficial in SLA research from the interactionist perspective to language learning (Long, 1996). In a longitudinal study, Jabbari and Peterson (2023) investigated the question whether playing *World of Warcraft* could improve English-as-a-foreign-language (EFL) gamers' speaking performance over time. The results suggested that playing the game collaboratively with peers and native English speakers could improve some aspects of the gamers' speaking performance in terms of complexity, accuracy and fluency. In a scoping review, Jabbari and Eslami (2019, p. 92) concluded that

> MMOGs provide socially supportive and emotionally safe (i.e., low-language anxiety) environments that afford multiple opportunities for L2 learning and socialisation, which, in turn, help L2 learners to enrich their L2 vocabulary repertoire and enhance their communicative competence in the target language.

Despite some scepticism regarding the effectiveness of massively multiplayer online games (MMOGs) as tools for language acquisition, these 'unorthodox language-learning tools' (Rankin et al., 2009, p. 162) are widely regarded as promising venues for both language learning and instruction.

In these gaming environments, players exhibit strong motivation to accomplish various in-game tasks to advance their avatars and achieve higher levels (see also ▶ 'Motivation and Affect in DGBLLT'). This intrinsic drive underscores the necessity for heightened awareness of any changes and updates within the game context (Jabbari & Eslami, 2023). For instance, players must remain vigilant regarding the positioning of enemies and availability of gaming resources in battle royale games, identifying crewmates and imposters in *Among Us* and coordinating attacks on other villages in *Clash of Clans*. As such, effective communication becomes essential as players engage in eliciting and disseminating information, strategising future in-game actions and assessing their collaborative outcomes.

Consequently, these interactions foster a dynamic and immersive communication milieu wherein language use is authentically contextualised and serves distinct meaningful purposes.

MMOGs are usually multinational, multilingual and multicultural communication settings (Thorne, 2008; Ryu, 2011; Zheng et al., 2009), where English or other languages can serve as a lingua franca. In such settings, an L2 gamer is exposed to different languages and probably different dialects and diverse micro- and macrocultures. These natural communication environments offer what language educators have long sought to provide in the instruction of oral and aural skills: authentically contextualised language input and output that accurately reflect the reality of language use today. Moreover, engaging in such communication contexts allows L2 gamers to develop their intercultural awareness and L2 communicative competence (Johnson, 2009; Peterson, 2010)—the other two competencies highly sought after in formal educational contexts (see also ▶ 'DGBLLT and Culture').

It is well documented that motivation and affective variables play critical roles in language learning behaviour and ultimately determine language learning achievements (Dörnyei, 2003). Researchers (e.g. Reinders and Wattana (2014, 2015)) have strongly argued that playing MMOGs help language learners develop some positive affective and motivational drives towards L2 learning and socialisation. Reinders and Wattana (2015) revealed that reduced communication anxiety, heightened self-efficacy beliefs regarding L2 competence and increased motivation to engage in L2 communication were the primary factors contributing to learners' greater willingness to use English during gaming activities. They stated that playing the game seems to have started 'a virtuous cycle of lowered anxiety, resulting in more L2 production, leading to greater self-satisfaction, and resulting in more motivation, which in turn led to a further lowering of affective barriers' (Reinders & Wattana, 2015, p. 50).

Applications of MMOGs in the Classroom

Prior to incorporating any resources or instructional tools, including technological advancements, into our classroom, it is essential for us, as language educators, to address several fundamental questions. Our answers to these inquiries determine both the extent and the method of utilising such tools in our educational setting; questions such as Why should we use this tool or resource in the class?, For what purpose(s)?, Under what circumstances? When? How?, What would be our roles as language teachers?, What would be the role of our students?, How can using this tool enhance our students' language learning experiences?, How can we design our lessons around this tool? and How can we monitor and assess our students' progress and engagement in the language learning activities mediated by this tool? and additional questions should be considered before integrating MMOGs into our teaching endeavours. These questions highlight the interplay among four fundamental elements that form the complex dynamics of a classroom: the student, the teacher, the content and the teaching method. MMOGs present promising, albeit unconventional, opportunities for language learning due to their capacity to engage students in diverse, meaningful interactions within a highly immersive environment (Peterson, 2010; Reinhardt & Kirby, 2022). However, their implementation must be carefully aligned with the specific characteristics of students, teachers, content and teaching methods. The following example elucidates this point.

Imagine an English for business communication teacher who utilises an MMORPG to enhance students' engagement and provide opportunities for practising various language functions, including negotiating, persuading and presenting. To create an engaging learning environment, the teacher selects *Eve Online*, a spacefaring MMORPG known for its complex social interactions and compelling narrative. Within *Eve Online*, players participate in authentic communication tasks, affording them numerous opportunities to practise meaningful language use. The

teacher, who has a positive attitude towards incorporating MMOGs in language instruction, undergoes training to become well versed in the game's mechanics and affordances for providing language learning opportunities. This preparation enables the teacher to effectively guide students through the game's virtual world, address any technical challenges and ensure the achievement of educational objectives.

Upon entering the game, students select specific roles, such as miners, manufacturers, traders and pirates, each requiring different language skills and strategies. These roles align with the course content, which focuses on relevant language functions, including negotiating, persuading and presenting information. The teacher guides students through creating in-game corporations and getting engaged in simulated market transactions. This in-game task allows students to practise negotiating prices, trade deals and delivery terms using in-game chat and email templates. The teacher also encourages students to collaborate in groups, create advertisements for their player corporations and persuade other gamers to join by highlighting specific benefits of joining. This task involves crafting persuasive messages underlining the corporation's missions, vision, achievements and training opportunities. To improve students' written and oral presentation skills, the teacher also asks students to collect in-game market data and generate a report, analysing trends, potential investment opportunities and resource availability. Then, the teacher prepares students to present their reports in class. This task simulates briefing sessions for investors in real-world business settings.

The teacher actively monitors the students' language learning behaviour and progress by participating in the game's virtual world, providing real-time feedback and scaffolding as needed. The teacher also implements a blended learning approach, supplementing the students' in-game experiences with classroom activities. The teacher encourages students to document their in-game interactions and discuss examples of successful negotiations, persuasive arguments or well-delivered presentations. By reinforcing and expanding upon traditional teaching methods, this approach ensures that the game becomes an integral part of the learning process. This method not only enhances language learning but also offers some 'positive impacts' (Chapelle, 2001, p. 57) beyond language skills. These positive impacts include fostering critical thinking, promoting collaboration and encouraging learner autonomy. This example exhibits how an MMOG can be effectively implemented in a language-learning context to create engaging and relevant learning experiences for language learners.

Challenges and Considerations

Engaging in MMOGs in a target language offers numerous advantages for L2 learners. Notably, they enhance L2 vocabulary knowledge and retention (Peterson, 2023) and help learners develop communicative strategies (Dixon & Christison, 2021; Jabbari & Eslami, 2019) through getting involved in L2 interactions (Reinhardt, 2021) and negotiations of meaning (Jabbari & Eslami, 2023). MMOGs offer non-threatening social environments that can significantly boost L2 learners' self-confidence and willingness to communicate in the target language (Reinders & Wattana, 2015). This enhanced confidence can facilitate L2 socialisation and, in turn, improve cross-cultural awareness among L2 learners within multicultural and multilingual gaming contexts (Jabbari & Eslami, 2023). Nevertheless, despite these advantages, MMOGs present several challenges.

Technical requirement can be a serious challenge. Accessing MMOGs requires technological devices such as PCs, mobile phones or tablets, along with stable Internet connection and headsets for communicating while gaming. Not everyone can afford to buy these devices. Besides, insufficient technology can lead to lags during gameplay, negatively impacting the gaming experience and causing frustration, anxiety or stress among players. Related to this challenge is the expertise teachers and students require when engaged with MMOGs. Lack of teachers' or students' expertise can also affect the efficacy of

MMOG-enhanced language lessons. Both sides should be trained and equipped with the skills and knowledge necessary to use MMOGs as mediums of instruction or avenues for language learning.

It can be difficult to manage a classroom in which students are engaged with individual or group activities within an MMOG. To enhance learning through gaming, teachers should be competent in monitoring students' in-game activities to ensure they are focused on language learning objectives rather than playing the game frivolously just for fun (Baierschmidt, 2013). Monitoring presents significant challenges, particularly because MMOGs frequently encompass extensive and often evolving virtual environments—due to periodic updates by developers—where language use and interactions may not consistently align with educational objectives. Additionally, educators have limited control over these contents and interactions. Moreover, balancing game time with more focused instructional activities presents an additional challenge. It is essential to ensure that students do not allocate excessive time to MMOGs at the expense of other critical language learning tasks. Another challenge related to classroom management pertains to novice players, or 'noobs', who lack basic skills in playing MMOGs. Noobs are often not welcomed by other gameplayers, as they can negatively impact the entire team's performance. Grouping noobs with skilled gameplayers can be a challenge for a teacher, as it can promote negative feelings such as anger and frustration among gameplayers (Barnett et al., 2010).

Using MMOGs as instructional materials is also challenging, as they must be integrated appropriately into an existing language learning curriculum. Aligning various in-game tasks with the curriculum necessitates meticulous planning and a high degree of creativity (Liu et al., 2012). This complexity is heightened, considering Reinhardt's (2021) observation that each gameplay session presents a unique interaction between game design mechanics and player actions and that these interactions result in varying dynamics regarding student engagement, in-game social interactions and ultimately language use and learning. Since MMOGs often feature open and dynamically changing environments, it is not easy to adapt them to fit a language curriculum. Ensuring that the contents within MMOGs are relevant and appropriate can be challenging and modifying MMOGs to meet educational expectations require careful consideration.

Assessing language learners' progress to evaluate the efficacy of MMOG-driven lessons is another challenge faced by teachers. Conventional language assessment methods, which prioritise the linguistic aspects of language, are not efficient in capturing language learners' progress in MMOG settings. Consequently, there is a need to develop more effective assessment methods that align with the foundational principles and frameworks of game-based language learning. Creating such assessment methods requires substantial investment in time, expertise and creativity.

Another significant challenge lies in motivating students to learn a language through playing MMOGs. Not all language learners have positive attitudes towards game-enhanced language learning practices. Some learners may not perceive gaming as relevant to their learning objectives or aligned with their learning styles. Additionally, certain content may be distressing or inappropriate, leading to disengagement from language learning activities. For example, shooter MMOGs, such as *Fortnite* and *Call of Duty*, frequently contain violent content that may be unsuitable for young players. Research suggests that exposure to violent video games can lead to increased anxiety and depression in some individuals (Chen et al., 2020). Ensuring all students are equally motivated and engaged with MMOG-enhanced language learning poses a continual challenge for educators. This challenge necessitates ongoing research and adaptation to integrate gaming principles into pedagogical practices effectively. Addressing these challenges will be crucial for advancing the efficacy of MMOGs in language instruction and achieving broader educational outcomes.

Cross-References

► DGBLLT and Culture
► DGBLLT and Entertainment Games

- ▶ Game Mechanics and Design Features for DGBLLT
- ▶ Game-related Social Practices
- ▶ Motivation and Affect in DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Baierschmidt, J. (2013). A principled approach to utilizing digital games in the language learning classroom. *JALT Call Journal, 9*(3), 307–315.

Barnett, J., Coulson, M., & Foreman, N. (2010). Examining player anger in World of Warcraft. In *Online worlds: Convergence of the real and the virtual* (pp. 147–160). Springer.

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge University Press.

Chen, A., Mari, S., Grech, S., & Levitt, J. (2020). What we know about massively multiplayer online role-playing games. *Harvard Review of Psychiatry, 28*(2), 107–112.

Chotipaktanasook, N., & Reinders, H. (2018). A massively multiplayer online role-playing game and its effects on interaction in the second language: Play, interact, and learn. In B. Zou & M. Thomas (Eds.), *Handbook of research on integrating technology into contemporary language learning and teaching* (pp. 367–389). IGI Global. https://doi.org/10.4018/978-1-5225-9618-9.ch023

Clement, J. (2024, February 29). *Topic: Online gaming*. Statista. https://www.statista.com/topics/1551/online-gaming/#topicOverview

Cole, D. A., Nick, E. A., & Pulliam, K. A. (2020). Are massively multiplayer online role-playing games healthy or not and why? Preliminary support for a compensatory social interaction model. *Computers in Human Behavior, 102*, 57–66.

Dixon, D. H., & Christison, M. A. (2018). The usefulness of massive multiplayer online role playing games (MMORPGs) as tools for promoting second language acquisition. In J. Perren, K. B. Kelch, J. Byun, S. Cervantes, & S. Safavi (Eds.), *Applications of CALL theory in ESL and EFL environments* (pp. 244–268). IGI Global. https://doi.org/10.4018/978-1-5225-2933-0.ch014

Dixon, D. H., & Christison, M. A. (2021). L2 gamers' use of learning and communication strategies in massively multiplayer online games (MMOs): An analysis of L2 interaction in virtual online environments. In K. B. Kelch, P. Byun, S. Safavi, & S. Cervantes (Eds.), *CALL theory applications for online TESOL education* (pp. 296–321). IGI Global. https://doi.org/10.4018/978-1-7998-6609-1.ch013

Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language Learning, 53*(S1), 3–32.

Freeman, G. (2018). *Multiplayer online games: Origins, players, and social dynamics*. AK Peters/CRC Press.

Ghobaei-Arani, M., Khorsand, R., & Ramezanpour, M. (2019). An autonomous resource provisioning framework for massively multiplayer online games in cloud environment. *Journal of Network and Computer Applications, 142*, 76–97.

Griffiths, M. D., Davies, M. N., & Chappell, D. (2003). Breaking the stereotype: The case of online gaming. *Cyberpsychology & Behavior, 6*(1), 81–91.

Gustafsson, V., Holme, B., & Mackay, W. E. (2020). Narrative substrates: Reifying and managing emergent narratives in persistent game worlds. In *Proceedings of the 15th international conference on the foundations of digital games* (pp. 1–12).

Jabbari, N., & Eslami, Z. R. (2019). Second language learning in the context of massively multiplayer online games: A scoping review. *ReCALL, 31*(1), 92–113.

Jabbari, N., & Eslami, Z. R. (2023). Negotiations for meaning in the context of a massively multiplayer online role-playing game. *Language Learning & Technology, 27*(1), 1–28. http://hdl.handle.net/10125/73517

Jabbari, N., & Peterson, M. (2023). Complexity, accuracy, and fluency improvements through massively multiplayer online gaming: A longitudinal mixed-methods case study. *The Language Learning Journal, 51*(4), 416–450. https://doi.org/10.1080/09571736.2023.2219713

Johnson, W. L. (2009). Developing intercultural competence through videogames. In *Proceedings of the 2009 international workshop on intercultural collaboration* (pp. 99–100).

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Linares, M., Gallego, M. D., & Bueno, S. (2021). Proposing a TAM-SDT-based model to examine the user acceptance of massively multiplayer online games. *International Journal of Environmental Research and Public Health, 18*(7), 3687.

Liu, K. Y., Yang, C. T., & Chang, K. H. (2012). Development of a multiplayer online role-playing game-based learning system for multiple curriculums. In *2012 IEEE fourth international conference on digital game and intelligent toy enhanced learning* (pp. 62–66). IEEE.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic.

Ma, L., Feng, Z., Feng, J., & Wang, L. (2019). Research on the factors affecting player stickiness in massively multiplayer online game (MMOG). In *2019 16th international conference on service systems and service management (ICSSSM)* (pp. 1–6). IEEE.

Peterson, M. (2010). Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning, 23*(5), 429–439.

Peterson, M. (2023). Digital simulation games in CALL: A research review. *Computer Assisted Language Learning, 36*(5–6), 943–967.

Peterson, M., White, J., Mirzaei, M. S., & Wang, Q. (2020). A review of research on the application of digital games in foreign language education. In M. Kruk & M. Peterson (Eds.), *New technological applications for foreign and second language learning and teaching* (pp. 69–92). IGI Global.

Rankin, Y., Morrison, D., McKenzie, M. C., Gooch, B., & Shute, M. (2009). Time will tell: In game social interactions that facilitate second language acquisition. In R. Michael Young (Ed.), *Proceedings of the 4th international conference on foundations of digital games* (pp. 161–168). ACM.

Reinders, H., & Wattana, S. (2014). Can I say something? The effects of digital game play on willingness to communicate. *Language Learning & Technology, 18*(2), 101–123.

Reinders, H., & Wattana, S. (2015). Affect and willingness to communicate in digital game-based learning. *ReCALL, 27*(1), 38–57. https://doi.org/10.1017/S0958344014000226

Reinhardt, J. (2021). Not all MMOGs are created equal: A design-informed approach to the study of L2 learning in multiplayer online games. In M. Peterson, K. Yamazaki, & M. Thomas (Eds.), *Digital games and language learning* (pp. 69–87). Bloomsbury.

Reinhardt, J., & Kirby, S. (2022). Second language acquisition and multiplayer gaming. In N. Ziegler & M. González-Lloret (Eds.), *The Routledge handbook of second language acquisition and technology* (pp. 286–296). Routledge.

Reinhardt, J., & Sykes, J. M. (2012). Conceptualizing digital game-mediated L2 learning and pedagogy: Game-enhanced and game-based research and practice. In H. Reinders (Ed.), *Digital games in language learning and teaching* (pp. 32–49). Palgrave Macmillan.

Ryu, D. (2011). Non-native English speakers' multiliteracy learning in beyond-game culture: A sociocultural study. *MERLOT Journal of Online Learning and Teaching, 7*(2), 231–243.

Statista. (n.d.). Video games – Worldwide: Statista market forecast. https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide

Tay, V. (2005). Massively multiplayer online game (MMOG)-a proposed approach for military application. In *2005 international conference on cyberworlds (CW'05)*. IEEE. https://doi.org/10.1109/CW.2005.61

Thorne, S. L. (2008). Transcultural communication in open internet environments and massively multiplayer online games. In S. S. Magnan (Ed.), *Mediating discourse online* (pp. 305–327). John Benjamins. https://doi.org/10.1075/aals.3.17tho

Thorne, S. L., Fischer, I., & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL, 24*(3), 279–301. https://doi.org/10.1017/S0958344012000158

Yahyavi, A., & Kemme, B. (2013). Peer-to-peer architectures for massively multiplayer online games: A survey. *ACM Computing Surveys (CSUR), 46*(1), 1–51.

Yee, N. (2006). Motivations for play in online games. *Cyberpsychology & Behavior, 9*(6), 772–775.

Zhang, K., Kemme, B., & Denault, A. (2008). Persistence in massively multiplayer online games. In *Proceedings of the 7th ACM SIGCOMM workshop on network and system support for games* (pp. 53–58).

Zheng, D., Young, M. F., Wagner, M. M., & Brewer, R. A. (2009). Negotiation for action: English language learning in game based virtual worlds. *The Modern Language Journal, 93*(4), 489–511. https://doi.org/10.1111/j.1540-4781.2009.00927.x

Zhong, Z. J. (2011). The effects of collective MMORPG (massively multiplayer online role-playing games) play on gamers' online and offline social capital. *Computers in Human Behavior, 27*(6), 2352–2363.

M

Motivation and Affect in Digital Game-Based Language Learning and Teaching

Nasser Jabbari
University of Essex, Colchester, UK

Abstract

This entry explores the emotional and motivational dimensions of digital game-based language learning and teaching. The analysis underscores the pivotal role of play in human learning and the significance of enjoyment in language acquisition. A review of studies examining various affective and motivational aspects in language learning and teaching revealed three key findings. First, digital game-based language learning enhances intrinsic and extrinsic motivation, enriching the learning experience. This heightened motivation improves engagement and outcomes within digital learning environments. Second, digital game-based language learning positively affects learners' enjoyment, a critical factor in sustaining engagement and interest in language learning tasks. Third, digital game-based language learning fosters collaboration and cooperation among learners, thereby enhancing teamwork skills and promoting positive attitudes through completing in-game tasks. In conclusion, future research should investigate the interplay between positive and negative emotions and the features and mechanics of educational digital games. Such exploration will elucidate the affordances of these games in enhancing the efficacy of language learning and teaching.

Keywords

Digital game-based language learning · Affect · Emotion · Motivation

The Role of Play in Human Learning

Learning through play is deeply ingrained in human psychology, serving as a fundamental mechanism through which we develop social, cognitive and sensi-motor skills (Piaget, 1951; Vygotsky, 1978). Play allows us to engage in learning activities eagerly, willingly and joyfully, leading to highly engaging and enjoyable learning experiences (Pellegrini, 2009; Hirsh-Pasek et al., 2009). Our natural inclination to learn through play provides a robust framework for educational strategies, particularly in language learning. As a structured form of play, games leverage this intrinsic motivation and present an effective means to facilitate learning (Gee, 2003). By tapping into the inherent enjoyment, excitement and engagement that play provides, games can offer a unique opportunity to enhance the learning process and make it more effective and enjoyable. This connection between play and learning shapes

© Crown 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_133-1

the basis for the growing interest in integrating games into educational contexts, particularly for second-language acquisition (Gee, 2003).

Digital Game-Based Language Learning and Teaching

The advent of Web 2.0 tools and digital technology has led to the proliferation of educational or 'serious' games, which gave rise to a distinct area of research known as Digital Game-Based Language Learning and Teaching (DGBLLT) (Reinhardt & Sykes, 2014; see ▶ DGBLLT). This research area focuses on using digital games specifically designed for educational purposes (e.g. *Scrabble*) to facilitate language learning. Reinhardt and Sykes (2014) differentiated DGBLLT from two frequently conflated concepts within the existing literature: game-enhanced and game-informed language learning and teaching. In digital game-enhanced language learning, commercial off-the-shelf (COTS) or 'vernacular' digital games, which are primarily designed for entertainment (e.g. *World of Warcraft*), are repurposed for educational purposes (e.g. see Jabbari and Eslami (2019) for review; see also ▶ DGBLLT and Entertainment Games). Game-informed language learning, on the other hand, involves applying game and play principles in both digital and nondigital contexts outside game boundaries, a practice also referred to as 'gamification' (Kapp, 2012) or 'gamefulness' (McGonigal, 2013) (see ▶ Gamification and LLT). This entry provides a brief overview of DGBLLT, specifically focusing on the emotional and motivational aspects of this innovative approach in language learning and teaching.

Digital games for educational purposes leverage their engaging, interactive and immersive nature to create effective language learning experiences (Mayer, 2014, 2019). These games incorporate elements such as fun, narrative, interaction, adventure, challenge and instant feedback to provide a dynamic and stimulating learning environment (Dickey, 2011). Engaging learners with meaningful and goal-oriented language use that is highly contextualised, DGBLLT aligns with language acquisition theories emphasising the importance of context and interaction in learning (Krashen, 1985; Swain, 2000). The playful environment of digital games enhances language learning by making the experience enjoyable and engaging, thereby supporting the development of language skills (e.g. Allen et al., 2014). Therefore, alongside the cognitive and sociocultural dimensions, the motivational and affective aspects of DGBLLT are essential for comprehending its effectiveness. Motivational factors, in conjunction with a diverse array of learning-relevant emotions—such as enjoyment, hope, pride, anxiety, anger, frustration, shame and boredom (Pekrun & Perry, 2014)—are widely recognised by scholars in second-language acquisition as significant factors in the language acquisition process. Various researchers in the field support this understanding (Dörnyei, 2001; Horwitz, 2001; Deci & Ryan, 2000; Gardner, 1985; MacIntyre, 1999). This entry explores how motivational and affective dimensions enhance learning experiences and outcomes in digital game-based language learning. We will review current literature on these factors, highlight effective strategies, identify research gaps and suggest future study directions.

Empirical Research on Motivation and Affect in DGBLLT

Sociocultural, cognitive, affective and motivational factors of learning are intricately interwoven and are mutually influential. With that in mind, it is critical to cast some lights on the emotional aspects of DGBLL, considering their impacts on the cognitive aspects of language learning through gaming. Research has shown that DGBLL profoundly impacts the emotional and motivational dimensions of language acquisition (e.g. Allen et al., 2014; Qiao et al., 2022). By creating an emotionally engaging and motivating learning environment, DGBLLT can harness the power of affective factors to facilitate language learning. The interactive and immersive nature of digital games can capture learners' attention and sustain their interest, the two critical components of effective language learning. Moreover, the

immediate feedback and rewards systems embedded in games help to build learners' confidence and reduce anxiety, which, in turn, culminates in creating a positive learning atmosphere (deHaan et al., 2010; see ▶ Feedback in DGBLLT). Digital games also foster a sense of achievement and progression through incremental challenges and milestones designed in the game setting, which can significantly boost learners' intrinsic motivation to explore and achieve more (e.g. Tsai et al., 2017). Research by Plass et al. (2015) highlighted that learners' emotional engagement and enjoyment derived from game-based learning can lead to higher levels of motivation and persistence in language learning tasks.

Our analysis of 21 empirical studies on motivation and affect in DGBLLT indicates that this approach can generate several positive outcomes for language learners. Here, we present an overview of these findings. The first key finding is that DGBLLT can stimulate both intrinsic and extrinsic motivation among learners (Iaremenko, 2017; Khalilian et al., 2021; Rajendran et al., 2019; Setiawan & Wiedarti, 2020; Sun & Hsieh, 2018; Tsai et al., 2017; Wichadee & Pattanapichet, 2018). This heightened motivation can, in turn, positively influence language learning outcomes (Berns et al., 2016; Liu & Chu, 2010). For example, Berns et al. (2016) demonstrated that using *VocabTrainerA1*, a hybrid game-based mobile application, significantly improved students' grammar and vocabulary scores. They contended that integrating individual learning tasks with engaging collaborative activities, which require students to use the target language for real-world negotiations, effectively motivated students and met their needs more efficiently than traditional learning approaches, positively impacting their learning outcomes.

Drawing on self-determination theory (Ryan & Deci, 2017), Qiao et al. (2022) found that game-based learning environments can enhance intrinsic motivation by satisfying learners' basic psychological needs for competence, autonomy and relatedness. The need for competence involves feeling effective and capable in one's activities. The need for autonomy refers to experiencing control and volition in one's actions and decisions. The need for relatedness pertains to feeling connected to others, belonging and experiencing caring relationships within one's community. Anisa et al. (2020) reported that using *Kahoot* boosted the learners' intrinsic motivation by promoting a sense of competence in them. They argued that *Kahoot* featured three critical factors related to perceived competence, according to Deci and Ryan (1985): optimally challenging in-game tasks, immediate feedback and feedback that serves as informational input for students. According to Anisa et al. (2020), *Kahoot* offered these critical elements by providing a challenging game, giving immediate feedback and informing learners if their responses were correct. Furthermore, Anisa and her colleagues discovered that using *Kahoot* significantly enhanced learners' sense of autonomy, as it empowered them to choose the games they wished to play, determine the challenges they aimed to tackle and set their own goals. The researchers further highlighted that applying *Kahoot* increased students' sense of relatedness and willingness to socialise, as they interacted with their friends more frequently and worked together on similar tasks and towards common goals. Huang and Huang (2015) also reported that utilising their handheld sensor-based vocabulary game increased low-achieving learners' motivation through the scaffolding mechanism implemented within the game, which helped the students tackle the challenges.

Research also indicates that DGBLLT can foster extrinsic motivation (Calvo-Ferrer, 2017) by incorporating elements such as points, levels and rewards (Zarzycka-Piskorz, 2016). In this context, Anisa et al. (2020) observed that external regulation, a type of extrinsic motivation, arises from external support and is associated with punishments and rewards. Their study found that students engaged with *Kahoot* to achieve high scores and win the game. The platform's rewarding system, including leaderboards and points and its competitive nature, stimulated learners' interest in playing. Additionally, their findings revealed that *Kahoot* contributed to two other types of extrinsic motivation: introjected regulation, which involves avoiding anxiety or guilt and

enhancing ego through feelings such as pride, and identified regulation, which entails endorsing goals. When asked whether *Kahoot* enabled them to showcase their intelligence and demonstrate their ability to succeed, most students reported feeling smart, as they could answer the questions accurately. The results also indicated that using *Kahoot* helped learners understand key concepts of the learning material, achieve learning objectives and be willing to learn the correct answers if they made mistakes.

The second key finding is that DGBLLT can positively contribute to learner's enjoyment (Rajendran et al., 2019). This enjoyment enhances language learners' engagement and motivation (Allen et al., 2014). It is important to reiterate that the primary aim of educational games is to create learning experiences that are both enjoyable and effective. Learners will likely disengage without enjoyment and avoid further interactions (Jackson & McNamara, 2013). When students enjoy the gaming experience, their persistent, attentive, competitive and explorative instincts, which facilitate learning, are amplified (Fithriani, 2021). In this context, a study by Allen et al. (2014) revealed that learners' level of enjoyment significantly predicted their motivation and engagement. They argued that despite the benefits of repetitive training, students can easily disengage from repetitive tasks within the game. Therefore, their enjoyment of the practice environment significantly influences second-language learners' engagement with learning activities. Allen et al. (2014) emphasised that leveraging L2 learners' inherent interest in games can enhance their motivation and engagement in training. They discovered that repeated writing strategy training using *Writing Pal* (W-Pal)—an intelligent tutoring system designed to provide explicit writing strategy instruction and practice in an engaging and adaptive environment—not only maintained learners' motivation and engagement levels but also positively influenced their writing performance and affect (i.e. engagement, motivation and perceived performance). Furthermore, the learners' attitudes towards the games impacted their perceived performance, as they rated the training sessions higher when they found the games beneficial, enjoyable and easy to play.

The third key finding highlights that educational games can significantly increase student cooperation and collaboration. The social aspects of many educational digital games provide opportunities for collaborative learning and interaction in the target language, which can help learners develop their communicative competence and cross-cultural understanding. For example, Connolly et al. (2011) reported that their developed Alternate Reality Game (ARG) for language learning fostered positive attitudes and increased teamwork skills, contributing to a more cooperative learning experience among learners. Both students and language teachers considered *The Tower of Babel* ARG a collaborative and stimulating environment that allowed students to communicate in different languages with their peers from across Europe. Utilising such technologies, typically absent in traditional classroom settings, creates opportunities for cooperation, collaboration and community-building among learners. This collaboration can enhance positive affective factors among language learners as they exchange ideas and work together to solve puzzles and complete various quests within the game.

In summary, incorporating digital games into language learning represents an innovative teaching method that aims to engage learners not just intellectually but also emotionally, socially and culturally. The interplay of cognitive, emotional, social and motivational aspects in DGBLLT can create a comprehensive learning experience, significantly improving language learning outcomes. The exploration of emotions in DGBLLT is still a relatively new field of study; hence, future research should continue to examine various emotional and motivational aspects of DGBLLT to understand how positive and negative emotions intersect with the mechanics of these games to provide optimum conditions for language learning and teaching. Furthermore, we recommend that researchers investigate the impact of using DGBLLT on other motivational aspects of L2 learning, such as grit, mindsets and goal orientations. Regarding emotional engagement in DGBLLT, we suggest that researchers explore

learners' psychological flow when using DGBLLT, as flow represents maximum emotional engagement in an activity. This can offer valuable insights on how to maximise the educational benefits of DGBLLT. Lastly, we advise that the examination of positive and negative emotions in DGBLLT encompass a broader range of emotions (e.g. disappointment, shame, regret) to more accurately depict students' emotional experiences in DGBLLT.

Cross-References

- ▶ DGBLLT
- ▶ DGBLLT and Entertainment Games
- ▶ Feedback in DGBLLT
- ▶ Gamification and LLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 writing practice: Game enjoyment as a key to engagement. *Language Learning & Technology, 18*(2), 124–150.

Anisa, K. D., Marmanto, S., & Supriyadi, S. (2020). The effect of gamification on students' motivation in learning English. *Leksika: Jurnal Bahasa, Sastra dan Pengajarannya, 14*(1), 22–28. https://doi.org/10.30595/lks.v14i1.5695

Berns, A., Isla-Montes, J. L., Palomo-Duarte, M., & Dodero, J. M. (2016). Motivation, students' needs, and learning outcomes: A hybrid game-based app for enhanced language learning. *Springerplus, 5*, 1–23. https://doi.org/10.1186/s40064-016-2971-1

Calvo-Ferrer, J. R. (2017). Educational games as standalone learning tools and their motivational effect on L2 vocabulary acquisition and perceived learning gains. *British Journal of Educational Technology, 48*(2), 264–278. https://doi.org/10.1111/bjet.12387

Connolly, T. M., Stansfield, M., & Hainey, T. (2011). An alternate reality game for language learning: ARGuing for multilingual motivation. *Computers & Education, 57*(1), 1389–1415. https://doi.org/10.1016/j.compedu.2011.01.009

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer. https://doi.org/10.1007/978-1-4899-2271-7

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268.

deHaan, J., Reed, W. M., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology, 14*(2), 74–94.

Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology, 42*(3), 456–469.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press.

Fithriani, R. (2021). The utilization of mobile-assisted gamification for vocabulary learning: Its efficacy and perceived benefits. *Computer-Assisted Language Learning Electronic Journal, 22*(3), 146–163.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave Macmillan.

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. G. (2009). *A mandate for playful learning in preschool: Presenting the evidence*. Oxford University Press.

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics, 21*, 112–126.

Huang, Y. M., & Huang, Y. M. (2015). A scaffolding strategy to develop handheld sensor-based vocabulary games for improving students' learning motivation and performance. *Educational Technology Research & Development, 63*(5), 691–708. https://doi.org/10.1007/s11423-015-9382-9

Iaremenko, N. (2017). Enhancing English language learners' motivation through online games. *Information Technologies and Learning Tools, 59*(3), 126–133. https://doi.org/10.33407/itlt.v59i3.1606

Jabbari, N., & Eslami, Z. R. (2019). Second language learning in the context of massively multiplayer online games: A scoping review. *ReCALL, 31*(1), 92–113. https://doi.org/10.1017/S0958344018000058

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*(4), 1036–1049. https://doi.org/10.1037/a0032580

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.

Khalilian, B., Hosseini, H., & Ghabanchi, Z. (2021). On the effect of employing the online Kahoot game-based app on Iranian EFL learners' structural ability and their motivation. *Journal of Language Teaching and Learning, 11*(2), 42–60.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on

learning outcomes and motivation. *Computers & Education, 55*(2), 630–643. https://doi.org/10.1016/j.compedu.2010.02.023

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24–45). McGraw-Hill.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology, 70*, 531–549.

McGonigal, J. (2013). *Reality is broken: Why games make us better and how they can change the world*. Penguin Press.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Pellegrini, A. D. (2009). *The role of play in human development*. Oxford University Press.

Piaget, J. (1951). *Play, dreams and imitation in childhood* (1st ed.). Routledge.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283.

Qiao, S., Yeung, S. S. S., Shen, X., & Chu, S. K. W. (2022). The effects of a gamified morphological awareness intervention on students' cognitive, motivational, and affective outcomes. *British Journal of Educational Technology, 53*(4), 952–976. https://doi.org/10.1111/bjet.13178

Rajendran, T., Naaim, N. A. B., & Yunus, M. M. (2019). Pupils' motivation and perceptions towards learning English using Quizvaganza. *International Journal of Scientific and Research Publications, 9*(1), 1–9. https://doi.org/10.29322/IJSRP.9.01.2019.p8529

Reinhardt, J., & Sykes, J. M. (2014). *Digital games and language learning: Theory, development and implementation*. Routledge.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publishing.

Setiawan, M. R., & Wiedarti, P. (2020). The effectiveness of Quizlet application towards students' motivation in learning vocabulary. *Studies in English Language and Education, 7*(1), 83–95. https://doi.org/10.24815/siele.v7i1.15359

Sun, J. C. Y., & Hsieh, P. H. (2018). Application of a gamified interactive response system to enhance the intrinsic and extrinsic motivation, student engagement, and attention of English learners. *Journal of Educational Technology & Society, 21*(3), 104–116. http://www.jstor.org/stable/26458511

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford University Press.

Tsai, C. H., Cheng, C. H., Yeh, D. Y., & Lin, S. Y. (2017). Can learning motivation predict learning achievement? A case study of a mobile game-based English learning approach. *Education and Information Technologies, 22*(5), 2159–2173. https://doi.org/10.1007/s10639-016-9542-5

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wichadee, S., & Pattanapichet, F. (2018). Enhancement of performance and motivation through application of digital games in an English language class. *Teaching English with Technology, 18*(1), 77–92.

Zarzycka-Piskorz, E. (2016). Kahoot it or not? Can games be motivating in learning grammar? *Teaching English with Technology, 16*(3), 17–36.

R

Research Practices in DGBLLT

Frederick J. Poole
Michigan State University, East Lansing
MI, USA

Abstract

This entry provides an overview of research methods applied in digital game-based language learning (DGBLL). Data collection approaches and analytical methods commonly used in DGBLL studies, including quantitative, qualitative and mixed methods approaches, are explored to illustrate current research trends in the field. The entry concludes with recommendations for future research.

Keywords

Research methods · Quantitative approaches · Qualitative approaches

Introduction

Compared to other subfields in applied linguistics, research on digital-game based language learning (DGBLL) is still in its infancy. This is evident by a recent a meta-analysis which was only able to identify 26 studies investigating DGBLL with the appropriate measures (reporting of means and SD to calculate effect sizes) to be applied to a meta-analysis (Dixon et al., 2022). Thus, while DGBLL is an exciting area of research full of potential, much work still needs to be done to improve the quality and breadth of research currently being done. In this entry, a survey of research methods that have been used to explore DGBLL will be reviewed.

Before going any further, it is important to first define key terms. Of importance to this entry are the terms 'research methods' and 'game-based language learning'. As McKinley (2019) notes, research methods are often interpreted in a variety of ways. Some may see research methods as the design of a study, others view it as analytical orientations (quantitative vs qualitative) and yet others focus more on data collection procedures. For this entry, research methods will be explored by identifying key data collection approaches and the data analyses applied to such data. The term 'game' within DGBLL has also been interpreted in various ways, with an important distinction between games and gamification. Gamification involves adding gaming elements to a nongame setting, like Kahoot and Quizlet, in which removing these elements leaves the original activity intact (e.g. multiple-choice questions) (see ▶ "Gamification and LLT"). In contrast, games involve rich contexts, goals and rules, in which removing such gaming elements also removes the gaming activity. Distinguishing between game-based language learning and gamification is important because they elicit different learning behaviours. DGBLL places students in

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_138-1

context-rich environments with roles and goals promoting authentic communication, while gamification adds competition mechanics to non-game scenarios to boost engagement with drill-based tasks. Researchers focus on enjoyment, motivation and persistence in gamification, while DGBLL focuses on how gaming environments promote authentic learning. This entry focuses on game-based language learning, not gamification (for more on this distinction, see York et al., 2022).

In the last 5 years, there have been several systematic reviews that have provided an overview of the field (e.g. Poole & Clarke-Midura, 2020; Yang et al., 2024). These systematic reviews have a few common themes. First, research on games is overly focused on English as a Foreign Language contexts, vocabulary learning in lieu of other language competencies (see ▶ Vocabulary and DGBLLT) and learners' enjoyment of games. Thus, it should not be surprising that one of the most common forms of data collection involve pre- and post-vocabulary tests (Yang et al., 2024). The second most widely used method involves the use of questionnaires. These questionnaires often involve a large battery of affective questions that target constructs like enjoyment, anxiety, motivation, difficulty and immersion, among many others (see Motivation and Affect in DGBLLT). The majority of these tend to focus on learners' enjoyment and/or perception of the game. Research on games has long sought to show that games are enjoyable, and they promote learning. Studies have taken quantitative, qualitative and mixed-method approaches. This entry will provide examples of quantitative and qualitative approaches to DGBLL, followed by some recent innovative approaches to investigating DBGLL.

Quantitative Approaches

Studies relying on a quantitative data and analyses often employ quasiexperimental approaches in which the experimental (treatment) group receives a digital gaming intervention and the comparison (control) group receives an alternative instructional approach. Similar to the applied linguistics field as a whole, much of this research relies on null hypothesis significance testing using one of the three main linear models: t-tests, ANOVA and correlations (Loewen & Godfroid, 2019). For example, Franciosi et al. (2016) used t-tests to compare learning gains between a gaming group and a control group. The authors found that while there was no difference in learning gains on an immediate vocabulary test, the gaming group demonstrated significant differences in a delayed test, suggesting that the gaming group had higher long-term vocabulary retention. Hitosugi et al. (2014) found similar results using a repeated measure ANOVA. However, instead of using multiple t-tests at each time point which increases the likelihood of a Type 1 error, the ANOVA allows all time points to be included in the analysis at once. Finally, Collentine (2011) used a linear regression in a nonexperimental study to explore the relationship between players in-game actions their language production complexity. The author found that learners who interacted with the non-player characters (NPCs) more tended to produce language with more syntactical complexity. Researchers tend to use quantitative approaches to establish differences between gaming and non-gaming interventions or to explore associations between gaming mechanics and learning outcomes.

Qualitative Approaches

Studies exploring DGBLL from a qualitative approach generally use student interviews, surveys and gameplay observations to explore both the affordances that games offer for language learning and student perspectives. Such data has been analysed using content analysis (Lee, 2019), conversation analysis (Marsh & Tainio, 2009) and discourse analysis (Ibrahim, 2019), among many others. These studies usually involve smaller samples but involve longer interventions. For instance, Marsh and Tainio (2009) followed two players for a year as they played *Final Fantasy*. Through conversation analysis, they illustrated

how two players engaged in dialogue around game and further how positive linguistic practices such as repetition and language play emerged naturally through gameplay. Similarly, Ibrahim (2019) used discourse an analysis to investigate gaming journals and think-aloud protocols. They used this data to identify how participants engaged with and decoded in-game foreign language discourse. Players were observed breaking down unfamiliar phrases into familiar words, using contextual clues to infer meanings and repeating phrases to internalize new vocabulary. Finally, Lee (2019), using a larger data set of 60 students, analysed student writing after playing the game 'Her Story' and was able to illustrate how the game inspired learners to engage in creative writing. Qualitative approaches have shown the rich interactions that games can promote and the language learning opportunities that they afford.

Mixed Methods and Other Innovative Approaches

The gold standard for research are mixed-methods approaches which involve the collection of quantitative and qualitative data to provide a more comprehensive understanding of the intervention impact. For example, Dourda et al. (2014) investigated a DGBLL intervention in a content and language integrated learning unit using pre- and postcontent knowledge and vocabulary assessments to measure learning quantitatively and then explored student learning journals and video observations to identify learning strategies employed during gameplay that may have resulted in learning. Similarly, Poole et al. (2022) explored the impact of using a digital game in a Chinese dual language immersion classroom by using pre- and postvocabulary and reading comprehension assessments to show learning occurred as a result of the gaming intervention. Then, qualitative analysis was used to investigate audio recordings of gameplay which identified the affordances of bringing digital games into classrooms including peer and teacher support that likely contributed to the observed learning gains. Mixed-method approaches are ideal because qualitative data can provide insight into observed quantitative differences.

In addition to mixed-methods approaches, in recent years, researchers have started to apply more complex statistical analysis to DGBLL research. Dixon (2022) used a discriminant analysis to compare the linguistic complexity between three major commercial-off-the-shelf games. Through this approach, the author was able to demonstrate that while linguistic complexity does not differ significantly across games, it does differ within game elements (e.g. objectives, dialogue, narrative; see Game Mechanics and Design Features for DGBLLT). Liu and Hwang (2024) used a lag sequential analysis to identify unique patterns that players implemented within a digital game designed for language learning. They used these patterns to illustrate how metalinguistic feedback provided opportunities for deeper learning within a game. Finally, Poole and Clarke-Midura (2023) applied data mining techniques to log data collected in a digital game to identify clusters of gaming behaviour and then illustrated how those clusters were associated with learning outcomes. These innovative approaches not only provide support for designing classes around games but also insight into how games impact learners at the individual level.

Look to the Future

Games are language-rich environments that are extremely engaging; thus it is unsurprising that researchers explore student interest in games. However, it is time that we move past simple comparative studies that measure vocabulary learning and enjoyment at two distinct points in time and explore approaches that allow for an understanding of how DGBLL effects learners at the individual level (Hiver & Nagle, 2024). This can be done through mixed-methods approaches, by using mixed-effects linear modelling to explore nested learning environments, and through data mining approaches that allow for unique research environments.

Cross-References

▶ Gamification and LLT
▶ Motivation and Affect in DGBLLT
▶ Vocabulary and DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Collentine, K. (2011). Learner autonomy in a task-based 3D world and production. *Language Learning & Technology, 15*(3), 50–67. http://llt.msu.edu/issues/october2011/collentine.pdf.

Dixon, D. H. (2022). The linguistic environments of digital games: A discriminant analysis of language use in game mechanics. *CALICO Journal, 39*(2), 150.

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. *Language Learning & Technology, 26*(1), 1–25. http://hdl.handle.net/10125/73464

Dourda, K., Bratitsis, T., Griva, E., & Papadopoulou, P. (2014). Content and language integrated learning through an online game in primary school: A case study. *Electronic Journal of e-Learning, 12*(3), 243–258.

Franciosi, S. J., Yagi, J., Tomoshige, Y., & Ye, S. (2016). The effect of a simple simulation game on long-term vocabulary retention. *CALICO Journal, 33*(3), 355–379.

Hitosugi, C. I., Schmidt, M., & Hayashi, K. (2014). Digital Game-Based Learning (DGBL) in the L2 classroom: The impact of the UN's off-the-shelf videogame, food force, on learner affect and vocabulary retention. *CALICO Journal, 31*(1), 19–39.

Hiver, P., & Nagle, C. (2024). Complex adaptive interventions: The challenge ahead for instructed second language acquisition research. *Annual Review of Applied Linguistics*, 1–16.

Ibrahim, K. (2019). Foreign language practice in simulation video games: An analysis of game-based FL use dynamics. *Foreign Language Annals, 52*(2), 335–357.

Lee, S. M. (2019). Her story or their own stories? Digital game-based learning, student creativity, and creative writing. *ReCALL, 31*(3), 238–254.

Liu, L. A., & Hwang, G. J. (2024). Effects of metalinguistic corrective feedback on novice EFL students' digital game-based grammar learning performances, perceptions and behavioural patterns. *British Journal of Educational Technology, 55*(2), 687–711.

Loewen, S., & Godfroid, A. (2019). Advancing quantitative research methods. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (1st ed., pp. 98–107). Routledge. https://doi.org/10.4324/9780367824471-9

Marsh, A. P., & Tainio, L. (2009). Other-repetition as a resource for participation in the activity of playing a video game. *The Modern Language Journal, 93*(2), 153–169.

McKinley, J. (2019). Theorizing research methods in the 'golden age' of applied linguistics research. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (1st ed., pp. 1–12). Routledge.

Poole, F. J., & Clarke-Midura, J. (2020). A systematic review of digital games in second language learning studies. *International Journal of Game-Based Learning (IJGBL), 10*(3), 1–15.

Poole, F. J., & Clarke-Midura, J. (2023). Applying educational data mining to explore individual experiences in digital games. *Language Learning & Technology, 27*(1), 1–26. http://hdl.handle.net/10125/73522

Poole, F., Clarke-Midura, J., & Ji, S. (2022). Exploring the affordances and effectiveness of a digital game in the Chinese dual language immersion classroom. *Journal of Technology and Chinese Language Teaching, 13*(1), 46–73.

Yang, L., Li, R., & Zhou, Y. (2024). Research trends of game-based language learning in K-12 education: A systematic review of SSCI articles during 2009–2022. *Journal of Computer Assisted Learning, 40*, 1218.

York, J., deHaan, J., Childs, M., & Collins, M. (2022). How is gamification like being trapped in the matrix? And what is the "real-world" of game-based learning. *Digital Culture & Education, 14*(3), 35–54.

S

Simulation Games in Language Learning & Teaching

Mark Peterson
Kyoto University, Kyoto, Japan

Abstract

In recent years, digital games have increasingly been investigated in computer-assisted language learning (CALL) research. Of the game types currently under investigation, digital simulation games have attracted significant interest from researchers. This entry will focus on this phenomenon. The discussion will first focus on defining the nature of this genre of digital game, and identifying distinguishing features. The hypothesized advantages of utilizing this type of game in language learning will be discussed from the perspective of contemporary SLA research. The discussion will then report noteworthy findings from the research base. Finally, areas of potential for future research will be identified.

Keywords

Digital simulation game · SLA · CALL · Game genres

Introduction

As technological advances continue, CALL researchers have taken the lead in exploring the potential of digital technologies to support the processes involved in second language acquisition (SLA). One significant aspect of this development concerns the use of contemporary digital games (Peterson & Jabbari, 2022). Although work in this area is long running (Peterson, 2013), the emergence of a wide range of accessible digital games in recent years has given fresh impetus to research work on their use as arenas for foreign language learning (Gao & Gee, 2023). One class of game that has been the focus of considerable attention is simulation games (Yang et al., 2022). This genre of game shares some features with other types of digital games that incorporate simulations. For example, simulation games provide players with access to persistent simulated virtual worlds that combine high-quality 3D graphics, avatars and multimedia content. In common with many digital games investigated in CALL research, these games are designed to elicit purposeful learner-centred activities that involve role play, by offering the player opportunities to engage in challenging problem-based scenarios. As with other types of games, players of simulation games can engage in content creation within the game and participate in game-focused online communities.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_139-1

However, there are also significant differences. Unlike many role-playing games that incorporate fantasy themes, digital simulation games provide a focus on real-world activities such as those encountered in business or everyday life (Peterson, 2023). Moreover, in contrast to many digital games, simulation games can be open-ended in nature and do not always have predetermined end states. The particular combination of features provided by simulation games has led to claims that this type of game has great potential as a tool language education (Purushotma, 2005). Drawing on SLA theory, a number of hypothesized benefits of the use of this type of game have been identified in the literature.

Use of Digital Simulation Games in CALL: Rationales

One influential rationale for the use of digital simulation games in CALL draws on psycholinguistic SLA theory. The learner-centred, highly interactive and purposeful nature of the game play in simulation games is perceived as providing a number of benefits. In the commercial simulation game, *The Sims*, the game rules compel individual players to solve problems encountered in everyday life (such as keeping a job) in order to prolong the life of their character. This type of purposeful target language (TL)-focused activity provides challenges and plentiful opportunities to engage in beneficial forms of TL practice. Game feedback that incorporates visual cues such as speech bubbles and annotations, offers possibilities for raising metalinguistic awareness regarding appropriate TL use (see ▶ Feedback in DGBLLT). Moreover, in many simulation games players are exposed to a wide range of everyday day vocabulary. Repeated exposure to this kind of TL input is seen as particularly valuable for learners of lower levels of language proficiency (Peterson, 2023). The everyday language found in this type of game appears well suited to the needs of this learner group as researchers note that learners find the complex game-specific registers often present in other types of digital games to be challenging (Thorne et al., 2009). The turn-based single-player nature of these games is beneficial for language learners as it provides extra time and a high degree of control. This later factor may foster agency and autonomy. In addition, the use of player avatars provides anonymity. This factor may enhance risk-taking, engagement, motivation and reduce affective barriers to learning, such as anxiety (see ▶ Motivation and Affect in DGBLLT).

An alternative justification for the use of simulation games in CALL is based on insights obtained from accounts of SLA that stress the social, contextualized nature of learning. As simulation games immerse players in TL-rich contexts where the TL must be used as a resource to solve problems, it is asserted that they can facilitate situated learning (Hitosugi et al., 2014). This rationale further emphasizes the opportunities for language use and learning presented by the online player communities associated with many types of digital games (Thorne et al., 2012). It is noted that game-specific fan sites and discussion boards are frequently international in nature providing learners with opportunities to access TL-based communities of practice (see ▶ Game-Related Social Practices). In these online social groups, learners can interact in the TL, collaborate, develop intercultural skills, and experience language socialization (Sykes et al., 2010). The apparent potential of simulation games as venues for language learning has attracted considerable interest from CALL researchers. In a growing body of work, researchers have investigated the potential of simulation games in a number of learner-based studies.

Research Findings

Research on the use of simulation games is currently focused on a variety of areas. In terms of vocabulary learning, a number of studies provide evidence that sustained game play in simulation games may facilitate the acquisition of new vocabulary. In an example of this work, a study by Chen and Hsu (2020) involved upper-intermediate-level ESL learners playing the simulation game *Slave Trade*. The researchers reported statistically significant gains in vocabulary on post-tests. In large-scale research conducted by Lin (2015), it was

found that when three groups of EFL learners played the multiplayer simulation game *Trade Ruler* each group outperformed a control in immediate vocabulary post-tests that involved 27 unique words used in the game. There is also evidence that the new vocabulary encountered during game play may be retained. In two projects that involved use of mixed methods carried out by Hitosugi et al. (2014), data analysis indicated that participants who played *Food Force* as part of a regular Japanese language course displayed enhanced retention of game vocabulary compared to when they used a textbook. A study reported by Franciosi et al. (2016) mirrors these findings. In this research, an experimental group that played *Third World Farmer* outperformed a control in a delayed post-test of game-related vocabulary administered 11 weeks after the main project was completed. In two further studies, Franciosi (2017) investigated the potential of vocabulary transfer to out-of-game contexts. This work involved game play in *Energy City*. The first project involved two learner groups. During the research, one group of participants learned target vocabulary using a drill application and then used the vocabulary in the game while the other group did not. Both groups then completed a writing task. In the second project, learners used targeted vocabulary while playing the above game and used this vocabulary in a writing task. Quantitative analysis revealed that in both studies, vocabulary used in the game appeared to be retained and transferred to an out-of-game context (see ▶ Vocabulary and DGBLLT).

A further significant area of investigation in research concerns the development of language skills. The findings of several leaner-based studies provide preliminary evidence suggesting that playing simulation games may have a positive impact on listening skills. deHaan (2005) reported that learners of Japanese as a foreign language who played the baseball simulation game *Jiikyoo Pawafuru Pro Yakkyu 6* over a one-month period claimed that their listening skills improved. In a study that involved the military simulation game *Americas Army,* Anderson and his associates (2008) found a limited improvement in learners' listening comprehension skills on a post-test. Researchers have also explored the potential of simulation games to enhance L2 writing skills. A study by Yang and Zapata-Rivera (2010) involved the use of a simulation game designed to elicit the production of pragmatic routines frequently found in academic contexts. The game provided opportunities for participants to interact in the L2 with a pedagogical agent embedded in the game. Data analysis suggested that the game provided valuable L2 writing practice relating to the production of pragmatically appropriate TL output (see ▶ Pragmatics and DGBLLT). The researchers further claimed that features of the game such as exposure to TL input, opportunities to produce L2 output, visual cues, feedback and a points system enhanced participation.

Affective factors represent another area that has been the focus of considerable research. Researchers in several studies report broadly favourable learner attitudes towards language learning with simulation games. Learner feedback indicates that many learners assert playing this type of game to be an enjoyable experience (Hitosugi et al., 2014; Miller & Hegelheimer, 2006). Researchers report that the repeated exposure to TL vocabulary and the provision of real-time feedback appear to assist learning (Chen & Hsu, 2020). These findings are mirrored in feedback regarding the use of this type of game in combination with supplementary materials and teacher intervention (Ranalli, 2008; Wang, 2019). Moreover, it is reported that the learner-centred nature of the gameplay appears to foster engagement, agency, and autonomy (Lorenset & Tumolo, 2019). Studies further indicate that playing this type of game frequently enhances learner motivation and reduces barriers to learning such as anxiety (Ibrahim, 2018; Yang et al., 2022). Researchers assert that a number of game features may account for this phenomenon. These include the provision of an engaging, realistic storyline, immersion, and multimodal TL feedback (Franciosi, 2017; Yang & Zapata-Rivera, 2010). Although simulation games are widely reported as providing learners with opportunities for learning, several issues with their use have been identified. Studies have found that a minority of learners report encountering difficulties when playing this type of game

(Anderson et al., 2008; Ranalli, 2008). Furthermore, in some projects, learners have claimed that teacher intervention is required as a learning support during gameplay (Wang, 2019; see ▶ Teacher Mediation in DGBLLT). In addition, as is the case with the wider research on the use of digital games in language learning findings are not definitive, and many areas remain unexplored. Moreover, it is noted in the literature the research base is subject to several limitations (Peterson, 2023). In this context, additional learner-based studies appear necessary.

Future Research Directions

Although current research is subject to limitations, the broadly positive findings reported in the literature draw attention to areas of potential in future work. One area of promise in future research concerns vocabulary learning. As noted previously, several studies suggest that the extensive exposure to everyday TL vocabulary provided by simulation games may enhance learning. Furthermore, there is limited evidence that the knowledge of this vocabulary may be transferred to real-world contexts. In this context, further longitudinal research involving larger and more diverse learner groups than have been explored to date appears warranted. Furthermore, implementing replication studies that adopt a wider range of data analysis approaches than have been implemented to previously offers the prospect of enhancing the research base, providing opportunities for new points of departure in research in this area.

A further area of likely interest in future research concerns the use of simulation games by language learners of lower levels of language proficiency. A noteworthy feature of the research on the use of games in CALL is the heavy emphasis on research involving intermediate and advanced-level learners (Peterson & Jabbari, 2022). Current studies suggest that the combination of features provided by simulation games, including the exposure to everyday vocabulary and authentic tasks, make them ideally suited to the needs of lower-level learners. Although research indicates that simulation games may be beneficial, further studies appear necessary given that learner-based studies involving this learner group are infrequent. Additional research may bring insights into the benefits of using simulation games with this significant learner group.

Another area of interest in future research concerns the role of the teacher. As noted at an earlier stage of this discussion, it has been reported that teacher interventions during game play may support learning. Although studies in this area are limited, promising findings have been reported regarding the application of teacher-designed learning scaffolds in projects involving simulation games (Miller & Hegelheimer, 2006; Ranalli, 2008). Moreover, it has also been reported that teacher interventions during game play are claimed to be beneficial (Wang, 2019). As these findings are encouraging, they suggest that further research in these areas may shed new light on the potential of both direct and indirect teacher interventions in learner-based CALL projects involving simulation games.

Cross-References

- ▶ Feedback in DGBLLT
- ▶ Game-Related Social Practices
- ▶ Motivation and Affect in DGBLLT
- ▶ Pragmatics and DGBLLT
- ▶ Task-Based Digital Game-Based Language Learning
- ▶ Teacher Mediation in DGBLLT
- ▶ Vocabulary and Digital Game-Based Language Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anderson, T., Reynolds, B. L., Yeh, X.-P., & Huang, G.-Z. (2008). Video games in the English as a foreign language classroom. In *2008 Second IEEE International Conference on Digital Game and Intelligent Toy*

Enhanced Learning (pp. 188–192). https://doi.org/10.1109/digital.2008.39

Chen, H.-J., & Hsu, H.-L. (2020). The impact of a serious game on vocabulary and content learning. *Computer Assisted Language Learning, 33*(7), 811–832. https://doi.org/10.1080/09588221.2019.1593197

deHaan, J. W. (2005). Acquisition of Japanese as a foreign language through a baseball video game. *Foreign Language Annals, 38*(2), 278–282. https://doi.org/10.1111/j.1944-9720.2005.tb02492.x

Franciosi, S. J. (2017). The effect of computer game-based learning on FL vocabulary transferability. *Educational Technology & Society, 20*(1), 123–133. https://doi.org/10.30191/ETS.201701_20(1).0011

Franciosi, S. J., Yagi, J., Tomoshige, Y., & Ye, S. (2016). The effects of a simple simulation game on long-term vocabulary retention. *CALICO Journal, 33*(3), 1–25. https://doi.org/10.1558/cj.v33i2.26063

Gao, Y. B., & Gee, E. R. (2023). Defining game-mediated second and foreign language teaching and learning: A review. *International Journal of Game-Based Learning, 13*(1), 1–15. https://doi.org/10.4018/IJGBL.323210

Hitosugi, C. I., Schmidt, M., & Hayashi, K. (2014). Digital game-based learning (DGBL) in the L2 classroom: The impact of the UN's off-the-shelf videogame Food Force, on learner affect and vocabulary retention. *CALICO Journal, 31*(1), 104–124. https://doi.org/10.11139/cj.31.1.19-39

Ibrahim, K. (2018). Player-game interaction: An ecological analysis of foreign language gameplay activities. *International Journal of Game-Based Learning, 8*(1), 1–18. https://doi.org/10.4018/IJGBL.2018010101

Lin, H. (2015). Effectiveness of interactivity in a web-based simulation game on foreign language vocabulary learning. *Procedia – Social and Behavioral Sciences, 182*, 313–317. https://doi.org/10.1016/j.sbspro.2015.04.772

Lorenset, C. C., & Tumolo, C. H. S. (2019). Vocabulary acquisition in English as a foreign language: Digital gameplaying the Sims. *Revista Linguagem & Ensino, 22*(4), 1002–1019. https://doi.org/10.15210/rle.v22i4.16642

Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education, 3*(4), 311–328. https://doi.org/10.1108/17415650680000070

Peterson, M. (2013). *Computer games and language learning*. Palgrave Macmillan. https://doi.org/10.1057/9781137005175

Peterson, M. (2023). Digital simulation games in CALL: A research review. *Computer Assisted Language Learning, 36*(5–6), 943–967. https://doi.org/10.1080/09588221.2021.1954954

Peterson, M., & Jabbari, N. (2022). Digital games and foreign language learning: Context and future development. In M. Peterson & N. Jabbari (Eds.), *Digital games in language learning and teaching: Case studies and applications* (pp. 1–13). Routledge. https://doi.org/10.4324/9781003240075-1

Purushotma, R. (2005). Commentary: You're not studying, you're just. . .. *Language Learning & Technology, 9*(1), 80–96. http://doi.org/10125/44010

Ranalli, J. (2008). Learning English with the Sims: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning, 21*(5), 441–455. https://doi.org/10.1080/09588220802447859

Sykes, J. M., Reinhardt, J., & Thorne, S. L. (2010). Multiuser digital games as sites for research and practice. In F. M. Hult (Ed.), *Directions and prospects in educational linguistics* (pp. 117–135). Springer. https://doi.org/10.1007/978-90-481-9136-9_8

Thorne, S. L., Black, R. W., & Sykes, J. M. (2009). Second language use, socialization, and learning in internet interest communities and online gaming. *The Modern Language Journal, 93*(S1), 802–821. https://doi.org/10.1111/j.1540-4781.2009.00974.x

Thorne, S., Fischer, I., & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL, 24*(3), 279–301. https://doi.org/10.1017/S0958344012000158

Wang, J. Q. (2019). Classroom intervention for integrating simulation games into language classrooms: An exploratory study with the SIMs 4. *CALL-EJ, 20*(2), 101–127.

Yang, H.-C., & Zapata-Rivera, D. (2010). Interlanguage pragmatics with a pedagogical agent: The request game. *Computer Assisted Language Learning, 23*(5), 395–412. https://doi.org/10.1080/09588221.2010.520274h

Yang, Y.-F., Hsieh, W.-M., Wong, W.-K., Hong, Y.-C., & Lai, S.-C. (2022). Reducing students' foreign language anxiety to improve English vocabulary learning in an online simulation game. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2039203

S

Student Game Design for LL

Torben Schmidt
Institute of English Studies, Leuphana University, Lüneburg, Germany

Abstract

This entry explores the transformative impact of game development activities (digital and non-digital) on diverse facets of student learning, with a special emphasis on foreign language learning. Focusing on creativity, problem-solving skills, motivation, attitudes, literacy, language skills, and digital capabilities, this review integrates findings from various studies to underscore the potential of game design activities in enriching both cognitive and affective dimensions of education. Across diverse contexts and age groups, the collective evidence illuminates the positive influence of student game design for learning, shedding light on its ability to engage and enhance students' overall learning experiences.

Keywords

Game design · Creativity · Design-thinking · Game literacies

Introduction

In recent years, the practice of learning through designing (digital) games has gained recognition as a promising method for enhancing educational outcomes among learners. Game design can be seen as a pedagogically significant activity that can actively engage students in cooperative or collaborative knowledge-creation processes. It is an endeavour with a dual focus, emphasizing both the game as a product and the design process leading to its development. In this entry, the terms "game design" and "game development" refer interchangeably to the "holistic, iterative process of game creation from initial ideas to computational implementation and production of the playable final product" (Laakso et al., 2021, p. 3). Game design as an innovative and interdisciplinary learning approach can potentially not only bolster literacy development, but also facilitate interdisciplinary learning (see Gaming Literacies and LLT). By engaging in the creative process of game design, learners develop critical thinking and problem-solving skills in a dynamic and interactive environment. The process of game design requires learners to understand and apply complex concepts in a practical context, thereby reinforcing their comprehension and retention. Through iterative design and

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_140-1

testing, students can also develop resilience and adaptability, learning to navigate challenges and improve their work based on feedback. Ultimately, the integration of (digital) game design in educational settings represents a forward-thinking approach that aligns with contemporary pedagogical goals, preparing students for the demands of the twenty-first century. Based on a literature review, the purpose of this entry is to present and discuss the potential and effects of utilizing game development as a method of learning in general and language learning in particular.

Student Game Development as a Design Process

The process of developing an analog or digital game by students in the classroom is multidisciplinary and involves several steps: It usually starts with the (1) *planning and conceptualization phase*, which includes brainstorming, goal setting, selecting game mechanics, defining requirements, and creating a narrative framework. Next comes the (2) *pre-production phase*, which involves creating prototypes and sketches, and the (3) *production phase*, which involves developing materials, User Interface (UI) and User Experience (UX) designs, level designs, graphics, audio, video, and programming. After that, (4) *playtesting and iterative improvement phases* follow based on user feedback until (5) *publication and distribution processes* can be addressed. Finally, a (6) *final evaluation of* the goals is completed, *and reflections* on the development process and the final product can take place. The summary of the steps involved, and the skills required to successfully complete these phases, suggest that the complexity and scope of relevant stages can bring about very diverse processes of learning and skills acquisition.

Student Game Design as a Learning Method in Diverse Educational Contexts

Studies on the use of game design activities in student learning encompass a diverse range of contexts, age groups, and educational settings (see also Game-based LLT Environments; Learner Variables in DGBLLT). Relevant studies provide a comprehensive overview of the significant benefits of engaging in educational game development activities for students' learning and skill development. Bulut et al. (2022), for instance, show that educational game design processes significantly enhance students' creativity. The study revealed an increase in students' creativity, comprising the dimensions of fluency, flexibility, originality and elaboration, through the game design process of 5th- and 6th-grade students in Istanbul, Turkey. Baytak and Land (2010), in their case study to investigate the processes, strategies and experiences of 5th-grade students during a computer game design project, find that game design can heighten engagement and collaboration skills, enhancing problem-solving abilities and content knowledge. Vos et al. (2011) demonstrate that students who actively worked on developing their own games exhibited a higher level of intrinsic motivation and a more profound use of strategy than their peers who only played games developed by others. Akcaoglu and Koehler (2014) assert that students' abilities in the domains of system analysis and design, troubleshooting, and decision-making exhibited notable growth, and that game development projects can potentially constitute an efficacious framework for developing problem-solving abilities and pertinent technical skills. Korkmaz et al. (2023) point out that game development exerts a positive influence on students' attitudes toward science and knowledge acquisition. Gallagher and Grimm (2018) demonstrate that game development enhances creativity, spatial skills, and positive attitudes toward science, technology, engineering, and mathematics (STEM) in women. A study by Hwang et al. (2013) shows that using peer assessment in game development can enhance learning outcomes and motivation. Robertson and Howells (2008) emphasize the importance of challenge and peer feedback in fostering motivation and independent learning.

Student Game Design for Developing Twenty-First-Century Skills

In addition to these results from various disciplines and with reference to various skills, a fairly large number of research papers are specifically concerned with the effects of the use of digital game development-based learning approaches on the development of IT- and job-related future skills (Bewer & Gladkaya, 2022). Romero and Lambropoulos (2015), for example, emphasize the particular importance of digital game development for learning as a creative process for the development of the twenty-first-century skills based on the 4Cs of communication, collaboration, creativity, and critical thinking. Numerous studies also show effects on the training of computer science principles, such as computational thinking, and on the development of programming skills in general (Dlab & Hoic-Bozic, 2021), as well as on the formation of positive attitudes towards programming and computer science (Dekhane et al., 2013). Additionally, participation in game development processes can also improve students' artistic digital competences (Laakso et al., 2021), such as game-art design and visualization techniques like 3D modelling and animation.

Student Game Design as a Language Learning Method and Literacy Practice

Although there are still very few studies that are specifically focused on language and foreign language learning through student game design, using student (digital and analog) game design approaches can offer a variety of additional opportunities and benefits for supporting diverse language learning processes. Game design as a language learning method is more in line with modern, communicative, task-based, and project-based foreign language learning pedagogy, as it shifts the focus to active engagement, creativity, problem-solving, language use, and language focus in meaningful contexts with a process-orientation and a clear outcome, in contrast to conventional language learning methods that rely on memorization and repetitive exercises and a focus on form(s) (see Task-Based DGBLLT). Students can develop critical thinking, collaboration skills, and (digital) game literacy with a critical understanding of games as cultural artifacts made up of complex narratives, systems, and game mechanics by immersing themselves in the process of creating games or just selected elements of games like a storyline, a character description, a mockup of gameplay, or just content for an existing game format (see Game Mechanics and Design Features for DGBLLT; DGBLLT and Narrative; Breakout Games and DGBLLT). For example, Owston et al. (2009) use game design as a pedagogical activity to engage students in curriculum-related literacy activities. In their study with 4th graders of nine public elementary schools in Canada, they find that the group of learners who had to develop content, questions, and answers for a trivia game achieved significantly better results in the areas of question formulation and logical sentence construction compared to a control group. Especially in more complex, multi-step student game development activities, iterative design processes facilitate work on complex problems, where critical thinking and problem-solving skills become crucial. Additionally, student game design can encourage knowledge sharing and peer-to-peer learning in the classroom. Students create, test, and refine their games in groups. They solve problems as a group, exchange insightful ideas, and offer and receive feedback—all activities that are highly beneficial for the development of communication skills. This cooperative dynamic has the potential to improve language competency while also developing communication abilities and creating a positive learning environment. DeHaan (2011) pursues an approach very strongly oriented towards the principles of constructivist learning and project-based foreign language learning in his study with students from a Japanese university who develop digital English-Language Role Playing Games and the design game magazines. In his report he concludes that through these projects students have gained a deeper understanding of games, game-related media, and creative

technologies alongside their acquisition of various aspects of the English language. In sum, student game design is a dynamic, multidimensional, integrated-skill method for learning a foreign language that enables students to take an active role in their education. Students can acquire language competence, critical thinking and problem-solving abilities, game skills, and (digital) literacy in an interesting and dynamic way.

Game Design in Teacher Training

Finally, game development as a learning method is not only a suitable approach for students, but also for student teachers (Schmidt, 2024). Initial studies that deal with game design in general clearly show potential: Zapušek and Rugelj (2021) describe the particularly positive influence of the diverse activities (specification, analysis, design, development of rules, implementation and evaluation, product iterations) on student teachers' motivation to learn and their ability to translate expertise and subject-specific pedagogical knowledge into well-founded design decisions. According to Jonassen (2011), creating playful learning environments calls for a high level of problem-oriented identification of learning objectives, competencies, and skills related to the subject matter, as well as the creation and assessment of learning strategies. Additionally, it involves various design considerations. It can therefore also provide an excellent, creativity-enhancing application context for prospective teachers in preparation for their future role as designers of learning environments and learning materials. Future educators can learn what makes learning interesting and enjoyable for pupils by creating games. They gain knowledge on how to include components that can hold students' interest in the material. They also learn how to keep refining their techniques and resources in response to input, results and learner/player feedback (see Teacher Mediation in DGBLLT).

Conclusion

Existing studies highlight the potential advantages of incorporating game development activities into student learning across different subjects, educational settings, learning objectives, and target audiences (from students in school to future teachers at universities). These advantages include opportunities to improve creativity, problem-solving skills, coding skills, literacy, motivation, and positive attitudes toward various subjects. Together, the studies highlight how student game design for learning can have a positive impact on cognitive and affective aspects of education. This information can be very helpful for educators who are trying to incorporate creative and interesting teaching strategies linked to the development of twenty-first-century skills into their curricula. The technique has novel prospects for achieving project-, task-, and product-oriented language learning that integrates several subject areas and language-related skill development in a purposeful and inspiring manner. However, more research is needed to determine the effects on language skills and the exact conditions for the effective use of (digital) game development processes in the language classroom, especially regarding the tasks, implementation features, and the necessary qualifications of teachers to initiate, accompany, and evaluate student game development.

Cross-References

- ▶ Breakout Games and DGBLLT
- ▶ DGBLLT and Narrative
- ▶ Game Mechanics and Design Features for DGBLLT
- ▶ Game-based LLT Environments
- ▶ Gaming Literacies and LLT
- ▶ Learner Variables in DGBLLT
- ▶ Task-Based DGBLLT
- ▶ Teacher Mediation in DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Akcaoğlu, M., & Koehler, M. J. (2014). Cognitive outcomes from the Game-Design and Learning (GDL) after-school program. *Computers & Education, 75*, 72–81. https://doi.org/10.1016/j.compedu.2014.02.003

Baytak, A., & Land, S. M. (2010). A case study of educational game design by kids and for kids. *Procedia: Social and Behavioral Sciences, 2*(2), 5242–5246. https://doi.org/10.1016/j.sbspro.2010.03.853

Bewer, N., & Gladkaya, M. (2022, January 24). *Game development-based approach for learning to program: A systematic literature review* [Video]. Wirtschaftsinformatik 2022 Proceedings. 3. https://aisel.aisnet.org/wi2022/digital_education/digital_education/3

Bulut, D., Samur, Y., & Cömert, Z. (2022). The effect of educational game design process on students' creativity. *Smart Learning Environments, 9*(1), 1. https://doi.org/10.1186/s40561-022-00188-9

deHaan, J. (2011). Teaching and learning English through digital game projects. *Digital Culture & Education, 3*(1), 46–55. http://www.digitalcultureandeducation.com/cms/wp-content/uploads/2011/04/dce1046_dehaan_2011.pdf

Dekhane, S., Xu, X., & Tsoi, M. Y. (2013). Mobile app development to increase student engagement and problem solving skills. *Journal of Information Systems Education, 24*(4), 299–308.

Dlab, M. H., & Hoic-Bozic, N. (2021). Effectiveness of game development-based learning for acquiring programming skills in lower secondary education in Croatia. *Education and Information Technologies, 26*(4), 4433–4456. https://doi.org/10.1007/s10639-021-10471-w

Gallagher, D. G., & Grimm, L. R. (2018). Making an impact: The effects of game making on creativity and spatial processing. *Thinking Skills and Creativity, 28*, 138–149. https://doi.org/10.1016/j.tsc.2018.05.001

Hwang, G., Hung, C., & Chen, N. (2013). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology Research and Development, 62*(2), 129–145. https://doi.org/10.1007/s11423-013-9320-7

Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.

Korkmaz, S., Çetin-Dindar, A., & Oner, F. K. (2023). Impact of educational game development on students' achievement and attitudes toward science. *The Journal of Educational Research, 116*(5), 268–279. https://doi.org/10.1080/00220671.2023.2265852

Laakso, N. L., Korhonen, T. S., & Hakkarainen, K. P. J. (2021). Developing students' digital competences through collaborative game design. *Computers & Education, 174*, 104308. https://doi.org/10.1016/j.compedu.2021.104308

Owston, R., Wideman, H., Ronda, N. S., & Brown, C. (2009). Computer game development as a literacy activity. *Computers & Education, 53*(3), 977–989. https://doi.org/10.1016/j.compedu.2009.05.015

Robertson, J., & Howells, C. (2008). Computer game design: Opportunities for successful learning. *Computers & Education, 50*(2), 559–578. https://doi.org/10.1016/j.compedu.2007.09.020

Romero, M., & Lambropoulos, N. (2015). Digital game creation as a creative learning activity. In *2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL)* (pp. 338–342). https://doi.org/10.1109/imctl.2015.7359615

Schmidt, T. (2024). Studierende als Spiele-Entwickler*-innen und Forscher*innen – Zur Integration von Produktdesign und forschendem Lernen in die Englisch-Lehrkräftebildung. In E. Burwitz-Melzer, C. Riemer, & L. Schmelter (Eds.), *Konzeption, Gelingensbedingungen und Qualität fremd- und zweitsprachendidaktischer Hochschullehre*. Narr.

Vos, N., Van Der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers & Education, 56*(1), 127–137. https://doi.org/10.1016/j.compedu.2010.08.013

Zapušek, M., & Rugelj, J. (2021). Game design-based learning for preservice and in-service teacher training. In C. V. de Carvalho & M. Bauters (Eds.), *Technology supported active learning: Student-centered approaches. Lecture notes in educational technology* (pp. 165–186). Springer. https://doi.org/10.1007/978-981-16-2082-9_10

V

Virtual Exchange and DGBLLT

Kristi Jauregi-Ondarra
Utrecht University, Utrecht, Netherlands

Abstract

This chapter explores the role of serious social games (SSGs) in enhancing virtual exchange (VE) encounters for linguacultural and intercultural learning. Emphasizing the integration of serious games, it draws on previous research conducted with university students and school learners who engaged in international social gaming using various platforms, including virtual reality (VR). The analysis delves into how SSGs, grounded in task-based language teaching principles, leverage game mechanics and social interaction features to facilitate immersive and collaborative learning experiences in virtual exchange settings. Additionally, the chapter discusses the benefits and potential obstacles associated with incorporating serious games within social apps and VR technology for VE initiatives. It concludes by providing recommendations for educators and researchers to effectively harness SSGs to enhance student engagement and motivation in VE projects.

Keywords

Virtual exchange · Telecollaboration · Serious games · Social games · Virtual reality · Language learning · Intercultural learning

Introduction

Virtual exchange (VE) (also referred to as telecollaboration or COIL by some scholars) has emerged as a vital educational practice in the twenty-first century, fostering meaningful intercultural dialogue and collaboration through digital communication technologies (Gijsen et al., 2024; Sadler & Dooly, 2016). Enabling learners from diverse backgrounds to engage with each other, VE activities promote understanding and cooperation beyond geographical boundaries. Research shows a significant impact on language learners' linguistic, communicative, and intercultural competences (Çiftçi & Savaş, 2018; Cunningham & Akiyama, 2018). Some studies even suggest that VE can provide a more effective learning environment than traditional in-class interactions (Canto & Jauregi-Ondarra, 2017; Zeng, 2017).

This chapter explores the potential of integrating serious social games (SSGs) within VE frameworks. SSGs are social or multiplayer game-based language learning environments designed with educational objectives at their core (see ▶ "Game-Based LLT Environments") that leverage game design elements to create engaging and

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_145-1

social-interactive learning experiences, promoting both linguacultural development (Risager, 2012) and intercultural competence (Byram, 2021). Within VE environments, SSGs can facilitate authentic communication and immersive cultural interactions by fostering a sense of enjoyment and social engagement during the learning process (Jauregi-Ondarra & Canto, 2022), particularly when cultural topics are thoughtfully and pedagogically embedded within the SSG design. Additionally, when played in a VR world that simulate real-world scenarios and cultural contexts (see ▶ "Extended Reality and DGBLLT"), SSGs not only enhance language proficiency but also foster critical intercultural skills such as empathy, perspective-taking, and cultural awareness (Canto et al., 2023; Gruber et al., 2023). By incorporating SSGs into VE, educators can transcend the limitations of traditional VE tasks and classroom boundaries. These games harness the motivational power of play, encouraging learners to engage deeply with the content and with each other.

This chapter builds on a body of research demonstrating the efficacy of SSGs in both university and school educational settings. The chapter is structured as follows: first, we delve into the primary theories and concepts related to the use of SSGs within VE projects, followed by analysis of various applications. We then address the benefits and challenges associated with using SSGs, followed by recommendations for their use. Finally, we provide ideas for future research and practice.

Theoretical and Conceptual Framework

Linguacultural and Intercultural Learning

The integration of SSGs in language and culture learning is grounded in sociocultural theories of language acquisition, particularly those influenced by Vygotsky (1986). These theories emphasize the social and cultural dimensions of learning, highlighting the importance of interaction with others. Aligned with sociocultural theory, Risager's (2012) framework for developing linguacultural skills recognizes the deep interconnection between language and culture, where language learning is inherently a cultural process as it carries cultural meanings and reflects cultural practices. SSGs can create environments that mirror real-life cultural and linguistic contexts, providing learners with opportunities to engage in meaningful and situated learning experiences (Dawley & Dede, 2014) that align with the principles of linguacultural learning (Risager, 2012), and VE contexts enriched with SSGs can provide rich linguacultural experiences by embedding language learning within culturally authentic scenarios (see ▶ "DGBLLT and Culture").

Intercultural learning, as articulated by Byram (2021), focuses on developing the skills necessary for effective and respectful interaction across cultural boundaries. This involves not only linguistic proficiency but also intercultural competencies such as cultural knowledge, attitudes of openness and curiosity, and the ability to interpret and relate cultural phenomena. SSGs, because of their interactive and immersive nature, are well-suited to fostering these competencies in a holistic and engaging manner.

Task-Based Task Design and SSGs for VE

Task-based language teaching (TBLT) is a pedagogical approach that centres on the use of meaningful tasks as the primary unit of instruction (Ellis, 2003; van den Branden, 2016). TBLT principles are highly relevant to the design of SSGs, as they emphasize learning through doing. In TBLT, tasks are designed to require learners to use language communicatively to achieve specific goals (see ▶ "Task-Based DGBLLT"). This approach aligns perfectly with the interactive and goal-oriented nature of games, providing a structured and engaging framework for language acquisition and intercultural development. While TBLT provides a theoretical framework for designing effective tasks, SSGs present the platform for implementing those tasks.

Key characteristics of a task that can be applied to the design of SSGs for VE (c.f. Hampel, 2006; Kurek & Müller-Hartmann, 2017) include:

1. Focus on communication: The primary goal is to encourage communication and interaction

between learners, requiring them to use the target language to complete the task (in our case to achieve game objectives).
2. Collaborative nature: Tasks should encourage teamwork and collaboration among VE participants by promoting negotiation, problem-solving, and shared decision-making.
3. Integration of language skills: Tasks should integrate various language skills such as listening, speaking, reading, and writing. Learners should use these skills to achieve the task/game goals.
4. Authentic context: Tasks should be relevant to the learners' interests and experiences, simulating real-life situations where they might use the language.
5. Contextualization in the partner culture: Tasks should reflect authentic situations and cultural practices relevant to the VE partners' backgrounds, making learning culturally relevant and engaging.
6. Task clarity and planning: Clear instructions and guidance should be provided for the task, outlining the goals, procedures, time constraints, and roles/responsibilities. This allows learners to engage with the task/games confidently and efficiently.
7. Differentiation: Tasks should be adaptable to different learning styles, proficiency levels, and cultural backgrounds of the VE participants.
8. Assessment and reflection: Tasks should be followed by opportunities for assessment and reflection. This allows learners to self-evaluate their performance, identify areas for improvement, and gain deeper understanding from the VE experience.

SSGs can capitalize on the power of game mechanics to make learning enjoyable and intrinsically motivating by translating TBLT principles into interactive experiences specifically designed for VE interaction (see ▶ "Game Mechanics and Design Features for DGBLLT"). This translation manifests in several ways, for example, as tasks and virtual scenarios that present challenges or missions requiring learners to communicate, collaborate, and problem solve. These tasks can be situated within rich narrative contexts that provide motivation and relevance for the learners (see ▶ "DGBLLT and Narrative").

Key Concepts in SSGs

To understand how SSGs function and create impactful learning experiences within VE, this section delves into three key concepts: game mechanics, social interaction features, and VR environments.

Game Mechanics and Dynamics

In game design, game mechanics are the rules and systems that guide gameplay (Figueroa, 2015), including elements such as challenges, rewards, or progression systems. Effective game mechanics are crucial for creating engaging and motivating learning experiences (Perttula et al., 2017; Reinhardt, 2019). In serious games, these mechanics are designed to support educational objectives, providing learners with clear goals, feedback, and a sense of achievement. Interaction with game mechanics leads to game dynamics that stimulate the students' emotions during the game experience and include senses of competition, challenge, self-expression, reward, achievement, fun, or satisfaction (Jauregi-Ondarra & Canto, 2022).

Social Interaction Features

Social interaction features are another key component of SSGs (Zhonggen, 2019). These features enable players to communicate and collaborate with each other, fostering a sense of community and shared purpose. In VE contexts, social interaction is essential for intercultural learning, as it allows learners to engage in meaningful dialogue and cultural exchange. SSGs can facilitate this interaction through multiplayer modes, chat functions, and collaborative tasks.

VR Environments

VR environments provide an additional layer of immersion, creating realistic and interactive settings for learning (Lee et al., 2024; see ▶ "Extended Reality and DGBLLT"). VR can transport learners to different cultural scenarios and allow them to experience and explore these

contexts in a way that is both engaging and authentic. The immersive nature of VR can enhance the impact of SSGs, making the learning experiences more vivid and memorable. VR environments offer a spectrum of immersion levels (Jauregi-Ondarra et al., in press) that can be leveraged by SSGs:

Low-Immersion VR: Often experienced via the computer screen, low-immersion VR provides a partially immersive environment. Learners can interact with 3D objects and navigate virtual spaces, but the sense of presence may be limited.

High-Immersion VR: Typically experienced through goggles, high-immersion VR offers a more realistic and immersive experience. Learners can feel a stronger sense of presence within the virtual world, with sights, sounds, and even physical sensations potentially simulated. This heightened level of immersion can significantly enhance the impact of SSGs.

The immersive nature of VR, particularly in high-immersion environments, can significantly enhance the impact of SSG games, making the learning experiences more vivid, memorable, and emotionally engaging. SSGs can be specifically designed to capitalize on the unique affordances of VR by integrating mechanics such as spatial interaction and promoting sensory dynamics through embodied experience.

Analysis of the Impact of SSGs on VE Projects

Many studies have explored the use of TBLT approaches to design tasks for VE projects. O'Dowd and Ware (2009) categorize these tasks into three main types: information exchange, establishing comparisons, and co-creating an artefact. Tasks requiring co-creation tend to be the most demanding but also yield the richest interactions in terms of the quality of communication stimulated. While research on TBLT in VE settings has primarily focused on information exchange and comparison tasks, there is a growing interest in leveraging SSGs to shape VE interactions.

Impact of SSGs in VE with University Students

Jauregi-Ondarra and Canto (2022) investigated the use of SSGs designed to promote Spanish language and intercultural learning with university students in a VE project. Working together to solve problems and complete missions using the target language, students participated in an international setting playing variations of "Snakes and Ladders", a cultural quiz, and a treasure hunt/escape room (Fig. 1). Half the students used a video-communication platform, while the other half used the low-immersion VR environment of Open Simulator to play the social games. The results showed high student engagement and motivation, with reported improvements in both language skills and intercultural understanding. Students described the social games as stimulating and enjoyable, a welcome change from "normal language exchange". Generally positive sentiment was expressed in statements such as "I had fun, met a new person, practiced Spanish and learned about a new culture", and "the games made learning fun and interactive. I felt more connected to my peers from a different culture" (Jauregi-Ondarra & Canto, 2022, p. 167).

VR environments, particularly those facilitating high-immersion experiences with goggles, add another dimension to the learning experience, providing immersive and realistic contexts for language use and cultural exploration (see ▶ "Extended Reality for DGBLLT"). Canto et al. (2023) conducted a study where university students participated in a VE project that involved interacting in a playful environment within high-immersion VR settings. The study found that VR enhanced the learners' sense of presence and immersion, making the learning experiences more impactful and engaging. Quotes from the students participating in the Canto et al. (2023) study revealed VR's potential to reduce speaking anxiety, for example: "This helped me a lot with my fear of speaking in front of large groups: I didn't have to look at anyone directly and yet I was speaking in front of a whole group of

Virtual Exchange and DGBLLT, Fig. 1 Dutch and British students playing the Treasure Hunt using Spanish as Lingua Franca

people. Because they were avatars, I was a lot less intimidated", and "The VR environment made the interactions feel real. It was like being in another country without leaving my home" (Canto et al., 2023, p.149).

The researchers argued that the VR environment contributed to an increased willingness to communicate in VE (Canto et al., 2023). However, some students reported experiencing a sense of impersonality when collaborating through avatars, even when real voices were used. Additionally, the lack of access to facial cues in the avatar format was noted by some students as a hindrance to comprehension, while others experienced some cybersickness.

Impact of SSGs in VE with Minors

While research on SSGs in VE settings with university students is growing, studies focusing on minors are less prevalent (see ▶ "Game-Mediated Practices for Young Language Learners"). Jauregi (2016a) explored the potential of gamified VE activities to enhance English language learning by younger students. The study investigated the impact of online intercultural exchanges in a low-immersion VR environment on student discussion skills and motivation. Finnish and Dutch students participated in five discussion tasks over 5 weeks, employing elements commonly found in serious games. These tasks included:

Role-playing: Students engaged in role-playing as they collaborated to solve a mystery within the shared virtual world. This activity incorporated purpose, challenge, and a sense of shared goals, mirroring mechanics used in serious games.

Scenario-based Learning: Learning tasks were not decontextualized conversation practice but were anchored in a specific scenario (e.g., a courtroom setting). This contextualization functioned similarly to game narratives, increasing student engagement and motivating them to utilize their English skills to progress through the scenario (Fig. 2).

Differentiated Learning: A variety of tasks, including information exchange, debate, and role-playing, potentially catered to different learning styles and preferences. This variety, akin to offering diverse challenges within a game, may have contributed to sustained student engagement by allowing individualized and differentiated learning.

Virtual Exchange and DGBLLT, Fig. 2 Dutch and Finnish learners playing the SSG using ELF to find the murderer

The study findings of increased motivation and improved discussion skills provide preliminary evidence suggesting that interactive, scenario-based tasks within a VE environment can mirror elements of SSGs and offer promising outcomes for student learning.

Building on the aforementioned study (2016a), Jauregi (2016b) explored the impact of VE games on the motivational aspects of foreign language acquisition in learners of German aged 14–15. The study involved Dutch and British students using the low-immersion VR platform OpenSim to participate in culturally-themed German language games. These games incorporated gamification elements such as quizzes, points, and time constraints, and were designed to encourage collaboration and communication in German (Figs. 3 and 4). The findings revealed a positive influence on the participant motivation for learning German. Participants found the VE game enjoyable and engaging and that it fostered communication and interaction with peers from another country, promoting intercultural awareness. Interestingly, the data suggested a specific growth in self-confidence and satisfaction among British students after the second game session, potentially indicating an accumulative benefit of the VE gaming experience.

These studies provide initial evidence that SSGs within VE settings can be a valuable tool for educators, promoting motivation, fostering more engaging and enjoyable learning, increasing intercultural awareness, and potentially improving language acquisition outcomes for both university students and minors. Further research is recommended to solidify these findings, explore the long-term impact of SSGs on foreign language acquisition, and investigate the optimal design of SSGs for different age groups and learning objectives within VE environments.

Benefits and Challenges of SSGs in VE

The incorporation of SSGs into VE initiatives offers a promising approach to language and culture learning, but it is not without its challenges. This section explores both the advantages and potential obstacles associated with implementing SSGs in VE settings.

Benefits of SSGs in VE

SSGs offer a range of benefits for language learning within VE environments. These games leverage their inherently engaging and goal-oriented nature to transform language learning from a passive activity into an active and enjoyable experience, fostering increased student motivation and enhanced engagement. Additionally, the social and collaborative aspects of SSGs align well with the principles of sociocultural theory, which emphasizes the importance of social interaction in

Virtual Exchange and DGBLLT, Fig. 3 Dutch and British learners playing the first part of the SSG, where they need to exchange information to jointly find the clue that will open the door to the next game

Virtual Exchange and DGBLLT, Fig. 4 Dutch and British learners are playing the Quiz Chair Game, where they need to answer quiz questions correctly to elevate the chair and gain points, which will open the door to the next game

learning (Lantolf & Thorne, 2006). Moreover, by providing opportunities for learners to communicate and collaborate in meaningful ways, SSGs can foster the development of intercultural communicative competencies. By working together to achieve common game goals, learners build relationships and may gain a deeper understanding of each other's cultures (Jauregi, 2016a; b; Jauregi-Ondarra & Canto, 2022). VR environments further enhance the learning experience by creating realistic and interactive settings for language use and cultural exploration (Gruber et al., 2023). They may also heighten the immersive impact of SSGs, making learning experiences more vivid and potentially anxiety-reducing, especially when interacting with native speakers.

Challenges of SSGs in VE

Despite the numerous benefits, there are also challenges associated with implementing SSGs in VE. These challenges can be broadly categorized into technical, pedagogical, and cultural obstacles.

Technical Challenges: Implementing SSGs, particularly those involving VR environments, requires a robust technological infrastructure with adequate resources. Schools and universities may need to invest in computing systems, VR goggles, specialized software, and reliable internet connections to support these advanced learning tools. Additional technical challenges can arise from game compatibility issues, user accessibility limitations (ensuring the technology is usable by all learners) and maintaining secure and stable platforms for VE activities. Overcoming these barriers necessitates investment and ongoing technical support, which may not be feasible for all educational institutions.

Pedagogical Challenges: From a pedagogical perspective, integrating SSGs into VE requires careful planning and alignment with curriculum goals. Educators need to be trained in the effective design, adaptation, and implementation of task-based SSGs, understanding how to facilitate and guide game-based learning activities, particularly when interacting with international peers. Designing and implementing SSGs that are both educational and engaging can be complex. Curriculum adaptations may also be necessary to integrate SSGs seamlessly into existing learning objectives.

Cultural Challenges: Cultural differences can also pose challenges in VE settings. Learners from different cultural backgrounds may have varying levels of familiarity and comfort with game-based learning and VR technologies. There can be differences in gaming cultures, preferences, and access to technology that need to be considered. Furthermore, cultural sensitivities and differences in communication styles can affect the dynamics of social interaction when using social games. It is essential to design serious games that are culturally inclusive and sensitive, providing equitable learning opportunities for all participants. While games have the potential to foster intercultural understanding, they can also reinforce stereotypes or present a biased view of cultures if not carefully designed.

Recommendations and Future Directions in Research and Practice

To maximize the potential of SSGs for enhancing student engagement and motivation in VE projects, several recommendations should be considered.

Educators require comprehensive training to effectively integrate SSGs into their teaching practices (Gee, 2017; Prensky, 2003). Professional development programs are essential, equipping educators with the necessary pedagogical principles of task-based gamified learning, technical skills for utilizing VR technology, and strategies for facilitating productive intercultural interactions within a virtual environment.

The cultural content embedded within SSGs should be accurate, respectful, and reflect diverse perspectives, avoiding the perpetuation of stereotypes (Byram, 2021). The design of SSGs must prioritize scenarios that necessitate meaningful interaction and collaboration among learners. Game mechanics and dynamics should promote active language use, problem-solving skills, and opportunities for cultural exploration (Jauregi-Ondarra & Canto, 2022; Tekinbaş & Zimmerman, 2003). Inclusivity and accessibility are paramount

considerations when designing and selecting SSGs for VE. Games should accommodate learners with diverse backgrounds and varying levels of technological experience and access, fostering a more enriching learning experience for all participants (Sousa et al., 2023).

To maximize the educational impact of SSGs, educators should integrate reflective activities throughout the learning journey, encompassing pre-game, in-game, and post-game phases. These activities can guide learners in processing their experiences, establishing connections between game-based learning and real-world contexts, and fostering deeper intercultural understanding (Blasco, 2012). The effectiveness of SSGs in achieving linguacultural learning objectives necessitates ongoing evaluation. Feedback from both learners and educators should be utilized to iteratively improve the SSGs and ensure alignment with educational goals (Arnab et al., 2015).

VR environments have the potential to significantly enhance the immersive nature of VE projects and stimulate co-presence in situated learning scenarios (Canto et al., 2023; Gruber et al., 2023; Jauregi-Ondarra et al., in press). Educators should explore the possibilities VR offers in creating realistic and engaging learning contexts that promote the development of linguacultural and intercultural competencies while being mindful of the technical challenges and accessibility limitations associated with VR technology (Gruber et al., 2023).

Looking ahead, the integration of SSGs in VE is a dynamic and evolving field. Future research and practice should continue to explore the potential of these tools, focusing on several key areas. Conducting longitudinal studies can provide deeper insights into the long-term effects of serious games on language and intercultural learning, helping to identify sustained impacts on learners' skills, attitudes, and competencies.

Developing scalable and sustainable models for implementing SSGs in VE can help make these tools more accessible to a broader range of educational institutions. Research should explore cost-effective solutions and best practices for widespread adoption. Further research is needed to understand the impact of SSGs on diverse learner populations, including those from different age, education, cultural, linguistic, and socioeconomic backgrounds. This research can inform the design of more inclusive and equitable learning experiences.

By following these recommendations and focusing on future research directions, educators and researchers can leverage the affordances of SSGs to create a more engaging and effective learning environment for language acquisition and intercultural development within VE projects.

Conclusion

This chapter has explored the pedagogical role of SSGs in enhancing VE encounters for language and intercultural learning. By integrating SSGs into VE settings, educators can create engaging, interactive, and immersive learning experiences that promote the development of language proficiency and intercultural competencies. Research shows that SSGs, when grounded in TBLT principles, can leverage game mechanics, social interaction, and VR features to facilitate collaborative and immersive learning. They can provide a structured yet flexible framework for learning, encouraging active participation and deep engagement.

The incorporation of SSGs into VE initiatives presents both opportunities and challenges. While there are significant benefits in terms of student motivation, engagement, and skill development, there are also technical, pedagogical, and cultural obstacles that need to be addressed. By investing in training and focusing on inclusivity and accessibility, educators and researchers can effectively harness the potential of SSGs to enhance VE projects.

The continued exploration and innovation in this field are essential for understanding and maximizing the educational impact of SSGs. As technology advances and educational paradigms evolve, immersive SSGs will likely play an increasingly important role in fostering meaningful intercultural dialogue and collaboration in VE contexts.

Cross-References

- ▶ DGBLLT and Culture
- ▶ DGBLLT and Narrative
- ▶ Extended Reality and DGBLLT
- ▶ Game Mechanics and Design Features for DGBLLT
- ▶ Game-Based LLT Environments
- ▶ Game-Mediated Practices for Young Language Learners
- ▶ Task-Based DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., De Freitas, S., Louchart, S., & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology, 46*(2), 391–411.

Blasco, M. (2012). On reflection: Is reflexivity necessarily beneficial in intercultural education? *Intercultural Education, 23*(6), 475–489. https://doi.org/10.1080/14675986.2012.736750

Byram, M. (2021). Teaching and assessing intercultural communicative competence: Revisited. *Multilingual Matters*. https://doi.org/10.21832/9781800410251

Canto, S., Jauregi-Ondarra, K., & Gruber, A. (2023). High-immersion virtual reality-mediated intercultural virtual exchange: A case study. In S. Di Sarno-García, S. Montaner-Villalba, & A. Gimeno-Sanz (Eds.), *Telecollaboration applications in foreign language classrooms* (pp. 136–159). IGI Global. https://doi.org/10.4018/978-1-6684-7080-0.ch007

Çiftçi, E. Y., & Savaş, P. (2018). The role of telecollaboration in language and intercultural learning: A synthesis of studies published between 2010 and 2015. *ReCALL, 30*(3), 278–298. https://doi.org/10.1017/S0958344017000313

Cunningham, J. D., & Akiyama, Y. (2018). Synthesizing the practice of SCMC-based telecollaboration: A scoping review. *CALICO Journal, 35*(1), 49–76. https://doi.org/10.1558/cj.33156

Dawley, L., & Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. In J. M. Spector, M. D. Merrill, J. Elen, & M. D. Bishop (Eds.), *Handbook of research on educational communications- and technology* (4th ed., pp. 723–734). Springer.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Figueroa, J. F. (2015). Using gamification to enhance second language learning. *Digital Education Review, 27*, 32–54. https://doi.org/10.1344/der.2015.27.32-54

Gee, J. P. (2017). Games, passion, and "higher" education. In W. G. Tierney, Z. B. Corwin, T. Fullerton, & G. Ragusa (Eds.), *Postsecondary play: The role of games and social media in higher education* (pp. 171–187). Johns Hopkins University Press.

Gijsen, L., Canto, S., & Jauregi-Ondarra, K. (2024). Intercultural virtual exchange for foreign language learning and teaching. In R. Helm & U. Stickler (Eds.), *Bloomsbury handbook of language learning and technologies*. Bloomsbury Academic. https://www.bloomsbury.com/uk/bloomsbury-handbook-of-language-learning-and-technology-9781350340329/

Gruber, A., Canto, S., & Jauregi-Ondarra, K. (2023). Exploring the use of social virtual reality for virtual exchange. *ReCALL, 35*(3), 258–273. https://doi.org/10.1017/S0958344023000125

Hampel, R. (2006). Rethinking task design for the digital age: A framework for language teaching and learning in a synchronous online environment. *ReCALL, 18*(1), 105–121.

Jauregi, K. (2016a). Interculturele discussies in een 3D virtuele wereld: impact op motivatie en discussievaardigheden. *Levende Talen Tijdschrift, 17*(1), 14–24.

Jauregi, K. (2016b). Telecollaborative games for youngsters: Impact on motivation. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture* (pp. 201–208). Research-Publishing.Net. https://doi.org/10.14705/rpnet.2016.eurocall2016.562

Jauregi-Ondarra, K., & Canto, S. (2022). Interaction games to boost intercultural communication in virtual worlds and video-communication. In M. Peterson & N. Jabbari (Eds.), *Digital games in language learning: Case-studies and applications* (pp. 158–182). Routledge. https://www.routledge.com/Digital-Games-in-Language-Learning-Case-Studies-and-Applications/Peterson-Jabbari/p/book/9781032145976

Jauregi-Ondarra, K., Melchor-Couto, S., & Canto, S. (in press). From low- to high-immersive virtual reality in language education. In G. Stockwell (Ed.), *Cambridge handbook of technology in language teaching and learning*. Cambridge University Press.

Kurek, M., & Müller-Hartmann, A. (2017). Task design for telecollaborative exchanges: In search of new criteria. *System, 64*, 7–20.

Lantolf, J. P., & Thorne, S. L. (2006). Sociocultural theory and second language acquisition. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 201–224). Erlbaum.

Lee, S. M., Wang, X., Park, I., & Lestiono, R. (2024). It feels so real! Situated authentic language learning in immersive virtual reality. *Education and Information Technologies*, 1–23.

O'Dowd, R., & Ware, P. (2009). Critical issues in telecollaborative task design. *Computer Assisted*

Language Learning, 22(2), 173–188. https://doi.org/10.1080/09588220902778369

Perttula, A., Kiili, K., Lindstedt, A., & Tuomi, P. (2017). Flow experience in game-based learning: A systematic literature review. *International Journal of Serious Games, 4*(1), 57–72. https://doi.org/10.17083/IJSG.V4I1.151

Prensky, M. (2003). Digital game-based learning. *ACM Computers in Entertainment, 1*(1), 1–4. https://doi.org/10.1145/950566.950567

Reinhardt, J. (2019). *Gameful second and foreign language teaching and learning: Theory, research, and practice*. Palgrave Macmillan.

Risager, K. (2012). Linguaculture. In C. A. Chapelle (Ed.), *Encyclopedia of applied linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0709

Sadler, R., & Dooly, M. (2016). Twelve years of telecollaboration: What we have learnt. *ELT Journal, 70*(4), 401–413. https://doi.org/10.1093/elt/ccw041

Sousa, C., Neves, P., & Luz, F. C. (2023). Barriers and hindrances to the effective use of games in education: Systematic literature review and intervention strategies. *European Conference on Games Based Learning*. https://doi.org/10.34190/ecgbl.17.1.1472

Tekinbaş, K. S., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. MIT Press.

Van den Branden, K. (2016). Task-based language teaching. In *The Routledge handbook of English language teaching* (pp. 238–251). Routledge.

Vygotsky, L. S. (1986). *Thought and language*. The MIT Press.

Zhonggen, Y. (2019). A meta-analysis of use of serious games in education over a decade. *International Journal of Computer Games Technology, 2019*(1), 4797032.

V

Vocabulary and DGBLLT

Barry Lee Reynolds
Faculty of Education, University of Macau, Taipa, Macau SAR, China
Centre for Cognitive and Brain Sciences, University of Macau, Taipa, Macau SAR, China

Abstract

This article explores the role of digital game-based language learning and teaching (DGBLLT) in second language (L2) vocabulary acquisition. It emphasizes the effectiveness of commercial entertainment games in promoting incidental vocabulary learning through engaging, multimodal interactions. The article distinguishes between receptive and productive knowledge, alongside vocabulary breadth and depth, utilizing an established taxonomy to underline the complexity of L2 vocabulary acquisition that can occur through digital gameplay. Various studies and systematic reviews are synthesized to demonstrate the positive impact of DGBLLT on L2 vocabulary learning. Despite the substantial evidence supporting DGBLLT for L2 vocabulary development, challenges remain, including the need for better access to and analysis of in-game language data and player interactions. Future research directions include refining game design principles, exploring long-term retention of word knowledge, and addressing cultural and linguistic diversity, among others. The article also calls for developing tools and methodologies for researchers to enhance our understanding and optimize the impact of digital games in L2 vocabulary learning and teaching.

Keywords

Commercial entertainment games · Digital game-based learning (DGBL) · Digital game-based language learning and teaching (DGBLLT) · Incidental vocabulary acquisition · Incidental vocabulary learning · Receptive and productive vocabulary knowledge · Vocabulary breadth and depth · Word knowledge

Introduction

Building a second language vocabulary is often considered arduous because it is impossible to teach and learn all the words of a second language (Nation, 2021). Research agrees that learning the most frequent words in a target second language through a teacher's direct instruction and learner-focused learning enables a language learner to become independent (Nation, 2022). After mastering frequent vocabulary through intentional methods—either in the classroom via language-focused instruction or independently through language-focused learning—learners should engage in meaningful input and output activities

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_146-1

that promote practical language use (Nation, 2021). Incidental exposure to unfamiliar words during language use gradually leads to the acquisition of new vocabulary (Hulstijn, 2012). This approach allows learners to concentrate on high-frequency words intentionally while using the language for meaningful purposes and picking up new words along the way, resulting in a gradual buildup of vocabulary knowledge over time (Nation, 2013). Digital game-based language learning and teaching provides such an opportunity (Anderson et al., 2008; Heidt et al., 2023; Reynolds, 2023b; Soyoof et al., 2024).

Commercial entertainment games are designed for mass market appeal and are intended to be sold to a wide audience (Whitton, 2014). These games are typically produced by established game development companies or independent developers aiming for broad distribution and high sales. In contrast, commercial educational games, often referred to as serious games, are specifically designed with educational objectives in mind, blending entertainment with purposeful learning. These games are custom-made to meet the specific needs and preferences of a particular client or a niche audience (Whitton, 2014) and are often developed for educational, training, promotional, or research purposes. Because commercial entertainment games offer more engaging and immersive environments that naturally expose players to a wide range of vocabulary in context, facilitating incidental learning of new words, they may be considered more useful for the gradual building of second language vocabulary over time (Dixon et al., 2022; Reynolds, 2017; Reynolds, 2023a; Soyoof et al., 2023; Sylvén & Sundqvist, 2012; Wrobetz, 2023).

Types of Vocabulary Knowledge That Can Be Learned Through Gaming

Vocabulary Breadth and Depth

When discussing a language learner's vocabulary knowledge, it is important to distinguish between how many words a learner knows and how well they know each of these words. This distinction is often referred to as vocabulary breadth and depth. Distinguishing between vocabulary breadth and depth is crucial because it provides a comprehensive understanding of a learner's language proficiency, allowing for targeted instruction and more effective communication. Vocabulary breadth refers to the number of words a learner has acquired in a language, while vocabulary depth refers to how well they have acquired them (Read, 2023). Different gauges exist for measuring vocabulary knowledge, but the measurement of breadth usually occurs through tests that measure vocabulary size, with the Vocabulary Size Test being the most commonly used in lexical studies (Nation & Beglar, 2007) and a popular option for certain aspects of depth have been the Word Associates Test (see Read, 2023 for commentary). Whether measuring vocabulary depth or breadth, tests often assess only one or a few aspects of vocabulary knowledge. This is because learning vocabulary is not a binary process (Nation, 2022); one does not simply know or not know a word. Instead, different aspects of vocabulary knowledge must be considered.

Receptive and Productive Vocabulary

The most fundamental approach to understanding word knowledge is to categorize vocabulary as either receptive or productive (Schmitt, 2010). Receptive knowledge refers to a language learner's ability to comprehend a word they hear or read, whereas productive knowledge involves their capability to use the word in speech or writing. A well-established taxonomy of word knowledge, developed by Nation (2022), categorizes it into form, meaning, and use, which can be either receptive or productive. Receptive form entails recognizing a word's spoken and written forms. Receptive meaning involves grasping the concept or idea a word represents. Receptive use includes understanding the grammatical functions and collocations of a word within context. Productive form requires accurately producing a word in both speech and writing. Productive meaning refers to correctly applying a word's meaning, while productive use involves appropriately using a word in grammatical structures and in combination with other words. Given that using a second language like English for most advanced purposes requires a vocabulary size of around

8000–9000 words (Nation, 2022), achieving this is an immense challenge even for the most dedicated learners and teachers. Instead, the acquisition of the majority of these words occurs incidentally when learners use the language for communicative purposes. Digital gaming stands out as one of the most effective methods for this purpose, providing opportunities for different aspects of vocabulary knowledge to be acquired (Reynolds, 2023a).

Benefits of Digital Game-Based Learning for Vocabulary Acquisition

Playing video games has been shown to positively impact language learning due to their structured and interactive nature (Sykes & Reinhardt, 2013). Many games exceed traditional language learning requirements by providing opportunities for language input, output, and interaction (Sykes & Reinhardt, 2013). Knight et al. (2020) outlined four benefits of informal gaming for language learning, particularly in vocabulary acquisition: (1) offering low-risk simulated language practice with meaningful consequences, (2) creating emotionally supportive environments, (3) introducing complexity to enhance learning readiness, and (4) fostering global competencies through online interactions. According to Lorenset and Piazza (2019), digital games offer numerous affordances, including inherent motivation and pleasure, which sustain engagement in contextualized language use over time (see Motivation and Affect in DGBLLT). Some genres of digital games also facilitate communication and interaction with other language users, providing authentic contexts for language use (Soyoof et al., 2023). Furthermore, the multimodal input provided by digital games require deep processing of new vocabulary (Rasti-Behbahani, 2023; Reynolds, 2017), thereby promoting development of vocabulary depth.

Research on Digital Game-Based Learning (DGBL) for Vocabulary Acquisition

Digital game-based learning (DGBL) has emerged as a promising approach for enhancing second language vocabulary acquisition, as evidenced by several systematic reviews and meta-analyses. Chen et al. (2018) highlighted the role of game design features in influencing vocabulary learning outcomes, suggesting that higher challenge levels can independently enhance learning regardless of learners' age or linguistic background. Conversely, Dixon et al. (2022) found a generally positive effect of DGBL on L2 development, noting that entertainment-focused games often prove more effective than educational ones. Thompson and von Gillern (2020) and Tsai and Tsai (2018) underscored significant overall benefits of DGBL in promoting English vocabulary acquisition, while Vnucko and Klimova (2023) emphasized the positive learning environment and emotional engagement fostered by DGBL. Wu et al. (2020) further identified critical factors such as game platform and teaching method that influence vocabulary learning outcomes. Additionally, Zou et al.'s (2021) synthesis revealed various types of digital games consistently enhance vocabulary learning, promote comprehension skills, and improve learner engagement.

Investigating the Language in Video Games

While substantial research has explored the potential of digital video games for vocabulary learning, less attention has been given to the actual language used in video games (Schmitt, 2019). As gameplay is dynamic, the words encountered may vary, affecting whether these encounters are superficial or lead to deeper processing. Additionally, some encounters may only occur in audio input or textual input, while others are presented in dual modalities with or without additional visual enhancement. However, most research has focused primarily on the written text that can be extracted from games. Recognizing these variations, some research has attempted to understand the potential of games for vocabulary development by examining how the exposure through input could lead to learning and how this knowledge is or is not retained over time (e.g., Wrobetz, 2023). Other studies have heavily relied on corpus

techniques, approximating gameplay by analysing the likely exposure to language using videos of game walkthroughs (e.g., Rodgers & Heidt, 2020). Initial research in this area has also led to the development of the Good Gaming List of Key Vocabulary in video games (Heidt et al., 2023), which is valuable for learners seeking to use games as a source of language input for incidental learning. Such lists are also potentially useful for teachers who want to incorporate vocabulary teaching for later independent gaming by their students. While these and similar lexical coverage studies have been instrumental in demonstrating the potentials of video games for vocabulary learning and teaching, they still do not provide a complete picture. A clearer understanding of the true potential of video games for language learning and teaching requires observation and analysis of dynamic gameplay.

Conclusion

The numerous meta-analytic and systematic reviews conducted suggest future research should focus on refining game design principles, investigating long-term retention effects, and addressing cultural and linguistic diversity to optimize the educational impact of DGBL in language learning contexts. While these are promising avenues to explore, a significant bottleneck remains: researchers' inability to easily access all the language content within a commercial entertainment game, both in written and spoken forms, and to extract analytics data on language learners' gameplay.

The crucial task ahead is to develop a deeper understanding of how interaction in the games, combined with frequency of exposure to unknown words through various modalities, leads to measurable gains in vocabulary knowledge. This research will require the collection of both online gaming data (i.e., gaming analytics and eye-tracking) and offline vocabulary assessment data to uncover a more solid understanding of the learning effects from digital gameplay. Additionally, there is a need to create tools and methodologies that can facilitate the extraction and analysis of in-game language data to support this research.

Cross-References

▶ Commercial Entertainment Games in DGBLLT
▶ Digital Game-Based Language Learning and Teaching
▶ Motivation and Affect in DGBLLT

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anderson, T. A., Reynolds, B. L., Yeh, X.-P., & Huang, G.-Z. (2008). Video games in the English as a foreign language classroom. Second IEEE international conference on digital game and intelligent toy enhanced learning, Banff Park Lodge Resort Hotel and Conference Centre, Banff, CA, USA. https://doi.org/10.1109/DIGITEL.2008.39

Chen, M. H., Tseng, W. T., & Hsiao, T. Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology, 49*(1), 69–77. https://doi.org/10.1111/bjet.12526

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. *Language Learning & Technology, 26*(1), 1–25. http://hdl.handle.net/10125/73464.

Heidt, J., Pinchbeck, G. G., & Rodgers, M. P. (2023). The good gaming (GG) list: Key vocabulary in videogames. In B. L. Reynolds (Ed.), *Vocabulary learning in the wild* (pp. 143–161). Springer. https://doi.org/10.1007/978-981-99-1490-6_5

Hulstijn, J. H. (2012). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–5). Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0530

Knight, S. W., Marean, L., & Sykes, J. M. (2020). Gaming and informal language learning. In M. Dressman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 101–115). John Wiley & Sons. https://doi.org/10.1002/9781119472384.ch7

Lorenset, C. C., & Piazza, P. T. (2019). Digital games, foreign language and vocabulary development. *Revista EducaOnline, 13*(3), 1–19. https://revistaeducaonline.eba.ufrj.br/edições-anteriores/2019-3/digital-games-foreign-language-and-vocabulary-development

Nation, P. (2013). *What every EFL teacher should know.* Compass Publishing.

Nation, P. (2021). Is it worth teaching vocabulary? *TESOL Journal, 12*(4), 1–9. https://doi.org/10.1002/tesj.564

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–11. https://openaccess.wgtn.ac.nz/articles/journal_contribution/A_vocabulary_size_test/12552197/1/files/23375153.pdf

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009093873

Rasti-Behbahani, A. (2023). Involvement load and vocabulary acquisition in digital game-based tasks. In B. L. Reynolds (Ed.), *Vocabulary learning in the wild* (pp. 163–189). Springer. https://doi.org/10.1007/978-981-99-1490-6_6

Read, J. (2023). Towards a new sophistication in vocabulary assessment. *Language Testing, 40*(1), 40–46. https://doi.org/10.1177/02655322221125698

Reynolds, B. L. (2017). Evidence for the task-induced involvement construct in incidental vocabulary acquisition through digital gaming. *Language Learning Journal, 45*(4), 466–484. https://doi.org/10.1080/09571736.2014.938243

Reynolds, B. L. (2023a). Introduction: A scoping review of second language vocabulary learning in the wild. In B. L. Reynolds (Ed.), *Vocabulary learning in the wild* (pp. 1–63). https://doi.org/10.1007/978-981-99-1490-6_1

Reynolds, B. L. (2023b). Vocabulary learning in the wild. *Springer Nature.* https://doi.org/10.1007/978-981-99-1490-6

Rodgers, M. P., & Heidt, J. (2020). Levelling up comprehensible input and vocabulary learning: The lexical profile of videogames. In V. Werner & F. Tegge (Eds.), *Pop culture in language education* (pp. 215–227). Routledge. https://doi.org/10.4324/9780367808334

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261–274. https://doi.org/10.1017/S0261444819000053

Soyoof, A., Reynolds, B. L., Chan, K. K., Tseng, W.-T., & McLay, K. (2023). Massive online multiplayer games as an environment for English learning among Iranian EFL students. *Computer Assisted Language Learning, 1–44*, 1. https://doi.org/10.1080/09588221.2023.2171065

Soyoof, A., Reynolds, B. L., Shadiev, R., & Vazquez-Calvo, B. (2024). A mixed-methods study of the incidental acquisition of foreign language vocabulary and healthcare knowledge through serious game play. *Computer Assisted Language Learning, 37*(1–2), 27–60. https://doi.org/10.1080/09588221.2021.2021242

Sykes, J. M., & Reinhardt, J. (2013). *Language at play: Digital games in second and foreign language teaching and learning*. Pearson.

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL, 24*, 302–321. https://doi.org/10.1017/S095834401200016X

Thompson, C. G., & von Gillern, S. (2020). Video-game based instruction for vocabulary acquisition with English language learners: A Bayesian meta-analysis. *Educational Research Review, 30*, 1–23. https://doi.org/10.1016/j.edurev.2020.100332

Tsai, Y. L., & Tsai, C. C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education, 125*, 345–357. https://doi.org/10.1016/j.compedu.2018.06.020

Vnucko, G., & Klimova, B. (2023). Exploring the potential of digital game-based vocabulary learning: A systematic review. *Systems, 11*(2), 1–18. https://doi.org/10.3390/systems11020057

Whitton, N. (2014). *Digital games and learning: Research and theory.* Routledge.

Wrobetz, K. R. (2023). Gaming in a foreign language: L2 vocabulary processing in a single-player role-playing game. In B. L. Reynolds (Ed.), *Vocabulary learning in the wild* (pp. 191–218). Springer. https://doi.org/10.1007/978-981-99-1490-6_7

Wu, Q., Zhang, J. W., & Wang, C. Y. (2020). The effect of english vocabulary learning with digital games and its influencing factors based on the meta-analysis of 2,160 test samples. *International Journal of Emerging Technologies in Learning, 15*(17), 85–100. https://doi.org/10.3991/ijet.v15i17.11758

Zou, D., Huang, Y., & Xie, H. R. (2021). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning, 34*(5–6), 751–777. https://doi.org/10.1080/09588221.2019.1640745

L

Learner's Attitudes, Styles, Strategies and Behaviours in Data-Driven Learning

Amelie Xiaohan Sun and Atsushi Mizumoto
Kansai University, Osaka, Japan

Abstract

This entry explores learner's attitudes, preferences (or styles), strategies and behaviours in data-driven learning (DDL) and the interconnections between them. Attitudes, representing learners' predispositions towards DDL, significantly influence their engagement and success. While many learners exhibit positive attitudes towards DDL, factors such as anxiety and scepticism also present challenges. Learning preferences, particularly the inductive-deductive spectrum, shape how students interact with corpus-based activities, yet this relationship is nuanced and influenced by task design. Learner strategies, including exploring, cross-checking and synthesizing, are essential for effectively navigating DDL, with their use often determined by both external factors and individual learner attributes. Behaviours, as the observable outcomes of these attitudes, preferences and strategies, provide crucial insights into how learners interact with DDL and its impact on language acquisition.

Keywords

Data-driven learning (DDL) · Learner's attitudes · Learning preferences · Learner strategies · Learner behaviours · Individual difference

Introduction

In the fast-changing world of language education, DDL has emerged as a revolutionary approach that utilizes the power of extensive digital corpora to enhance language acquisition. By providing learners with direct access to authentic linguistic data, DDL empowers them to independently discover language patterns and usage through interactive data analysis. DDL not only promotes a deeper, more nuanced understanding of language but also encourages learners to become active participants in their own learning process. The success of DDL, however, is not solely dependent on the availability of data or tools. It is significantly influenced by how learners interact with and utilize these resources.

Many factors that influence the learning process should be taken into account in order to completely comprehend how DDL affects students. These dimensions include learner's attitudes, preferences, strategies and behaviours. Each of these elements plays a crucial role in how effectively learners engage with data and what outcomes they achieve.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_148-1

In DDL, these connections become particularly significant. A learner's attitude towards DDL methods can influence their engagement with corpus-based activities (behaviour), which may be more or less aligned with their learning preferences. Their choice of strategies for navigating and analysing corpora will also depend on both their preference and attitude. In essence, these four elements form a dynamic system where each component influences and is influenced by the others, contributing to the overall effectiveness of the learning process. Understanding and addressing these interconnections is key to creating a supportive and effective learning environment.

Learner's Attitudes in DDL

Learner's attitudes are one of the individual characteristics of second-language (L2) learners. Attitude is defined as the extent to which an individual assesses a psychological object along bipolar dimensions, including good-bad, pleasant-unpleasant and like-dislike (Ajzen, 2001). Based on these bipolar dimensions, learner's attitudes can generally be categorized as positive or negative. Learners' attitudes can play a crucial role in determining their success in L2 acquisition (Tsunemoto & McDonough, 2021).

In DDL, where learners actively engage with authentic language data, their positive attitudes, such as curiosity and openness to new methods, can lead to higher levels of engagement and motivation, allowing learners to make the most of the DDL approach. Conversely, negative attitudes, such as anxiety or scepticism towards technology-based learning, may hinder their capacity to fully benefit from DDL.

A great number of studies have examined learner's attitudes towards DDL, with varying results. Recent syntheses and meta-analyses (Boulton & Cobb, 2017; Chen & Flowerdew, 2018; Lee et al., 2019; Mizumoto & Chujo, 2015; Pérez-Paredes, 2019; Boulton & Vyatkina, 2021) demonstrated that overall, a variety of students held positive attitudes towards DDL activities, especially in terms of the nature of some of its features, such as authenticity and discovery learning. However, the number of studies reporting the presence of negative reactions is not insignificant. Many concerns have been identified, such as participants feeling anxiety when managing technology and handling substantial amounts of data (Gordani, 2012). Some studies have found students maintaining a sceptical attitude to unreliable information being presented on the Internet when they use Google search or Google Scholar as a corpus (Chang, 2014; Charles, 2014). Others report frustration with the complexity of tasks (Chen et al., 2016; Marcos Miguel, 2021) and a few note a lack of motivation due to the perceived difficulty of the activities (Moon & Oh, 2018).

Expanding on the mixed findings, Boulton (2017) cautioned against overinterpreting positive findings, noting potential biases such as the "novelty factor" and the "Hawthorne effect" (p. 185), where learners may respond positively due to the newness of the method or the enthusiasm of researchers.

While learners' attitudes towards DDL are generally positive, there is a need for further research to explore how these attitudes evolve over time, rather than focusing solely on short-term perceptions. Future studies should aim to provide a more nuanced understanding of how learner's attitudes, motivation and self-efficacy interact with DDL in the long term and how teachers can best support learners in this context. By addressing these gaps, teachers can better tailor DDL to meet the needs of diverse learners and maximize its potential as a language learning tool.

Learning Preferences in DDL

Learning preferences, often referred to as learner styles, are the preferred ways in which individuals receive, process and understand information. In this entry, we intentionally avoid using the term "learning styles" because the concept has been heavily scrutinized over the past decade. Some researchers even argue that learning styles are a myth (e.g. Lethaby and Harries (2016)).

In DDL, learning preferences play a significant role in determining how students interact with and

benefit from corpus-based activities. DDL involves learners actively engaging with authentic language data, often through tasks that require pattern recognition, hypothesis testing and independent analysis of linguistic phenomena. The success of DDL can be influenced by whether a learner's preference aligns with the inductive or deductive nature of the DDL tasks.

Research in DDL often categorizes learners along an inductive-deductive spectrum. Inductive learners tend to prefer discovering rules and patterns themselves through observation and analysis. They are generally more receptive to DDL, where exploration and self-directed learning are key components. Studies by Chan and Liou (2005) and Turnbull and Burston (1998) suggest that inductive learners may find DDL particularly engaging and effective.

However, not all research supports a strong link between a learner's inductive-deductive preference and their success in DDL. Mizumoto and Chujo (2016) found that the specific nature of the DDL task might be more critical than the learner's overall preference. For example, learners who prefer a deductive approach—starting with explicit rules before analysing data—might still benefit from DDL if the tasks are designed to accommodate this preference.

Aside from the inductive-deductive spectrum, Poole (2006) investigated learning styles by utilizing the Learning Styles Questionnaire (LSQ) developed by Honey and Mumford (1992). Poole found that "active" learners (those who are activists or pragmatists) were more receptive to DDL, particularly in online components of blended courses. This finding aligns with the notion that learners who engage actively with their learning process—whether inductively or deductively—tend to benefit more from DDL activities.

Similarly, Boulton (2009) expanded the discussion by examining learner's preferences that extend beyond the traditional inductive-deductive framework, employing the Index of Learning Styles (ILS) for analysis. This tool categorizes learners into four distinct categories: active-reflective, sensing-intuitive, visual-verbal and sequential-global. Boulton's study revealed that learners with a visual preference were slightly more inclined to use corpora, but overall, the differences between styles were minimal.

While learning preferences do influence how students engage with DDL, the relationship is complex and not entirely straightforward. Other dimensions of cognitive preferences, such as visual-verbal preferences or active-reflective tendencies, can also play a role, although the impact may be subtle. Therefore, teachers should consider a variety of learning preferences when designing DDL activities, ensuring that tasks are flexible enough to accommodate different preferences and maximize the effectiveness of this innovative approach to language learning.

Learner Strategies in DDL

Learning strategies are "specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations" (Oxford, 1990, p. 8). In DDL, learner strategies relate to the particular techniques and approaches that students employ to engage with and learn from corpus data.

Various studies have identified key learner strategies in DDL. Hu and Nassaji (2014) identified nine lexical strategies categorized into three distinct groups: form-focused strategies, which include analysing, associating and repeating; meaning-focused strategies, comprising the use of textual clues, prior knowledge and paraphrasing; and evaluating strategies, which encompass making inquiries, confirming or disconfirming information and providing commentary. Furthermore, Lee et al. (2020) pinpointed three specific strategies tailored to DDL: exploring, cross-checking or double-checking and synthesizing. According to Lee et al. (2020), exploring involves navigating corpora to uncover patterns, word collocations or examples of specific grammatical structures. This involves utilizing a concordancer and analysing the search results to gain insights into authentic language use. Cross-checking or double-checking is another important strategy in that students compare multiple examples from the

corpus to confirm their hypotheses about language rules or usage. This strategy ensures that initial observations are accurate and that incorrect conclusions are avoided by examining additional data points. Synthesizing is also crucial, as learners integrate information from different corpus searches to construct a broader understanding of language patterns. This could involve combining findings from various queries to form a comprehensive view of how certain words or structures function in different contexts.

These strategies are also influenced by various external environmental factors. Sun (2003) identified four key factors that influenced L2 learner strategies in DDL: learner's existing knowledge of the language, concordancer skills, cognitive skills and the teacher's intervention. Learners with a solid linguistic background are better equipped to interpret corpus data, while strong cognitive skills aid in processing and analysing large volumes of information. Proficiency in concordancer skills allows learners to navigate and utilize corpus tools effectively. Lastly, teacher's intervention plays a crucial role by guiding learners, offering feedback and helping them interpret complex data, thereby enhancing the overall effectiveness of DDL activities.

The choice of strategies in DDL is also influenced by learners' attitudes and preferences. For instance, a learner with a positive attitude towards experimentation may be more inclined to adopt a trial-and-error exploring strategy, testing different hypotheses against the data until they find a pattern that holds true. Similarly, a learner with an analytical cognitive preference may prefer strategies such as cross-checking, where they compare multiple examples from the corpus to confirm their hypotheses, ensuring the accuracy of their initial observations. The interconnection between strategies, attitudes and preferences suggests that successful DDL engagement requires a harmonious alignment of these elements.

Learner Behaviours in DDL

Learner behaviours can be understood as the external and observable actions learners take during learning, such as visual orientation or attentiveness, which reflect their engagement and contribute to academic success (Goldberg et al., 2021).

Researchers observe and summarize learner behaviours through various methods to draw meaningful conclusions. Boulton and Vyatkina (2021) synthesized 489 empirical DDL studies published in years 1989–2019 and found that most studies employ multiple instruments to capture learner's behaviours, with an average of 1.9 instruments per study. While questionnaires are commonly used to gather learners' self-reported perceptions and actions during DDL (see Vyatkina (2016) and Zhang (2018)), they are often supplemented by more behaviour-focused methods to better understand how learners interact with corpus data in real time such as teacher observations, field notes and various self-reports, including journals, diaries and think-aloud protocols.

Several studies found that learners engage in exploratory behaviours when using corpus data. These behaviours include frequently consulting corpus examples to verify hypotheses about language use, as seen in Boulton (2012), where learners cross-checked their writing choices against corpus entries to confirm correct word usage. Learners also demonstrate selective focus behaviours, as noted by Smith (2011), where they tend to prioritize certain language forms or structures that are immediately relevant to their learning goals, such as focusing more on verb collocations in writing tasks.

In addition to exploration and selective focus, learners exhibit self-regulatory behaviours. Hadley and Charles (2017) observed that learners often monitor and adjust their strategies in real time, such as switching from simple word look-ups to more complex searches involving collocations or grammatical structures, depending on their task requirements. This aligns with the findings of Naismith (2017), who noted that learners sometimes shift their attention from isolated vocabulary learning to more holistic language patterns.

Tracking studies, such as Crosthwaite et al. (2019), revealed behavioural patterns where

learners spent varying amounts of time navigating between corpus entries, often revisiting specific examples multiple times. These behaviours suggest iterative processing, where learners use repetition and revisiting as strategies to internalize language forms more effectively. Similarly, in Pérez-Paredes et al.'s (2012) study, screen recordings showed that learners often compare multiple sources or linguistic structures in a single session, indicating a pattern of comparative analysis in their DDL activities.

In terms of production-related behaviours, Vyatkina (2016) found that learners made self-corrections after consulting corpus data, demonstrating an active engagement with the learning material and a behaviour of self-repair during DDL-based writing tasks.

These examples of learner behaviours—exploration, selective focus, self-regulation, iterative processing, comparative analysis and self-repair—illustrate the complexity and richness of how learners interact with DDL resources. Additionally, for theorists and practitioners, the choice of instruments is crucial in capturing these behaviours accurately and comprehensively. As DDL continues to evolve, so too will the methods used to study learner's behaviours, offering new insights into the ways in which data-driven approaches can enhance language acquisition. These findings underscore the importance of using a wide range of instruments to fully capture the nuances of learner's behaviours and the impact of DDL on language learning outcomes.

Interconnections Between Learner's Attitudes, Preferences, Strategies and Behaviours

Imagine a learner as a sailor navigating the seas. The ocean represents the vast amount of language data available in DDL. Attitude is like the sailor's mindset or approach to the journey. If the sailor is enthusiastic, curious and confident, they are more likely to embrace challenges and explore new waters. On the other hand, if the sailor is anxious or sceptical, they might hesitate, avoid unfamiliar territories or even give up when the seas get rough.

The cognitive preference of a sailor is similar to the kind of ship they choose to sail. Some sailors, with a preference for quick processing and adaptability, might choose a speedy, agile boat that allows for rapid navigation and quick changes in direction. Other sailors, perhaps those who prefer thorough, methodical approaches, might opt for a sturdy, reliable vessel that moves steadily through the water. The type of ship they choose, influenced by their cognitive preference, then affects how they navigate the sea of information in DDL.

Strategy is the sailor's plan or method for reaching their destination. It is how they decide to chart their course, what tools they use for navigation and how they respond to obstacles along the way. A well-thought-out strategy helps the sailor reach their destination efficiently, just as a good learning strategy helps learners achieve their goals.

The sailor's behaviours are the actions they take during the journey—how they steer the ship, adjust the sails and respond to changing weather conditions. In DDL, behaviours are the observable actions that learners take, such as how often they engage with the data, their level of participation in activities and how they apply what they have learned. Consistent, purposeful behaviours can help the sailor stay on course and make steady progress, just as they can help a learner succeed in DDL.

By understanding these elements and how they interconnect—just as a sailor must balance their mindset, choice of vessel, navigational strategy, and actions—a learner can more effectively navigate the learning process in DDL and reach their educational destination.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52*(1), 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Boulton, A. (2009). Testing the limits of data driven learning: Language proficiency and training. *ReCALL, 17*(1), 47–64. https://doi.org/10.1017/S0958344009000068

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261–291). John Benjamins.

Boulton, A. (2017). Data-driven learning and language pedagogy. In *Language, education and technology: Encyclopedia of language and education* (Vol. 3, pp. 181–192). Springer. https://doi.org/10.1007/978-3-319-02328-1_15-1

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning and Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Chan, T., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18*, 231–251. https://doi.org/10.1080/09588220500185769

Chang, J. Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL, 26*(2), 243–259. https://doi.org/10.1017/S0958344014000056

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30–40. https://doi.org/10.1016/j.esp.2013.11.004

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics, 23*(3), 335–369. https://doi.org/10.1075/ijcl.16130.che

Chen, H. H. J., Wu, J. C., Yang, C. T. Y., & Pan, I. (2016). Developing and evaluating a Chinese collocation retrieval tool for CFL students and teachers. *Computer Assisted Language Learning, 29*(1), 21–39. https://doi.org/10.1080/09588221.2014.889711

Crosthwaite, P., Wong, L., & Cheung, J. (2019). Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. *ReCALL, 31*(3), 255–275.

Goldberg, P., Schwerter, J., Seidel, T., Müller, K., & Stürmer, K. (2021). How does learners' behavior attract preservice teachers' attention during teaching? *Teaching and Teacher Education, 97*, 103213. https://doi.org/10.1016/j.tate.2020.103213

Gordani, Y. (2012). The effect of the integration of corpora in reading comprehension classrooms on English as a Foreign Language learners' vocabulary development. *Computer Assisted Language Learning, 26*(5), 430–445. https://doi.org/10.1080/09588221.2012.685078

Hadley, G., & Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Language Learning & Technology, 21*(3), 131–152. https://doi.org/10125/44624

Honey, P., & Mumford, A. (1992). *The manual of learning styles*. Ardingley House.

Hu, H. M., & Nassaji, H. (2014). Lexical inferencing strategies: The case of successful versus less successful inferencers. *System, 45*, 27–38.

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal, 104*(2), 345–362. https://doi.org/10.1111/modl.12634

Lethaby, C., & Harries, P. (2016). Learning styles and teacher training: Are we perpetuating neuromyths? *ELT Journal, 70*(1), 16–27. https://doi.org/10.1093/elt/ccv051

Marcos Miguel, N. (2021). Exploring tasks-as-process in Spanish L2 classrooms: Can corpus-based tasks facilitate language exploration, language use, and engagement? *International Journal of Applied Linguistics, 31*(2), 211–228. https://doi.org/10.1111/ijal.12314

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies, 22*(1), 1–18.

Mizumoto, A., & Chujo, K. (2016). Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System, 61*, 55–64. https://doi.org/10.1016/j.system.2016.07.010

Moon, S., & Oh, S. Y. (2018). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL, 30*(1), 48–67. https://doi.org/10.1017/S0958344017000246

Naismith, B. (2017). Integrating corpus tools on intensive CELTA courses. *ELT Journal, 71*(3), 273–283. https://doi.org/10.1093/elt/ccw076

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Newbury House.

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pérez-Paredes, P., Sánchez-Tornel, M., & Alcaraz Calero, J. M. (2012). Learners' search patterns during corpus-based focus-on-form activities: A study on hands-on concordancing. *International Journal of Corpus Linguistics, 17*(4), 483–515. https://doi.org/10.1075/ijcl.17.4.02par

Poole, J. (2006). E-learning and learning styles: students' reactions to web-based language and style at Blackpool and the Fylde college. *Language and Literature, 15*(3), 307–320. https://doi.org/10.1177/0963947006066129

Smith, S. (2011). Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning, 24*(4), 291–316. https://doi.org/10.1080/09588221.2011.557024

Sun, Y. C. (2003). Learning process, strategies and web-based concordancers: A case study. *British Journal of Educational Technology, 34*, 601–613. https://doi.org/10.1046/j.0007-1013.2003.00353.x

Tsunemoto, A., & McDonough, K. (2021). Exploring Japanese EFL learners' attitudes toward English pronunciation and its relationship to perceived accentedness. *Language and Speech, 64*(1), 24–34. https://doi.org/10.1177/0023830919900372

Turnbull, J., & Burston, J. (1998). Towards independent concordance work for students: Lessons from a case study. *On-call, 12*(2), 10–21.

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159–179. https://doi.org/10125/44487

Zhang, L. (2018). Effectiveness of college student English writing teaching based on data-driven learning. *International Journal of Emerging Technologies in Learning (iJET), 13*(4), 106–116. https://doi.org/10.3991/ijet.v13i04.8474

R

Research Synthesis in Data-Driven Learning

Clarence Green
Faculty of Education, University of Hong Kong, Hong Kong, China

Abstract

Research synthesis has been widely applied in Data-Driven Learning (DDL) and will continue to play an important role as research progresses within CALL. This entry provides an overview of research synthesis in DDL, with a review of several key research syntheses. The entry explains how research synthesis is used in DDL for evidence evaluation, summarisation, critical appraisal and articulating the scope of research in the field. The most common synthesis research designs are outlined, guiding readers through the methods, motivations and findings of selected studies that illustrate best practice and point toward areas for future research synthesis.

Keywords

Research synthesis · DDL · Data-Driven Learning · Corpus linguistics

Introduction

This entry provides an overview of research synthesis in Data-Driven Learning (DDL), with a focus on its applications, methods and findings. Research synthesis is important to DDL, as it supports evidence evaluation, summarisation, critical appraisal and articulation of the scope of research in the field, and what is needed to move it forward. Since primary studies will increase, research synthesis will continue to play an important role. This entry outlines common synthesis research designs, with illustrative examples from DDL, guiding readers through their methods and motivations. An additional goal of this entry is to highlight areas of development for future DDL syntheses.

Synthesis in Data-Driven Learning: Research Designs, Motivations and Methods

There has been increasing use of research synthesis across most subfields of applied linguistics and education. Reasons motivating this include a trend toward researching the research, which provides insights into what has been done and found within an ever-increasing volume of research. Further, 'what works' in education and the evidence for pedagogical practices have become increasingly important to policy, curriculum and teacher training. While not all research synthesis is effectiveness-

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_149-1

oriented, a good deal is, and since DDL is a pedagogy, we are interested in knowing which educational outcomes it improves (Crosthwaite & Boulton, 2024; O'Keeffe, 2020).

There are different methods for synthesizing educational research. Traditional synthesis consists of narrative literature reviews. These reviews are non-systematic, that is, no pre-specified methodology guides the selection of studies to be synthesised, nor the way the primary studies are analysed to draw conclusions. Though they can be valuable, narrative reviews can introduce bias, such as excluding primary studies that might affect synthesis conclusions. Educational research has therefore largely shifted to systematic reviews. Several methods for systematic synthesis exist, depending on one's goals. Evidence evaluations are typically accomplished by metanalysis (Boulton & Cobb, 2017) and Narrative Synthesis Systematic Reviews/NSSR (Green et al., 2024; Popay et al., 2006), though the latter is yet to be employed with a DDL focus. Metanalysis can evaluate both general effects of DDL on any outcomes (Ueno & Takeuchi, 2023), or specific outcome effects, such as on L2 vocabulary (Lee et al., 2019). Scoping reviews are a method for synthesising what has been done and, typically, highlighting gaps (Boulton & Vyatkina, 2021). Further, there are an increasing number of Qualitative Systematic Reviews (Pérez-Paredes, 2022). These can provide scoping reviews, (re)code primary data from previous qualitative research (Chong & Plonsky, 2021), or code primary studies themselves for themes from which a critical review can be conducted to inform theory and new research avenues (Chen & Flowerdew, 2018). What all research synthesis methods have in common is the goal of reducing bias in the representation of the research. They are systematic because they are (typically) exhaustive summaries, methodologically transparent and reproducible.

Once the research goal is set, the synthesis process is typically similar across subtypes of systematic reviews. First, a pool of search terms is developed that will adequately capture relevant research. These search terms should be reported in full. A selection of research databases, and perhaps specific journals, are searched to extract bibliographic records. Researchers make decisions about what types of research to include, for example, restricted to peer-reviewed journal articles and books, or inclusive of 'grey' literature such as unpublished research. After a pool of bibliographic records is obtained, titles and abstracts are screened for inclusion, often within a research synthesis tool (e.g. Rayyan, https://www.rayyan.ai/; Covidence, https://www.covidence.org/). Screening against inclusion criteria is typically done blindly by multiple research team members. This results in the final study pool for synthesis. Often, ancestry searches are conducted, that is, additionally screening studies cited in the references of the final pool for inclusion.

Inclusion criteria depend on the research motivations and type of review. In DDL, metanalysis and scoping have been the most applied systematic review methods. For DDL effectiveness, metanalysis computes average effect sizes from the data reported in prior studies, so inclusion criteria require studies to have reported measured outcomes, pre- and post-test data, with control groups and/or paired samples; in other words, experimental or quasi-experimental interventions. Statistical data is extracted from included studies and an average effect size calculated. Variables that influence the outcome are also often coded, such as proficiency, outcome type, effects measured against controls or pairwise comparisons. Robustness is evaluated through techniques such as funnel plots to check for publication bias. Conclusions about pedagogical effectiveness can be evaluated across studies through metanalytic synthesis, particularly useful when studies report conflicting or heterogenous results. Such syntheses are powerful in research-practice translation, offering teachers an expected impact on students from implementation of the pedagogy.

Scoping reviews typically have different inclusion criteria, chosen by researchers depending on their research goals. For example, they might include research designs broader than (quasi) experimental studies, or, on the hand, potentially limit scope to particular contexts (e.g. English for Academic Purposes/EAP). In DDL, typically

scoping reviews have included both numerical information and qualitative analyses of themes generated from coding primary studies. Coding may aim to synthesise the scope of research areas, the main theories tested, sampling practices, and establish meta-themes from the research record as a whole.

Specific Applications and Findings of Research Synthesis in DDL

In their review of 21 previous syntheses, Boulton et al. (2024) suggest four periods of DDL research synthesis. The first period, ranging from 2007 to 2012, consisted exclusively of traditional narrative literature reviews. The second period, 2013–2018, expanded to include critical reviews and metanalysis. The third period, 2019–2024, was characterized almost exclusively by advanced systematic reviews. Using these periods as the framework, the following section guides readers through illustrative applications of research synthesis in DDL, some important findings, and traces the development of research synthesis as a method within the field over time.

One of the earliest DDL syntheses was conducted by Chambers (2007). This was a traditional literature review of 12 papers, chosen by the researcher as being influential since 1990. Chambers synthesized narratively what these studies reported about corpus consultation by learners and identified what might facilitate research translation to classrooms. Chambers concluded, amongst other things, that her primary study synthesis revealed the rising popularity of DDL at the time, that corpus consultation was generally favourably viewed by learners, but that the research record contained limited testing of language learning.

Boulton (2010), another first period synthesis, conducted an early scoping review. He reviewed 39 DDL evaluation studies narratively to identify the main research themes since 1991. Boulton found the scope of the research at the time focused on attitudes toward corpus use, practices, corpus tools, and (less prominently) learning outcomes. This review reflects research synthesis in a transitional stage toward more systematic quantification, with Boulton (2010) also extracting data from the primary studies and computing, for example, an average sample of 38.97 students across studies, ranging from $N > 100$ to $N = 1$. He further tabulated that only 8 studies included controls, 16 were qualitative, and of 23 quantitative studies, only 12 reported more than descriptive statistics. Boulton concluded the synthesis demonstrated DDL was positively perceived by learners, that they were capable of learning from corpora and using tools, but that the scope of research evidence had not yet extended to younger learners or large-samples.

Addressing evidence of learning, and falling within the second wave of research synthesis, Boulton and Cobb (2017) conducted a metanalysis on all second language outcomes. They generated bibliographic records from peer-reviewed articles, books, book chapters, and grey literature (e.g., PhD theses) indexed in LLBA, ERIC, MLA, ProQuest, JSTOR, and DOAJ. They also conducted Google Scholar and ancestry searches. Search terms included the following: 'corpus/corpora', 'DDL', 'data-driven', 'concordance/concordance/concordancing' and 'Johns', combined with 'language' and/or 'learning'. Boulton and Cobb found 205 DDL evaluations. These were screened for studies that were L2-focussed, used a linguistic tool (e.g. concordancing), reported participant numbers and measured outcomes. This resulted in 64 meta-analysed studies. Boulton and Cobb computed the Cohen's *d* statistic for each study and averaged these for an overall effect size. Cohen's *d* is a standardized measure that can be used to evaluate the impact of interventions by measuring the difference between two means (e.g. pre- and post-test), taking into account pooled standard deviation. For paired research designs, the effect size for DDL was $d = 1.50$ ($SD = 0.91$). For comparisons against control groups, the effect was $d = 0.95$ ($SD = 0.99$). These are large effect sizes. Boulton and Cobb concluded that DDL had been effective in most contexts, for most learners, and for most outcomes.

Chen and Flowerdew (2018), also falling within the second period of research synthesis, applied critical review methods to DDL. They aimed to synthesize publications missed by metanalysis (e.g. by not reporting numbers) but would nonetheless add value to understanding the complex processes of applying DDL to EAP, particularly for academic writing. Their goal was to critically evaluate what was needed to move DDL forward in EAP. Chen and Flowerdew conducted targeted searches of journals (e.g. *Computer Assisted Language Learning, English for Specific Purposes,* etc.), library database searches of target words (e.g. 'DDL', 'data-driven learning', 'EAP', 'corpora', etc.) and ancestry searches. This resulted in 37 articles between 2000 and 2018. The researchers coded primary studies for themes, supported by concordance analysis of keywords. From this process emerged several critical recommendations, including the need for more research in different geographical and classroom contexts, more complex interventions that included DDL within broader writing instruction, more research into self-directed corpus use and more studies of teacher training.

The latest period of research synthesis, as noted, has been dominated by methodologically advanced systematic reviews. Ueno and Takeuchi (2023), for example, updated the DDL evidence of effectiveness in a metanalysis. They used the same inclusion and screening criteria as Boulton and Cobb but synthesised research after 2014 to add 144 more publications and applied advanced statistical methods (e.g. mixed effects models). Their synthesis came to the same conclusion, however, that DDL positively impacts L2 language outcomes. For studies measuring within-group DDL effects, they found an overall average effect of $d = -1.12$ ($p < 0.001$) (please note the negative value reflects the way the effect size was computed, but in fact represents a positive impact of DDL). For studies measuring between group comparison, a moderate effect was found $d = 0.47$ ($p < 0.001$).

Recent DDL synthesis has applied qualitative methods, as with Pérez-Paredes' (2022) scoping review. Pérez-Paredes was interested in the extent of DDL in CALL, and which factors played a role in facilitating, or hindering, DDL's normalisation in the classroom. For example, normalisation may be helped or hindered by the extent of stakeholder training, syllabus uptake and logistical integration. He included research published 2011–2015 in five journals, namely, *Computer Assisted Language Learning; System; CALICO Journal; ReCALL* and *Language Learning & Technology.* From these, 32 papers were sourced via search terms: 'corpus', 'corpora', 'DDL', 'data-driven learning' and 'corpus-based'. Studies were qualitatively coded for research focus, questions and factors related to DDL normalisation. Pérez-Paredes found prominent themes related to the lack of support and syllabus integration of DDL. He concluded that DDL is still perceived as developing within CALL, and this perception as 'developing' might limit its normalisation in the classroom.

The increasing historical span of DDL has allowed for new kinds of synthesis. For example, Boulton and Vyatkina (2021) conducted a scoping review of how DDL research changed from 1989 to 2019, synthesising what previous studies proposed should be future research avenues and analysing the extent to which these were pursued in subsequent research. Boulton and Vyatkina included all empirical (e.g. experimental, survey and observational) peer-reviewed journal articles, book chapters and conference proceedings. Using the search terms of Boulton and Cobb (2017) and similar bibliographic databases, 489 studies were identified. The authors created a corpus out of these articles and, using AntConc freeware (https://www.laurenceanthony.net/software/antconc/), computed keywords and ngrams (word sequences) occurring in the conclusion sections of primary studies. This process identified four main themes: 'theory and methods', 'learning contexts' 'implementation' and 'technology'. Boulton and Vyatkina found change over time, for example, early DDL research included keywords related to generativist theory, while later periods shifted to constructivism and sociocultural theory. Methods

increased in thematic prominence around 2016 onward. Further, themes emerged related to the need for larger sample sizes, longer treatments, more diverse research on learner-corpus interactions, more longitudinal research and more elaboration of the theory underpinning DDL. However, the researchers conclude that their synthesis found little evidence of a historical trend toward addressing these calls, thus remaining important areas for future research.

DDL Research Synthesis Going Forward

As the primary research record grows, there will be more DDL research synthesis needed. Second-order syntheses, sometimes called tertiary reviews, become possible with multiple prior research syntheses. These are systematic reviews of systematic reviews. For example, a second-order metanalysis computes an overall effect size from multiple metanalyses, essentially applying the standard methods above but using prior metanalyses as the input data rather than primary studies. Future syntheses might also use metanalysis variants such as NSSRs, which can include statistical metanalysis and individual study quality assessments. Thick descriptions of interventions with quality assessment make the evidence of impact more robust in terms of what works, for whom, for which outcomes and in which contexts.

Cross-References

- ▶ Corpus Literacy and Data-Driven Learning
- ▶ Learning Outcomes from Data-Driven Learning
- ▶ Pedagogical Innovation in CALL
- ▶ The History of Data-Driven Learning
- ▶ The Scope of Data-Driven Learning
- ▶ Tools for Data-Driven Learning

References

Boulton, A. (2010). Learning outcomes from corpus consultation. In M. Moreno Jaén, F. Serrano Valverde, & M. Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching* (pp. 129–144). Equinox.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology, 25*(3), 66–89.

Boulton, A., Vyatkina, N., & Cobb, T. (2024). *Classroom applications of corpus analysis*. [Manuscript submitted for publication].

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3–16). Rodopi.

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics, 23*(3), 335–369.

Chong, S. W., & Plonsky, L. (2021). A primer on qualitative research synthesis in TESOL. *TESOL Quarterly, 55*(3), 1024–1034.

Crosthwaite, P., & Boulton, A. (2024). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In H. Tyne, M. Bilger, L. Buscail, M. Leray, N. Curry, & C. Pérez-Sabater (Eds.), *Discovering language: Learning and affordance*. Peter Lang.

Green, C., Giblin, I., & Mulder, J. (2024). A systematic narrative synthesis review of the effectiveness of genre theory and systemic functional linguistics for improving reading and writing outcomes within K-10 education. *The Australian Journal of Language and Literacy, 47*, 203–223.

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753.

O'Keeffe, A. (2020). Data-driven learning: A call for a broader research gaze. *Language Teaching, 54*(2), 259–272.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme. *Version*, *1*(1), b92.

Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics, 3*(3), 100076.

N

Notable Influential Researchers in Data-Driven Learning

Claudia Theresia Wunderlich
Faculty of Applied Natural Sciences and Humanities, Technical University of Applied Sciences, Würzburg-Schweinfurt, Germany

Abstract

In this entry, notable influential researchers who shaped the field of data-driven learning (DDL) are identified. The term DDL itself was coined by Tim Johns in 1990, and since then, this area has given rise to much research activity. The group of eminent DDL researchers is determined with the help of recent bibliometric data and a list of leading language researchers based on a cocitation ranking by the academic publishing company Elsevier as the most objective sources available. These scholars are grouped according to the four developmental stages of a discipline (Shneider, Trends in Biochemical Sciences 34:217–223, 2009) during which they proved to be the most impactful. How the final stage of this relatively young discipline has been manifesting in the work of these academics is also outlined, as well as the most recent impact they have made. This particularly regards the theoretical underpinnings of DDL, the advent of generative artificial intelligence (AI) tools such as ChatGPT as well as the hitherto underdeveloped area of LOTEs (Languages Other Than English) in DDL.

Keywords

Data-driven learning · DDL · Classroom concordancing · Influential researchers · Notable researchers · Second-language acquisition

Introduction

With the rise of corpus linguistics since the 1960s and the advent of microcomputers in the 1980s, data-driven learning (DDL) emerged as a new approach to language teaching. Typically, DDL research has been situated at the intersection of applied linguistics, corpus linguistics and computer-assisted language learning (CALL, Vyatkina, 2024, p. 5). In the past, a frequently cited definition has been that by Gilquin and Granger (2010) stating that "Data-Driven Learning (DDL) consists in using the tools and techniques of corpus linguistics for pedagogical purposes" (p. 430). Since then, the leading researchers have proposed to widen the notion of DDL (see also Section "Stage IV: Most Recent High-Impact Researchers").

Tim Johns, a leading pioneer in the field, originally coined the term DDL in 1990 (Johns, 1990, 1991). He taught English for Specific Purposes

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_150-1

(ESP) to international students at the University of Birmingham when John Sinclair was working there on the COBUILD (Collins Birmingham University International Language Database) project.

Sinclair provided principal theoretical underpinnings to DDL by proposing the idiom principle stressing that words tend to cooccur as multiword expressions in a language unlike the open-choice principle where words combine freely, and he actively promoted DDL (Sinclair, 1991). Sinclair based his ideas on Firth's contextual theory of meaning according to which "You shall know a word by the company it keeps!" (Firth, 1957/1968, p. 179).

DDL researchers not only draw on the work of Sinclair. Also influenced by Firth is M.A.K. Halliday's term lexicogrammar, frequently encountered in DDL, emphasising that grammar and lexis should not be viewed separately (Halliday, 1966). Other language theories DDL exploits are Hunston and Francis's (2000) Pattern Grammar as a Lexical Grammar, also originating in the COBUILD project, Hoey's (2005) concept of lexical priming to explain statistically significant cooccurrences of words as well as Goldberg's (1995) Construction Grammar (CxG) as another approach to describe language patterns and most recently Taylor's (2012) Mental Corpus, that is, collections of word combinations which are more or less frequently reused in language. Overall, however, all eminent corpus linguists have contributed to DDL, at least indirectly.

Here, the most eminent DDL researchers are identified on the basis of the data and names provided in the main sections of the bibliometric study by Dong et al. (2023). It measures the scholarly impact of 126 research articles and 3297 references from the period between 1994 and 2021 from the Web of Science database analysed with CiteSpace 6.1.R2. Dong et al. outlined four (not always clear-cut) evolutionary stages of a discipline identified by Shneider (2009): (1) conceptualising DDL (1980–1998), (2) a developing field (1998–2011), (3) a growing field (2011–now) and (4) the emerging stage. It thus recognises key DDL topic clusters and researchers through a cocitation and structural variation analysis (Dong et al., 2023, pp. 341 ff.). Of these individuals, the main researchers were chosen who are well known in the community. Additionally, the list of the 93 most eminent language researchers from the New Slator 2022 Language Industry Market Report—Language AI Edition based on the author-citation dataset by Elsevier—was used to identify potential additional top-ranked researchers (Txabarriaga, 2022). These are the leading 93 language experts featured among the 100,000 scientists most frequently cited by other researchers.

Despite possible limitations that may be inherent to the methodologies of these rankings, these sources were chosen in an attempt to be as objective as possible to identify the leading researchers. Additionally, how some of these most eminent researchers in DDL have influenced the latest research and discussion since 2022 is also briefly reviewed. This particularly regards bridging the research-practice gap (Chambers, 2019; Crosthwaite, 2024), the wider application and scope of DDL and the tools it entails, the arrival of ChatGPT, theoretically linking DDL with Second Language Acquisition (SLA) and Languages Other Than English (LOTEs).

Stage I: Tim Johns and the Conceptualisation of DDL (1980–1998)

While groundbreaking work was conducted, and key methods and concepts introduced during this initial period, none of the researchers from that period appeared in the later stages identified by Dong et al. (2023, pp. 345 ff.). At this stage, Tim Johns is recognised as the most eminent pioneer of DDL (e.g. Johns (1990, 1991)). He regarded students as language detectives, considering "every student a Sherlock Holmes" (Johns, 1997, p. 101). A collection of his DDL activities called Kibitzers is available on Mike Scott's page (https://lexically.net/TimJohns/#gsc.tab=0). Additionally, also with Mike Scott, he created the first concordancer for DDL, MicroConcord (Johns, 1986).

Stage II: Researchers of a developing field (1998–2011)

Stage two "tends to display a primary focus on the principal techniques and tools" (Shneider, 2009, p. 341). Some of the researchers who started publishing during this stage have remained among the most prominent DDL researchers until present, i.e. also at later stages, and are mentioned in all relevant sections.

Designers of Corpora and Corpus Tools

Lou Burnard is well known as the creator of the BNC Baby corpus, a four-million-word BNC subcorpus and software on CD-ROM specifically designed for classroom use (Burnard, 2004). Further, Mark Davies is best known for compiling the Corpus of Contemporary American English (COCA, Davies, 2008), although he also contributed various corpora to english-corpora.org and to LOTEs, for example, the Corpus del Español and the Corpus do Português. Also, Laurence Anthony is a researcher not only known as an ESP specialist (e.g. Anthony (2018)) but as one who first reached eminence during the second stage of DDL on account of the AntConc software he developed (Anthony, 2004). This freeware corpus analysis toolkit (https://www.laurenceanthony.net/software/antconc/)—and later a suite of various other corpus tools—became the most popular software replacing Wordsmith by Mike Scott, which was based on MicroConcord (Johns, 1986; Scott, 2024). Finally, Adam Kilgariff must be included here on the basis of the Elsevier ranking (Txabarriaga, 2022). In 2003, he founded the Czechia-based company Lexical Computing Ltd. behind Sketch Engine, the corpus manager and text analysis software (https://www.sketchengine.eu). Additionally, he coined the expression "word sketches", a key component of the Sketch Engine software that visualises and summarises a word's collocational behaviours. It can also be used for DDL, currently in more than 100 languages. As a spin-off, Sketch Engine for Language Learning (SkELL), a free, simplified version of Sketch Engine for classroom use, was created offering examples, word sketches and synonyms (https://skell.sketchengine.eu/#home?lang=en, Kilgarriff et al., 2015).

Focus on Learners

The second category of researchers of this period focused on the learners. One such researcher is Angela Chambers, who described how individual factors such as learning style and motivation impact the success of DDL (Chambers, 2005) and conducted the first ever synthesis of 12 DDL studies, demanding more quantitative studies (Chambers, 2007). Íde O'Sullivan is another often-cited DDL researcher who investigated how corpus consultation literacy impacts process-oriented learning and identified the skills learners acquire through DDL: "predicting, observing, noticing, thinking, reasoning, analysing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorising, hypothesising, and verifying" (O'Sullivan, 2007, p. 277). Furthermore, Claire Kennedy and Tiziana Miceli were among the first to focus on a LOTE: they tested the effectiveness of corpora as guides to pattern usage and as a writing aide for intermediate learners of Italian in Australia (Kennedy & Miceli, 2001, 2010). Last but not least, Alex Boulton concluded in his learner-centred research that DDL is suitable for low-level learners despite widespread beliefs to the contrary and that paper-based activities can be beneficial while pursuing the overall goal of popularising DDL (2009a, b, 2010).

Stage III: Eminent Researchers in a Growing Field (2011–now)

Investigating the Effectiveness of DDL

To begin with, Ana Frankenberg-Garcia not only researched the effectiveness of DDL noting a positive correlation between corpus-based examples and learners' heightened perception of grammatical form even at lower levels when scaffolding was provided (Frankenberg-Garcia, 2014) but is also known as the leader of the ColloCaid project (https://collocaid.uk/). The ColloCaid tool does

not expose concordance lines but suggests potential collocates to the words typed by the user that are typical in academic writing. The tool has a potential for future integration into standard text editors. Tom Cobb is the author of The Compleat Lexical Tutor (https://www.lextutor.ca/), a freely accessible software suite for the study of language and DDL. Additionally, he has published widely in the field of DDL, including the first meta-analysis of corpus use in second-language learning and teaching with Alex Boulton for the period between 1989 and 2014, proving the effectiveness of DDL and calling for longitudinal studies as well as delayed posttesting (Boulton & Cobb, 2017). A leading linguist, Mark Warschauer, too, has been researching DDL (Txabarriaga, 2022) also with his colleagues Hansol Lee and Jang Ho Lee. For example, they conducted a meta-analysis on DDL in L2 vocabulary learning and factors influencing its outcome calling for linking DDL with SLA theories (Lee et al., 2019). These researchers with their first meta-analyses also signalled the advance of stage four of the discipline.

Researchers with a Focus on Pedagogy

A researcher with a strong focus on pedagogy is Lynne Flowerdew. She not only wrote an extensive introduction to corpus linguistics and DDL discussing pedagogical questions and making suggestions for how to bring DDL into classrooms but is otherwise known for her work on DDL for academic writing and for calling out the lack of theoretical underpinnings (Flowerdew, 2012, 2015).

Furthermore, Pascual Pérez-Paredes has also been a key player in DDL. He investigated how English as a Foreign Language (EFL) students pursuing a degree in English at a Spanish university approach corpus searches in focus-on-form tasks and what the shortfalls are (e.g. Pérez-Paredes et al. (2012)). More recently, he has conducted a meta-study on corpus use and DDL in CALL for the period between 2011 and 2015, called for more theoretical underpinnings and pleaded for Broad Data-Driven Learning (BDDL) "particularly in informal contexts outside the university classroom" (Pérez-Paredes, 2024, p. 211; cf. also Pérez-Paredes, 2022).

Another researcher in this category is Maggie Charles, who focuses on English for Academic Purposes (EAP) and is particularly known for promoting the implementation of DDL in the classroom, evaluating learners' attitudes towards DDL and conducting a longitudinal analysis of the effectiveness of subject-specific DIY corpora with international postgraduate students in the UK (Charles, 2014, 2015).

Also, Anne O'Keeffe is not only an eminent corpus linguist generally (O'Keeffe et al., 2007) but has also made significant contributions to DDL research directly. She prominently called for a "broader research gaze" and more theory in DDL including the usage-based (UB) model, not limiting oneself to constructivism and sociocultural theory and joining forces with SLA more generally (O'Keeffe, 2021, p. 259), reflecting features of stage four of the discipline. At this point, too, Angela Chambers continued her prominent contributions (see also Section "Focus on Learners"), calling for attention to the degree of explicitness in teaching and the cognitive processes involved, and formulated the desire for ordinary teachers to research into DDL, coining the expression "bridging the research-practice gap" (Chambers, 2019, p. 460), which is frequently cited. Chambers's research thus also displays features of stage four.

Stage IV: Most Recent High-Impact Researchers

John Flowerdew is a researcher with a main focus on DDL in academic writing. With Meilin Chen and Laurence Anthony, he introduced PhD students and in-service teachers of English in Hong Kong to DDL through workshops stressing the relevance of teacher training for popularising DDL and reporting successes in this training (e.g. Chen and Flowerdew (2018), Chen et al. (2019)). Laurence Anthony's pertinent contributions to the field of DDL have thus been ongoing (see also Section "Designers of Corpora and Corpus Tools").

Another particularly frequently mentioned academic with versatile contributions to DDL is Peter

Crosthwaite. He, for example, explored the benefits of corpus use for thesis writing (Crosthwaite et al., 2019), investigated what degree of explicitness is necessary for effective error feedback in DDL (Crosthwaite et al., 2020) and evaluated the efficacy of a preservice secondary teacher-training programme (Crosthwaite et al., 2023). He also organised a series of international DDL webinars online during the COVID-19 epidemic to broaden the definition and application of DDL, developed the tool CorpusMate for younger learners (Crosthwaite, 2024) and called for judicious integration of generative artificial intelligence (AI) applications such as ChatGPT into corpus tools (Crosthwaite & Baisa, 2023).

Overall, Alex Boulton is the most longstanding high-impact researcher in DDL (see also Sections "Focus on Learners" and "Investigating the Effectiveness of DDL"). He has been increasingly involved in meta-studies on DDL, in which he charts the field and uncovers the directions taken. At stage IV, he conducted a meta-study of 489 empirical studies written in English between 1989 and 2019 with Nina Vyatkina. Boulton and Vyatkina (2021) proposed that the way forward is to offer open-access materials and training programmes for educators, readily available teaching materials, work with multimodal corpora and, echoing O'Keeffe (2021), demanded greater theoretical underpinnings focusing on L2 English DDL studies.

Most recently, Boulton and Vyatkina (2024) updated their research synthesis with an analysis of another 148 papers (limited to L2 English) published between 2018 and 2022. They called to widen the traditional view of DDL beyond corpora and concordancers by incorporating web-as-corpus applications and search engines as well as other tools students are familiar with, to pay more attention to instruction below university level, to conduct mixed-method studies and to broaden the research approaches of the field. Additionally, both Alex Boulton and Nina Vyatkina have shaped the research of DDL in LOTEs.

DDL in LOTEs Across the Timeline

Overall, the influence of the eminent DDL researchers has meant that LOTEs have recently begun to be studied more systematically, which led to further studies and textbook production, as another sign of stage four in Shneider's (2009) sense. The topic of LOTEs was addressed for French, German, Italian and Spanish by a group of researchers in a paper presented at the EuroCALL CorpusCALL symposium in 2020 (Jablonkai et al., 2020).

Concerning French, Alex Boulton had a direct impact on DDL for LOTEs, demystifying and popularising the French term "apprentissage sur corpus" (ASC) for data-driven learning with Henry Tyne in a French language publication (Boulton & Tyne, 2014). Before, Angela Chambers had already conducted research into teaching French as a foreign language in Irish university contexts (e.g. Chambers (2010)). Following Kennedy and Miceli's (2001, 2010) work (see Section "Focus on Learners"), Luciana Forti published the first full-length monograph on DDL for Italian that encompassed both research and pedagogical applications (Forti, 2023). With regard to Spanish, the term "aprendizaje basado en datos" was adopted after the publication of an influential article by Yuly Asención-Delaney, Joseph G. Collentine, Karina Collentine, Jersus Colmenares and Luke Plonsky (Asención-Delaney et al., 2015). As for German, Nina Vyatkina is the leading researcher into DDL in North America and has recently published a monograph as well (Vyatkina, 2024). She had previously conducted research into DDL teaching university-level German recommending scaffolded, open access teaching resources to broaden the application of DDL, particularly with regard to LOTEs, primary and secondary school contexts and beginners (Vyatkina, 2016, 2020). She first appeared in stage three and her most recent work displays signs of a beginning stage four. Further, Eva Schaeffer-Lacroix is the main researcher of DDL for German in francophone countries (Schaeffer-Lacroix, 2019).

In-depth investigation of DDL for LOTEs is also underway in the languages in question such as German. While the German term "datengesteuertes Lernen" for DDL was first introduced by Anke Lüdeling and Maik Walter (2009, p. 6), DDL is not widespread in the German-speaking world. However, in 2021, Carolina Flinz with Britta Hufeisen, also an eminent researcher into multilingualism, launched the primarily German-language journal KorDaF (https://kordaf.tujournals.ulb.tu-darmstadt.de/) that publishes leading research in the field.

Overall, however, LOTEs are still a niche area, and more researchers investigating additional languages are still needed, especially regarding other widespread languages such as Arabic, Russian and Chinese (cf. also Jablonkai et al., 2020, 136).

Summary

On the basis of the bibliometric study by Dong et al. (2023) and a list of 93 notable language researchers based on the 2022 Elsevier cocitation index (Txabarriaga, 2022), the most notable, influential researchers in DDL along with main research contributions were identified. They were grouped according to the four stages of development identified by Shneider (2009): Conceptualising DDL (1980s-1998) as the first stage with Tim Johns as the main protagonist, a developing field (1998–2011) as stage two and a growing field (2011–now) as stage three with the final stage four only just emerging. It was also shown how the most recent transformative researchers have been shaping the field lately, especially with regard to the theoretical underpinnings of DDL, the advent of AI tools such as ChatGPT and the hitherto little-developed area of LOTEs.

Website Links

AntConc: https://www.laurenceanthony.net/software/antconc/

BNC Baby: http://www.natcorp.ox.ac.uk/corpus/babyinfo.html

ColloCaid: CollCaid: https://collocaid.uk/

The Compleat Lexical Tutor: https://www.lextutor.ca/

Corpus del Español: https://www.corpusdelespanol.org/

Corpus do Português: https://www.corpusdoportugues.org/

English Corpora: https://www.english-corpora.org/

Tim John's Kibitzers: https://lexically.net/TimJohns/#gsc.tab=0.

MicroConcord: https://lexically.net/software/

Sketch Engine: https://www.sketchengine.eu//

Sketch Engine for Language Learning (SkELL): https://skell.sketchengine.eu/#home?lang=en

Wordsmith: https://lexically.net/wordsmith/

Cross-References

▶ Data-Driven Learning and Languages other than English
▶ Notable Publications in Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anthony, L. (2004). *AntConc (Version 3.0.1)*. Waseda University. http://www.antlab.sci.waseda.ac.jp/

Anthony, L. (2018). *Introducing English for Specific Purposes*. Routledge.

Asención-Delaney, Y., Collentine, J. G., Collentine, K., Colmenares, J., & Plonsky, L. (2015). El potencial de la enseñanza del vocabulario basada en corpus: optimismo con precaución. *Journal of Spanish Language Teaching, 2*(2), 140–151. https://doi.org/10.1080/23247797.2015.1105516

Boulton, A. (2009a). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37–54. https://doi.org/10.1017/S0958344009000068

Boulton, A. (2009b). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics, 35*(1), 81–106.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60*(3),

534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Tyne, H. (2014). Des documents authentiques aux corpus. *Démarches pour l'apprentissage des langues*. s. Didier.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. https://doi.org/10125/73450

Boulton, A., & Vyatkina, N. (2024). Expanding methodological approaches in DDL research. *TESOL Quarterly, 58*(3), 1193–1204. https://doi.org/10.1002/tesq.3269

Burnard, L. (Ed.). (2004). *BNC Baby* [CD-ROM]. Oxford University Research and Technology Service. http://www.natcorp.ox.ac.uk/corpus/babyinfo.html

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology, 9*(2), 111–125. https://doi.org/10125/44022

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3–16). Rodopi. https://doi.org/10.1163/9789401203906_018

Chambers, A. (2010). L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé. *Revue française de linguistique appliquée, 15*(2), 9–20. https://doi.org/10.3917/rfla.152.0009

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching, 52*(4), 460–475. https://doi.org/10.1017/S0261444819000089

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30–40. https://doi.org/10.1016/j.esp.2013.11.004

Charles, M. (2015). Same task, different corpus: The role of personal corpora in EAP classes. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 131–154). https://doi.org/10.1075/scl.69.07cha

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics, 23*(3), 335–369. https://doi.org/10.1075/ijcl.16130.che

Chen, M., Floweerdew, J., & Anthony, L. (2019). Introducing in-service English language teachers to data-driven learning for academic writing. *System, 87*, 102148. https://doi.org/10.1016/j.system.2019.102148

Crosthwaite, P. (2024). *Bridging the research-practice divide for data-driven learning*. Routledge. https://doi.org/10.4324/9781003413301

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics, 3*(3), 100066. https://doi.org/10.1016/j.acorp.2023.100066

Crosthwaite, P., Wong, L. L. C., & Cheung, J. (2019). Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. *ReCALL, 31*(3), 255–275. https://doi.org/10.1017/S0958344019000077

Crosthwaite, P., Storch, N., & Schweinberger, M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing, 49*, 100729. https://doi.org/10.1016/j.jslw.2020.100729

Crosthwaite, P., Luciana, L., & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392–1420. https://doi.org/10.1080/09588221.2021.1995001

Davies, M. (2008) *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Dong, J., Zhao, Y., & Buckingham, L. (2023). Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL, 35*(3), 339–355. https://doi.org/10.1017/S0958344022000222

Firth, J. R. (1968). A synopsis of linguistic theory, 1930-55. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth* (pp. 168–205) (Original work published 1957).

Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan. https://doi.org/10.1057/9780230355569

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.69.

Forti, L. (2023). *Corpus use in Italian language pedagogy: Exploring the effects of data-driven learning*. Routledge. https://doi.org/10.4324/9781003137320

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL, 26*(2), 128–146. https://doi.org/10.1017/S0958344014000

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 430–442). Routledge. https://doi.org/10.4324/9780203856949

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago University Press.

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of F. R. Firth* (pp. 148–162). Longman.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins. https://doi.org/10.1075/scl.

Jablonkai, R., Forti, L., Abad Castelló, M., Salengros Iguenane, I., Schaeffer-Lacroix, E., & Vyatkina, N. (2020). Data-driven learning for languages other than English: the cases of French, German, Italian, and Spanish. In K.-M. Frederiksen, S. Larsen, L. Bradley, & S. Thouësny (Eds.), *CALL for widening participation: short papers from EUROCALL 2020 (pp. 132–137)*. Research-Publishing.Net. https://doi.org/10.14705/rpnet.2020.48.1177

Johns, T. (1986). Micro-concord: A language learner's research tool. *System, 14*(2), 151–162. https://doi.org/10.1016/0346-251X(86)90004-7

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria, 10*, 14–34.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), Classroom Concordancing. *English Language Research Journal, 4*, 1–16.

Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100–115). Longman.

Kennedy, C., & Miceli, T. M. (2001). An evaluation of intermediate students' approaches to corpus Investigation. *Language Learning & Technology, 5*(3), 77–90. https://doi.org/10125/44567.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology, 14*(1), 28–44. https://doi.org/10125/44201.

Kilgarriff, A., Marcowitz, F., Smith, S., & Thomas, J. (2015). Corpora and language learning with the Sketch Engine and SKELL. *Revue française de linguistique appliquée, 20*(1), 61–80. https://doi.org/10.3917/rfla.201.0061

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Lüdeling, A., & Walter, M. (2009). Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung, 1–37. https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-innen/anke/pdf/LuedelingWalterDaF.pdf

O'Keeffe, A. (2021). Data-driven learning—A call for a broader research gaze. *Language Teaching, 54*(2), 259–272. https://doi.org/10.1017/S0261444820000245

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*(3), 269–286. https://doi.org/10.1017/S095834400700033X

O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge University Press.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pérez-Paredes, P. (2024). Data-driven learning in informal contexts? Embracing Broad Data-driven learning (BDDL) research. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 211–221). Routledge. https://doi.org/10.4324/9781003413301

Pérez-Paredes, P., Sánchez-Tornel, M., & Calero, J. M. A. (2012). Learners' search patterns during corpus-based focus-on-form activities: A study on hands-on concordancing. *International Journal of Corpus Linguistics, 17*(4), 483–516. https://doi.org/10.1075/ijcl.17.4.02par

Schaeffer-Lacroix, E. (2019). *Corpus et didactique de l'allemand*. Lambert-Lucas.

Scott, M. (2024). *WordSmith Tools version 9 (64 bit version) [software]*. Lexical Analysis Software. https://lexically.net/wordsmith/

Shneider, A. M. (2009). Four stages of a scientific discipline; four types of scientist. *Trends in Biochemical Sciences, 34*(5), 217–223. https://doi.org/10.1016/j.tibs.2009.02.002

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford University Press.

Txabarriaga, R. (2022, October 21). Over 300 Language Experts Made the 'Most Cited Scientists List'. https://slator.com/over-300-language-experts-made-the-most-cited-scientists-list/

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL, 28*(2), 207–226. https://doi.org/10.1017/S0958344015000269

Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals, 53*(2), 359–370. https://doi.org/10.1111/flan.12464

Vyatkina, N. (2024). *Corpus applications in language teaching and research: The case of data-driven learning of German*. Routledge. https://doi.org/10.4324/9781003223627

C

Corpus Representativeness in Data-Driven Learning

Tove Larsson[1] and Tülay Dixon[2]
[1]Northern Arizona University, Flagstaff, AZ, USA
[2]Emory University, Oxford College, Atlanta, GA, USA

Abstract

Data-driven learning (DDL) is an approach to language learning and teaching that relies on the identification of linguistic patterns in authentic language use. Therefore, corpora, which are collections of spoken or written texts, are key to DDL practices. For corpora to yield useful findings, however, they need to be representative of the target language domain. This chapter introduces the concept of *corpus representativeness* and discusses its current role in DDL research and practice.

Keywords

Data-driven learning · Corpora · Representativeness

Introduction

Just like researchers in other fields collect samples in an attempt to represent a population of interest, corpus linguists collect texts, or *corpora* (see Egbert et al., 2022). Any corpus that is used either for research or for teaching purposes would therefore need to be treated as a sample and evaluated accordingly. Data-driven learning (DDL) is an approach to language learning and teaching that enables students to discover patterns of language in authentic data, that is, in corpora (e.g., Flowerdew, 2015). The number of corpora that are available for students to use, and thus the number of choices, is ever-increasing. This means that teachers have to weed through them to find one that suits their purposes and students. This chapter introduces the concept of corpus representativeness and discusses some key considerations for teachers and researchers to ponder in relation to choosing corpora for teaching or research. It also provides a brief overview of how the concept has been approached in the DDL literature.

Corpus Representativeness

Since it is typically impossible to obtain access to the full population of interest, corpus compilers have to try to obtain a representative *sample*, that is a smaller version of the *population*. A population is a technical term used to refer to the entire group (e.g., of people, texts, plants, or animals) that researchers want to draw conclusions about (Egbert et al., 2022). For example, if someone is interested in learning about how all second-language (L2) English speakers use

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_154-1

language, they will need to collect a representative sample (i.e., a corpus) of language produced by such speakers. 'Corpus representativeness' denotes the degree to which a corpus accurately represents the population of interest. To be able to assess this, it is important to learn as much as possible about the domain (Egbert et al., 2022).

Corpus representativeness can be assessed in terms of *situational* and *linguistic representativeness* (Biber, 1993; Egbert et al., 2022). Both of these need to be considered to ensure accurate generalizations about language use and variation in a given language domain. Situational representativeness can be achieved when a corpus represents the full range of text types that exist within a target population. For example, if a study targets university student writing, then text types such as argumentative essays, emails, and lab reports, among many others should be included. A corpus that only encompasses one of these (e.g., argumentative essays) is not situationally representative of university student writing, given the high number of different text types produced in a university setting that are not included. However, it is not sufficient to only represent the range of *text types*; the range of *linguistic* variability within these would also need to be considered. Linguistic representativeness can be achieved when a corpus is large enough to capture the full distributional variability of linguistic features within and across texts. For example, if a linguistic feature such as attributive adjectives does not occur in a corpus, it may be an artefact of the fact that the corpus is not large enough (or well-designed enough) to reflect the true frequency and variability of that feature that one would expect to see in the population.

Although the aforementioned considerations should guide the compilation of a corpus, corpus compilers are bound by practical constraints. For example, it is often much easier and less resource-intensive to compile written data than spoken data, meaning that available corpora tend to capture written discourse (Paquot & Plonsky, 2017). Given these practical constraints, corpora will represent the population at hand to a greater or lesser extent and should thus always be evaluated for representativeness to ensure that they capture what the DDL teacher or researcher hopes to capture (e.g., the intended mode: writing vs. speech; the intended register: academic prose vs. newspaper writing). At times, however, the dataset that DDL teachers or researchers decide to use might be difficult (if not impossible) to evaluate in this manner. This is often the case for large-language models (LLMs) such as BERT (Devlin et al., 2018), GPT-4 (OpenAI, 2023), and LLaMa (Touvron et al., 2023). Such models often use the web as training data. Thus, if learners use the standard LLM interfaces, they do not have the option of taking mode or register into consideration. A possible solution to this issue may be to instead use LLMs to help analyse a corpus with known domain and register boundaries. This solution is exemplified in Anthony's (2024) newly released beta version of AntConc (4.3.0), which is a corpus tool that can be used to identify patterns of language use. With the help of LLMs GPT 3.5 and 4, this version of AntConc allows users to analyse a user-uploaded corpus with prompts written in natural language.

Discussions of Corpus Representativeness in the DDL Literature

In the DDL literature, it is not uncommon for corpus representativeness to be mentioned, and it is clear that the term carries weight in the field (Casal & Kessler, 2024). Readers are often urged to evaluate any corpus they are considering and choose accordingly. For example, Fuster Márquez and Clavel Arroitia (2010) critically assess the extent to which the design of the British National Corpus actually captures all of British English, given, for example, that it only contains 10% spoken data. In a similar vein, Gilquin (2022) and Friginal and Cox (2022) recommend users of corpora in the classroom to evaluate the extent to which a corpus is representative of the variety and student population of interest.

However, in practice, it is still uncommon for DDL empirical studies to report on evaluations of corpus representativeness (though see Vandeweerd & Keijzer, 2018). Instead, it is

sometimes mentioned in relation to a methodological choice made or as a limitation of a study. An example of the former can be seen in Özer and Özbay (2022): "This corpus was called the Alternate Corpus of Academic Texts (the ACAT), later combined with a corpus of graded readers previously compiled by the researchers to serve a similar purpose and to increase the **representativeness** [emphasis added] of the database" (p. 40). An example of the latter can be found in Yao (2019) who writes that "[t]he scope of this study is limited regarding the sample size and the **representativeness** [emphasis added]" (p. 41).

In addition, representativeness tends to be treated as a dichotomous concept: either a corpus is representative or it is not: "authentic language in the context of language learning differs from the range of attested uses recorded in **representative** [emphasis added] L1 corpora" (Pérez-Paredes, 2019, p. 71). In this usage, the concept is often mentioned in relation to a definition of a corpus. For example, Lusta et al. (2023) cite McEnery and Wilson (2001) and Boulton (2015) as follows:

> McEnery and Wilson [[[13,164]], p. 32] define a corpus as 'a finite-size body of machine readable text, sampled in order to be maximally **representative** of the language variety under consideration'. This denotes that a corpus is characterised by sampling, **representativeness**, finite size, machine readability, and a standard reference. Similarly, Boulton [[14], p.3] describes a corpus as 'a large collection of real-life texts in electronic format, chosen to be **representative** of a language variety'. (p. 2, all emphases added)

While McEnery and Wilson (2001) acknowledge that representativeness is gradable, Lusta et al.'s (2023) comment, together with Boulton's (2015) definition, not only suggests that it is a dichotomous construct, but also that it is one part of the definition of a corpus. Since, by definition, no sample (and, thus, no corpus) can ever fully represent the population, there are no corpora out there that achieve full representativeness (see Leech, 2007). Instead, researchers and teachers who wish to use corpora in their classrooms need to evaluate the corpus that they are interested in and make their choice accordingly (Egbert et al., 2020).

Conclusion

With the ever-increasing number of available corpora, it is imperative that future DDL initiatives focus on training language teachers, learners, and researchers in assessing corpus representativeness. In line with this goal, this chapter introduced the concept, provided considerations for evaluating it, and provided a brief overview of how it has been approached in the DDL literature.

Cross-References

- ▶ Corpora as a Reference Resource in Data-driven Learning
- ▶ Corpora in Language Assessment
- ▶ Corpus-informed Materials in Language Teaching
- ▶ Corpus Literacy and Data-driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anthony, L. (2024). *AntConc* (Version 4.3.0) [Computer Software]. https://www.laurenceanthony.net/software

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Boulton, A. (2015). Applying data-driven learning to the web. In A. Lénko-Szymánska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 267–295). John Benjamins.

Casal, J. E., & Kessler, M. (2024). New theoretical and practical horizons in data-driven learning: Introduction to the special issue. *TESOL Quarterly, 58*(3), 1027–1045.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding., arXiv*, 1–16. https://doi.org/10.48550/arXiv.1810.04805

Egbert, T., Larsson, L., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge University Press.

Flowerdew, L. (2015). Data-driven learning and language learning theories. In *Multiple affordances of language corpora for data-driven learning* (Vol. 69, pp. 15–36). John Benjamins Publishing Company.

Friginal, E., & Cox, A. (2022). Corpus uses in language teaching. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp. 161–172). Routledge.

Fuster Márquez, M., & Clavel Arroitia, B. (2010). Corpus linguistics and its applications in higher education. *Revista Alicantina de Estudios Ingleses, 23*, 51–67.

Gilquin, G. (2022). Written learner corpora to inform teaching. In *The Routledge handbook of corpora and English language teaching and learning* (pp. 281–295). Routledge.

Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 133–150). Rodopi.

Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon, 9*, 1–9.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics*. Edinburgh University Press.

OpenAI. (2023). *GPT-4 technical report, arXiv*, 1–100. https://doi.org/10.48550/arXiv.2303.08774

Özer, M., & Özbay, A. (2022). Exploring the data-driven approach to grammar instruction in the ELT context of Turkey. *Australian Journal of Applied Linguistics, 5*(2), 35–63.

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research, 3*(1), 61–94.

Pérez-Paredes, P. (2019). The pedagogic advantage of teenage corpora for secondary school learners. In *Data-driven learning for the next generation* (pp. 67–87). Routledge.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and efficient foundation language models., arXiv*, 1–27. https://doi.org/10.48550/arXiv.2302.13971

Vandeweerd, N., & Keijzer, M. (2018). J'ai l'impression que: Lexical bundles in the dialogues of beginner French textbooks. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée, 21*(2), 80–101.

Yao, G. (2019). Vocabulary learning through data-driven learning in the context of Spanish as a foreign language. *Research in Corpus Linguistics, 7*, 18–46.

W

Wordlists and Data-Driven Learning

Tom Cobb
Didactique des langues, Université du Québec à Montréal, Montreal, QC, Canada

Abstract

While concordances and word lists are the two most usable forms of corpus data, word lists are largely absent from both practitioner awareness and ongoing research in data-driven learning (DDL). This is mainly because word lists work behind the scenes in the design of DDL instruction rather than up front as concordancers do in the hands of learners. Both are instances of DDL in the sense of language patterning exposed with computer software. Examples of these points are elaborated and a proposal is made to integrate word list research within DDL.

Keywords

Data-driven learning · Language learning · Corpus · Concordance · Frequency list · Profiler

Introduction

If data-driven learning (DDL) is broadly "the use of a corpus in language learning" (Boulton & Cobb, 2017), then, since corpora of any size are unapproachable as totalities, DDL refers to the use in language learning of the two most common break-outs of a corpus, word lists and concordances. Word lists are normally frequency lists, count-ups of every word-type in a corpus (Table 1 shows the first 30 types of the Brown Corpus, 1969, sorted by frequency), or else concordances, a virtual list of every word token in a corpus, with a few words of context to either side (Fig. 1 shows the first 10 of 26,327 instances of "to" in the same corpus).

Note that a frequency list is a complete, existing artefact, composed of every word type in a corpus along with the number of its occurrences, while a concordance is typically a virtual list of every token, some part of which is created dynamically on user request. Despite this difference, it is useful to see concordances and lists as aspects of the same phenomenon, i.e. concordances as expanded word lists with a few context words added to each side. The original purpose of the context words was to make part of speech (POS) parsing of the keyword possible, thereby creating a more accurate list (e.g. with *"run"* noun and *"run"* verb counted separately); a more recent purpose is to improve the comprehensibility of corpus data for language teachers and learners. Seeing lists and concordances as similar corresponds to how concordances are programmed in software, i.e. the same as frequency lists but with extra characters.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_155-1

Wordlists and Data-Driven Learning, Table 1 Start of the Brown Corpus frequency list

1. the 69,992	7. that 10,783	13. with 7290	19. at 5382	25. but 4381
2. of 36,472	8. is 10,102	14. as 7252	20. by 5348	26. from 4370
3. and 28,931	9. was 9815	15. his 6999	21. this 5146	27. or 4227
4. to 26,237	10. he 9766	16. on 6766	22. had 5130	28. have 3942
5. a 23,541	11. for 9499	17. be 6387	23. not 4618	29. an 3748
6. in 21,419	12. it 9071	18. i 5620	24. are 4394	30. they 3648

```
001. ☐ delegation in Washington who would like TO see it (the resolution) passed". But
002. ☐ ia's congressmen specifically asked him TO offer the resolution. The resolution,
003. ☐ e Board of Education should be directed TO "give priority" to teacher pay raises
004. ☐ n should be directed to "give priority" TO teacher pay raises. _COLQUITT_ - Afte
005. ☐ enerally accepted practices which inure TO the best interest of both governments
006. ☐ s "these two offices should be combined TO achieve greater efficiency and reduce
007. ☐ es". It urged that the city "take steps TO remedy" this problem. Implementation
008. ☐ State Welfare Department "has seen fit TO distribute these funds through the we
009. ☐ lable funds", the jurors said. "Failure TO do this will continue to place a disp
010. ☐ said. "Failure to do this will continue TO place a disproportionate burden" on F
```

Wordlists and Data-Driven Learning, Fig. 1 Start of the concordance for "to" in Brown Corpus. (From https://www.lextutor.ca/conc/eng)

The Low Profile of Lists

But if there is a connection between concordances and word lists, it is not obvious to many trainee English as a Second Language (ESL) teachers I have known. In successive rounds of (unpublished) classroom research, I have asked classes, after the DDL part of their course in Computer-Assisted Language Learning (CALL), to brainstorm in twos, then fours, then as a class, the top 10 keywords of DDL. These were typically "corpus/corpora", "specialist/generalist", "discovery/inductive/problem-based learning", "authentic/simplified", "hands-on/printout approaches", and "collocate/collocation". Word lists never came up, though their role and uses had been mentioned in sessions. I occasionally distributed my own set of DDL keywords which included "word lists", and asked groups to rank-order them, and lists were never far from the bottom.

The picture is not so different in the work of DDL researchers. Word lists are hardly unknown to researchers, since the main DDL toolkits available (Anthony's AntConc, (https://laurenceanthony.net/software/antconc) Cobb's LexTutor, (https://lextutor.ca) Davies's COCA, (https://English-corpora.org) Kilgariff's Sketch Engine, (https://sketchengine.eu) Smith's WordSmith Tools (https://lexically.net/wordsmith)) all include both corpus-based lists and tools for building lists from texts and corpora. But do these lists get research attention? They should, inasmuch as word lists function behind the scenes or in the "back-ends" of many types of DDL software and hardware and come bearing variables that involve judgments that may affect learning. One is whether word lists require parsing and/or grouping into some type of larger unit, and if so which type, to be most useful to learners in, say, corpus-based dictionaries or some of the software to be discussed below. Nonetheless, in a collection of 790 research abstracts produced between 1988 and 2021, assembled by Boulton and Vyatkina (2021) and updated thereafter, there are just 21 occurrences of "word list" (including "list", "vocabulary list" and "frequency list") compared to 354 of "concordance" (with "concordances" and "concordancing"). It is as if the concordancer were the only tool in the DDL tool shed.

The equation of DDL and concordance on the part of teachers and researchers is probably due to the fact that concordances (being readable pieces of sentences linked to their original texts) are immediately comprehensible to learners in ways that lists are not, and to the fact that concordance

output contains information that a list does not and cannot, notably about grammar and collocation. But lists contain important information that concordances do not, and can do work in DDL that concordances can not.

Two Corpus Outputs for Two Jobs

To explore the different roles of lists and concordances in DDL, it may help to expand the definition of DDL slightly. To "the use of a corpus in language learning. . .", we should add an answer to the implicit question the definition poses, "to do what?" DDL is the proposition that language, when amassed and transformed into text data, can be squeezed, sliced, and diced by computer software to (1) reveal its underlying patterns—to linguists, certainly, but also with some injection of pedagogy, to language learners; and to (2) raise the comprehensibility and learnability of particular instances of language through features like parsing, exhaustiveness, simplification, the increased motivation of knowing a feature is frequent, the comprehension of a new word by seeing several examples of it grouped together, and others. With some overlap, concordances work up front to make language patterning available for deliberate learning, while lists work behind the scenes to make language comprehensible for implicit learning. These different loci of operation may explain the neglect of lists in the DDL conversation.

Patterns Revealed in Concordances

There are cases where lists and concordances can both reveal the same pattern but concordances can do it far better. Wordlists can certainly display the frequency, the number of times a word or phrase appears in a text, corpus, or lifetime, which is largely opaque to language users and even teachers (McCrostie, 2007). But such a list reveals the frequency only of single words and adjacent-word phrases. Disjunctive collocations (with keyword and collocate separated by intervening words), which are typically the last thing learned by unassisted L2 learners (Boers et al., 2014), can be revealed only by a lifetime of experience in a language, or by a concordancer. Even linguists discovered the extent of collocation in language only with the help of concordancers, and this is a clear case where the "learner as linguist" (Seliger, 1983) can follow in their footsteps (Cobb, 2018).

Similarly, grammatical patterning can be found only in concordances, not in lists. All the individual words of "He plays badly the piano*" may feature in a particular frequency list, but they will never appear in this sequence in any corpus. Indeed, the ultimate revelation of concordances that can never be provided by word lists is what it does *not* contain. Concordances are an answer to the generative linguists' famous "negative evidence" (e.g. Marcus, 1993), the lack of which in natural input argues for an innatist acquisition theory. By this, they mean that no one points to a sentence like the example just above and tells you, "This is ungrammatical". But a skilled DDL coder can set up a concordance that makes the absence of such a structure clear to learners. Gaskell and Cobb (2004) show ways of responding to learners' writing errors with negative-evidence concordances.

To summarize, the sheer amount and variety of language patterning that can be revealed by concordances, shortly to be augmented by LLM (large language model) artificial intelligence, will always make the concordancer the primary learning tool of DDL. One might ask what place is left for lists.

Up-Front Word Lists

Concordances pre-date word lists in the history of corpus research. The Bible was probably the first corpus, or the first multi-part text treated as a corpus, with just under 800,000 words in 66 sub-corpora. The concordancing of its main concepts (not every word) was performed by hand and eye by monks, and no attempt is recorded to render the corpus as a frequency list (which would have been an unimaginable labour). But in modern digital corpus research, the word list predates the concordancer. The first digital corpus, large for the time at just over one million words, was the Brown Corpus (Francis & Kučera, 1967),

of 500 texts of 2000 words each in three broad subdivisions, and like the Bible large enough to be uninterpretable in itself without some sort of breaking into pieces. The Brown's first usable break-out was an unparsed frequency list of its 41,745 word types (270 printed pages at four columns/page) which, though rudimentary, produced at least two notable contributions: (1) the first large-scale confirmation of Zipf's law, that rank and frequency are inverse quantities in natural distributions like lexicons; and (2) the selection of items for the innovative American Heritage Dictionary (1969), probably to date the most widely used corpus resource that has ever existed. It was a huge advance in lexicography to ground a dictionary in the words of current and recent usage rather than mixing historical and current items randomly or with voluminous annotation.

Concordances of the Brown came only a few years later, in the form of a very large index of every word in the corpus, written in code to conserve space, and used mainly as a means to hand-parsing the words of the list into 80 parts of speech (POS). Without the codes, one million concordance lines at 35 lines/page are almost 29 thousand pages. The concordances were basically a broadening out of the tokenized frequency list with a few words to either side, not enough to be read continuously but enough to identify recurring immediate collocations and inform a guess at a POS. Parsing, even though problematic in the beginning and done largely by hand, eventually allowed the refinement of the frequency analysis to include separate counts for "run" (verb), "run" (noun), and the many more unmarked POS distinctions of English. Thus, the concordance lines were a means to refining the list; it was probably never imagined that the concordance would largely replace the list.

Frequency lists retained a degree of prominence even in the early DDL era, roughly 1970–2000, probably because frequency software was readily available but concordance software less so, especially before it could be run over the Internet. One application, adapted from Sinclair (1991), was to use a frequency list of words in a text as a measure of the text's lexical density and thus its suitability for use with language learners of various proficiency levels. Sinclair noticed that the most frequent words in a frequency list were invariably function words (articles, prepositions, pronouns, helping verbs, conjunctions) with content words appearing only well down the list, and, further, that the ranks of the first content words varied by text topic and complexity. The first content word in the general-English Brown corpus list is "said" at rank 55; in a corpus of graded stories it is "Mr" at rank 57. But in a combined corpus of physics, engineering, and math (from the BAWE; Nesi et al., 2009), "voltage" is at rank 29 with "current" and "circuit" at ranks 31 and 32. The reason for the earlier appearance of content words is that complex texts employ a greater number of them—e.g. in series lists, compound phrases, etc.

This first-content-word method of text selection was widely used in the 1980s era of English for Specific Purposes (ESP) and English for Academic Purposes (EAP), which sought to give learners authentic but more or less comprehensible reading texts. A weakness of the method was its sensitivity to text length; a short text about Donald Trump may well make "Trump" its most frequent word. Skilled practitioners were nonetheless able to use the procedure effectively.

Lists at this time were rarely given directly to learners, as they had been in the grammar-translation and audio-visual periods. Such use of lists had come into disrepute, owing to questions about where they came from (one list circulating where I worked had "brook" and "ye" as frequent words, suggesting a Biblical origin) and to abundant research showing that simply memorizing word lists did not improve comprehension, even for the texts containing the same words (e.g. Nagy, 1988). There is only one instance of giving lists to learners in DDL work that I am aware of, and that was in one of my own studies (1999a), where learners were given explicit access to appropriate corpus-based vocabulary lists, in a computer programme hyperlinked to a corpus of their own course materials. The idea was to get the best out of both lists and concordances, combining "breadth and depth" of learning, in the words of a title that came out of this work (1999b). The list provided the breadth, a guarantee that every word of a certain

frequency would be met and considered, while the concordance provided the depth, the numerous rich contextualizations for each word that would reveal its meaning and pattern of use. It is not clear that this strategy should be called "list-based learning", since the words on the lists were effectively suggested inputs to the concordance programme, and there was no suggestion of trying to learn list words out of context.

Backend Word Lists

Since then, lists have largely disappeared behind the scenes of other DDL software, a disappearance so total that users are often unaware it is a word list that is performing the magic on their screens. A good example is the lexical frequency profiler (LFP; with just four occurrences out of 790 in the research abstracts cited above). A profiler reads a text one word at a time, comparing every word against a set of frequency lists and recording the result, and at the end yielding a profile or summary of the frequency level of the whole text. Profilers include Anthony's AntProfiler, (https://laurenceanthony.net/software/antwordprofiler) Browne's NGLSPProfiler, (https://ngslprofiler.com) Cobb's VocabProfile, (https://lextutor.ca/vp) Finlayson et al.'s MultilingProfiler, (https://www.multilingprofiler.net) and Nation et al.'s Range (https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs)—all of which are widely used in language instruction, forensics, and other branches of the language industry.

The lists running inside a profiler are normally organized by frequency but can also be organized by subject area, etymological origin, or in combinations of primary and secondary lists (say, frequency then etymology, so that for a given frequency level it is clear which particular words are cognate). A typical way of organizing a profile is by the number of 1000 lemma or family sets needed to achieve 95% and 98% coverage in the input text, these being key points in the development of reading comprehension (Laufer & Ravenhorst-Kalovski, 2010). If just 1000 word families achieve 98% coverage in a text, then the text is (other things being equal) easy to read for most learners. Profiles are widely used in text selection, text simplification, examination control, and identification of items for pre-teaching.

Such judgments can be made more accurately if teachers know the lexical profile of their learners as well as the profiles of their texts. If they know this, their learners have probably taken a vocabulary "size" or "levels" test, probably one produced by Nation and colleagues (e.g. Nation & Beglar, 2007) or Meara and colleagues (e.g. Meara & Buxton, 1987) or their successors (e.g. Sasao & Webb, 2018). All such tests are based on items drawn from a sequence of corpus-based frequency lists (or phrase lists; Martinez & Schmitt, 2012). For example, Nation and Beglar's Vocabulary Size Test samples 10 items from each of the first 14 thousand levels of Nation's BNC/COCA word families lists (Described and available at https://www.victoria.ac.nz/lals/about/staff/paul-nation) and provides a size estimate by multiplying the percentage score at each level times 1000 and totalling the levels. The procedure is simple but effective. Since in most cases, the same lists are behind both test and profiler, it is possible to tune texts to learners in a variety of useful configurations. For learners with receptive knowledge of the first 2000 word families, a 3000 level text provides an intensive reading activity, while 1000 provides a fluency activity, and so on.

Another DDL tool that is list based is keyword, which inputs a text and shows the words that are significantly more frequent than in some reference corpus, i.e. the words or phrases that truly characterize its "aboutness" (Scott, 2001; Bondi & Scott, 2010). If teachers know what these keywords are, they can draw learners' attention to them as an aid to comprehending the text. This is especially important to do if the keywords are also beyond the learners' current vocabulary level, as is usual, since texts' lower-frequency words tend to carry their main ideas (Kučera, 1982). For this reason, it is logical for keyword and profiler to work together. (e.g. at https://lextutor.ca/key) The keywords of a text are determined by breaking the text into a frequency list and then comparing it to the frequency list of a standard non-specialist corpus. Words that are 25 times (a commonly used but arbitrary figure) more frequent in the text than in

the corpus are probably its keywords. For example, in the Oxford Bookworms graded story *Dracula,* the word "coffin" occurs 29 times in 7613 words, or, scaled up to a corpus of ten million words, 38,000 times. By comparison, in a general corpus of 10 million words, "coffin" occurs 73 times, giving the word a keyness in *Dracula* of 520. And further, "coffin" is in the fourth 1000 of Nation's BNC/COCA word list, as are its fellow keywords "hammer", "garlic", "diary", and "carriage"; these are all mid-frequency rather than basic items, which is important in a text where the first 2000 families provide 98% coverage. Keyword analysis is thus important even in simplified texts, but it is crucial to the establishment of domain-specific word lists. And again, the count for "keyword" in the research abstracts is just 2 out of 790 (though the search is error-prone in that "keyword" also appears in the unrelated phrase "keyword in context", KWIC).

Lists are also at work within concordance routines themselves. For example, when the search input space offers the options "lemma" or "family", that means the programme can access a list of all the fleshed-out lemmas or families of the language which will be incorporated into the search term through a regular expression that matches not just "see" but any of "see|sees|seeing|seen|saw". Lists are also employed in tokenizing texts prior to concordance analysis, for example reducing contractions to separate words rather than piling up separate counts of "can" and "can't" as if they were unrelated items. On a much larger scale, Davies' 1 billion word Corpus of Contemporary American English (COCA) is "really" just an unreadable list of numbers until it is formatted into comprehensible language following data assembly. Davies explains that this is the only way truly massive corpora can be searched or the results assembled in a reasonable amount of time. (From a piece entitled 'Architecture' at https://www.english-corpora.org/help/architecture.pdf)

Back at the user level, one visible integration of lists and concordances is the ability of some concordancers to use resident frequency lists to work out the average frequency of every line in a concordance output and then sort it from high to low, or, normally, readable to challenging. Thus, the lines at the top of the output are composed mainly of common words and have a chance of being readable despite the relative difficulty of the overall corpus. This feature is widely used in ESP courses as a way to give learners input that is authentic yet comprehensible without being simplified. For instance, an electrical engineering student can meet the word "current" in a sentence like "The heater was connected to a heat monitor and power supply and was switched on while a note was made of its CURRENT and voltage" (first 1000 rating) rather than, "In 1888 Tesla demonstrated his brushless alternate-CURRENT induction motor to the Institute of Electrical Engineers (IEEE). . ." (sixth 1000). The main concordancers offering this service are Sketch Engine and Lextutor, the latter employing the 1000-families scheme of the BNC/COCA thereby integrating concordancer and profiler.

A final example involves a typical DDL tutorial activity, in this case a cloze passage builder in which a teacher puts in a text and gets out a cloze passage with every n^{th} word removed for the learner to replace. Cloze is a comprehension-oriented activity, focusing on comprehension of the text as a whole though the task is to replace single words or adjacent phrases. There is magic that can be added to this routine, and it involves recourse to word lists. For example, using a short list of prepositions, or articles, the programme can choose just those classes of words to remove for replacement. A more interesting example involves focusing the gaps on morphology instead of lexis, without requiring the teacher to parse or tag the text, changing all the inflectable items in the text (nouns, verbs, adjectives in the case of French, for example) to a pulldown menu of inflection options for the learner to choose from. For "The boy [run, runs, ran, running] home when he heard his mother calling": to find the inflectables within the chosen n-range, the programme first determines programmatically any proper nouns and disqualifies them, then through list comparison that the word is not present in the lists of function words and interjections but *is* present in the list of inflected nouns or verbs and builds the gap accordingly from its lemma lists. One routine that can do this is N-Word Close Builder. (https://lextutor.ca/cloze/n) This is just one example of list magic.

So, to summarize, word lists do a lot of the lifting in DDL.

What Would List Research Look Like?

It is not quite true that no one is doing research on word lists and their role in language learning. There is strong ongoing research on these topics, just categorized under vocabulary research within applied linguistics, and to some extent CALL, rather than DDL. One possible reason for this is the relative recentness of DDL research; the compilation of research abstracts referenced above (790 studies) is probably pretty close to all the empirical studies published in English. Another is the lack of a dedicated DDL journal; some of the list work in applied linguistics is probably being done by DDL researchers.

Word list research within applied linguistics is rich and varied and has compared and categorized lists on questions like their text coverage (Cobb & Laufer, 2021; Dang, 2020); their usability by teachers (Dang et al., 2020); their role in exam writing (Marsden et al., 2023); whether the lemma or the family is the more useful unit of word counting and list organization (McLean, 2017); or whether token lists or family lists provide the better predictor of writing proficiency (Crossley et al., 2013). All these issues have a potential bearing on which lists should be used within DDL. It seems implicit in this research that the best lists, as progressively identified, will be used mainly in behind-the-scenes software, i.e. in tests, profilers, and cloze passage builders, rather than given up-front to learners out of context.

Is it a problem that mainstream DDL confines itself to concordancing while applied linguistics/CALL gets the rest of DDL? In other words, is it a problem that DDL researchers and applied linguists appear to know each other so little? It could be a problem. DDL's main successes so far have been, as shown in meta-analyses by Boulton and Cobb (2017) and Ueno and Takeuchi (2023), mainly in vocabulary and collocation, yet largely without input from or engagement with the vocabulary research on the same topics. Conversely, many of the recent successes of vocabulary research have been dependent on corpus work, yet possibly without sufficient engagement with DDL expertise, where the possibilities of the technologies that these involve may be better known. The problem of "siloing" exists in many branches of applied linguistics, not just this one, and the relative youth and evident energy level of DDL researchers probably means this work will eventually be incorporated as a branch of CALL, and may even provide its long-awaited theoretical basis.

Conclusion

Word lists have been left out of the DDL conversation. However, they are as much a part of DDL as concordances, and it is a limitation in the development of DDL not to include list research within our research agenda. Doing so will involve working with and learning from our colleagues in applied linguistics where word list research is flourishing.

Cross-References

- ▶ Concordancers for Data-Driven Learning
- ▶ Corpus-Informed Materials in Language Teaching
- ▶ Future Challenges and Opportunities for Data-Driven Learning
- ▶ Pattern-Hunting and Pattern-Defining in Data-Driven Learning
- ▶ Pedagogical Guides for Data-Driven Learning
- ▶ Teacher Education for Pedagogical uses of Corpora
- ▶ The History of Data-Driven Learning
- ▶ Tools for Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

American Heritage Publishing Company. (1969). *The American Heritage dictionary of the English language*. Houghton- Mifflin-Harcourt.

Boers, F., Lindstromberg, S., & Eychmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics, 11*, 41–62.

Bondi, M., & Scott, M. (Eds.). (2010). *Keyness in texts*. John Benjamins.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 65*(2), 1–46. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology, 25*(3), 66–89. https://doi.org/10125/73450

Cobb, T. (1999a). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research & Development, 47*(3), 15–33. https://doi.org/10.1007/BF02299631

Cobb, T. (1999b). Breadth and depth of vocabulary acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*, 345–360. https://doi.org/10.1076/call.12.4.345.5699

Cobb, T. (2018). From corpus to CALL: The use of technology in teaching and learning formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 192–211). Taylor & Francis. https://doi.org/10.4324/9781315206615-11

Cobb, T., & Laufer, B. (2021). A nuclear word family list: The most frequent family members, base and affixed words. *Language Learning, 71*(3), 834–871. https://doi.org/10.1111/lang.12452

Crossley, S., Cobb, T., & McNamara, D. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System, 41*(4), 965–981. https://doi.org/10.1016/j.system.2013.08.002

Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 288–303). Routledge. https://doi.org/10.4324/9780429291586

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high- frequency words: Teachers' and learners' perspectives. *Language Teaching Research, 26*(4), 1–25. https://doi.org/10.1177/1362168820911189

Francis, W., & Kučera, H. (1967). *Computational analysis of present-day American English*. Brown University Press.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://doi.org/10.1016/j.system.2004.04.001

Kučera, H. (1982). The mathematics of language. In *The American Heritage dictionary*. Houghton Mifflin.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*, 15–30. https://doi.org/10125/66648

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition, 46*(1), 53–85. https://doi.org/10.1016/0010-0277(93)90022-N

Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *The Modern Language Journal, 107*(3), 669–692. https://doi.org/10.1111/modl.12866

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320. https://doi.org/10.1093/applin/ams010

McCrostie, J. (2007). Investigating the accuracy of teachers word frequency intuitions. *RELC Journal, 38*(1), 53–66. https://doi.org/10.1177/0033688206076158

McLean, S. (2017). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics, 39*, 823–845. https://doi.org/10.1093/applin/amw050

Meara, P., & Buxton, B. (1987). An alternative to multiple choice will vocabulary testing. *Language Testing, 4*(2), 142–154. https://doi.org/10.1177/026553228700400202

Nagy, W. (1988). *Teaching vocabulary to improve reading*. NCTE.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Nesi, H., Gardner, S., Thompson, P., & Wickens, P. (2009). *British academic written English corpus*. Warwick University.

Sasao, Y., & Webb, S. (2018). The guessing from context test. *International Journal of Applied Linguistics, 169*(1), 115–141. https://doi.org/10.1075/itl.00009.sas

Scott, M. (2001). Comparing corpora and identifying key words, collocations, and frequency distributions through the Word Smith Tools suite of computer programs. In *Small corpus studies and ELT* (pp. 47–67). John Benjamins. https://doi.org/10.1075/scl.5.07sco

Seliger, H. (1983). The language learner as linguist: Of metaphors and realities. *Applied Linguistics, 4*(3), 179–191. https://doi.org/10.1093/applin/4.3.179

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Ueno, S., & Takeuchi, O. (2023). Effective corpus use in second language learning: A meta-analytic approach. *Applied Corpus Linguistics, 3*, 1–11. https://doi.org/10.1016/j.acorp.2023.100076

C

Corpora in Language Assessment

Agnieszka Leńko-Szymańska
University of Warsaw, Warsaw, Poland

Abstract

Corpora have had a significant impact on language assessment, contributing to the creation of valid instruments for evaluating second language (L2) proficiency. Different types of corpora provide empirical data that support language assessment theory and research as well as enhance the design, development, and validation of assessment instruments. Corpus linguistics facilitates the theoretical evolution of language proficiency constructs to encompass communicative competence and its components. It helps to identify key linguistic elements and skills necessary for effective language use across various domains, ensuring that assessment instruments align with authentic language practices. Furthermore, corpora prove invaluable resources in the development of test items and the refinement of scoring rubrics. They are also employed to validate existing tests against native speaker and L2 learner benchmarks. Finally, corpora and corpus analytical methods underpin the creation of automated scoring systems that enable valid, reliable and efficient assessment of test-takers' written and spoken responses.

Keywords

Automated assessment · Construct definition · Construct validity · Exam corpora · L2 assessment · Learner corpora · Scoring rubric · Test item · Test validation

Introduction

Corpora—large electronic collections of written or spoken discourse—have made a significant impact on language pedagogy and, more broadly, language education, with their influence extending into the field of language assessment. The availability of diverse corpora, along with corpus-based analytical methods and tools, has proven to be conducive to enhancing and supporting evaluation of second language (L2) competence and proficiency. Large reference corpora of native-speaker language and smaller collections of specialised registers provide a wealth of information on typical language use in different contexts and situations. Learner corpora, which include L2 written and spoken production, make it possible to chart detailed trajectories of L2 development. Finally, exam corpora, such as the Cambridge Learner Corpus (CLC, Cambridge University Press & Assessment, n.d.-b) and the Trinity Lancaster Corpus (TLC; Gablasova et al., 2019), which consist of test takers' responses to open-ended test items in proficiency exams, provide insights into learners' performance at different proficiency levels and across specific exam

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_157-1

tasks (Gablasova, 2020). Outcomes of corpus-based analyses, such as frequency lists, statistical collocations, keywords or concordance lines, have proven invaluable to assessment researchers and exam developers.

The effects of corpora and corpus linguistics are evident on two levels: the theoretical and research-oriented aspects of language assessment, as well as the practical application in the design, development, and validation of assessment instruments. Examination boards utilise corpora to create standardised high-stakes proficiency exams, while language teachers benefit from these resources to develop tailored tests for their students.

Defining the Constructs

At a theoretical level, the influence of corpora on language assessment emerged well before researchers explicitly acknowledged such connections (Alderson, 1996). The rapid advancements in corpus linguistics in the final three decades of the twentieth century have contributed to challenging traditional views of language proficiency, which had previously focused solely on grammatical accuracy and vocabulary range. Communicative language ability, as conceptualised for language assessment by Bachman (1990), came to be understood as the capacity to use language effectively and appropriately in social contexts, integrating various competences: linguistic, textual, illocutionary, sociolinguistic, and strategic. Corpus linguistics has helped to refine these concepts by providing insights into actual language practices. It has demonstrated that communicative competence extends beyond adherence to grammatical rules and the use of sophisticated vocabulary; it also involves the ability to employ appropriate discourse strategies, manage pragmatic functions, be sensitive to sociolinguistic norms, and use language with fluency.

One of the significant contributions of corpus linguistics to language assessment has been highlighting the interdependence of lexis and grammar, as well as the prominence of phraseology in language. Sinclair's (1991) idiom principle emphasised the crucial role of formulaic expressions in building proficiency. Moreover, corpora have drawn researchers' attention to variations in language according to register, genre, and modality. Corpus-based studies have revealed that effective language use differs substantially across contexts. For example, comparisons of spoken and written corpora have demonstrated differences in grammatical complexity, formality, and cohesion strategies specific to each modality (Biber et al., 1999). This indicates that language proficiency cannot be fully captured by a single, unified construct. Instead, it involves a range of skills and strategies, including the ability to adapt language to various communicative situations.

These findings have led researchers and practitioners in language assessment to redefine and re-operationalise the constructs of language proficiency and its components. The field has moved beyond generalised standards toward more nuanced definitions that consider how proficient speakers navigate diverse language tasks, making the construct of proficiency more precise and grounded in real-world language use (Xi, 2017).

Describing Target Language Use Domains

Corpora have played a crucial role in providing precise descriptions of target language use domains, offering empirical insights into how language is employed in specific contexts and for particular purposes. By analysing large reference corpora of real-world language data as well as carefully curated collections representative of well-defined linguistic populations (Egbert, 2017), test developers can identify linguistic features and patterns characteristic of relevant communicative situations. This enables more accurate descriptions of essential language elements such as vocabulary, phrases, grammatical structures, as well as skills and strategies, including discourse, pragmatic, and rhetorical strategies, that learners need to master. This approach ensures that assessment instruments align with authentic language use, reflecting the demands of real-life communication.

Corpora have proven equally valuable in professional and occupational language testing. Specialised collections, including those of academic, medical, and business registers, help delineate target language use domains by identifying key terminology, discourse patterns, and interactional features relevant to specific fields. A notable example is the Pearson International Corpus of Academic English (PICAE; Ackermann et al., 2011), which was compiled to inform the development of the Pearson Test of English Academic. On a smaller scale, Mell (2004) analysed an aviation corpus to identify communicative functions and target language structures that should be included in a test of English for aeronautic professionals.

Writing Test Items

Another significant application of corpora in language evaluation is the development of test items. Corpora offer a data-driven approach to creating assessment tasks that accurately reflect real-world language use and are tailored to different proficiency levels. By analysing authentic language data, test developers can ensure that assessment tasks are not only relevant and contextually appropriate but also systematically aligned with the language features they aim to evaluate, thereby enhancing the validity and authenticity of the tests. This application encompasses various aspects of item development.

One key area where corpora have made substantial contributions is the selection of lexis and phraseological units to be targeted in a test. Frequency analyses from large general and small specialised corpora enable test developers to identify commonly used words, collocations, and phrases, ensuring that the selected items reflect actual usage patterns. Similarly, corpora assist in the creation of test items that focus on phraseological, grammatical, discourse, or pragmatic features by providing empirical data on their actual distribution in natural language. Additionally, error-annotated learner and exam corpora offer insights into frequent mistakes made by candidates at various proficiency levels, indicating which items should be targeted in a language test. Frequent errors can also be suitable distractors for multiple-choice items (Cushing, 2021). Corpora can also act as a source of appropriate contexts in which to present target language elements, ensuring that the material aligns with authentic usage (Leńko-Szymańska, 2024).

This corpus-based approach is particularly relevant in the realm of testing languages for specific purposes. Specialised corpora enable quick identification of domain terminology, including multiword units. For instance, Pacheco et al. (2023) demonstrated how a corpus of Aeronautical English can be utilised to create test items for developing items in a test of English for pilots and air traffic controllers.

Refining Scoring Rubrics

By providing empirical evidence on L2 use across different proficiency levels, learner and exam corpora are invaluable for refining scoring rubrics in the assessment of candidates' open-ended responses. Their analyses enable test developers to identify specific linguistic features that correlate with varying performance across different scoring bands, thus helping to create more precise, data-driven descriptors within the rubrics. This process ensures that rating guidelines accurately reflect the range of language abilities being assessed and provide clear, objective, and transparent criteria for evaluating learners' performance in speaking and writing tasks, thereby enhancing the validity and reliability of language tests (Wisniewski, 2017).

For instance, in assessing spoken language, corpora contribute to refining scoring criteria by offering insights into features of L2 production such as fluency, pronunciation, and pragmatic competence. Spoken learner corpora help to identify patterns in fluency, such as the frequency of hesitations, self-corrections, and speech rate across different proficiency levels. This information can be utilised to define rubric criteria that distinguish fluent speech from less fluent speech. Römer (2017) highlighted the importance of incorporating criteria related to lexicogrammar

in rubrics for high-stakes speaking tests, recommending that scoring guidelines should better reflect the role of phraseology in language use.

Corpora also facilitate the calibration of rubrics for specialised assessments, such as those targeting academic or professional language use. By analysing corpora of expert language in fields such as law, business, or healthcare, rubrics can be refined to include criteria that reflect the specialised vocabulary, discourse structures, and communicative functions typical of those contexts. This ensures that scoring criteria are aligned with the real-world language demands of specific domains.

Validating Tests

Corpora play a crucial role in the process of test validation, offering a robust, data-driven approach to evaluating the extent to which language tests accurately measure the intended constructs, which helps in verifying construct validity. Test developers can analyse the linguistic features present in test items and compare them with authentic native-speaker language data. This comparison helps to ensure that test content reflects the range of vocabulary, grammar, discourse structures, and functions appropriate for the targeted language variety. For instance, LaFlair and Staples (2017) employed this approach to investigate the extrapolation interference of the Michigan English Language Assessment Battery (MELAB) Oral Proficiency Interview (OPI), which assesses L2 English speaking ability in academic, professional, and social domains. Among other applications, this exam is often required for certifying L2-speaking medical professionals. To support the extrapolation of test scores to real-world performance, the researchers compared linguistic features of a small corpus of exam responses with reference corpora representing different target language use domains, including office-hour academic consultations and nurse-patient interactions.

Moreover, corpora support test validation by facilitating the analysis of test-takers' responses to assess the consistency and accuracy of scoring practices. By comparing language produced by test takers with language samples in learner and exam corpora, researchers can evaluate whether the criteria used for scoring adequately capture the quality and range of actual language use by L2 speakers. This approach ensures that test tasks are appropriately challenging and that scoring practices are transparent and reliable, thus enhancing the overall validity of the assessment instrument. For instance, test tasks' written and oral responses can be validated against learner and exam corpora to confirm that they align with the linguistic capabilities expected at a specific level of the Common European Framework of Reference for Languages (CEFR).

Automated Assessment

One of the key corpus-related advancements in the field of language assessment, encompassing both research and practical applications, is the automated evaluation of candidates' responses to open-ended test items that elicit extended samples of writing or speech. Test takers' written and spoken production has been traditionally evaluated by human raters using carefully designed scoring rubrics that describe performance characteristics across different score bands. However, this assessment method has its limitations, including reliability issues as well as the demands on time and financial resources. Corpora have played a pivotal role in the development of automated language assessment systems that can support and, in some cases, replace human raters providing accurate, consistent and efficient evaluations of candidates' unstructured responses.

Automated scoring (AS) systems employ analytical methods from corpus linguistics and natural language processing to evaluate candidates' output. For each written or spoken response, a system computes a large number of metrics that capture both surface-level and deep linguistic features, for example response length, number of pauses, lexical diversity and sophistication, spelling, punctuation or pronunciation errors as well as semantic meaning, coherence, syntactic structure, and prosodic variations (Foltz et al., 2020). In the

realm of spoken language assessment, corpora also facilitate automatic speech recognition (ASR). Such corpora must include various accents and speech patterns, which can improve the system's ability to handle diverse user inputs.

In automated scoring systems, generated metrics are calibrated and transformed to compute a score based on a sophisticated model built from a corpus of past exam responses that were evaluated by human raters. This model establishes a complex statistical relationship between the generated measures and the human-assigned scores. Thus, exam corpora serve as benchmarks for interpreting measures of language features, allowing automated systems to assess candidates' responses in relation to established norms. Such systems can also offer a detailed breakdown of a test-taker's strengths and weaknesses by comparing the candidate's linguistic patterns to those observed in native and non-native corpora, identifying deviations in language use. This comparison enables a nuanced evaluation, highlighting specific areas where the test-taker's language use diverges from established norms and benchmarks.

The representativeness of corpora is crucial in the development of automated scoring systems, as the data must accurately reflect the diversity of language varieties and contexts being assessed to ensure fair and unbiased scoring. A well-balanced corpus enables the scoring model to accommodate a wide range of language use, enhancing the system's ability to deliver valid and reliable assessments across different language levels and contexts.

A prominent example of such automated scoring system is *e-rater*, developed by educational testing service (ETS), which is used in high-stakes examinations such as test of english as a foreign language (TOEFL) and Graduate Recod Examinations (GRE) for the evaluation of written production (Attali & Burstein, 2004). Another ETS-developed tool, *SpeechRater*, evaluates spoken responses, analysing features such as fluency, pronunciation, and language use. Beyond standardised tests, corpora also support the development of automated feedback systems in language learning platforms like *Write & Improve* (Cambridge University Press & Assessment, n.d.-a). This tool analyses learner input, compares it with data from learner corpora (e.g., Cambridge Learner Corpus), and provides targeted feedback on common errors. By drawing on error-tagged learner corpora, the system can automatically detect issues such as article misuse or verb agreement errors, providing corrective suggestions based on patterns found in the data.

Overall, the application of corpora in L2 assessment has revolutionised language testing, enhancing its validity, authenticity, and data-driven nature. As artificial intelligence continues to advance, we can expect further developments in the application of corpus data and methodologies, ushering language assessment into a new era marked by increased precision and adaptability.

Cross-References

- ▶ Automated Scoring in CALL
- ▶ CALL Glossaries and Vocabulary Assessment
- ▶ LSP Assessment and Technology
- ▶ Validity and Reliability of CALL Assessments

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Ackermann, K., De Jong, J. H. A. L., Kilgarriff, A., & Tugwell, D. (2011). The Pearson International Corpus of Academic English (PICAE). In *Proceedings of corpus linguistics*. Retrieved October 24, 2024, from https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2011/Paper47.pdf

Alderson, J. C. (1996). Do corpora have a role in language assessment? In J. Thomas & M. H. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leech* (pp. 248–259). Longman.

Attali, Y., & Burstein, J. (2004). Automated essay scoring with e-rater® v.2.0. *ETS Research Report Series, 2004*(2), i–21. https://doi.org/10.1002/j.2333-8504.2004.tb01972.x

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.

Cambridge University Press & Assessment. (n.d.-a). Write & improve with Cambridge. Retrieved October 24, 2024, from https://writeandimprove.com/

Cambridge University Press & Assessment. (n.d.-b). *Cambridge Learner Corpus*. Retrieved October 24, 2024, from https://www.cambridge.org/elt/corpus/learner_corpus2.htm

Cushing, S. T. (2021). Corpus linguistics and language testing. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed., pp. 545–560). Routledge.

Egbert, J. (2017). Corpus linguistics and language testing: Navigating uncharted waters. *Language Testing, 34*(4), 555–564. https://doi.org/10.1177/0265532217713045

Foltz, P. W., Yan, D., & Rupp, A. A. (2020). The past, present, and future of automated scoring. In D. Yan, A. A. Rupp, & P. W. Foltz (Eds.), *Handbook of automated scoring* (pp. 1–9). Taylor & Francis.

Gablasova, D. (2020). Corpora for second language assessments. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 44–52). Routledge.

Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research, 5*(2), 126–158.

LaFlair, G. T., & Staples, S. (2017). Using corpus linguistics to examine the extrapolation inference in the validity argument for a high-stakes speaking assessment. *Language Testing, 34*(4), 451–475. https://doi.org/10.1177/0265532217713951

Leńko-Szymańska, A. (2024). Corpora in L2 vocabulary assessment. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 157–173). Routledge.

Mell, J. (2004). Language training and testing in aviation needs to focus on job-specific competencies. *ICAO Journal, 59*, 12–14.

Pacheco, A., Garcia, A. C. de M., Monteiro, A. L. T., Prado, M. C. de A., & Tosqui-Lucks, P. (2023). Using corpus linguistics to create tasks for teaching and assessing Aeronautical English. *Applied Corpus Linguistics, 3*(3), 100075. https://doi.org/10.1016/j.acorp.2023.100075

Römer, U. (2017). Language assessment and the inseparability of lexis and grammar: Focus on the construct of speaking. *Language Testing, 34*(4), 477–492. https://doi.org/10.1177/0265532217711431

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Wisniewski, K. (2017). Empirical learner language and the levels of the Common European Framework of Reference. *Language Learning, 67*(1), 232–253. https://doi.org/10.1111/lang.12223

Xi, X. (2017). What does corpus linguistics have to offer to language assessment? *Language Testing, 34*(4), 565–577. https://doi.org/10.1177/0265532217720956

P

Pronunciation and Data-Driven Learning

Kevin Hirschi[1] and Okim Kang[2]
[1]The University of Texas at San Antonio, San Antonio, TX, USA
[2]Northern Arizona University, Flagstaff, AZ, USA

Abstract

Data-driven learning (DDL) provides second-language (L2) learners with numerous samples of natural language from which they can detect patterns and understand usage. While largely restricted to the textual domain, the increasingly available number of multimedia corpora and concordance tools provide opportunities for learners to search and listen to natural speech. This entry reviews the theoretical underpinnings for applying DDL to pronunciation as well as the empirical work that provides guidance for its implementation. It also outlines a research agenda for future investigations of DDL for L2 pronunciation that is likely to enhance our understanding of how to best implement multimedia corpora considering design factors in a wide range of learning contexts.

Keywords

Data-driven learning · Second-language pronunciation

Introduction

Data-driven learning (DDL) for pronunciation is an emerging focus of second-language teaching and research in which learners and/or potentially teachers or material developers exploit corpora of natural language to promote learners' pronunciation skills. While DDL has gained empirical support and diversified its approaches for several decades, the incorporation of pronunciation is a relatively recent innovation, likely due to both the focus of DDL on grammar and vocabulary as well as the technological constraints of working with multimedia corpora. With the advent of multimedia corpus search engines such as YouGlish (https://youglish.com/), as well as programmes for compiling and searching multimedia corpora, the literature incorporating and supporting DDL for pronunciation is a nascent yet promising area of research which is supported by leading theories of language learning.

Theoretical Underpinnings for DDL for Pronunciation

DDL draws from theoretical groundings in language acquisition that primarily centre on usage-based approaches to language acquisition (see Römer (2024), for an overview). In one explanation of the usage-based model, Langacker (2008) structures learning though patterns of features that are common across language units and

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_161-1

encountered in events that are produced by actual language users. Elements that comprise these features include vocabulary items, grammatical structures and 'sound patterns' (Langacker 2008, p. 81). He further posits that learning occurs when learners are given sufficient, representative and meaningful exposure to units. While such exposure is a natural part of acquisition, it can be bolstered by engaging with corpora as patterns become more salient when language units are collected and presented together. The relationship between saliency and learning is also supported by the Noticing Hypothesis (Schmidt, 1990), which posits that learners cannot learn elements of input, such as sound patterns, without paying attention to these elements.

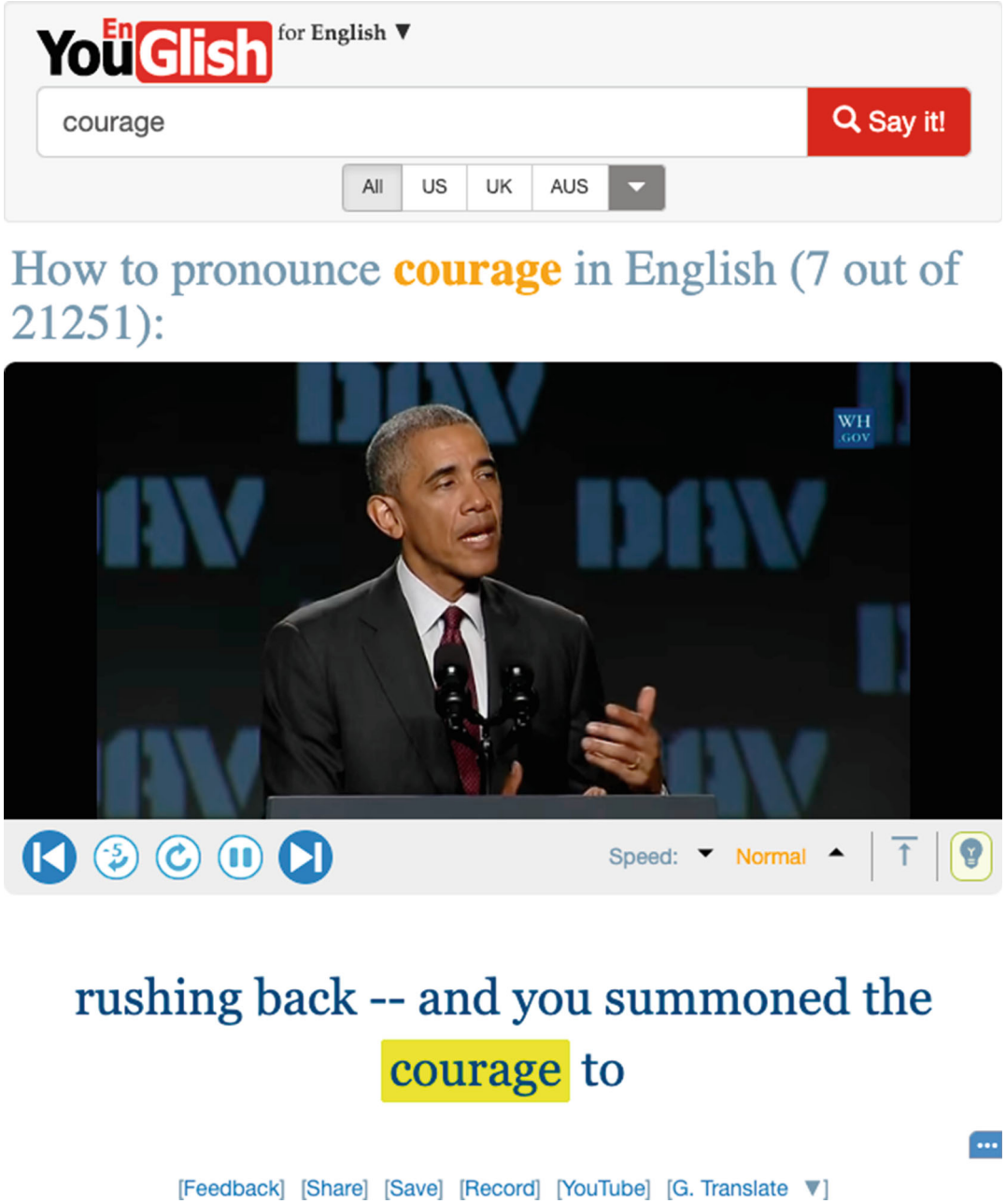
Models of learning phonology also support the use of input, in general terms, and contextually varied input to improve perceptive and productive skills. These include the Revised Speech Learning Model (Flege & Bohn, 2021), which theorizes that perception and production coevolve and that input can be effective for improving L2-speaking abilities, particularly considering individual sounds. Furthermore, research supports the use of High Variability Phonetic Training (HVPT), in which the phonological contexts and speakers are varied intentionally so that learners can generalize individual sounds to a variety of contexts in the perceptual systems (Thomson, 2018). While these theories focus on segmental (i.e. individual phoneme) learning, the use of DDL to bolster input, when coupled with productive practice, is also likely to improve suprasegmentals (i.e. intonation, pitch, fluency) because it can allow for targeted yet generalizable input across a variety of factors both inside and outside the language unit.

Previous Research on DDL for Pronunciation

Recent research has reported on efforts to investigate using corpora to create DDL instructional materials (Chen & Han, 2020; Staples, 2019) or investigated the effects on learner pronunciation when students engage in DDL directly by searching a corpus (Fu & Yang, 2019; Hirschi & Kang, 2024; Kartal & Korucu-Kis, 2020; Sardegna & Jarosz, 2022).

Several researchers have used corpora in innovative ways to inform pronunciation instruction and enhance learning outcomes. Chen and Tian (2022) report on training teachers to use spoken corpora in their classroom which resulted in positive perceptions and improved outcomes in an accompanying mock demonstration lesson. In a related study, Chen and Han (2020) report on a corpus-based systematic analysis of problematic phonological features for Hong Kong learners of Mandarin. In their study, the feature list which was derived from their corpus was then used to create an online practice programme for learners to self-assess their pronunciation, learn from model speakers and study phonology. These two studies highlight the potential of DDL forms in which learners may not directly engage with a multimedia corpus but can benefit from language data in their pronunciation development. In the same vein, Staples (2019) reports on a comprehensive corpus analysis of internationally educated nurses which in turn resulted in corpus-based materials in a pronunciation course for nurses. The results of the course indicated improved pronunciation, especially pitch range and sentence stress, due to the DDL intervention.

To date, four empirical studies are known to have investigated the effects of allowing learners to engage directly with a multimedia corpus for the express purpose of practising pronunciation, all of which indicated advantages when aligned with theory and pedagogy. Three studies used YouGlish (Fu & Yang, 2019; Kartal & Korucu-Kis, 2020; Sardegna & Jarosz, 2022), a search tool for YouTube videos and allows for playback manipulation (e.g. repeating, slowing down and viewing captions) of a search token across several YouTube examples (see Fig. 1). In the Taiwanese context, Fu and Yang (2019) guided 60 English learners in engaging with YouGlish to practise specific words and intonation. Their study found positive effects for overall pronunciation and specifically for intonation. In a study that focused on lexical stress among twelve L1 high school-aged Polish learners of English, Sardegna and Jarosz (2022) also used YouGlish, finding an improvement in



Pronunciation and Data-Driven Learning, Fig. 1 Example YouGlish search result for the word 'courage'. *Note*. YouGlish searches across YouTube videos for a target word or phrase. The variety of English, or other languages, can be selected, and the resulting videos that contain the search string can be advanced through using the blue forward and backwards buttons. The playback speed can be adjusted, allowing for careful listening. This screen shot was captured in September 2024

both perception and production of stress but only if they first introduced lexical stress rules. Beyond the typical classroom context, Kartal and Korucu-Kis (2020) included YouGlish in English teacher skill development efforts in Türkiye for production of lexical stress in English with positive outcomes.

Taken together, these studies find a positive impact of search tools such as YouGlish among a wide range of participant characteristics and support word-level sound learning through DDL but caution that implicit approaches to teaching phonological features of the target language may not be as effective as explicit instruction.

Beyond the use of YouGlish and outside of the typical classroom context, Hirschi and Kang (2024) report on a novel corpus concordance search tool that includes audio playback of highly intelligible L2 speakers of English and compare its use by learners with teacher-created materials in a self-access learning study. The findings indicate that DDL has an advantage for learning patterns between words, such as sentence stress. However, for lexical stress, participants gained equally from DDL as they did from textbook materials. While other investigations using purpose-built programmes that allow learners to engage with multimedia corpora are certain to be forthcoming, the initial studies reported above suggest that DDL can have a positive impact on learner's pronunciation with an advantage for sound patterns that extend beyond the single word.

Future Research

While the studies outlined above investigated DDL for pronunciation with various tools and designs, several threads emerge for practitioner and future researchers interested in expanding our understanding of the area. These can be categorized into two broad areas: *design and implementation* (e.g. design of the intervention, teacher involvement, duration and intensity) and *target features* (e.g. segmental/suprasegmental, prosody, fluency).

Several variables related to the *design and implementation* of DDL for pronunciation have yet to be explored. These include, but are not limited to, determining the amount, pacing and duration of exposure to a specific feature that results in successful and sustained learning. These design elements could be explored through a variety of studies that focus on a single feature but vary the number of times, targets, concordance lines, linguistic context, instructional components or activities associated with the DDL activity. Furthermore, these elements could be studied in the classroom or in a self-access context. Additional efforts and studies on preparing teachers to use corpora in the classroom for pronunciation, especially given the constraints of tests and curricula, are also needed. Furthermore, investigations of DDL tool designs are required to fully investigate the visual and interactive components of the corpora and how learners (and/or their teachers) can interact with language data. While resources such as YouGlish are certain to remain a mainstay of incorporating pronunciation in DDL, the visual presentation, accessibility and interactive features of a purpose-built multimedia concordance tool are certain to impact learning outcomes and as of yet are rarely grounded in theory (Römer, 2024). Finally, the design of spoken corpora can serve as another area of study. Dimensions unique to using multimedia corpora in DDL include a collecting clear recordings of speakers' voices while not compromising the naturalness of the language, considering the phonological layers of discourse when determining the representativeness of the language domain and compiling enough data to be able to provide multiple language events for pattern building. In sum, there are a multitude of ways one could use a multimedia corpus for language learning, and only few of which have been validated empirically.

Following emerging research on second-language phonology, investigations of the features that are most critical for effective communication should drive future research in promoting pronunciation through DDL. For English as a second- or foreign language learners, suprasegmental features such as lexical stress, sentence stress, intonation, pauses and other fluency-based features, as well as high functional load segmentals, are known to be critical for intelligibility (Kang et al., 2020). In Spanish, intonation and the final /o/ are known to impact intelligibility and comprehensibility (Nagle et al., 2023). Although research on phonological features of intelligibility and comprehensibility is rare in other languages, adopting a communicative competence-focused corpus investigation to inform pedagogy, such as Staples (2019) and Chen and Han (2020), provides a fruitful path forward for a wide range of purposes and learning contexts.

Conclusion

This encyclopaedia entry discusses the theories that support the emerging empirical research on DDL for second-language pronunciation. While there are certainly promising reports on improving pronunciation skills with DDL with classroom and self-access learners, from schools to professional contexts, there remain a significant number of unexplored possibilities. Following the theoretical underpinnings of pattern making and noticing salient features of input (Langacker, 2008; Schmidt, 1990), and expanding upon work that uses novel and available Internet tools, the future of bolstering pronunciation development through the use of DDL is bright.

Cross-References

- ▶ Communicative Data-Driven Learning
- ▶ Concordancers for Data-Driven Learning
- ▶ Listening and Data-Driven Learning
- ▶ Multimodal Tools and Corpora for Data-Driven Learning
- ▶ Speaking and Data-Driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Chen, H. C., & Han, Q. W. (2020). Designing and implementing a corpus-based online pronunciation learning platform for Cantonese learners of Mandarin. *Interactive Learning Environments, 28*(1), 18–31. https://doi.org/10.1080/10494820.2018.1510422

Chen, H. C., & Tian, J. X. (2022). Developing and evaluating a flipped corpus-aided English pronunciation teaching approach for pre-service teachers in Hong Kong. *Interactive Learning Environments, 30*(10), 1918–1931. https://doi.org/10.1080/10494820.2020.1753217

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Fu, J. S., & Yang, S. H. (2019). Exploring how YouGlish facilitates EFL learners' speaking competence. *Journal of Educational Technology & Society, 22*(4), 47–58.

Hirschi, K., & Kang, O. (2024). Data-driven learning for intelligibility in the academic spoken context. *TESOL Quarterly, 58*(3), 1081–1107. https://doi.org/10.1002/tesq.3302

Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics, 41*(4), 453–480. https://doi.org/10.1093/applin/amy053

Kartal, G., & Korucu-Kis, S. (2020). The use of Twitter and Youglish for the learning and retention of commonly mispronounced English words. *Education and Information Technologies, 25*(1), 193–221. https://doi.org/10.1007/s10639-019-09970-8

Langacker, R. W. (2008). Cognitive grammar as a basis for language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 66–88). Routledge.

Nagle, C., Huensch, A., & Zárate–Sández, G. (2023). Exploring phonetic predictors of intelligibility, comprehensibility, and foreign accent in L2 Spanish speech. *The Modern Language Journal, 107*(1), 202–221. https://doi.org/10.1111/modl.12827

Römer, U. (2024). Usage-based approaches to second language acquisition vis-à-vis data-driven learning. *TESOL Quarterly, 58*(3), 1235–1245. https://doi.org/10.1002/tesq.3278

Sardegna, V. G., & Jarosz, A. (2022). Exploring how YouGlish supports learning English word stress: A perception study. In V. G. Sardegna & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training: Studies in honour of Ewa Waniek-Klimczak* (pp. 165–184). Springer International Publishing. https://doi.org/10.1007/978-3-030-98218-8_10

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158. https://doi.org/10.1093/applin/11.2.129

Staples, S. (2019). Using corpus-based discourse analysis for curriculum development: Creating and evaluating a pronunciation course for internationally educated nurses. *English for Specific Purposes, 53*, 13–29. https://doi.org/10.1016/j.esp.2018.08.005

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation, 5*(1), 208. https://doi.org/10.1075/jslp.17038.tho

T

Tim Johns and Data-Driven Learning

Mike Scott
Lexical Analysis Software and Aston University, Birmingham, UK

Abstract

This entry considers the role played by Tim Johns mainly in the 1980s and 1990s creating an environment which would promote data-driven learning in English for academic purposes and in attaching the term *data-driven learning* to such educational endeavours. It takes a number of key ideas that Johns promoted in this connection and outlines how they operated.

Keywords

Tim Johns · Data-driven learning · CALL · English for academic purposes

Introduction

Although research into data-driven learning (DDL) still has many unanswered questions (O'Keeffe, 2021), the term *DDL* is now well established and highly relevant to language teaching (Crosthwaite, 2024).

My aim herein is to provide some reflections on the role Tim Johns (1936–2009) played in this, some 40 or more years ago, as someone who worked with him fairly extensively in the 1980s and 1990s when his ideas were developing. Tim was at Birmingham university and I was based in Brazil in the 1980s and then at Liverpool university, but he spent some months in Brazil with several visits, one of those as our visiting lecturer for the Brazilian National ESP Project in a 1982 conference in Natal RN, and at another in 1989 in Porto Alegre and then São Paulo where he lodged in our apartment for a few days. We also acted as lecturers in a summer school for the Catholic University.

Our collaboration came about essentially because in his 1989 stay in Brazil, Tim had asked me to help users at the University of São Paulo with his new program, *MicroConcord*. However, providing backup to another person's program proved beyond my capabilities so I decided to write a separate concordance program based on Tim's description and design. I sent it to him on a floppy disk through the post (!) and thus the collaboration that would eventually be MicroConcord (OUP, 1993) was born.

In this entry I have benefitted considerably from the advice and memories of John Higgins, who collaborated with Tim over many years and wrote a moving testimonial to him. The structure of the entry is derived from a number of important DDL principles which Johns outlines in his publications.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_239-1

Data-Driven Learning

Tim Johns brought the term *DDL* formally into language teaching and into corpus linguistics with his 1991 *ELR Journal* article. However, his pedagogical insights and his enthusiasm for computer-assisted language learning (CALL) go back at least to the start of the previous decade. In 1982 he had already written:

> . . . a refinement of CAI which we should like to look into. . . [is] a blurring of the distinction between 'research program' and 'teaching/learning program'. What I have in mind here is an interactive program which would give the language-learner a 'user-friendly' tool for investigating problems in the target language – for example, the use in context of a problem word, or the vocabulary of a certain type of text. (Johns, 1981, p. 105)

John Higgins, in his testimonial to Tim Johns, saw this as "the source of DDL, the spring from which flows the river and from it the wide estuary I am looking at now" (no page).

1980s: A Ferment of CALL Activity

With the new decade in 1980, personal computers were beginning to be readily available and affordable by schools and individuals in the USA and its sphere of influence. I worked in Mexico in the late 1970s and saw Commodore Pet machines—but had no easy access to one. (The computer processing I did there involved buying punch-hole cards.)

Johns was excited by this nascent technology; he wanted access, he wanted to puzzle out how to use it effectively. The way to do that in those days was to teach yourself, to read as widely as you could, picking up tips in amateur enthusiast publications, by trial and error, by testing a plan and seeing where it failed then struggling on to find a better path through to get it to work, by discussing it with others. This was his experience and became mine, partly fired by his enthusiasm and knowledge of the hardware and software of the early 1980s. (Tim suggested I should get the NewBrain computer when we met in 1982, having seen its possibilities. I bought one (for £199) later that year when in the UK and took it back to Brazil. There was no software available to me there, so I had to use the excellent manual supplied with the NewBrain and guess my way forward using Basic and a few peeks and pokes.) He built a bridge between his profession as a teacher and the new technology. This led him to design computer tools for *investigating* language, for trying to understand words in context in a new language. The learner as researcher. Crucial ideas in Johns' philosophy include

- Rethinking the notion of play
- Student as researcher
- Beat the machine
- Analysis v. synthesis CALL activities
- Avoid one right answer or closed versus open issues
- Trialling
- Pattern spotting

The Notion of Play

John Higgins had a year working with Tim in 1982–1983. This gave rise eventually to Higgins and Johns (1984). In 2009 in his tribute to Tim, Higgins wrote:

> Tim had completely rethought the notion of play. . . Other people working in CALL at the time thought of play purely for motivation. The student has worked through some dull exercises; now they can play a game as reward. Tim put the play into the learning. Learning occurred because of the play. He also introduced a novel idea, that of asking the student not just to find answers but to wager counters on their confidence that the answer was right. (Higgins, 2009, n.p.)

Play is central to discovery. This is not to suggest that for Johns play was all: working one's way through his materials required concentration and effort and it is clear that his materials stretched students' capabilities. He was well aware that initially they were perplexed by concordance lines which contained only 80 or 100 characters, often beginning and ending with an incomplete word—never mind no further context. He was also a stickler for authentic text. At that time in the 1970s and 1980s, most language teaching

materials incorporated simplified text and teachers would question whether authentic text could be used in class. So Johns' activities were not child's play even though they were fun to tackle. Higgins (2009, n.p.): "Tim described himself as a kidult, someone who pursues childish-seeming activities with childish enthusiasm".

Teacher and Student

For Johns (1991b, p. 21), the role of the teacher is as an organiser of research rather than imparter of knowledge. The student is not to absorb the teacher's knowledge but to learn with the teacher from educationally relevant materials, with the aid of corpus software tools. Even outside the classroom, the Internet provides a way to promote learning:

> I have long been aware that points of interest not covered in standard works of reference come up in the course of this work, and that unless they are recorded in some way they are 'lost on the air'. In 1996 I started on my web site a series based on some of the points of lexis, syntax and discourse that have cropped up in one-to-one consultations... As an ex-chess player I called the points Kibbitzers, a term from the chess cafés of Central Europe around 1900, The Kibbitzer is not the chess player, but the man who watches two other people playing, and comments (often critically) on their play. In the context of the Web, I wanted it to suggest the possibility of observing and possibly commenting on the 'game' between student and consultant. (Johns, 2002, p. 111)

Beat the Machine

Discussing his Contexts program of 1994, Johns refers to a wish students have to "beat the machine", to find out what it cannot do as well as what it can, to get it if possible to generate an ungrammatical structure. For they

> ...are intelligent, and respond to challenges to their intelligence; and most of them are accustomed to the idea of research and finding things out for themselves. They are also tolerant, and even if they do not share my belief that one of the best ways to learn a foreign language is to fool around with it and in it, they are prepared to go along with the more improvisatory activities. (Johns, 1991c, p. 29)

CALL Analysis and Synthesis

The constraints of the programming environment forced programmers like Johns and Higgins to think hard about what could be achieved and what educational purposes could be met. In his 1981 text, Johns distinguishes between generation (of a CALL program) by synthesis and by analysis:

> The analysis group... took pieces of text and mutilated them by re-ordering or masking elements; students had to restore the original text by using guesswork and collaborative discussion. The synthesis group had the computer displaying language, while learners were asked to discover the rules that the machine was using and, if possible, force it to make a mistake. (Higgins, personal communication)

The constraints were considerable:

> Tim was an excellent programmer. His software didn't crash. The front ends were unglamorous but easy to use. Above all, the activities were riveting. In the days of the 1K ZX81, conciseness was the supreme virtue, and Tim achieved more in 1K than one can imagine now. He wrote S-ENDING which would inflect any verb with a 3rd-person –s, real or nonsense. (Higgins, 2009, n.p.)

Not a simple task with the computer memory available at that time. To hold a couple of pages of text 1 K is about the total space required. (A mobile phone in 2024 typically has at least 128GB. That is 1 K times 128 times 1000 times 1000.) That program needed to store any text it needed as well as the instructions in that tiny space. The aim at all times was to find tasks which would respect the user's intelligence and promote learning despite these constraints.

Closed and Open

For Johns, closed activities are those with a correct answer "in the sense that the result is known to the teacher in advance" (Johns, 1997, p. 101). What he discovered with DDL was that very often contemporary grammar based on the linguist's intuition was over-simplified or misleading, and that the activity of trying to work out how the language does work is informative for both teacher and student. His Kibbitzers arose from this and are still DDL models today.

Trialling

In a paper on his 1994 Contexts program, Johns analysed the responses of users (i.e. "trialling"), with the aim of improving it educationally. One user reported:

> (At first) I felt cheated. The program only accepts the exact words that have been programmed, although the words that I have chosen fit the sentences. As I progress through the other files, this feeling wears off and I found that it's more challenging to find a word that fits several contexts. It made me look 'deeper' into the word, how it's used in different contexts. (Johns, 1997, p. 111)

Another user reported focussing on the fact that the program reports on how much time one spent on a task, inferring that this easily-programmed feature may be pedagogically questionable.

It is clear that DDL was for Johns not a straightforward educational principle, still less a solution to classroom problems, but a struggle, requiring the teacher to keep learning in mind at all times.

Pattern Spotting

Some of the time his students had access to the software and could generate their own listings, but a favourite Johns' strategy exemplified in his other materials such as his Kibbitzers and his program *Contexts*, was for the teacher to generate a listing and indeed select the best 20 entries for the pedagogical purpose. The idea was for learners to infer from these lines a patterning which would give them insight into the language they were learning. Sources were usually study materials in the student's field.

In 1986, Johns published a description of MicroConcord, calling it a research tool for learners. (This is the program which we eventually published together with OUP in 1993.) What he shows and discusses there is the use of sets of 20 concordance lines with students. With 20 lines one begins to spot patternings, reading "vertically", seeing how a node term is used repeatedly with certain kinds of collocate.

> By concentrating and making it easy to compare the contexts within which a particular item occurs, it organizes data in a way that encourages and facilitates inference and generalization. Such generalizations may leap out of the contexts in an obvious fashion (for example, the word before 'same' is almost always 'the'), or may require a good deal more work on the part of the user. (Johns, 1986, p. 160)

This procedure he contrasts with much language teaching of the time and nearly all early CALL materials: a "conventional constructed exercise in which the learner is searching for a single 'correct' answer fixed in advance" (Johns, 1986, p. 161). With his students he discovered that studying patterns of colligation using a concordancer was "far more helpful than the usual exercise involving 'filling in the missing prepositions'" (pp. 161–162).

Team Teaching

This view of the learner as a potential researcher was indubitably strengthened by Johns' work with postgraduate learners in Birmingham. He worked extensively with English of Academic Purposes students, who typically knew their field already to degree level but needed help especially with academic writing of term papers and theses. He practised (and told us enthusiastically about his experience) team-teaching, where he and the subject-specialist would work together, liaising to ensure face validity of the materials for students and ease of inference because the field was generally familiar.

DDL Perspective

Discussing the concordance:

> Without questions given in advance, it leads the learner to generate his or her own questions, and to test them out against the evidence. There is at least a prima facie case for thinking that the early exposure to authentic text and the skills of observation and inferencing developed in working with concordance output may be transferred to language learning away from the computer and outside the classroom: this is one of the many aspects of the approach that merit further investigation. (Johns, 1986, p. 160)

DDL is often not a solitary matter of the student with their corpus and concordance, as it involves one-to-one consultation with the teacher.

> To employ a fashionable distinction, what is important in the one-to-one consultations is the process of revision – the finding out, the testing of various possibilities, and not the product of revision – i.e. the maximum amount of 'corrected' text. There does, in fact, seem to be an inverse relationship between the amount of text revised and the week-by-week progress made by the student: it is difficult, if not impossible, for the average student to learn much from large quantities of corrected text. (Johns, 1996, Kibbitzer 3)

Of course in this consultation process, one of the delights for the teacher is finding an aspect of patterning that they didn't know about before. It's data-driven learning for both parties.

Cross-References

- ▶ Inductive and Deductive Approaches to Data-Driven Learning
- ▶ Notable Influential Researchers in Data-Driven Learning
- ▶ Notable Publications in Data-Driven Learning
- ▶ The History of Data-Driven Learning

References

Crosthwaite, P. (2024). Data-driven learning. CorpusCast hosted by Robbie Love (Aston). https://www.youtube.com/watch?v=WiYZi8mnHDA&ab_channel=Aston Originals

Higgins, J. (2009). *Tribute to Tim Johns*. Unpublished text for Tim Johns Memorial held at Birmingham University, September 2009.

Higgins, J., & Johns, T. F. (1984). *Computers in language learning*. Collins.

Johns, T. F. (1981). The uses of an analytic generator: The computer as teacher of English for specific purposes. British Council: *ELT* Documents 112.

Johns, T. F. (1986). Micro-Concord: A language learner's research tool. *System, 14*(2), 151–162. https://doi.org/10.1016/0346-251X(86)90004-7

Johns, T. F. (1991a). Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal, 4*, 1–16.

Johns, T. F. (1991b). Data-driven learning and the revival of grammar. *Eurocall* (pre-conference booklet, pp. 20–21). 21–23 August, Finland.

Johns, T. F. (1991c). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *ELR Journal, 4*, 27–45.

Johns, T. F. (1994). *Contexts.* Unpublished but widely distributed free CALL program running under DOS. On floppy disk. Revived and usable at https://vocab.today/history/contexts.

Johns, T. F. (1996). *Contexts 2.* Free CALL program running under DOS, 1994. Revived and usable (Windows) at https://vocab.today/history/contexts2

Johns, T. F. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100–115). Addison Wesley Longman.

Johns, T. F. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107–117). Rodopi. https://doi.org/10.1163/9789004334236_010

Johns, T. F. (n.d.). Kibbitzers. https://www.lexically.net/TimJohns/index.htm

Johns, T.F., & King, P. (Eds.) (1991). Classroom concordancing. ELR Journal, 4, 178.

O'Keeffe, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching, 54*, 259–272. https://doi.org/10.1017/S0261444820000245

Scott, M. (2009). In memory of Tim Johns. *International Journal of Corpus Linguistics, 14*(3), 271–273. https://doi.org/10.1075/ijcl.14.3.01sco

Scott, M., & Johns, T. (1993). *MicroConcord*. Oxford University Press. ISBN 0-19-458322-8.

N

New Literacies: A Historical Perspective

Richard Kern
University of California, Berkeley, CA, USA

Abstract

The notion of New Literacies (and the interdisciplinary field of New Literacy Studies) developed in the 1980s and 1990s in reaction to cognitive skill-based models of reading and writing that focused largely on language processing and knowledge development at the level of the individual. Literacy increasingly came to be understood not just in terms of cognitive processes but also as a dynamic set of social practices that vary across contexts and communities and that are imbued with cultural values (with this plurality of forms, functions, and goals reflected in the term *literacies*). At the same time, there was growing recognition of how non-linguistic dimensions of texts contribute to the making of meaning. This set the stage for discourse-oriented research on the Internet, digital texts, hypertext, and multimodal electronic communication—which revealed new demands on people's ability to produce and interpret texts and created new imperatives for literacy education. This entry traces approaches to literacy in language education from Michael Halliday's notion of language as social semiotic, Brian Street's distinction between ideological and autonomous models of literacy, James Gee's work on ideology in discourse, and Gunther Kress's work on multimodality to the New London Group's 1996 multiliteracies manifesto and its impact over the past 25 years, especially in the area of second/foreign language education. It concludes with a discussion of evolving new literacies related to the introduction of generative artificial intelligence applications.

Keywords

New literacies · Multimodality · Discourse · Semiotics · Cognitive processes · Mediation · Artificial intelligence · New London Group · Digital literacies

Definition

New Literacies are the wide-ranging social practices associated with composing, interpreting, and talking about texts in new media (digitally-mediated) environments. They represent a shift away from cognitivist and universalist understandings of literacy towards recognition of the multifarious and socioculturally situated ways people create and interpret texts for various purposes in a range of different mediums and environments (and are therefore essential for language learners). They address the broader palette of semiotic sensitivities needed to deal with

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_241-1

multimodal communication that relies on non-linguistic as well as linguistic dimensions of meaning.

Introduction

Literacy involves familiarity with the conventions of texts and the ability to use those conventions to understand and produce texts. Historically, technologies of writing (from clay tablets to hieroglyphs to paper, print, and computers) have influenced how we define literacy (Kern, 2015). Up through the 1980s, the dominant paradigm was cognitive and process oriented, emphasising cognitive/metacognitive skills and strategies that focused largely on language processing and knowledge activation at the level of the individual reader/writer. However, the advent of ethnographic studies of literacy in various cultures, increasing emphasis on social theories of learning, and the development of personal computers and network-based communication technologies in the 1980s gradually led to expanded notions of literacy ("new literacies") that encompassed the production and interpretation of multimodal representations of information (▶ "Multimodal Literacy"), new genres, and new sets of beliefs and customs related to communication in electronic environments.

Historical Overview

The groundwork for rethinking literacy was already being established as early as the 1970s. In the light of societal change brought about by an explosion in information and the modes in which it was presented, Gordon (1971) revisited the fundamentals of mediation, perception, and communication, and Foster (1979) urged a visual literacy curriculum in schools, both authors promoting the term "new literacy". In the field of linguistics, Michael Halliday's notion of language as social semiotic laid the groundwork for a perspective of language and literacy situated in a sociocultural context "in which the culture itself is interpreted in semiotic terms—as an information system" (Halliday, 1978, p. 2). Language offered "an open-ended range of semantic choices" and material and social contexts influenced what specific choices were made in speech acts. Context was important for Halliday not only because it shaped what people said, but also because it was itself *shaped by* what people said as they acted out social roles and statuses, and communicated not just information but also systems of values, knowledge, and ideologies (p. 2).

Meanwhile, a similar social turn was happening in literacy studies. In their study of Vai literacy in Liberia, Scribner and Cole (1981) characterised literacy as a social practice that involved "not just simply knowing how to read and write a particular script but applying this knowledge for specific purposes in specific contexts of use" (p. 236). Brian Street (1984) conceptualised a distinction between "autonomous" and "ideological" models of literacy. The autonomous model, which had dominated literacy research, isolates literacy as an independent variable that can be detached from social context and studied in terms of its consequences for individual cognition, social mobility, and societal progress. It is an "essay-text" brand of literacy associated with schooling that Street qualified as "a narrow, culture-specific literacy practice" (p. 1) whose beneficial effects tend to be broadly overgeneralised. In contrast, the ideological model of literacy recognises the culturally embedded nature of all literacy practices and highlights the importance of the processes by which people are socialised into them. It is a critical model that analyses literacy teaching as a form of social control and hegemony of a ruling class (pp. 2–3). With it, Street introduced the idea that it is more appropriate to talk about "literacies" in the plural because practices are not universal but particular to cultures and social institutions. Pluralising literacy also raised consciousness of the need to ask "'whose literacies' are dominant and whose are marginalised or resistant" (Street, 2003, p. 77). Like Street, James Gee (1990) emphasised the sociocultural embeddedness of "new literacy" in terms of what he called capital-D Discourses, which he defined as:

> a socially accepted association among ways of using language, of thinking, feeling, believing, valuing, and of acting that can be used to identify oneself as a member of a socially meaningful group or 'social network', or to signal (that one is playing) a socially meaningful 'role'. (p. 143)

Gee stressed that the focus of literacy studies must be on social practices (p. 137). Harking back to Halliday, just as literacy practices shape how people read and write, they are themselves shaped by broader systems of values, knowledge, and ideologies.

In 1994, James Gee along with Courtney Cazden, Bill Cope, Norman Fairclough, Mary Kalantzis, Gunther Kress, Allan Luke, Carmen Luke, Sarah Michaels, and Martin Nakata met in New London, New Hampshire (USA), to discuss new paradigms of literacy education. The manifesto they published in the *Harvard Educational Review* (New London Group, 1996) introduced the notion of ▶ "multiliteracies" and remains a foundational publication in New Literacy Studies. The New London Group broadened the notion of literacy to include the negotiation of a "multiplicity of discourses" (p. 61), referring back to Gee's notion of Discourses, and they posited two aspects of this multiplicity. The first was related to linguistic and cultural diversity and the second was related to "the burgeoning variety of text forms associated with information and multimedia technologies" (p. 61). Cope and Kalantzis have since updated the theory and practice of multiliteracies in a series of publications (Cope & Kalantzis, 2000, 2009; Kalantzis et al., 2016).

Gunther Kress, having long promoted social theories of language and literacy (Kress, 1989), contributed significantly to the New London Group's theorisation of multimodality (Kress, 2003, 2010). For Kress, communicating via images, video, music, sound, or writing involved distinct logics and distinct affordances. Because meaning-making increasingly drew on modes other than speech and writing (e.g. gesture, colour, moving and still images, 3-D objects, sound and music), semiotics provided a more adequate theoretical base than linguistics for the study of new literacies. Key concepts in Kress's semiotic theory of multimodality were *transduction* (reshaping semiotic material from one mode to another, such as from speech to image or from writing to film) and *transformation* (re-ordering the elements of a text or other semiotic object, but all within the same mode). Both of these processes were related to Kress's concept of *design*, which has to do with individuals' and communities' agency in shaping communication, conventions, and the social and semiotic world. When people make meaning, they make design choices according to the environment and function of the communication, the identity and social role of participants, their power relations, the desired rhetorical effects, and the semiotic resources available to them.

The New London Group formalised this notion of design by positing three constituent elements: *Available Designs* (grammars of languages and other semiotic systems and conventions of their use in various contexts), *Designing* (the creative meaning-making process involving Available Designs and their re-presentation and recontextualisation), and *the Redesigned* (new meaning, yielding Available Designs transformed through use, which in turn become new Available Designs). In this framework, then, every act of designing (e.g. reading, writing, speaking, listening, creating) is an active, dynamic process of using existing resources in fresh ways and in that process establishing new resources for future meaning-making.

L2 Learning

In the context of second/foreign language education, the notion of multiliteracies has been taken up in both theory and practice and has become a major instructional framework. Janet Swaffar, Katherine Arens, and Heidi Byrnes (1991) anticipated new literacy approaches by shifting the focus of instruction away from sentence-level grammar and information retrieval to making connections between details of texts, cultural notions, and pragmatic implications. Richard Kern (2000) drew on work by the New London Group and others to develop a literacy-based heuristic framework to integrate the teaching of critical reading

and writing with face-to-face verbal interaction. Heidi Byrnes (Byrnes & Kord, 2002; Byrnes & Maxim, 2004; Pfeiffer & Byrnes, 2009) designed a 4-year content-oriented, genre- and task-based German curriculum entitled "Developing Multiple Literacies" that gained international recognition. Kate Paesani, Heather Allen, and Béatrice Dupuy (2016) developed a widely influential multiliteracies framework for language teaching that includes a systematic scope and sequence as well as a wealth of practical activities.

Digital Technologies

How "new literacies" have been understood with respect to digital technologies has evolved dramatically over the past 30 years. What began as knowing how to access web pages or follow hyperlinks evolved into knowing how to create personal websites, use social media platforms, create digital stories, participate in fan fiction, play multi-user virtual reality games, and contribute content to Wikipedia, YouTube, Instagram, and TikTok, among many other practices. "Writing" in these multimodal environments is often reconceived as "authoring" and sometimes involves programming skills. Of course, the definition of being digitally "literate" is not uniform and monolithic but highly individualised and tied to a person's particular online practices. For example, sharing and commenting on images in X or Instagram involves a different set of knowledge, skills, and practices than producing a video for TikTok or designing a personal website.

Media pundit Marshall McLuhan popularised the idea that we shape our technological tools and then they shape us (Culkin, 1967, p. 70). From an educational viewpoint, if our literacy practices are being shaped by new technologies, then it is important to understand each form of technology in terms of its dominant metaphors and ideological assumptions—how it wants us to "be" in relation to it—as well as the technical nuts and bolts of how to access and use it and the various communicative genres associated with it.

Recently, generative artificial intelligence (GenAI) applications such as ChatGPT have introduced new challenges related to digital literacies (▶ "AI Literacy Practices"). GenAI systems create significant epistemic and pedagogical problems that call for a rethinking of goals and methods of language teaching. New London Group members Bill Cope and Mary Kalantzis (Peters et al., 2024) argue that Large Language Models combined with chatbots are harmful to a social understanding of knowledge and learning in that they obscure their sources of information, they have no notion of empirical truth and often present fabricated information, they have no conception of a theoretical frame or disciplinary practice, they are indifferent to ethical concerns, and they cannot engage in critical dialogue with their human users. Nevertheless, Cope and Kalantzis recognise the potential that GenAI has for education and they suggest ways of "recalibrating" GenAI systems by prompting them with epistemic, empirical, and ontological frames to make them more useful for learning.

Conclusion

New Literacies highlight the educational imperative to foster sophisticated abilities to manipulate and interpret a broad range of symbol systems—linguistic and nonlinguistic—and to promote critical reflection on the tools and cultural practices we engage with when we make and take meaning. Over the last 50 years, the notion of New Literacies has evolved with societal changes such as globalisation, technological changes that have engendered new social practices, and academic disciplinary changes that have increasingly come to view cognition as not only mental but also embodied and integrated with social practices. The notion will undoubtedly continue to evolve with future societal and social changes and shifting disciplinary perspectives.

Cross-References

▶ AI Literacy Practices
▶ Multiliteracies
▶ Multimodal Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Byrnes, H., & Kord, S. (2002). Developing literacy and literary competence: Challenges for foreign language departments. In V. M. Scott & H. Tucker (Eds.), *SLA and the literature classroom: Fostering dialogues* (pp. 35–74). Heinle & Heinle.

Byrnes, H., & Maxim, H. (Eds.). (2004). *Advanced foreign language learning: A challenge to college programs*. Heinle & Heinle.

Cope, B., & Kalantzis, M. (Eds.). (2000). *Multiliteracies: Literacy learning and the design of social futures*. Routledge.

Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies, 4*(3), 164–195.

Culkin, J. M. (1967). A Schoolman's guide to Marshall McLuhan. *The Saturday Review, 50*, 51–53; 70–72.

Foster, H. M. (1979). *The new literacy: The language of film and television*. National Council of Teachers of English.

Gee, J. P. (1990). *Social linguistics and literacies: Ideology in discourses*. Falmer Press.

Gordon, D. R. (1971). *The new literacy* (1st ed.). University of Toronto Press. https://doi.org/10.3138/9781487571764

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. University Park Press.

Kalantzis, M., Cope, B., Chan, E., & Dalley-Trim, L. (2016). *Literacies* (2nd ed.). Cambridge University Press.

Kern, R. (2000). *Literacy and language teaching*. Oxford University Press.

Kern, R. (2015). *Language, literacy, and technology*. Cambridge University Press.

Kress, G. (1989). *Linguistic processes in sociocultural practice*. Oxford University Press.

Kress, G. (2003). *Literacy in the new media age*. Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–92.

Paesani, K., Allen, H. W., & Dupuy, B. (2016). *A multiliteracies framework for collegiate foreign language teaching*. Prentice Hall.

Peters, M. A., Jackson, L., Papastephanou, M., Jandrić, P., Lazaroiu, G., Evers, C. W., Cope, B., Kalantzis, M., Araya, D., Tesar, M., Mika, C., Chen, L., Wang, C., Sturm, S., Rider, S., & Fuller, S. (2024). AI and the future of humanity: ChatGPT-4, philosophy and education – Critical responses. *Educational Philosophy and Theory, 56*(9), 828–862. https://doi.org/10.1080/00131857.2023.2213437

Pfeiffer, P. C., & Byrnes, H. (2009). Learning, and identity of majors: A case study of program outcomes evaluation. In J. Norris, J. M. Davis, C. Sinicrope, & Y. Watanabe (Eds.), *Toward useful program evaluation in college foreign language education* (pp. 183–208). University of Hawai'i Press.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Harvard University Press.

Street, B. V. (1984). *Literacy in theory and practice*. Cambridge University Press.

Street, B. (2003). What's "new" in new literacy studies? Critical approaches to literacy in theory and practice. *Current Issues in Comparative Education, 5*(2), 77–91.

Swaffar, J. K., Arens, K. M., & Byrnes, H. (1991). *Reading for meaning: An integrated approach to language learning*. Prentice Hall.

P

Pluriliteracies

Fanny Meunier
University of Louvain (UCLouvain), Ottignies-Louvain-la-Neuve, Belgium

Abstract

The concept of pluriliteracies (commonly used in the plural) builds on earlier work in biliteracy, new literacies, multilingual and multimodal literacies, and plurilingualism. It emphasises the hybridity of literacy practices in situated sociocultural contexts and highlights the need for merging (rather than separating) literacy practices between languages. The first part of the entry presents the key features of a pluriliteracies approach, including the continuous interplay of multiple languages, scripts, discourses, dialects, registers, and technologies. It also explores the efforts to translate theoretical insights into major pedagogical innovations and changes, stressing the role of pluriliteracies in critical pedagogies that promote students' and teachers' cognitive, emotional, and/or behavioural engagement and agency. The second part provides concrete examples of how pluriliteracies can be used to promote deeper learning (e.g. plurilingual focus on disciplinary literacies, enhancement of textual fluency). The final section emphasises the role of pluriliteracies in supporting students and teachers with the competencies needed to become responsible global citizens.

Keywords

Pluriliteracies · Plurilingualism · Hybridity · Sociocultural contexts · Critical pedagogies · Engagement · Deep learning · Global citizenship · Digital literacies

Definition

Pluriliteracies can be defined as approaches that recognise and promote the hybrid nature of literacy practices rooted in situated contexts. These approaches advocate for the integration of plurilingual and multimodal literacy practices across various disciplines, highlighting the continuous interplay of multiple languages, scripts, discourses, dialects, and registers. Pluriliteracies go hand in hand with pedagogical innovations that enhance the cognitive, emotional, and behavioural engagement of both students and teachers to equip them to actively contribute to responsible global citizenship.

Introduction: The Origins of Pluriliteracies: From *Multi-* to *Pluri-*

It is not always easy to distinguish between terms that appear similar—such as multiliteracies and

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_244-1

pluriliteracies—especially when both prefixes relate to the idea of 'many'. One way to grasp the difference is to remember that *multi-* emphasises the number of separate elements, while *pluri-* emphasises the interconnectedness among those elements. The prefix pluri- has been adopted in various fields to highlight the dynamic and integrated nature of concepts and has increasingly been used to replace prefixes like 'bi-' or 'multi-'. For instance, bilingualism refers to the ability of an individual to use two languages rather proficiently, and multilingualism refers to the ability of an individual to communicate effectively in more than two languages. As for plurilingualism, it refers not just to the ability of an individual to use multiple languages, but more importantly, to interrelate and interconnect these languages as part of a unified communicative repertoire (Council of Europe, 2020).

Similarly, the term pluriliteracies, introduced by García et al. in 2007, builds on earlier concepts such as biliteracy, multiliteracies (see ▶ "Multiliteracies"), new literacies (see ▶ "History of New Literacies"), multi- and plurilingualism, multimodal literacies (see ▶ "Multimodal Literacy"), and critical literacies (see ▶ "Critical Digital Literacy"). The term highlights the importance of having numerous, hybrid, dynamic, and adaptive literacies as tools to navigate an increasingly complex social, cultural, and technological era.

Being pluriliterate today thus means being able to juggle numerous literacy types in a fluid way, depending on the context, situation, and needs. It goes beyond producing or understanding multimodal communication; it also implies being critical—that is, questioning, challenging, or transcending power structures, social norms, and ideologies that are embedded in communication and in the languages and tools used to communicate.

Given the number and types of literacies that have emerged (due in part to the rapid technological advancements and the immense volume of information individuals must deal with to date), supporting students on the path to developing pluriliteracies should become an essential part of teachers' agendas throughout the educational curriculum. Although the term pluriliteracies remains relatively new and is seldom found in today's curricula, its emerging importance suggests that we should definitely give it more attention.

Pioneering Pluriliteracies: Insights from CLIL Research and Practice

A group of researchers known as the Graz Group (including among others Oliver Meyer, Do Coyle, Ana Halbach, Kevin Schuck, and Teresa Ting) and who are working for the European Center for Modern Languages at Graz have proposed a framework called Pluriliteracies Teaching for Learning (or PTL). PTL puts subject literacy development in more than one language at the core of learning. This approach focuses on helping learners become literate in content subjects and also communicate that knowledge across cultures and languages (see, e.g. Meyer & Coyle, 2017). The members of the Graz Group stress the fact that language is the key to developing and increasing conceptual understanding and that the focus on language—and plurilingualism—will ultimately lead to deeper learning. Acquiring subject literacy in more than one language helps learners construct knowledge, leads to shared learning and interactions with others, and contributes to developing creative, responsible global citizens (see, e.g. Coyle & Meyer, 2021, 2023). The Group has also worked on helping teachers put pluriliteracies into practice. *A deeper learning companion for CLIL: Putting pluriliteracies into practice* (Coyle et al., 2023) offers concrete examples and lesson plans for 'deeper learning episodes' in disciplines such as chemistry, physics, or geography. All the examples are illustrative of the three following key principles: (1) an explicit focus on disciplinary literacies in all subjects of schooling (focusing on subject-specific ways of constructing and communicating knowledge), (2) pluriliterate language use (communicating knowledge across cultures and languages), and (3) textual fluency (being able to communicate across plurimodal channels and semiotic systems and being able to critically evaluate sources).

Navigating a Multiliterate, Multicultural, Multilingual, and Hypertechnological Era: Supporting Pluriliteracies

Single narratives—be they linguistic, cultural, disciplinary, methodological, technological, etc.—create blind spots. They limit our understanding, hinder innovation, and narrow the path to inclusion. And despite pluriliteracies being a rather complex construct, it may be considered an imperative one to prioritise in educational reform efforts. It is essential for individuals to 'make meaning' across multiple languages, cultures, and modes of communication. Digitisation today enables plurimodal and pluriliterate learning and invites us to reconsider our learning spaces and biases. Larsen-freeman (2018) invites us, for instance, to see classrooms as porous. Our learning spaces are welcoming many different types of learners (school children, university students, asylum seekers, seniors, refugees, distance learners, etc.) with varied linguistic backgrounds and learning goals. This diversity reflects the realities of current multilingual societies. It highlights the need for inclusive and adaptive approaches that can also be used, for instance, to support refugee education, to help heritage speakers, and to contribute to the revitalisation of indigenous languages. The boundaries between classroom and society are porous and, as suggested at the start, this entails the following needs: (1) to juggle numerous literacy types in a fluid way, depending on the context, situation, and needs; (2) to be critical and challenge power structures, norms, and ideologies; and (3) to transcend monolingual biases. Many people today grow up and use multiple languages for a variety of purposes. In addition, numerous educational and language policies invite us to promote and implement linguistic and cultural diversity, to be more equitable and inclusive and to integrate technologies in a human-centred way that helps safeguard human cognition (see, e.g. Rehm & Way, 2023; UNESCO, 2023a, b). Fostering pluriliteracies is one promising path to meeting these key recommendations.

To concretely illustrate how pluriliteracies can take shape or be trained, imagine a classroom where students are working on a project that involves exploring environmental issues across different regions of the world. Students can begin by reading an article on local environmental initiatives (in the language of schooling). They can then be guided by the teacher to identify key ideas and terms, linking these to their knowledge of other contexts (other environmental issues, other key terms related to these issues, other texts, or testimonies in other languages). In doing so, the teacher can highlight similarities and differences between contexts, cultures, and languages by encouraging students to recognise those similarities and differences in terminology, concepts, problems, and initiatives that are taking shape worldwide. Students can then be invited to work collaboratively on a poster, a blog, or a video to share with peers from different linguistic backgrounds. This task would call for linguistic mediation (reformulating, paraphrasing, illustrating complex ideas to make the content accessible and relevant to others). This could be followed by a plurilingual-friendly discussion about how environmental practices differ across cultures, drawing from their own linguistic and cultural backgrounds (see ▶ "Intercultural Literacy"). The discussion would foster social and cultural mediation, helping to bridge gaps in understanding between peers, irrespective of their origin. Students could also synthesise their findings into a multilingual digital presentation to raise awareness of environmental issues in their community—and beyond—thereby making students active global citizens. This final production would require mediation between modes (e.g. written texts, visuals, speech, and all this in different languages) and also require cognitive flexibility (clarifying ideas, adapting them for different audiences).

In all the tasks listed above, a range of generative AI technologies could also be used, when relevant. For instance, *ChatGPT* can provide suggestions for organising a digital presentation. It can also translate a text excerpt into another language or simplify the content of a technically complex section to make it more accessible to less specialised audiences. *Copilot* could be used

to summarise key ideas in language A from a text written in language B. Similarly, an AI subtitle generator like *Happy Scribe* can create a first draft of subtitles in various languages, which can later be proofread and edited as needed to ensure accuracy. *DALL-E* could generate images tailored to the specific needs of the presentation. In all cases, however, it is crucial to critically assess the results generated by these AI tools to ensure quality and appropriateness.

For their final presentation, learners could also be encouraged to analyse and critique possible inequality, injustice, and power dynamics in the issues addressed and sources covered; acknowledge how social, historical, and cultural contexts have played a role in environmental issues; and advocate for change.

The pluriliteracies example suggested above could be implemented in a disciplinary context (science class, or language class), could be co-supervised by teachers from different disciplines, or could be carried out in community contexts other than classrooms (local community event, intergenerational project, digital or virtual communities, etc.). The activities and tasks proposed would not only develop participants' pluriliteracies by enabling them to read, write, and mediate across languages and contexts. They also align with modern pluralistic approaches recommended notably in Europe and actively train mediation, one of the four modes in which the CEFR model organises communication (see Piccardo, 2023). Participants are seen as true social agents who engage in a combination of receptive, productive, interactive, and mediation activities to co-construct meaning/knowledge and communicate beyond linguistic or cultural barriers. Various literacy types are intertwined and agency is activated at every stage (be it individual agency, collective agency or by-proxy agency —i.e. when relying on human or technological help or to complete the task).

Knowledge building and sharing today often requires mediation among diverse actors, sources, tools, and spaces—including digital spaces—to facilitate collaboration, integrate multiple perspectives, and co-construct understanding across contexts. These new ways of learning and teaching naturally foster pluriliteracies: users are being exposed to multilingual and multimodal content and are given opportunities to interact and collaborate in spaces that blend text, visuals, and audio and that integrate technological affordances.

In a way, pluriliteracies both reflect the complexities of globalised communication and knowledge and serve as essential pathways for accessing and contributing to it. Pluriliteracies can be challenging to navigate because they require adapting to constantly evolving languages, tools, and cultural contexts, as well as integrating multiple modes of communication and perspectives. They are nonetheless essential for fostering deeper learning and developing richer, more nuanced knowledge and communication, making them worth actively exploring in education.

Cross-References

- ▶ Critical Digital Literacy
- ▶ History of New Literacies
- ▶ Intercultural Literacy
- ▶ Multiliteracies
- ▶ Multimodal Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume*. Council of Europe Publishing. https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-companion-volume

Coyle, D., & Meyer, O. (2021). *Beyond CLIL: Pluriliteracies teaching for deeper learning*. Cambridge University Press.

Coyle, D., & Meyer, O. (2023). An introduction to the special issue 'towards pluriliteracies approaches to language teaching and learning: Theories and principles for transformative practices'. *English Teaching & Learning, 47*, 263–270. https://doi.org/10.1007/s42321-023-00157-6

Coyle, D., Meyer, O., & Staschen-Dielmann, S. (2023). *A deeper learning companion for CLIL: Putting pluriliteracies into practice*. Cambridge University Press.

García, O., Bartlett, L., & Kleifgen, J. A. (2007). From biliteracy to pluriliteracies, heritage language maintenance 184. In P. Auer & L. Wei (Eds.), *Handbook of applied linguistics* (Vol. 5, pp. 207–228). Mouton de Gruyter.

Larsen-freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals, 51*(1), 55–72

Meyer, O., & Coyle, D. (2017). Pluriliteracies teaching for learning: Conceptualizing progression for deeper learning in literacies development. *European Journal of Applied Linguistics, 5*(2), 199–222. https://doi.org/10.1515/eujal-2017-006

Piccardo, E. (2023). Mediation and the plurilingual/pluricultural dimension in language education. *Italiano LinguaDue, 14*(2), 24–45. https://doi.org/10.54103/2037-3597/19568

Rehm, G., & Way, A. (2023). *European language equality: A strategic agenda for digital language equality*. Springer International Publishing. https://library.oapen.org/handle/20.500.12657/63568

UNESCO. (2023a). *Global education monitoring report, 2023: Technology in education: A tool on whose terms?* https://doi.org/10.54676/UZQV8501

UNESCO. (2023b). *Guidance for generative AI in education and research*. https://doi.org/10.54675/EWZM9535

Z

Zero Literacy

Nicky Hockly
The Consultants-E, Swansea, UK

Abstract

The development of teachers' and learners' digital literacies is a goal in many educational curricula, including in language teaching and learning. However, access to hardware, data and reliable internet connections remains challenging for a significant proportion of language learners and practitioners, who are often excluded from computer-assisted language learning (CALL) studies. Zero literacy recognises users whose lack of access to digital technologies has excluded them from opportunities to develop the preliminary skills necessary for digital engagement, and this exclusion is frequently symptomatic of wider systemic inequalities. Reflected in UNESCO's Digital Literacy Global Framework as a level '0. Devices and software operations', zero literacy involves basic operational and technological abilities required for digital communication, such as turning on a computer, typing or using a mouse. These skills are fundamental yet cannot be presumed universally. Zero literacy precedes more advanced digital literacy levels, focusing on the initial, essential steps needed to interact with digital devices.

Keywords

Zero Literacy · Technological Skills · Operational Skills · Global South · Digital Exclusion · Design Justice · Design-based Research · Digital Literacies

Definition

Zero literacy refers to the basic operational and technological skills needed by learners and teachers, including language learners and teachers, for engaging with digital technologies, and forming the foundation for all other levels of digital literacy. Zero literacy is often neglected in CALL research agendas.

Introduction

The need for learners of all ages to develop a range of ▶ digital literacies is axiomatic, and reflected in the multitude of frameworks designed for this purpose. However, despite widespread acceptance of digital literacies as a key twenty-first- century skill set, and the increasing ubiquity of (particularly mobile) digital technologies, access to and experience with these technologies remains uneven.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_245-1

Digital Exclusion

A lack of opportunities to interact with digital technologies is frequently framed in terms of digital exclusion, which can exacerbate existing inequalities, with learners unable to develop their 'digital capital' (Ragnedda & Ruiu, 2020). Digital exclusion—reflected in the widely referenced older term 'digital divide'—can be conceptualised on three interlinked levels (Ragnedda & Ruiu, 2017). Firstly, learners need *access* to digital devices and internet connectivity. Secondly, they need to know how to *use* these devices and online resources, both from an operational perspective (e.g., knowing how to turn on a device, how to connect to the internet, and how to use a browser) and from a digital literacies perspective (e.g., knowing how to search for trustworthy information online, or how to engage with social media effectively). Finally, from their digital interactions, learning and experiences, learners can *derive benefits* (e.g., increased social capital, or improved employment opportunities that impact on financial capital). Exclusion can exist at any of these three interlinked levels, with obvious negative consequences for excluded learners at each stage.

Researchers have suggested for some time that digital exclusion can be reframed as "adverse incorporation", which exists "when powerful, connected people command resources from which they draw significantly increased returns by coordinating the effort of outsiders whom they exclude from the full value added by that effort" (Phillips, 2013, p. 176, citing Tilly, 1998, p. 10). Essentially, adverse incorporation underpins unequal digital access and use, and serves to maintain and exacerbate existing systemic inequalities; it is therefore helpful to understand digital exclusion as a 'digital justice' issue (Heeks & Shekhar, 2019).

Much CALL research is concerned with the second level (i.e., use) of digital literacies described above, with the access and skills implicit in the first level taken for granted. Those learners who either have no access to digital technologies, or have very limited or non-existent technological and/or operational skills, are often ignored when it comes to research and practice in the field of language teaching. Operational skills can be characterised as the surface-level or basic skills required to use specific devices and software (e.g., turning a device on or off), while technological skills suggest an understanding of how technology systems work (e.g., understanding what a browser does); operational skills are typically device or software-specific, whereas technological skills are more broadly applicable across different platforms and technologies. A lack of these skills, then, can lead to digital exclusion. The reasons for digital exclusion are complex, and, as suggested by the concept of adverse incorporation explored above, frequently indicative of wider systemic inequalities; for example, although often associated with learners in underprivileged and/or isolated communities in the Global South, the digitally excluded can also include immigrant and refugee populations (Pegrum, 2025) as well as older citizens (Ngiam et al., 2022) or underserved minority populations with lower socioeconomic status in the Global North (Helsper & Reisdorf, 2017; Heponiemi et al., 2023). As an example of the latter, a US study that aimed to develop older African American adults' digital literacies, found that even regular computer users "lacked knowledge and perspectives on basic issues and functions of the computer. [. . .] a few participants said they had never used a computer before" (Seo et al., 2019, p. 141).

Clearly then, digitally excluded learners are a heterogeneous group, and their individual levels of technological competence, access to digital devices and digital experiences, and levels of digital literacies—or lack of these—are equally multifaceted and varied. Considering these learners as individuals who have "super jagged [digital] literacy profiles", that is, a complex assemblage of skills and knowledge based on their varied life experiences (including their experiences with technologies) is a useful construct here; a nuanced understanding of the complexity and uniqueness of learners' digital literacy profiles enables educators to recognise both areas for improvement and

existing strengths that can be built upon for digital empowerment (Pegrum, 2025).

UNESCO's Digital Literacy Global Framework (Law et al., 2018) highlights the importance of including zero ('0') literacy in descriptions and subsequent evaluations of digital literacies, defining zero literacy as referring to 'devices and software operations', and providing the following examples among others:

- "turning on/off device, charging"
- "login, user account & password management, locking devices"
- "searching for, understanding & approving privacy policy statements" (p. 59)

Pegrum (2025, with ref. to Hockly, 2022) defines zero literacy as "the functional or operational literacy, or even pre-literacy, skills needed to open the door to the universe of digital communications: examples include the ability to switch on a computer, type on a keyboard, move a mouse, connect to the internet, or open a browser…", and points out that zero literacy "must precede all other digital literacies, and must therefore be in place before … learners … can proceed to develop more complex skills".

From the language teacher's perspective, helping learners develop specific zero literacy skills will depend on their context and needs. The zero literacies needed to operate basic mobile 'feature phones', smartphones or computers vary, and although the use of mobile devices for daily tasks is widespread, including in low-resource contexts, the use of computers is frequently associated with workplace skills, "often in the context of white-collar jobs" (Law et al., 2018, p.18). Exploring what devices learners will use, and for what purposes, is therefore an important prerequisite in supporting learners' development of zero literacy.

Implications for educators and learners:

- **Conduct a needs analysis.** Start by exploring what operational and technological skills your learners already have, and what skills they need. If your learners need computer skills to increase their employment opportunities, this will be an important area to address. A needs analysis can be conducted as a survey or by simply asking learners.
- **Provide plenty of practice.** To develop basic skills with computers, learners need access to the relevant hardware (e.g., monitor, mouse and keyboard). First steps then involve switching a computer on and off, moving the mouse, opening a browser, and typing. For learners completely new to the computer interface, lots of practice with these basic operations needs to be provided before more complex operations are undertaken. A 'computer dictation' activity that gives learners this practice might consist of the teacher verbally giving a series of repeated commands that learners need to follow ('Turn the computer on', 'Move the cursor to the top right of the screen', 'Turn the computer off', 'Turn the computer on', 'Open the browser', 'Move the cursor to the bottom left of the screen', 'Close the browser'…). As learners become more familiar with the operations, the teacher increases the speed of the commands to create a sense of fun. The same activity, which combines language practice with the key terminology and structures needed to discuss digital operations, can be repeated over several classes, and until learners have mastered these basic operational skills.
- **Address the skills gaps.** Your learners may already have strong operational skills for their feature phones or smartphones, but they may have gaps in their technological skills as regards language learning—for example, they may be unaware of language learning opportunities (such as apps) that can be accessed via their mobile devices. Suggesting contextually relevant language learning apps (which might include translation apps), and showing learners how to use them for independent study, is therefore one strategy educators can implement.
- **Integrate skills into language learning tasks**. Make the development of zero literacy skills part of your language class, for example by pairing learners with stronger operational skills

with learners with less developed skills, and asking them to complete a task together in the target language. Taking the example above, this might involve downloading and installing a language learning app onto both learners' devices, with the more skilled learner talking the less skilled learner through the process; the pair then work together to complete a few tasks in the app itself. In this way, language learning and digital literacy development can be combined in a simple classroom task.

Future CALL research might usefully address the gap in the literature around zero literacy, both in low-resource contexts, and in underserved populations in medium- and high-resource contexts. Addressing the implications of adverse incorporation, for example by drawing on approaches rooted in design justice and design-based research, is one promising area for both researchers and educators (see Pegrum et al., 2022, pp. 70–72, 285–287; Hockly, 2025). Classroom-based research of this type is both contextually appropriate and deeply rooted in learners' lived experiences; it therefore provides opportunities for language teachers, researchers and learners to work together to gain more nuanced understandings of zero literacy and the wider contexts that perpetuate it.

Cross-References

► Digital Literacies

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Heeks, R., & Shekhar, S. (2019). Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society, 22*(7), 992–1011.

Helsper, E. J., & Reisdorf, B. C. (2017). The emergence of a 'digital underclass' in Great Britain and Sweden: Changing reasons for digital exclusion. *New Media & Society, 19*(8), 1253–1270.

Heponiemi, T., Gluschkoff, K., Leemann, L., Manderbacka, K., Aalto, A.-M., & Hyppönen, H. (2023). Digital inequality in Finland: Access, skills and attitudes as social impact mediators. *New Media & Society, 25*(9), 2475–2491.

Hockly, N. (2022). *What's up with WhatsApp? Supporting language learning and teacher training in low-resource contexts*. Keynote presentation at TISLID, Madrid, Spain, 27–28 May.

Hockly, N. (2025). Researching mobile technologies in language teaching. In A. Burns & K. Dikilitas (Eds.), *The Routledge handbook of language teacher action research*. Routledge.

Law, N., Woo, D., de la Torre, J., & Wong, G. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2* (Information Paper No. 51, UIS/2018/ICT/IP/51). UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf. Accessed 2 Oct 2024.

Ngiam, N. H. W., Yee, W. Q., Teo, N., Yow, K. S., Soundararajan, A., Lim, J. X., Lim, H. A., Tey, A., Tang, K. W. A., Tham, C. Y. X., Tan, J. P. Y., Yinn, S. Y., Yoon, S., Ng, K. Y. Y., & Low, L. L. (2022). Building digital literacy in older adults of low socio-economic status in Singapore (project wire up): Non-randomized controlled trial. *Journal of Medical Internet Research, 24*(12), 1–12.

Pegrum, M. (2025). Super jagged literacies: Super-diversity, jagged profiles, and digital literacies in refugee and migrant education. In E. Tour, E. Creely, P. Waterhouse & M. Henderson (Eds.), *Digital empowerment for refugee and migrant learners: Applying strengths-based practice to adult education*. Routledge. https://doi.org/10.4324/9781003422891-4

Pegrum, M., Hockly, N., & Dudeney, G. (2022). *Digital literacies* (2nd ed.). Routledge.

Phillips, N. (2013). Unfree labour and adverse incorporation in the global economy: Comparative perspectives on Brazil and India. *Economy and Society, 42*(2), 171–196.

Ragnedda, M., & Ruiu, M. L. (2017). Social capital and the three levels of digital divide. In M. Ragnedda & G. Muschert (Eds.), *Theorizing digital divides* (pp. 21–34). Routledge.

Ragnedda, M., & Ruiu, M. L. (2020). *Digital capital. A Bourdieusian perspective on the digital divide*. Emerald Publishing.

Seo, H., Erba, J., Altschwager, D., & Geana, M. (2019). Evidence-based digital literacy class for low-income African-American older adults. *Journal of Applied Communication Research, 47*(2), 130–152.

Tilly, C. (1998). *Durable inequality*. University of California Press.

C

Critical Digital Literacy

Mirjam Hauck
The Open University, Milton Keynes, UK

Abstract

Critical digital literacy (CDL) is a convergence of both digital and critical literacies. It encompasses the skills and competencies required to critically engage with digital media and technologies. It goes beyond basic digital literacy, i.e. the ability to use digital tools and platforms, and foregrounds critical thinking, analysis, and evaluation in digital contexts. In recent years, its scope has broadened to include awareness of how power operates in digital spaces and how this relates to how knowledge, identities, and social relations are shaped in ways that privilege some and marginalise others (Darvin R, Language, ideology, and critical digital literacy. In: Thorne S, May S (eds) Language education and technology, encyclopedia of language and education, vol 9. Springer, pp 17–30, 2017). With the arrival of generative AI, the remit of CDL has widened even further to include critical AI literacy, framed as a subset of CDL skills and concerned with bias and perpetuation of epistemological injustices through large language models (LLMs). Engaging with CDL, then, can help us understand how technologies are used in situated and encultured ways (Darvin R, Language, ideology, and critical digital literacy. In: Thorne S, May S (eds) Language education and technology, encyclopedia of language and education, vol 9. Springer, pp 17–30, 2017), akin to what Thorne (Lang Learn Technol 7(2):38–67, 2003, Lang Learn Technol 20(2):185–191, 2016) refers to as cultures-of-use in CALL contexts. This entry considers the nuances and implications of different conceptualisations of CDL including, most recently, the leveraging of digital technologies for social justice-oriented action and change (Jiang L, Gu MM, TESOL Q 56(3): 1029–1040, 2022), which intersects with critical approaches to CALL.

Keywords

Critical literacy · Power · Social justice · Meaning-making · AI literacy · Digital literacies

Definition

Critical digital literacy "equips learners with the tools to examine the linguistic and non-linguistic features of digital media, to identify their embedded biases and assumptions, in order to access the truth" (Darvin, 2017) and leverages digital technologies for social justice-oriented action and change (Jiang & Gu, 2022).

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_246-1

Introduction

Critical digital literacy (CDL) combines both digital and critical literacies and examines "how the operation of power within digital contexts shapes knowledge, identities, social relations, and formations in ways that privilege some and marginalize others" (Darvin, 2017, p. 2). Putting a critical lens on digital literacy helps us understand how technologies are used in situated and encultured ways (Darvin, 2017) akin to Thorne's (2003, 2016) notion of "cultures-of-use" in CALL, which relates to the cultural dimension of tools and culturally specific communicative norms and modes online. Darvin (2017) suggests that the material dimensions of online spaces—the modes and affordances available to make meaning and communicate—can be indicative of dominant ideologies, economies, and institutions and that critical digital literacy can raise our awareness to this effect.

This has been echoed by Helm (2019) in relation to virtual exchange (VE), which originates in the fields of applied linguistics and language education and which offers participants (semi-) authentic language interactions with peers, mediated by technology (Helm & Hauck, 2022). Helm (2019) points out that the online environments used for the exchanges are not ideologically neutral nor are they inherently equitable. They steer learners to normative behaviours and meanings and shape how learners position each other, how they perform identities, and how information is legitimated and distributed (Helm & Hauck, 2022). This also applies to CALL environments beyond VE settings. Being critically digitally literate, then, is what enables us to interrogate and—when necessary—to also challenge these processes.

Thus, critical digital literacy (CDL) "presupposes a degree of critical consciousness in relation to technology use" (Satar & Hauck, 2021, p. 279), achieved by providing learners with the tools to examine the linguistic and nonlinguistic features of digital media, to identify their embedded biases and assumptions (Darvin, 2017), and, most importantly, to take action (Jones, 2022).

Engaging with the critical dimension of digital literacy also addresses Kern's (2014) call for a "relational pedagogy" which encourages (language) learners to consider past as well as present literacy practices, offering them perspectives that allow them "to engage critically with today's media but also to help shape the language and literacy practices that will develop with new technologies of the future" (p. 353). Kern (2014) wants to develop among students "a disposition for paying critical attention to relations among forms, contexts, meanings, and ideologies" (p. 353). This points to the link between CDL and meaning-making in (inter-) action which is the remit of social semiotics.

CDL, Social Semiotics, and CALL

A growing body of CALL studies is informed by social semiotics (e.g. Hauck & Satar, 2018; Hauck et al., 2021); Bilki et al. (2022) conceptualise CDL as learners' awareness of the impact of meaning-making potentials—affordances and limitations—of digital media on exercising agency and establishing symbolic power relationships online. Their approach is inspired by Darvin (2017), who sees "the examination of how meanings are represented in ways that maintain and reproduce relations of power" (p. 5) as core to CDL. Bilki et al. (2022) also draw on Kramsch (2006), who—in turn—extends Bourdieu's (1991) concept of symbolic power to second language learning and teaching and argues that language learners need to acquire symbolic competence, that is, the ability to interpret and manipulate signs and symbolic systems for powerful meaning-making so that they can produce meanings that are likely to be listened to.

Satar and Hauck (2021) argue that (language) learners' varied levels of semiotic skills can influence, if not determine, their learning experiences online, because—in digital spaces—they need to understand the semiotic potentials of a much wider range of resources for meaning-making, communication, and identity creation than offline. They refer to Bilki et al. (2019), who demonstrate that those learners who can represent their

meanings and identities effectively online by capitalising on the affordances of the modes and media available (i.e. with symbolic competence; Kramsch, 2016) have comparatively more symbolic power than their peers. This confirms Hauck's (2019) reasoning that engaging in interactions such as they happen in VE, for example, can help learners become critically digitally literate, i.e. aware of how operating in digital spaces shapes ways of thinking and doing and thus meaning-making.

In the same vein, Bilki et al. (2019) suggest that VE interactions and exchanges can be a sandpit for learners to experiment with online meaning-making and identity representation and—in this way—become aware of power dynamics implicated in digital social practices which, in turn, is a characteristic feature of CDL. They propose that such awareness will enable learners to address inequalities that stem from varying levels of symbolic power and to consciously practise inclusiveness by supporting others in having a presence and a voice online and thus in exercising agency and symbolic power.

Finding your voice online and encouraging others to do the same are another core dimension of CDL. "While some can express their identities, emotions, and thoughts effectively achieving intended impact", Bilki et al. (2022) observe, "others' voices may not be listened to" (p. 60). This is likely to lead to social inequalities by privileging and empowering some while marginalising others (Bourdieu, 1991; see also Darvin, 2017).

Framing CDL as the ability to have a voice online and exercise agency hints at yet another nuance in conceptualising CDL, namely, the leveraging of digital technologies for social justice-oriented action and change (Jiang & Gu, 2022), which intersects with critical approaches to CALL. It pushes our understanding of CDL beyond meaning-making and agency in virtual contexts towards civic agency and participation, that is, towards doing things in the world.

CDL Beyond Online Meaning-Making and Online Agency

Jones (2022) highlights that "teaching students to 'think for themselves', to be 'skeptical' about what they read or watch online, to search for the 'hidden agendas' of the people who create content, and to 'do their own research'" is only part of what it means to be critically digitally literate (p. 1074). CDL is not just about how and what we think but also about what we do. This correlates with Helm's (2015) understanding of critical CALL (Helm, 2015) as the systematic "engagement with issues of power and inequality and an understanding of how our classrooms and conversations are related to broader social, cultural and political challenges" (p. 4).

Similarly, Brown (2017) posits that if digital literacies are core to what it means to be an educated person in the twenty-first century, our thinking needs to go beyond preparing people to fit the type of inequitable and socially unjust societies we have created over the past century. He wants CDL to challenge us to reshape and reimagine our societies. "As educators we will fail to serve future generations", Brown (2017) holds, "if our definition of digital literacies does not help to produce a sense of agency both with and without new technologies to disrupt 'a world where 1% of humanity controls as much wealth as the bottom 99%' of the population (Oxfam, 2017, p. 1)" (n.p.). For Brown (2017), CDL provision needs to be anchored in real-life settings and take account of the wider sociopolitical context.

Such a larger activist agenda where CDL is perceived as being instrumental to changing the world is the backdrop to Jiang and Gu's (2022) study which also illustrates how CDL interfaces with online ▶ participatory literacy as defined by Pegrum et al. (2022), i.e. the leveraging of digital networks "in the service of goals which may be online and/or offline and which may be personal and/or collective" (p. 41). Jiang and Gu's (2022) professional development model for second language (L2) teachers to teach critical digital literacies in their classrooms has three essential elements: CDL conceived as participatory, digital

multimodal composing, and teacher inquiry. They see their model as an attempt to support Teaching English to Speakers of Other Languages (TESOL) practitioners and L2 teachers to enact CDL.

Ehret and Becerra Posada (2022) report on teachers helping EFL students to enact political change within their communities. They are interested in the role of affect in CDL and language learning and teaching at different sociopolitical levels and across digital and physical spaces including the immediate feelings, online and offline, triggered by policies and ideologies developed in governmental institutions. The latter refer to the Colombian Ministry of Education's National Plan of Bilingualism and accompanying guidelines, which focus on developing communicative competences in EFL to serve the government's goal of integrating Colombia into the global economy. The policy has been criticised by teachers and researchers for ignoring the needs, priorities, interests, and knowledges of local communities. They underscore that "a global, neoliberal-focused bilingual policy may sometimes pose barriers for developing critical literacies that empower teachers and learners to value their place and communities, recognize social inequity, and appreciate foreign language learning for purposes other than economic development" (Ehret & Becerra Posada, 2022, p. 1044). This example also shows how CDL can contribute to challenging the hegemony of English in global education contexts.

Successful CDL classroom interventions, such as the one described by Ehret and Becerra Posada (2022), are usually based on principles of affinity which allow students to think together and collaborate on group projects on topics they feel strongly about (Jones, 2022). Yet, there can also be a dark side to affinity, especially online "where social media algorithms amplify particular ways of thinking and even push like-minded groups toward more extreme viewpoints" (Jones, 2022, p. 1077). In CALL, Dooly and Darvin (2022) also acknowledge the benefits of social media, which allow instant communication and collaboration across time zones and geographical distance. At the same time, they remind us how these perceived benefits can also be pitfalls, leading both locally and globally to polarisation between individuals, communities, and nations and—consequently—also raising issues of agency and power in language education (Hauck, 2025), hence the relevance of CDL to promote positive affinity, social justice, and inclusion in CALL and beyond.

The aforementioned power of algorithms also underpins concerns regarding large language models (LLMs) and generative AI (GenAI) chatbots more generally and their potential to create and perpetuate bias. This leads us to the nexus between CDL and the urgent need for critical AI literacy education ("▶ AI Literacy Practices").

CDL and Critical AI Literacy

GenAI tools use artificial intelligence to produce content such as text, images, music, or computer code. The tools are trained on large datasets which stem predominantly from the Global North and therefore carry an inherent risk of promoting epistemic injustices and colonial perspectives in their outputs. There is general agreement among educators on a need for a critical approach to their use for learning and teaching, including in CALL. The dominance of English in LLMs raises particular concerns. However, there are also benefits to the use of GenAI such as the potential to tailor the way we teach and assess to individual (language) learner needs. GenAI tools can be used as a personal tutor, a conversation partner (to practise speaking skills), a study buddy, a motivator, and a dynamic assessor (see Sharples (2023) for a comprehensive GenAI usage typology in education). Sharples (2023) goes even as far as suggesting that "every powerful pedagogy could be augmented by AI" (n.p.).

GenAI also offers opportunities for student agency and authorship as long as challenges related to its ethically responsible use and access are systematically addressed. This is the remit of critical AI literacy, understood as a subset of CDL and with a focus on equity, diversity, and inclusion. In the spirit of Kalluri (2020) of the Radical AI Network, critical AI literacy does not ask if AI

is good or fair; it rather asks how AI can contribute to shifting power. Considerations regarding power and agency are also at the heart of CDL and critical CALL. Power relationships preserve inequality within our society in real and material terms. (Language) educators will have to consider how AI exacerbates those inequalities and find opportunities to foster new balances of power through the critical and self-reflective use of GenAI.

Critical AI literacy expands the critical evaluation dimension of current AI literacy definitions such as the one put forward by Long and Magerko (2020), that is, "a set of competencies that enables individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool online, at home, school and in the workplace" (p. 8). Critical AI literacy involves the intentional use of GenAI tools for social justice and inclusion purposes in education and life in general, i.e. for the social good.

Critical AI literacy also addresses the criticism that AI actually decreases agency, notably by only ever presenting us with limited, algorithmically predetermined sets of options and recommendations, and also by treating humans as disaggregated/reaggregated datasets rather than viewing us as 'whole' people (Pegrum, 2024). It shifts the focus to an informed and intentional usage of AI to counteract this trend.

Finally, critical AI literacy also responds to Chiu et al.'s (2024) call for "self-reflective mindsets" (n.p.) to be included in AI literacy definitions, which they consider essential for lifelong learning about emerging AI technology. Their AI competency definition includes learners' "confidence and ability to self-reflect on their AI understanding for further learning" and focuses on "how well individuals use AI in beneficial ways" (n.p.). They see community engagement as an essential learning experience in AI education for the social good. The integration of student learning with the community is a purposeful pedagogical strategy used by educators to connect what is taught in the classroom with the students' local communities. Chiu et al. (2024) see it as a vehicle for more inclusive and diverse AI education.

All of the above represent the tenets of critical AI literacy which moves away from asking whether AI is good or fair to asking how it can contribute to the greater good.

Implications for Educators

- Foster critical awareness of affordances and constraints of modes in online contexts.
- Establish whether the design of a virtual (language) learning environment is indicative of a dominant institutional, educational, or corporate ideology and opt in or out accordingly.
- Draw attention to symbolic power relations online and how they manifest themselves.
- Support learners in developing symbolic competence through appropriate activity design.
- Promote the use of digital technologies for social justice and inclusion purposes (e.g. through project work).

Implications for Learners

- Pay attention to the modes—verbal, nonverbal, written, visual—that are available in a virtual (language) learning environment and the possibilities they offer for meaning-making and communication as well as their limitations in this respect.
- Note how the features of a virtual (language) learning environment promote or hinder the way you and others can participate, create an identity, and position yourselves in relation to others.
- Make a conscious effort to find your voice online and encourage others to do the same. Listen to their voices!
- Seek out opportunities to use digital technologies—including GenAI—for the social/greater good.

In sum, this entry has presented various understandings of CDL and how they interrelate via the concept of "agency", from awareness of the impact of digital media's meaning-making potential on agency and symbolic power relationships online to drawing on digital technologies—including AI tools and applications—for social justice-oriented action and change. It has also

demonstrated the alignment of these understandings of CDL with critical approaches to CALL.

Cross-References

- ▶ AI Literacy Practices
- ▶ Critical CALL
- ▶ Participatory Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Bilki, Z., Satar, M., & Hauck, M. (2019). *Virtual exchange for teacher education: A focus on critical digital literacy*. EUROCALL 2019: CALL and Complexity.

Bilki, Z., Satar, M., & Sak, M. (2022). Critical digital literacy in virtual exchange for ELT teacher education: An interpretivist methodology. *ReCALL, 35*(1), 58–73. https://www.cambridge.org/core/journals/recall/article/critical-digital-literacy-in-virtual-exchange-for-elt-teacher-education-an-interpretivist-methodology/5D2E4AD77D2937F8B6460D91EF0575A1

Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press.

Brown, M. (2017). Exploring the underbelly of digital literacies. https://oeb-insights.com/exploring-the-underbelly-of-digital-literacies

Chiu, T. K. F., Zubair, A., Murod, I., & Ismaila, T. S. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open, 6*, 100171. https://doi.org/10.1016/j.caeo.2024.100171

Darvin, R. (2017). Language, ideology, and critical digital literacy. In S. Thorne & S. May (Eds.), *Language education and technology, encyclopedia of language and education* (Vol. 9, pp. 17–30). Springer.

Dooly, M., & Darvin, R. (2022). Intercultural communicative competence in the digital age: Critical digital literacy and inquiry-based pedagogy. *Language and Intercultural Communication, 22*(3), 354–366. https://doi-org.libezproxy.open.ac.uk/10.1080/14708477.2022.2063304

Ehret, C., & Becerra Posada, T. (2022). Considering affective geographies in critical digital literacy pedagogies: How Colombian, English language teachers mitigated "feeling at a distance" during remote instruction. *TESOL Quarterly, 56*(3), 1041–1051. https://doi-org.libezproxy.open.ac.uk/10.1002/tesq.3139

Hauck, M., & Satar, M. (2018). Learning and teaching languages in technology-mediated contexts – The relevance of social presence, co-presence, participatory literacy, and multimodal competence. In R. Kern & C. Develotte (Eds.), *Screens and scenes: Multimodal communication in online intercultural encounters* (pp. 133–157). Abingdon: Routledge Studies in Language and Intercultural Communication. ISBN 9780367589271.

Hauck, M. (2019). Virtual exchange for (critical) digital literacies skills development. *European Journal of Language Policy, 11*(2), 187–210. https://doi.org/10.3828/ejlp.2019.12

Hauck, M., Satar, M., & Kurek, G. (2021). Where multimodal literacy meets online language learner autonomy: "Digital resources give us wings". In C. Fuchs, C. M. Hauck, & M. Dooley (Eds.), *Language education in digital spaces: Perspectives on autonomy and interaction* (pp. 85–111). Springer Nature. ISBN 9783030749583.

Hauck, M. (2025). Critical virtual exchange: At the interface of critical CALL, critical digital literacy, and critical global citizenship education. In E. Britton, A. Kraemer, T. Austin, H. Liu, & X. Zuo (Eds.), *Advancing critical CALL across institutions and borders*. Equinox.

Helm, F. (2015). Critical CALL. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds.), *Proceedings of the 2015 EuroCALL conference* (pp. 1–6). Padova, Italy: Research-publishing.net.

Helm, F. (2019). EMI, internationalisation, and the digital. *International Journal of Bilingual Education and Bilingualism, 23*(3), 314–325. https://doi.org/10.1080/13670050.2019.1643823

Helm, F., & Hauck, M. (2022). Language, identity and positioning in virtual exchange. In L. Klimanova (Ed.), *Identity, multilingualism and CALL: Responding to new global realities. Advances in CALL research and practice* (pp. 24–48). Equinox.

Jiang, L., & Gu, M. M. (2022). Toward a professional development model for critical digital literacies in TESOL. *TESOL Quarterly, 56*(3), 1029–1040. https://doi-org.libezproxy.open.ac.uk/10.1002/tesq.3138

Jones, R. H. (2022). Commentary: Critical digital literacies as action, affinity, and affect. *TESOL Quarterly, 56*(3), 1545–7249. https://doi.org/10.1002/tesq.3153

Kalluri P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature, 583*(7815), 169–169. https://doi.org/10.1038/d41586-020-02003-2

Kern, R. (2014). Technology as pharmakon: The promise and perils of the internet for foreign language education. *The Modern Language Journal, 98*(1), 340–357. https://doi.org/10.1111/j.15404781.2014.12065.x

Kramsch, C. (2006). From communicative competence to symbolic competence. *The Modern Language Journal, 90*(2), 249–252. https://doi.org/10.1111/j.1540-4781.2006.00395_3.x

Kramsch, C. (2016). The multiple faces of symbolic power. *Applied Linguistics Review, 7*(4), 517–529. https://doi.org/10.1515/applirev-2016-0023

Long, D., & Magerko B. (2020) What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

Pegrum, M. (2024). From revolution to evolution: What generative AI really means for language learning. Keynote presentation at English Australia Conference, Perth, 11–13 September.

Pegrum, M., Hockly, N., & Dudeney, G. (2022). *Digital literacies*. Routledge.

Satar, M., & Hauck, M. (2021). Exploring digital equity in online learning communities. In A. M. Sousa Aguiar de Medeiros & D. Kelly (Eds.), *Language debates: Digital media* (pp. 270–290). John Murray Learning.

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning. Research and Practice, 9*(2), 159–167. https://www.tandfonline.com/doi/full/10.1080/23735082.2023.2261131

Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology, 7*(2), 38–67.

Thorne, S. L. (2016). Cultures-of-use and morphologies of communicative action. *Language Learning & Technology, 20*(2), 185–191.

M

Media Literacy

Jihyun Lee
Kookmin University, Seoul, South Korea

Abstract

With the expansion in media volume, the importance of media literacy has grown, underscoring the significance of media literacy education. Encouraging media literacy in educational contexts, particularly in language education, is essential for helping learners navigate a vast and often overwhelming domain with greater comprehension and discernment. Embedding critical thinking skills into language learning activities empowers learners to evaluate accuracy and bias and make informed decisions regarding the media that they consume, create and share. Media literacy education is supported by constructivism, which emphasises social interactions and the construction of meaning. It fosters deeper engagement with media messages and promotes critical analysis. This entry illustrates what media literacy entails and provides an example of how to integrate media literacy with language learning based on critical thinking and constructivism.

Keywords

Media literacy · Media literacy education · Media literacy in language learning · Critical thinking · Constructivism · Digital literacies

Definition

Media literacy refers to the ability to access, analyse, and evaluate messages from media and the ability to act through media. *Access* involves locating media content through available sources, and *analyse* means decoding media messages to understand their meaning. *Evaluate* refers to scrutinising the source, purpose, credibility, and authenticity of information while *act* entails expressing oneself and communicating with others through media.

Introduction

Technological advancements have led to the proliferation of diverse media, creating a society saturated with information from such channels. Unless individuals deliberately avoid media consumption, these outlets inevitably expose them to their effects in their daily lives. Individuals frequently encounter advertisements for a specific brand, which can foster positive perceptions among consumers and, consequently, lead to purchases of that brand's products. Similarly, a celebrity's photo with a particular politician can influence the political orientation of the celebrity's social media followers. Engaging with media reflects its significance in modern life.

In this context, media literacy has become of great importance. Media can spread false news, misinformation and disinformation, but it also

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_248-1

plays a crucial role in informing, empowering and entertaining users. As media channels have expanded from traditional forms such as television, radio and newspapers to digital media such as social media, blogs, and streaming services, the abundance of available information has made it increasingly important for users to develop the ability to discern and make informed decisions. This capability defines media literacy and underscores the necessity for comprehensive media literacy education.

The goal of media literacy education is to enable individuals to critically analyse media messages, evaluate the reliability of information, understand the impact of media on their personal beliefs and attitudes, and engage as active democratic citizen.

Media literacy should be integrated into subjects as a component of education (Jolls, 2015; Korona & Hutchison, 2023), equipping students with essential skills to navigate an increasingly complex information landscape. By embedding media literacy across the curriculum, educators can foster critical thinking, enhance students' ability to construct knowledge based on credible information, and prepare them to engage thoughtfully in society. This comprehensive approach not only empowers students as informed consumers of media but also cultivates responsible citizenship. In terms of language teaching, the integration of media literacy into language instruction is highly applicable. Media serves as both a natural source of language input and a model for authentic language use. By engaging with the media, understanding its messages, and analysing them to drive action, students can significantly enhance their language practice.

Nature of Media Literacy

Since the concept of media literacy was developed at the 1992 National Leadership Conference on Media Literacy, researchers have defined it in various ways (Potter, 2022a). This entry elaborates on media literacy through four key components: access, analyse, evaluate, and act. The following example elaborates on how media literacy works in daily life.

Sarah is a fan of animated films. While travelling to work, she learns about the debut of a new animation on the radio and then searches for reviews by well-known cinema reviewers (access). The animation has received mixed reviews, but she decides to watch it at the cinema (access). As she enjoys the film, she understands the relationships among the characters; comprehends the background, plot and twist; and becomes emotionally engaged with the story (analyse). Throughout the film, she notices that it follows a clichéd formula and features specific cultures and social classes among the main characters (evaluate). She worries that her little sister might internalise such stereotypes about race and class (evaluate). Subsequently, Sarah takes a selfie in front of the film poster to post on social media with the caption, 'It's intriguing, but still clichéd' (act). Later, when her friend Jack sees her post, he decides to book a cinema ticket, while another friend, Emily, chooses to watch it on an over-the-top platform instead of at the cinema (▶ "Multimodal Literacy" and ▶ "Social Media Literacy").

Media literacy is the ability to critically evaluate media messages and decide, based on personal judgement, whether to accept or reject them. Additionally, factors such as individual experiences, economic conditions and social infrastructures significantly influence how media literacy is practised. For instance, Jack and Emily each have their own experiences communicating with Sarah. Although they access the same media, they analyse, evaluate and respond to this information differently. A friend without a budget for cultural activities might skip streaming services or cinema, while another friend with limited internet access may miss Sarah's post entirely. Contextual factors can influence the practice of media literacy.

Media Literacy Education

Media literacy education has two orientations: protection and empowerment. The former is about safeguarding people from the potential

harmful effects of media, while the latter is about instilling a sense of power in individuals by cultivating informed participants. In both approaches, teachers play a vital role by helping students critically analyse the media and navigate it in ways that foster positive experiences (Potter, 2022b).

From the perspective of learning theories, media literacy education is underpinned by constructivism (Hobbs & Jensen, 2009; Share & Mamikonyan, 2020). Constructivism is characterised by the incorporation of linguistic, psychological and sociological perspectives into education and focuses on social interaction and cooperative learning (Brown, 2014). It emphasises the active construction of learners' own understanding and knowledge of the world (Vygotsky, 1978). Even if students interpret the same message through the same media channel, they may understand it differently and take varied actions as a result. Constructivism engages learners by allowing them to explore information and participate in decision-making, thereby promoting empowerment.

One of the key aspects of media literacy education is critical thinking (▶ "Information Literacy"). Critical thinking is the process of questioning, analysing, evaluating, and judging the value of the information you encounter. It involves not only understanding the information but also assessing its credibility, relevance, and bias. Critical thinking contributes to the development of one's own knowledge, enabling individuals to cope with misinformation, disinformation, and fake news. Critical framing allows learners to view their knowledge from a theoretical perspective, consider its social and cultural context, and both critique and build upon it (Buckingham, 2007).

Media Literacy Education in Language Learning

Teachers should integrate media literacy into language instruction, considering three essential dimensions: linguistic aspects, critical thinking, and constructivist principles. Input activities that facilitate understanding of media content should be implemented, followed by output activities that encourage students to express their opinions. The processes of comprehension and expression should demand critical thinking from students while simultaneously aligning with a learner-centred approach and collaborative practices. The following activities can be utilised in lessons that integrate media literacy with language practice through various media forms. These activities foster critical thinking and encourage students to engage with media on multiple levels. *Analysis* activities identify the themes and techniques used. *Evaluation* activities assess purpose and impact. And *Creation & Sharing* activities build on analysis and evaluation, developing students' ability to communicate through these media forms.

***Analysis* Activities**

1. Themes: Students analyse the themes addressed in a film and discuss the intended messages that the filmmakers aim to convey.
2. Characters: Students examine what each character symbolises in relation to the film's themes.
3. Film Techniques: Students investigate the cinematographic techniques, editing, music and colour schemes used in the film, discussing how these elements reinforce its messages.

***Evaluation* Activities**

4. Target Audience and Intent: Students consider the target audience for the film and analyse the emotions or social messages that the work seeks to communicate.
5. Teaser Trailer: Students analyse the messages emphasised in the film's teaser trailer and discuss the techniques used to capture the audience's interest.
6. Context: Students explore the relationship between the film and the social and political contexts of its time, reflecting on how the film mirrors or impacts society.
7. Film Critic Opinions: Students gather opinions from various film critics' blogs and discuss how these perspectives might influence future audiences.

8. Role-Playing: Students engage in role-playing by assuming the roles of key characters and posing questions to one another.

***Creation and Sharing* Activities**

9. Design a Poster: Students select a line and a scene that encapsulates the film's theme to design a digital poster.
10. Reimagine a Script: Students rewrite the film's ending or a specific scene in a preferred manner and draft a new script.
11. Film a Scene: Students re-enact a scene or scenes and film them using the script they rewrote.
12. Write a Review: Students write a personal review of the film and share it on social media, engaging with audience comments on their reviews.

Conclusion

Media literacy is the ability to access, understand, and evaluate media messages, enabling individuals to interact effectively with various media platforms. In a world filled with diverse media and messages, media literacy is a crucial skill for the twenty-first century (UNESCO, n.d.). Individuals with media literacy skills can effectively filter information and act as responsible media users.

The importance of media literacy underscores the necessity for media literacy education. Media literacy should be a foundational element of all curricula, helping learners assess the validity of information and understand diverse perspectives. Incorporating critical thinking and constructivist approaches encourages learner participation and autonomy, enabling students to grasp the context of information and form their own opinions.

In language classrooms, it is essential to integrate media literacy and language education around the two key elements of critical thinking and constructivism. From a linguistic perspective, activities that involve understanding the language provided by media and using language through media are crucial. In terms of media literacy, activities that enhance the ability to select and accept media information by considering the intentions, purposes, and impacts of media are also necessary. This empowers students to develop their understanding and lays the groundwork for active social participation.

Cross-References

▶ Information Literacy
▶ Multimodal Literacy
▶ Social Media Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Brown, H. D. (2014). *Principles of language learning and teaching* (6th ed.). Pearson Education.

Buckingham, D. (2007). Digital media literacies: Rethinking media education in the age of the Internet. *Research in Comparative and International Education, 2*(1), 43–55. https://doi.org/10.2304/rcie.2007.2.1.43

Hobbs, R., & Jensen, A. (2009). The past, present, and future of media literacy education. *Journal of Media Literacy Education, 1*(1), 1–11. https://doi.org/10.23860/jmle-1-1-1

Jolls, T. (2015). The new curricula: How media literacy education transforms teaching and learning. *Journal of Media Literacy Education, 7*(1), 65–71. https://doi.org/10.23860/jmle-7-1-7

Korona, M., & Hutchison, A. (2023). Integrating media literacy across the content areas. *Reading Research Quarterly, 58*(4), 601–623. https://doi.org/10.1002/rrq.517

Potter, W. J. (2022a). Analysis of definitions of media literacy. *Journal of Media Literacy Education, 14*(2), 27–43. https://doi.org/10.23860/JMLE-202214-2-3

Potter, W. J. (2022b). Analyzing the distinction between protectionism and empowerment as perspectives on media literacy education. *Journal of Media Literacy Education, 14*(3), 119–131. https://doi.org/10.23860/JMLE-2022-14-3-10

Share, J., & Mamikonyan, T. (2020). Preparing English teachers with critical media literacy for the digital age. *Contemporary Issues in Technology and Teacher Education, 20*(1). https://citejournal.org/volume-20/issue-1-20/english-language-arts/preparing-english-teachers-with-critical-media-literacy-for-the-digital-age

UNESCO. (n.d.). Media and information literacy. https://www.unesco.org/en/media-information-literacy

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

A

Attentional Literacy

Agnieszka (Aga) Palalas[1] and Mark Pegrum[2]
[1]Open, Digital & Distance Education, Athabasca University, Athabasca, AB, Canada
[2]The Graduate School of Education, The University of Western Australia, Perth, WA, Australia

Abstract

The growing trend towards digital learning, situated within a wider context of hyperconnectivity and digital disarray, places significant attentional demands on all students, including language learners. To succeed in online education, learners need to develop attentional literacy, a concept which draws on insights from the mindfulness and contemplative pedagogy traditions to address very contemporary needs. As both a skillset and a disposition, attentional literacy involves learners developing awareness of how their attention is solicited, honing the ability to focus intentionally, remaining open to new perspectives, and considering their roles and impact within the larger attentional ecosystem. After developing their own attentional literacy holistically, educators can help students develop this skillset in a pedagogically staged way. As a result, learners can benefit from greater control over their attentional resources and better understandings of the self and others, allowing them to engage more successfully in digital learning. The development of attentional literacy, as part of the wider digital literacies skillset, can and should occur in tandem with the development of language and intercultural skills, to support learners with the cognitive load and socio-emotional aspects of language usage as well as the building of intercultural competence.

Keywords

Attentional literacy · Attention · Awareness · Contemplative pedagogy · Mindfulness · Intercultural competence · Digital literacies

Introduction

Attentional literacy is "the ability to notice and understand how one's attention is solicited; to intentionally focus and sustain one's attention; to become and remain open to new perspectives; and to consider one's impact on the attentional ecosystem" (Palalas et al., 2024). It can improve the effectiveness of language learning and intercultural experiences.

Attentional Literacy in Education

The rapid growth of online and blended learning, taking place in a wider context of hyperconnectivity

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_250-1

and digital disarray, places significant attentional demands on learners, including language learners (Pegrum & Palalas, 2021; Reinders et al., 2024). In digital environments learners are challenged by information overload, an abundance of competing stimuli, a need to navigate multiple, often unintuitive technologies, and individualised but depersonalised AI algorithms that seek to capture, manipulate and direct their attention, "often against [their] conscious will" (Jeffrey et al., 2023, p. xii), and ultimately in the service not of users themselves but of technology providers. These challenges can lead to cognitive overwhelm, stress and fatigue, occasioning distractibility and an attention deficit (Mark, 2023; Palalas & Doran, 2024). While the issues are partly systemic and structural, personal habits and choices often compound learners' struggles to pay attention deeply and intentionally to learning activities and opportunities, which by their nature require the application of cognitive and emotional resources.

Digital literacies are indispensable to address the issues of hyperconnectivity and digital disarray and the challenges faced by online learners (Pegrum & Palalas, 2021). Attentional literacy has recently emerged as a key literacy within the digital literacies skillset (Reinders et al., 2024). Conceptually, it draws on insights from the mindfulness and contemplative pedagogy traditions (e.g. Ergas, 2019; Goleman & Davidson, 2017), builds on past work on "attention literacy" (e.g. Palalas, 2019; Rheingold, 2009; Wenger, 2019) and overlaps with parallel concepts like agentive literacy (Murray et al., 2020) as well as intersecting literacies such as critical, ethical and intercultural literacy (Pegrum et al., 2022). Attentional literacy can support both teachers and learners in staying present and engaged in meaningful ways so as to make the most of their online learning experiences (Jeffrey et al., 2023; Palalas et al., 2024).

Attentional literacy, as defined above, encompasses a number of focus points linked to contemplative pedagogy. Educators should attend to their own needs first, ideally developing their attentional literacy through interwoven personal practices (e.g. meditation, walking) and professional practices (e.g. academic research, collegial discussions), some of which may be offline and some online. They can then guide students through activities which support the development of the following focus points (Palalas et al., 2024):

- *Awareness and noticing* of how one's attention is drawn to the self, others and the environment; e.g. digital habit journals to promote learner consciousness
- *Focus and concentration* coupled with *intention and choice,* exercised through intentional focus which can eventually be sustained at will; e.g. mindful listening
- *Openness and curiosity,* involving growing awareness of and a non-judgemental attitude towards new perspectives, multiple viewpoints and shifting contexts; e.g. online discussions inquiring into diverse global and local perspectives
- *Consideration of the attentional ecosystem,* including one's own place within this fluid, reciprocal and relational complex; e.g. co-creation of engagement and interaction guidelines that promote co-presence in learning environments

While the focus points are all interrelated, educators could begin with the first two sets before expanding their pedagogical approach to include the final two. As learners become aware of their own behavioural patterns and practise intentionally exercising different options, they can build their "attentional muscle" (Kabat-Zinn, 2013; Levy, 2016). As with gym-based physical training, this requires willingness and perseverance: it may be a time-consuming and sometimes frustrating process, and some initial learner training may be helpful to outline its benefits for learning in general and language learning in particular. In time, students are likely to see that with greater control over their attentional resources, they can engage more effectively in online learning, while the "relational mindfulness" (Gruber & Henriksen, 2024, p. 59) that comes from attending to oneself in relation to others can help to foster more connected online learning communities.

Importance for language learning. Attentional literacy affords language learners enhanced *awareness and noticing* of

information coming at them and being solicited from them. It allows them to be selective, actively processing specific information while effectively tuning out other details (James, 2007). *Intentionally focusing* in this way helps to lighten potential cognitive overload by controlling the volume of information, minimising other distractions and balancing interactivity. Attention can be directed to various aspects of language input and output, from the target language structures and rules to the nuances of its usage in different contexts. Learners are also able to witness their own social and emotional reactions to communications, promoting understanding of the self, others and the environment along with "the ability to convey those understandings effectively and meaningfully to others" (Delphi study participant cited in Palalas et al., 2024), and opening up space for the development of intercultural competence (see below). By attending to their understandings with wise discernment, learners can strengthen their capacity for navigating and shaping their learning processes online (Jeffrey et al., 2023).

Importance for intercultural competence. The later focus points—*openness and curiosity* and *consideration of the attentional ecosystem*—are particularly relevant to the development of intercultural competence, or intercultural literacy, long considered an important corollary of language learning (Pegrum, 2019). Openness and non-judgement towards new perspectives, and an awareness of one's embeddedness in a relational ecosystem, are integral to honing understanding not only of others but the self, as promoted in the collaborative online international learning (COIL) or virtual exchange (VE) initiatives that are now becoming a core part of internationalisation-at-home strategies (e.g. Helm et al., 2020). Crucially, it is not just a case of intellectual appreciation but of the development of empathy (e.g. Goleman & Davidson, 2017). In this way, attentional literacy and its promotion of a mindful connection to the self and others can help learners develop into digital citizens who are ready to contribute to collaboratively addressing today's global challenges (Pegrum & Palalas, 2021).

Implications for Educators

- **Relate attentional literacy to learners' own lives and learning goals.** Help learners to become aware of their current attentional habits and to experience the benefits of developing more attentional control in everyday online (and indeed offline) settings. Where possible, integrate attentional activities with language learning activities to emphasise their relevance to the development of language skills.
- **Begin by teaching awareness and noticing, followed by intentional focus.** Help learners to develop these fundamental skills and dispositions, for example through journaling practices to highlight the drawbacks of multitasking and the benefits of deep engagement or *monotasking*; centring practices, like breathing meditations, to calm the mind and reduce distractions; and concentration-based practices, like mindful listening or reading, to develop the "attentional muscle". These can support learners' focus in online (and offline) language usage.
- **Continue by teaching openness to new perspectives and active consideration of one's place within the attentional ecosystem.** Help learners to develop these more sophisticated skills and dispositions, for example when preparing to engage in conversation in intercultural settings, starting with carefully scaffolded COIL initiatives. These can help to support the growth of the intercultural competence necessary for global digital citizenship.

Implications for Learners

- **Become aware that developing attentional literacy is like a training muscle** and requires perseverance, time and sustained, regular practice. As with any untrained muscle, this may initially feel challenging.
- **Develop attentional literacy** iteratively, building awareness and intentionally trying out available options to identify the practices that best support your own learning.

- **Employ attentional literacy** to improve focus in your digital language learning activities and opportunities and to build connectedness in intercultural conversational settings.

In summary, attentional literacy is emerging as a key digital literacy of our time, with important implications for language learners. Its benefits can potentially extend to all learners, both online and offline, and in both formal and informal learning contexts.

Cross-References

▶ Critical Digital Literacy
▶ Digital Literacies
▶ History of New Literacies
▶ Intercultural Literacy

Competing Interests Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Ergas, O. (2019). A contemplative turn in education: Charting a curricular-pedagogical countermovement. *Pedagogy, Culture & Society, 27*(2), 251–270. https://doi.org/10.1080/14681366.2018.1465111

Goleman, D., & Davidson, R. J. (2017). *The science of meditation: How to change your brain, mind and body.* Penguin.

Gruber, N., & Henriksen, D. (2024). Bridging the distance: Relational mindfulness practices for connection and compassion in online education. *Journal of Technology and Teacher Education, 32*(1), 59–97. https://www.learntechlib.org/primary/p/223931/

Helm, F., Guth, S., Shuminov, E., & Van der Velden, B. (2020). Erasmus+ virtual exchange. *Internationalisation of Higher Education, 1*, 92–109. https://www.handbookinternationalisation.com/media/02818d033e0d8e74995aed9e16e1bb8970060360/r7aklh/p df_5e81ffb452439.pdf

James, W. (2007). *The principles of psychology* (Vol. 1). Cosimo.

Jeffrey, L., Palalas, A., Robert, K., & Wong, Y.-L. R. (2023). Introduction. In L. Jeffrey, A. Palalas, K. Robert, & Y.-L. R. Wong (Eds.), *Presence in the online world: A contemplative perspective and practice guide for educators* (pp. ix–xxxii). Rowman & Littlefield.

Kabat-Zinn, J. (2013). *Full catastrophe living: How to cope with stress, pain and illness using mindfulness meditation* (revised ed.).

Levy, D. M. (2016). *Mindful tech: How to bring balance to our digital lives*. Yale University Press.

Mark, G. (2023). *Attention span: A groundbreaking way to restore balance, happiness and productivity.* Harlequin.

Murray, L., Giralt, M., & Benini, S. (2020). Extending digital literacies: Proposing an agentive literacy to tackle the problems of distractive technologies in language learning. *ReCALL, 32*(3), 250–271. https://doi.org/10.1017/S0958344020000130

Palalas, A. (2019). Mindfulness for human-centred digital learning. *Argentinian Journal of Applied Linguistics, 7*(2), 110–125.

Palalas, A., & Doran, M. (2024). Digital wellness framework for online learning. *Canadian Journal of Learning and Technology, 49*(3), 1–25. https://cjlt.ca/index.php/cjlt/article/view/28581

Palalas, A., Pegrum, M., & Dell, D. (2024). Regaining focus: Promoting attentional literacy in digital higher education.

Pegrum, M. (2019). *Mobile lenses on learning: Languages and literacies on the move*. Springer.

Pegrum, M., & Palalas, A. (2021). Attentional literacy as a new literacy: Helping students deal with digital disarray. *Canadian Journal of Learning and Technology, 47*(2). https://cjlt.ca/index.php/cjlt/article/view/28037

Pegrum, M., Hockly, N., & Dudeney, G. (2022). *Digital literacies* (2nd ed.). Routledge.

Reinders, H., Kukulska-Hulme, A., & Wilson, A. (2024). *Humanizing technology in language learning and teaching*. Oxford University Press. https://elt.oup.com/feature/global/expert/humanising-tech

Rheingold, H. (2009, April 20). *Attention literacy*. SFGate. http://www.sfgate.com/cgibin/blogs/rheingold/detail?entry_id=38828

Wenger, C. I. (2019). Teaching attention literacy: A case for mindfulness in the composition classroom. *Pedagogy, 19*(1), 53–78. https://doi.org/10.1215/15314200-7173752

P

Participatory Literacy

Mirjam Hauck[1] and Sylvia Warnecke[2]
[1]The Open University, Milton Keynes, UK
[2]The Open University in Scotland, Edinburgh, UK

Abstract

Participatory literacy (PL) has two dimensions: being able to contribute to the collective intelligence of digital networks and being able to draw on this intelligence to pursue personal goals as well as collective endeavours (Pegrum et al., Digital literacies. Routledge, 2022, p. 41). In the early 2000s, technology became increasingly social, and a literacy shift could be observed from the acquisition and interpretation of online content to its production and to active participation in digital culture and society. As a result, a new subset of digital literacies emerged: PL, also referred to as "digital communicative literacy", a foundation for online interactions and facilitating collaborative processes (Pegrum, 2009). It is considered fundamental for successful involvement in participatory cultures (Jenkins et al., MacArthur Foundation. http://digitallearning.macfound.org/atf/cf/%7B7E45C7E0-A3E0-4B89-AC9C-E807E1B0AE4E%7D/JENKINS_WHITE_PAPER.PDF, 2006). Starting with Giger's (Participation literacy: Part I: constructing the Web 2.0 concept series. Blekinge Institute of Technology, Karlskrona, Sweden, 2006) conceptualisation of PL as "the ability to create and share knowledge and content collectively through the use of online tools and the completion of collaborative tasks in online environments", this entry traces the way our understanding of participatory literacy has evolved over the past two decades and considers its constituent elements—social presence, multimodal communicative competence, contextual knowledge, and collective intelligence—with a focus on CALL contexts.

Keywords

Collective intelligence · Agency · Participatory cultures · Social presence · Multimodal communicative competence · Contextual knowledge · Digital literacies

Definition

Participatory literacy is "the ability to contribute to the collective intelligence of digital networks and to leverage the collective intelligence of those networks in the service of goals which may be online and/or offline and which may be personal and/or collective" (Pegrum et al., 2022, p. 41).

Introduction

In the early 2000s, Internet technologies became increasingly social, and a literacy shift could be

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_251-1

observed from the acquisition and interpretation of online content to its production and to active participation in digital culture and society. In parallel, the so-called social turn in second language acquisition (SLA) had started (Block, 2003), manifesting itself in a general shift in second language (L2) learning theory "from behaviourist, to cognitive, to situated, social, and distributed views of human cognition" (Johnson, 2006, p. 236). It led to an increasing recognition of language as social practice, whereby learning a new language entails learning to participate in new communities and to negotiate identities. This challenged the deficit view of language learners as striving for native speaker competence and emphasised concepts such as agency and identity. SLA identity theory made "the convincing case that additional language learning is much more than just learning language [. . .]. It is about power and social transformation" (Ortega, 2011, p. 172).

This turn to the social also manifested itself in computer-assisted language learning (CALL), with the computer providing alternative contexts for social interaction and facilitating access to existing discourse communities as well as enabling the creation of new ones (Kern & Warschauer, 2000). In alignment with the movements towards a socially informed SLA research agenda, the term "social turn" was used in CALL to capture the transition from cognitive to socio-culturally oriented theoretical approaches.

However—similar to what Ortega (2011) observes regarding the impact of the social turn in SLA—access to existing and new discourse communities, online groups, or communities dedicated to language learning was not restricted to language learning. It allowed users to establish social ties, to exercise agency, and to experience the power dynamics implicated in digital social practices.

As a result, a subset of digital literacy became relevant in computer-mediated communication (CMC) in general and in CALL in particular: participatory literacy (PL), also referred to as "digital communicative literacy", providing a foundation for online social interactions and facilitating collaborative processes (Pegrum, 2009) and becoming essential for successful involvement in participatory cultures (Jenkins et al., 2006). However, Jenkins et al. (2006) point to the "awareness dilemma" or "transparency challenge" regarding such online participatory cultures, i.e. the assumption that users, including online language learners, are "actively reflecting on their media [. . .] experiences and can thus articulate what they learn from their participation" (p. 12). The systematic integration of PL skills into CMC and CALL, Hauck (2010) suggests, is required to address Jenkins et al.'s (2006) concern and promote such reflection.

The Rise and Relevance of Participatory Literacy in Online (Language) Learning and Teaching

Giger (2006) provided an early conceptualisation of PL, focusing on its relational aspects in the context of social software use. He defined "participation literacy" as the ability to create and share knowledge and content collectively through the use of online tools and the completion of collaborative tasks in online environments. "You are participation literate", Giger (2006) posits, "if you work actively to invite everyone into a discussion and consider everyone's discourse valuable" (p. 84). Apart from his focus on equity and inclusion, he also sees it as the precondition for developing the ability to dialogue and to argue one's own ideas online. More recently, Pegrum et al. (2022) have highlighted the added benefit to those who are able to actively participate online. Their definition of PL has two dimensions: being able to contribute to the collective intelligence of digital networks and being able to draw on this intelligence to pursue personal goals as well as collective endeavours (Pegrum et al., 2022, p. 41).

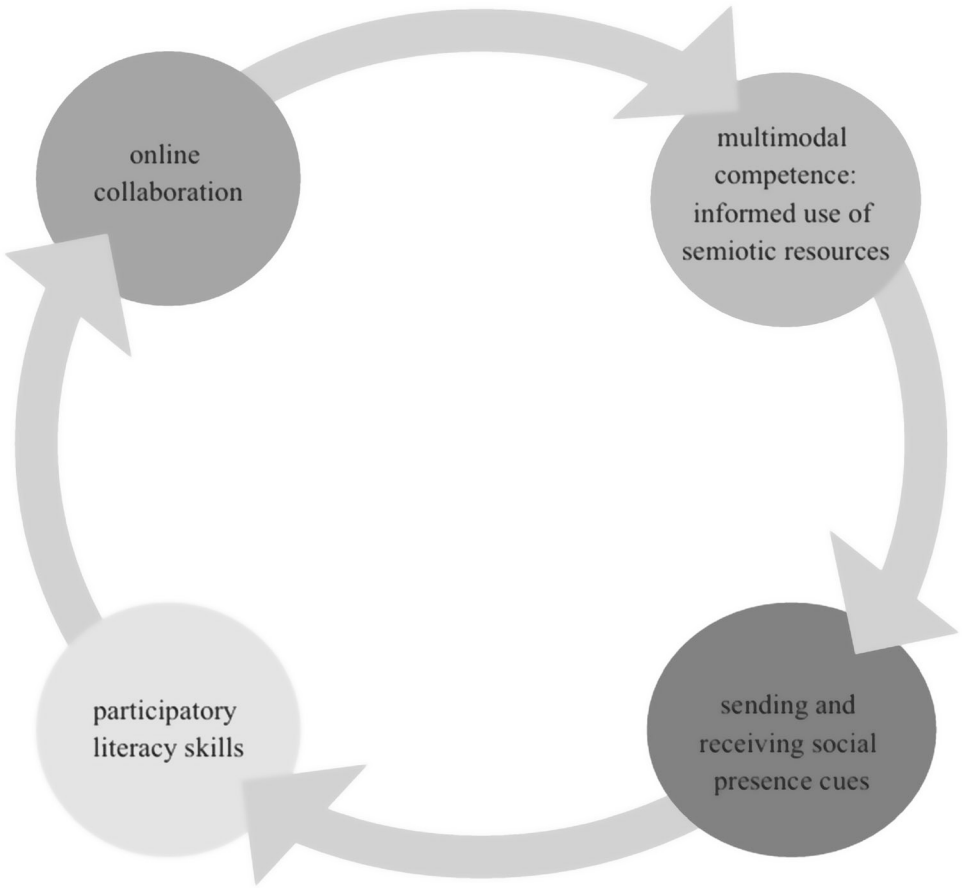
Yet, successful participation online is context-dependent (Kern et al., 2004). Hence, Hauck and Warnecke (2012) argue that participation and social agency online presuppose, among other capabilities, social presence (SP) online: the means by which online participants inhabit virtual spaces and indicate not only their presence in the online environment but also their availability and willingness to engage in the communicative exchanges which constitute learning activity in

these environments (Kehrwald, 2008), i.e. in short, demonstration of availability for, and willingness to participate in, interaction. The concept of SP was first introduced by Garrison et al. (2000) as a constituent element of their Community of Inquiry (COI) model which distinguishes between cognitive presence, teaching presence, and SP with all three types of presence being required for successful online learning in general.

Hauck and Warnecke (2012) illustrate how increasing awareness of the affordances of a given online environment—their contextual knowledge—allowed the language teachers who took part in their study to consciously develop SP, i.e. "through seeing and experiencing how others project themselves into the environment, how others interact with one another and how others react to [one's] personal efforts to cultivate a social presence" (Kehrwald, 2010, p. 47).

Similarly, Pegrum (2010) reasons that the best way to develop participatory literacy is by actually participating in the social web. Accepting Pegrum's reasoning allowed Hauck and Satar (2018) to construct the following circular argument: it is through online participation and collaboration that multimodal communicative competence (multimodal literacy), participatory literacy skills, and thus the ability to send and receive social presence (SP) cues are developed (see Fig. 1).

This approach to PL is informed by Hauck and Warnecke (2012) who first pointed to the close interrelationship between contextual knowledge,



Participatory Literacy, Fig. 1 Participatory literacy development (Hauck & Satar, 2018). (Screens and Scenes—Online Multimodal Communication and Intercultural Encounters by R. Kern and C. Develotte (Eds.), Copyright (2018) by Imprint. Reproduced by permission of Taylor & Francis Group)

PL and SP in online language learning and teaching, and Hampel and Hauck's (2006) understanding of contextual knowledge as awareness of online modes and their potential for meaning-making, communication, and online participation. Hampel and Hauck (2006), in turn, drew inspiration from Kress' (2003) definition of multimodal competence—the ability to express ideas across a wide range of representational systems or modes including words, spoken or written; images, still and moving; musical; and 3D models. Royce (2007) adapted multimodal competence to language learning highlighting the receptive as well as the productive dimension of the competence in question:

> The ability to understand the combined potential of various modes for making meaning (Royce, 2002), that is the ability to process intersemiotic relationships and also the ability to produce these kinds of relationships. (p. 375)

Hauck and Warnecke (2012) also show how PL and SP can be learned through a task-based approach in CALL and through "experiential modelling" (Hoven, 2006) which immerses (language) teacher trainees "in the use of technologies, while at the same time providing them with the freedom and framework within which to experience the practical application of theory in their own learning" (n.p.). The data for their study comes from a training programme for Open University UK language educators which—at the time—was innovative considering that it was designed to:

- Raise the educators' awareness of the relevance of PL and SP in online language learning and teaching
- Allow the educators to find out how they could promote PL skills and SP in their learners through task design for online environments

Hauck et al. (2016) propose to conceive of participatory behaviours as "positioning", which signals a move beyond Kehrwald's (2008) notion of "projecting oneself" towards a notion of shifting patterns in the way we relate to others online, in direct response to the learning context, its modes, and affordances. The authors illustrate how participant engagement in tasks designed to trigger exchanges on the very topic of participation and positioning led to the kind of active reflection Jenkins et al. (2006) wanted to see more systematically promoted to address the aforementioned "awareness dilemma".

All of the above speak to the interface of PL and social agency as outlined in the introduction. Recently, however, another type of agency has become associated with PL: civic agency, which is a concept based on the principles of an equitable and democratic society.

Participatory Literacy for the Greater Good

Holmes and McGlinn Mafra (2020) have framed the systematic integration of PL in education as a means of supporting students in taking informed (civic) action. They remind us that in the era of fake news and misinformation, PL has taken on new importance as "educators and policymakers are increasingly concerned with the ability of students to navigate a complex and often deliberately misleading online landscape" (p. 40). Drawing on Wohlwend's (2017) framing of PL as ways of interpreting, making, sharing, and belonging in increasingly globally and digitally mediated cultures across multiple media and platforms, Holmes and McGlinn Mafra (2020) argue that students must be given opportunities to develop the required skills as well as the ability to discern credible information from web-based resources. This includes the opportunity to "contribute to blogs, wikis, social networking and sharing sites, virtual worlds, and gaming environments, which rely upon creativity, reasoning, focus, critical thinking, and capacity to collaborate" (Wang & Torrisi-Steele, 2023). With reference to the College, Career, and Civic Life (C3) Framework by the US National Council for the Social Studies, Holmes and McGlinn Mafra (2020) explore strategies for supporting the development of PL with a specific focus on civic engagement, that is, ways to impact the world beyond educational institutions.

This approach to PL is also reflected in Pegrum et al.'s (2022) understanding of PL as the ability to leverage the collective intelligence of digital

networks in the service of personal or collective endeavours. It foregrounds an important dimension of the wider concept of critical digital literacy (CDL) of which PL is a defining element: critical consciousness and critical agency in the spirit of the Brazilian pedagogue and educational theorist Paulo Freire and the American–Canadian cultural critic Henry Giroux, that is, the ability to read the world critically and the ability to act in the world to change it, which starts with an understanding of one's power in relation to others' power. Drawing on Buchholz et al.'s (2020) conceptualisation of CDL as participatory, Jiang et al. (2020) and Mirra and Garcia (2020), for example, talk about critical digital literacies in terms of leveraging digital technologies for social justice-oriented action and change.

Similarly, critical CALL (critical approaches to CALL; Helm, 2015) has been framed as the systematic "engagement with issues of power and inequality and an understanding of how our classrooms and conversations are related to broader social, cultural and political challenges" (p. xiii). Thus, critical CALL also presupposes a sense of civic agency on behalf of the learner and shares a concern for social equity.

PL is core to exercising such agency and for acting upon such concerns. PL skills as outlined above allow students to resist predetermined literacy and offer new avenues and opportunities for literacy expression through the creation of vlogs, websites, online games, etc. These opportunities are not limited to formal educational settings though and happen wherever there is a commitment to learner agency, including where educators see students as literate agents for resistance and creativity.

Implications for Educators

- Promote active participation, collaboration, and reflection on the learning processes in digital environments.
- Design activities that help learners develop contextual knowledge and the critical navigation of digital spaces.
- Facilitate the development of skills related to social presence, multimodal communication, and civic agency online.

Implications for Learners

- Engage actively in online communities to develop multimodal communicative competence and participatory skills.
- Be aware of the power dynamics at play in online spaces.
- Reflect actively on your role in digital environments, and use your digital/social presence for both personal and collective goals.
- Leverage your online participation, where appropriate, for civic engagement and social change.

Summary

Participatory literacy (PL) encompasses the ability to contribute to and utilise the collective intelligence of digital networks, facilitating both personal and collective goals. Emerging in the early 2000s alongside a shift toward social technologies, PL reflects a move from merely consuming online content to actively engaging in digital culture. This evolution aligns with the so-called social turn in second language acquisition (SLA)—also mirrored in CALL—which emphasises language learning as a social practice with a focus on agency and identity.

PL, also termed "digital communicative literacy", serves as a foundation for online interactions and collaborative processes. It comprises several key elements: social presence, multimodal communicative competence, contextual knowledge, and collective intelligence. These elements are essential for effective participation in online environments, which have become central for language learning and broader social engagement.

Subsequently, scholars have pointed to the importance of context in online participation, arguing that successful engagement depends on awareness of social dynamics and the affordances of digital spaces. Recent perspectives on PL highlight its role in fostering civic agency, particularly in combating misinformation and enabling informed action in digital contexts.

As PL is integrated into educational frameworks, it empowers students to critically engage with digital content and participate meaningfully

in their communities. This critical dimension of PL underlines the need for learners to understand power dynamics online so that they can leverage digital tools for social justice and change. In sum, PL is vital for developing the skills necessary to navigate and contribute to an increasingly complex digital landscape.

Cross-References

▶ Critical Approaches to CALL
▶ Critical Digital Literacy
▶ Multimodal Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Block, D. (2003). *The social turn in second language acquisition*. Edinburgh University Press. ISBN: 0748615520.

Buchholz, B. A., DeHart, J., & Moorman, G. (2020). Digital citizenship during a global pandemic: Moving beyond digital literacy. *Journal of Adolescent & Adult Literacy, 64*(1), 11–17. https://doi-org.libezproxy.open.ac.uk/10.1002/jaal.1076

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105.

Giger, P. (2006). *Participation literacy: Part I: constructing the Web 2.0 concept series*. Blekinge Institute of Technology.

Hampel, R., & Hauck, M. (2006). Computer-mediated language learning: Making meaning in multimodal virtual learning spaces. *JALT CALL Journal, 2*, 3–18. https://doi.org/10.29140/jaltcall.v2n2.23

Hauck, M. (2010). Telecollaboration: At the interface between multimodal and intercultural communicative competence. In S. Guth & F. Helm (Eds.), *Telecollaboration 2.0: Language, literacies and intercultural learning in the 21st century* (pp. 219–244). Peter Lang. https://doi.org/10.5565/rev/jtl3.498

Hauck, M., & Satar, M. (2018). Learning and teaching languages in technology-mediated contexts: the relevance of social presence, copresence, participatory literacy and multimodal competence. In R. Kern & C. Develotte (Eds.), *Screens and scenes – online multimodal communication and intercultural encounters* (pp. 133–157). Routledge. https://doi.org/10.4324/9781315447124

Hauck, M., & Warnecke, S. (2012). Materials design in CALL: social presence in online environments. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary computer-assisted language learning* (pp. 95–115). Bloomsbury. https://doi-org.libezproxy.open.ac.uk/10.1111/modl.12195

Hauck, M., Galley, R., & Warnecke, S. (2016). Researching participatory literacy and positioning in online learning communities. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 71–87). Routledge. https://doi-org.libezproxy.open.ac.uk/10.4324/9781315657899

Helm, F. (2015). The practices and challenges of telecollaboration in higher education in Europe. *Language Learning & Technology, 19*(2), 197–217. http://dx.doi.org/10125/44424

Holmes, C., & McGlinn Mafra, M. (2020). Participatory literacy and taking informed action in the social studies. In J. S. Mitchell, E. N. Vaughn, & E. N. (Eds.), *Participatory literacy practices for P-12 classrooms in the digital age*. IGI Global. https://doi-org.libezproxy.open.ac.uk/10.4018/978-1-7998-0000-2

Hoven, D. (2006). *Designing for disruption: Remodeling a blended course in technology in (language) teacher education*. In Presented at the 23rd Annual Ascilite Conference: Who's learning? Whose technology? University of Sydney, Sydney, Australia.

Jenkins, H., Clinton, K., Purushotma, R., Robison, A. J., & Weigel, M. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century*. MacArthur Foundation. http://digitallearning.macfound.org/atf/cf/%7B7E45C7E0-A3E0-4B89-AC9C-E807E1B0AE4E%7D/JENKINS_WHITE_PAPER.PDF

Jenkins, H., Clinton, K. A., Purushotma, R., Robinson, A. J., & Weigel, M. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. MIT Press. https://doi.org/10.7551/mitpress/8435.001.000

Jiang, L., Yang, M., & Yu, S. (2020). Chinese ethnic minority students' investment in English learning empowered by digital multimodal composing. *TESOL Quarterly, 54*(4), 954–979. https://doi-org.libezproxy.open.ac.uk/10.1002/tesq.566

Johnson, K. E. (2006). The sociocultural turn and its challenges for second language teacher education. *TESOL Quarterly, 40*(1), 235–257. https://doi.org/10.2307/40264518

Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education, 29*(1), 89–106. https://doi.org/10.1080/01587910802004860

Kehrwald, B. (2010). Being online: Social presence and subjectivity in online learning. *London Review of Education, 8*(1), 39–50. https://doi.org/10.1080/14748460903557688

Kern, R., & Warschauer, M. (2000). Introduction: Theory and practice of network-based language teaching. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 1–19). Cambridge University Press. https://doi.org/10.1017/CBO9781139524735

Kern, R., Ware, P., & Warschauer, M. (2004). Crossing frontiers: New directions in online pedagogy and research. *Annual Review of Applied Linguistics, 24*(1), 243–260. https://doi.org/10.1017/S0267190504000091

Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

Mirra, N., & Garcia, A. (2020). "I Hesitate but I Do Have Hope": Youth speculative civic literacies for troubled times. *Harvard Educational Review, 90*(2), 295–321. https://doi-org.libezproxy.open.ac.uk/10.17763/1943-5045-90.2.295

Ortega, L. (2011). SLA after the social turn: Where cognitivism and its alternatives stand. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 167–180). Routledge. https://doi.org/10.4324/9780203830932

Pegrum, M. (2009). *From blogs to bombs: The future of digital technologies in education*. Perth: University of Western Australia Press.

Pegrum, M. (2010). I link, therefore I am: Network literacy as a core digital literacy. *E-Learning and Digital Media, 7*(4), 346–354. https://doi.org/10.2304/elea.2010.7.4.346

Pegrum, M., Hockly, N., & Dudeney, G. (2022). *Digital literacies*. Routledge.

Royce, T. (2007). Multimodal communicative competence in second language contexts. In T. Royce & W. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 361–390). Routledge.

Wang, V., & Torrisi-Steele, G. (2023). LMS course design for adult learning: Heutagogy, andragogy, pedagogy. In V. Wang (Ed.), *Handbook of research on andragogical leadership and technology in a modern world* (pp. 1–17). IGI Global. https://doi-org.libezproxy.open.ac.uk/10.4018/978-1-6684-7832-5.ch00

Wohlwend, K. (2017). Chasing literacies across action texts and augmented realities: e-Books, animated apps, and Pokémon Go. In *The Case of the iPad* (pp. 49–66). Springer. https://doi.org/10.1007/978-981-10-4364-2_4

S

Social Media Literacy

Jessie S. Barrot
National University, Manila, Philippines

Abstract

In the digital age, social media literacy (SML) is crucial for understanding and navigating modern communication. This entry examines the evolving concept of SML and its integration into language pedagogy, focusing on its nature, application, and associated challenges. It explores how SML extends beyond traditional literacy and differs from broader media literacy while also addressing the practical aspects of SML, particularly within second language (L2) learning. Strategies for incorporating SML into language teaching are outlined, including analysing social media language, creating digital content, and fostering intercultural competence. Additionally, challenges such as unequal access to social media, content curation, and integrating digital citizenship are discussed. Despite these challenges, incorporating SML remains vital for equipping students to engage with and contribute to the digital communication landscape effectively.

Keywords

Social media · Social networking sites · Media literacy · Social media literacy · Digital literacies

Definition

Social media literacy is the ability to critically understand and manage one's interactions with social media, incorporating content choices, consumption habits, engagement patterns, and platform technologies, in order to use social media effectively and purposefully.

Introduction

In today's digital era, literacy has broadened to include social media literacy (SML), an integral aspect of ▶ digital literacies that signal the significant influence of social media on communication, information sharing, and societal discourse (Cho et al., 2024; Kalorth et al., 2020). Key platforms such as Facebook, X, Instagram, and TikTok have become central to communication. As social media continues to shape communication patterns, computer-assisted language learning (CALL), and language use (Barrot, 2022), understanding and clarifying SML becomes essential. Thus, this entry explores SML by examining it from theoretical and practical perspectives, discussing its integration into language teaching,

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_252-1

and addressing challenges and considerations in language pedagogy.

Social Media Literacy

As media technologies have evolved, literacy has expanded beyond traditional text-based forms to include audiovisual, digital, and information literacies (Lee et al., 2015; ▶ "Information Literacy"). New media literacy includes the ability to consume, create, and share media, critically evaluate content, and understand its social, political, and cultural contexts (Jenkins, 2011; Koc & Barut, 2016; Luo et al., 2022; ▶ "Media Literacy"). With the rise of social media, the focus has shifted to the skills needed to select and navigate these platforms, including understanding online interactions and managing online identities and privacy concerns. A distinct subset of new media literacy, social media literacy (SML) refers to one's ability to critically understand and manage one's interactions with social media, incorporating content choices, consumption habits, and engagement patterns and effectively operating within the network environment while also comprehending the technological features and underlying structures of social media platforms (Festl, 2021; Polanco-Levicán & Salvo-Garrido, 2022). While broader media literacy frameworks cover general digital competencies, SML specifically targets the skills required for effective social media engagement.

Cho et al. (2024) describe SML as an interactive process where user choices and social media environments influence each other, shaping users' reality. Users select what content to engage with, which shapes their exposure to information, beliefs, and perceptions. Emotions and personal values often guide these choices, creating a feedback loop between user preferences and the content they encounter. Thus, they propose the social media literacy model (SoMeLit), which highlights the role of family, peers, and educators in developing cognitive and emotional processing abilities for social media engagement. This model addresses distinct features of social media, such as algorithm-driven content, participatory culture, and real-time interaction, which are not fully considered in traditional media literacy models.

Meanwhile, Manca et al. (2021) argue for a "glocal" approach to SML, combining global and local social media skills. This approach aligns with ▶ intercultural literacy, which involves understanding and adapting to diverse cultural norms and communication styles (Wang et al., 2021). Just as intercultural literacy helps individuals engage across cultural contexts, a glocal approach to SML ensures users can engage in global practices while remaining sensitive to local dynamics. Manca et al. (2021) further emphasise the need for both transversal skills across platforms and platform-specific competencies, reflecting the variety of social media landscapes.

To cultivate SML, individuals need to understand the features and purposes of different platforms, enabling them to engage effectively and critically evaluate content. Critical consumption skills, such as verifying sources and identifying biases, are essential for assessing the credibility of shared information (Solmaz, 2017; Tandoc Jr et al., 2021). Sensationalised content and clickbait (content designed to attract attention and encourage clicks, often at the expense of accuracy or quality) require particular caution. Integrating new media literacy into language curricula can further enhance SML development.

Social Media Literacy for L2 Learners

SML requires understanding and producing multimodal content that combines text, images, videos, and other media, where meaning is conveyed through both linguistic and nonlinguistic elements (Lotherington & Jenson, 2011; Yi, 2014). Creating multimodal content demands linguistic proficiency, design awareness, and cultural sensitivity, aligning with ▶ multimodal literacy, which integrates various semiotic modes into coherent communication (Walsh, 2010). Moreover, effective engagement with multimodal content requires not only technical skills but also the capacity to critically analyse how different modes interact to shape meaning. This analysis includes

understanding how visuals, sound, and textual elements work together to influence interpretation and audience perception.

L2 learners must not only grasp how different modes interact but must decode the language as well as the cultural and symbolic nuances embedded in the content. SML involves recognising diverse language inputs and understanding their differences from formal language use, offering exposure to authentic communication. Learners must also interpret cultural cues, which shape their digital identity and online presence. This process includes creating posts or videos that balance linguistic proficiency with cultural awareness. SML also enhances language acquisition and fluency, as learners use interactive tools like online communities and forums to practise skills, engage with credible sources, and participate in supportive platforms.

While L1 learners are typically familiar with cultural references in social media content, allowing them to focus on critical evaluation, L2 learners face the dual challenge of developing language skills while navigating unfamiliar cultural norms. As a result, SML instruction for L2 learners should incorporate both ▶ intercultural literacy and critical thinking, along with digital citizenship skills. Introducing digital citizenship training that emphasises responsible behaviour in social media further supports SML development within CALL. Such training should cover privacy protection, online etiquette, and cyberbullying prevention, ensuring learners become informed digital participants. By fostering critical engagement with social media content through discussions and educational resources, educators can deepen students' understanding and promote ethical participation in digital spaces.

Integrating SML into Language Pedagogy

Integrating SML into language pedagogy represents a contemporary approach to language education that acknowledges the profound influence of digital communication on language use and cultural exchange. First, language teachers can incorporate SML into their teaching practices by designing activities that require students to analyse the use of different language registers—such as formal, informal, and colloquial language—across platforms like TikTok, Instagram, or YouTube. Through critical analysis of how language content adapts to various digital environments and how these environments influence language in a feedback loop, students gain valuable insights. Students learn to recognise how language shifts across different digital platforms, such as the informal tone of social media versus the structured language of academic forums. They also gain insights into how digital environments shape communication styles, which enables them to adapt their writing for specific audiences and purposes. Teachers can also curate authentic language materials from social media, including articles, videos, podcasts, and memes (▶ "Meme Literacy"), and incorporate them into classroom activities to expose students to real-world language use, idiomatic expressions, and cultural references prevalent in digital communication.

Second, social media platforms offer language learners opportunities to create and share their own content, enhancing language proficiency and communication skills in diverse digital contexts. Assignments involving the production of written posts, videos, or podcasts in the target language for platforms like Facebook, TikTok, or LinkedIn encourage effective communication, while tasks designed in the style of these platforms—without actual posting—address privacy concerns. Additionally, language pedagogy benefits from promoting autonomous language use on social media. Teachers can guide students to follow accounts or channels in the target language, participate in online discussions, and engage with media that interests them, such as films, music, or blogs.

Finally, social media platforms facilitate interactions with individuals from diverse cultural backgrounds. Language teachers can leverage these platforms to promote intercultural competence among students. By engaging in online discussions, collaborating on language projects, or participating in language exchange programmes with speakers of the target language from different

cultural backgrounds, students develop cross-cultural communication skills and gain insights into cultural nuances. However, these interactions need to be carefully scaffolded and guided to avoid misunderstandings or entrenchment of stereotypes (Varmazyar & Cardama, 2023).

Challenges and Considerations

Developing SML within the context of language teaching and learning poses several challenges. Not all students may have equal access to social media platforms, leading to disparities in digital literacy skills. Additionally, the vast amount of content on social media makes it challenging for teachers to curate high-quality and relevant materials for language learning while ensuring student privacy and safety. To address these concerns, there is a range of content creation apps that allow students to engage in similar activities without directly using public social media platforms. For example, apps like Adobe Spark and Canva offer controlled environments for producing videos, visual presentations, and multimedia content. Integrating digital citizenship education into language teaching requires additional time and expertise, as teachers manage privacy settings, online safety protocols, and potential risks such as cyberbullying and inappropriate content. Furthermore, effectively integrating social media into language teaching requires pedagogical planning and alignment with learning objectives, as well as ongoing professional development for teachers to build confidence and competence in using social media tools. Assessing SML skills within language learning contexts presents another challenge, as traditional assessment methods may not capture students' ability to critically evaluate information, engage responsibly online, and communicate effectively in digital environments. Addressing these challenges requires collaboration among teachers, administrators, policymakers, and technology experts to ensure that social media integration in language teaching and learning is inclusive and effective.

Conclusion

As society increasingly relies on social media platforms for information dissemination and interpersonal connection, integrating SML into language teaching and learning becomes not just advantageous but imperative. Equipping students with the ability to critically evaluate information, practise creativity, and engage responsibly in online environments enables them to effectively manage the complexities of the social media environment with greater proficiency. Emphasising SML in language pedagogy fosters responsible digital citizenship and prepares learners to actively shape and contribute to the evolving landscape of communication in an increasingly digital society.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Barrot, J. S. (2022). Social media as a language learning environment: A systematic review of the literature (2008–2019). *Computer Assisted Language Learning, 35*(9), 2534–2562.

Cho, H., Cannon, J., Lopez, R., & Li, W. (2024). Social media literacy: A conceptual framework. *New Media & Society, 26*(2), 941–960.

Festl, R. (2021). Social media literacy & adolescent social online behavior in Germany. *Journal of Children and Media, 15*(2), 249–271.

Jenkins, H. (2011). Convergence culture. Where old and new media collide. *Revista Austral de Ciencias Sociales, 20*, 129–133.

Kalorth, N., Verma, M., & Sagar, M. (2020). Information and user: Social media literacy in digital societies. *Journal of Content, Community and Communication, 12*, 263–269.

Koc, M., & Barut, E. (2016). Development and validation of New Media Literacy Scale (NMLS) for university students. *Computers in Human Behavior, 63*, 834–843.

Lee, L., Chen, D. T., Li, J. Y., & Lin, T. B. (2015). Understanding new media literacy: The development of a measuring instrument. *Computers & Education, 85*, 84–93.

Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics, 31*, 226–246.

Luo, Y. F., Yang, S. C., & Kang, S. (2022). New media literacy and news trustworthiness: An application of importance–performance analysis. *Computers & Education, 185*, 104529.

Manca, S., Bocconi, S., & Gleason, B. (2021). "Think globally, act locally": A glocal approach to the development of social media literacy. *Computers & Education, 160*, 104025.

Polanco-Levicán, K., & Salvo-Garrido, S. (2022). Understanding social media literacy: A systematic review of the concept and its competences. *International Journal of Environmental Research and Public Health, 19*(14), 8807.

Solmaz, O. (2017). Adapting new media literacies to participatory spaces: Social media literacy practices of multilingual students. *Journal of Media Literacy Education, 9*(1), 36–63.

Tandoc Jr, E. C., Yee, A. Z., Ong, J., Lee, J. C. B., Xu, D., Han, Z., . . . & Cayabyab, M. Y. (2021). Developing a perceived social media literacy scale: Evidence from Singapore. *International Journal of Communication, 15*, 2484–2505.

Varmazyar, R., & Cardama, S. M. (2023). Stereotyped-information behavior caused by social media: a case study of tiktok. *Revista EDICIC, 3*(3), 1–20.

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *The Australian Journal of Language and Literacy, 33*(3), 211–239.

Wang, X., Jiang, L., Fang, F., & Elyas, T. (2021). Toward critical intercultural literacy enhancement of university students in China from the perspective of English as a lingua franca. *SAGE Open, 11*(2), 1–12.

Yi, Y. (2014). Possibilities and challenges of multimodal literacy practices in teaching and learning English as an additional language. *Language and Linguistics Compass, 8*(4), 158–169.

H

Hashtag Literacy

Yeong-Ju Lee
Macquarie University, Sydney, NSW, Australia

Abstract

This entry explores hashtag literacy, highlighting its role in facilitating collaborative language learning, especially on popular multimodal social media platforms like Instagram and TikTok. Hashtag literacy entails using hashtags to navigate, discover, and categorise digital content on specific topics. It enables language learners to access targeted educational content and engage in digital communications and community building with other learners. However, the contemporary perspective on hashtag literacy extends beyond these functionalities to include users' active participation in collaborative content creation and trend-setting. This advancement in hashtag literacy practices empowers language learners not only to access but also to contribute co-creatively to educational content, thus facilitating more dynamic and interactive learning experiences. This entry demonstrates how hashtag literacy can enhance pedagogical strategies and promote learner engagement in both formal curricula and informal autonomous learning environments.

Keywords

Hashtag literacy · Collaboration · Interactive language learning · Multimodal social media · Digital content creation · Digital literacies

Introduction

Hashtags are words or phrases prefaced by the hash symbol # to mark digital content on specific topics or with specific perspectives. Hashtag literacy refers to the skills and knowledge required to effectively use hashtags for enhancing digital communication and interaction. It involves understanding how to use hashtags to categorise and access information on social media platforms, engage with specific communities, and participate in content creation and trend-setting.

Hashtag literacy is a key aspect of → digital literacy that involves comprehending how to use hashtags to categorise and access information, engage with specific communities, and participate in digital interaction and collaboration on social media platforms. Since hashtags were introduced as a means to group related messages on Twitter in 2007, hashtag literacy has expanded its role from basic content categorisation and search to facilitating digital communications and community building on specific topics or perspectives (Sykes, 2019). However, the contemporary perspective on hashtag literacy extends beyond these functionalities to include users' active participation in collaborative content creation

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_253-1

and trend-setting, especially on popular multimodal social media platforms like Instagram and TikTok (Lee, 2023). Therefore, this entry focuses on this advancement towards such contemporary hashtag literacy practices, suggesting how they can enhance pedagogical strategies and promote learner engagement in computer-assisted language learning (CALL).

In line with the 'social turn' in second language acquisition (SLA) that views language development as a social phenomenon (Block, 2003), hashtag literacy has emphasised the social aspect of hashtag use, which fosters community-based digital communication and interaction (Reinhardt, 2019; Sykes, 2019). Initially, on platforms like Twitter, hashtags were mainly used for categorisation and as 'searchable' tags, allowing users to find specific content, including language learners who can find targeted educational content and engage in community discussions with other learners or educators (Zappavigna, 2015). Developments on user-generated content-sharing platforms such as YouTube and blogs, which promote 'a participatory culture' encouraging users' active participation in creating, sharing, and collaborating on content, have shifted the focus towards fostering more interactions among users, including language learners with their particular focus and needs (Androutsopoulos, 2013). This shift has expanded the role of hashtag literacy, enabling language learners not only to create and propagate hashtags but also to shape discourse and content in relation to their learning on social media platforms. As a result, hashtag literacy has encouraged language learners to take ownership of their learning process through building communities around shared interests, engaging in meaningful interactions, and contributing to the collective knowledge (Androutsopoulos, 2013; Reinhardt, 2019).

The growing impact of participatory hashtag literacy is particularly evident on popular multimodal platforms like Instagram and TikTok (Lee, 2023). On Instagram, users often share language learning resources by creating image posts incorporating hashtags (e.g. #Spanishlearning and #learnEnglish). Given that Instagram affords visual rather than textual communication, hashtags on this platform indicate community participation or provide context for images, rather than linking posts to text-based conversations as on Twitter (Laestadius, 2017). On Instagram, hashtag literacy thus requires strategic hashtag use in multimodal content creation, promoting collaborative learning with users of similar interests. On TikTok, the collaborative aspect is even more prominent, with hashtags like #challenge used to create and participate in trends involving competition or achievement (e.g. #language-challenge and #accentchallenge). These challenges involve language learning tasks such as vocabulary learning and pronunciation practice (Vazquez-Calvo et al., 2022). The tagging also enables these challenges to be widely shared on the platform, encouraging broader participation, including those who may have never considered learning a language but may find such challenges interesting (Lee, 2023). In this sense, using hashtags for trend-setting enables language learners not just to access but also to contribute co-creatively to educational content within a wider community. Therefore, these contemporary hashtag literacy practices facilitate more dynamic, collaborative learning experiences.

Hashtag literacy has significant pedagogical implications to enhance student engagement in both formal curricula and informal autonomous language learning environments.

Implications for formal curricula:

- **Analysing hashtags:** Teachers can implement an activity to analyse popular hashtags related to students' target language or culture by exploring tagged content. In a Korean language class, for instance, students could examine hashtags like #KpopKorean. They could categorise posts by identifying themes and discuss linguistic insights from these posts. This activity would help students understand the role of hashtags in shaping online content and obtaining useful learning resources.
- **Incorporating hashtags into assignments:** Teachers can ask students to use hashtags when posting their language learning assignments on social media platforms. For instance, students could share visual materials such as representations of idioms in everyday settings or digital posters on particular topics using

relevant hashtags on Instagram or TikTok. This approach would encourage students to deeply engage with the material, knowing that their work can be possibly viewed by other users.

- **Creating class-specific hashtags:** Teachers can create unique hashtags for their classes to facilitate resource sharing and peer interaction, shaping a controlled online environment. For example, a Korean language teacher could create a hashtag like #SeoulScholars. Students could use this hashtag when discussing language topics or posting questions on a designated social media platform. This method can foster a sense of community among students as they can view and comment on each other's work. It can also help teachers track and evaluate the students' progress by searching for the specific hashtag.

Implications for informal autonomous learning:

- **Participating in online language challenges:** Learners can join language learning challenges or tasks on platforms like TikTok and Instagram, using hashtags such as #languagechallenge, #accentchallenge, and #dailyvocab to connect with other learners and track their progress. These challenges would promote interactive learning and consistent practice, with self-directed tasks such as daily vocabulary exercises or weekly pronunciation practice.
- **Engaging with language teaching/learning content creators:** Hashtags can help learners find and interact with language teachers and other learners who share linguistic knowledge and learning experiences or tips. By building a supportive community with users of similar interests, they can gain feedback and encouragement. Through community interactions, they can also learn authentic language use and experience cultural exchanges.

As social media platforms evolved, so did the use of hashtags, encouraging users not only to access digital content but also to contribute to community-based interaction and collaboration. By understanding this advancement in hashtag literacy, it is now important to take a more engaging approach to language education by leveraging hashtags to contribute to digital content and enhance learners' experiences on social media platforms.

Cross-References

► Digital Literacies

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Androutsopoulos, J. (2013). Participatory culture and metalinguistic discourse: Performing and negotiating German dialects on YouTube. In D. Tannen & A. M. Trester (Eds.), *Discourse 2.0: Language and new media* (pp. 47–71). Georgetown University Press.

Block, D. (2003). *The social turn in second language acquisition*. Georgetown University Press.

Laestadius, L. (2017). Instagram. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 573–592). SAGE Publications.

Lee, Y.-J. (2023). Language learning affordances of Instagram and TikTok. *Innovation in Language Learning and Teaching, 17*(2), 408–423. https://doi.org/10.1080/17501229.2022.2051517

Reinhardt, J. (2019). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching, 52*(1), 1–39. https://doi.org/10.1017/S0261444818000356

Sykes, J.-M. (2019). Emergent digital discourses: What can we learn from hashtags and digital games to expand learners' second language repertoire? *Annual Review of Applied Linguistics, 39*, 128–145. https://doi.org/10.1017/S0267190519000138

Vazquez-Calvo, B., Zhang, L.-T., & Shafirova, L. (2022). Language learning hashtags on TikTok in Chinese, Italian, and Russian. In L. Klimanova (Ed.), *Identity, multilingualism and CALL: Responding to new global realities* (pp. 104–134). Equinox Publishing Ltd.

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics, 25*(3), 1–18. https://doi.org/10.1080/10350330.2014.996948

Intercultural Literacy

Florence Le Baron-Earle and
Marie-Thérèse Batardière
University of Limerick, Limerick, Ireland

Abstract

This entry explores the notion of intercultural literacy—a lifelong process which involves competencies enabling cross-cultural understanding and positive experiences between people from diverse backgrounds. Intercultural literacy is not merely about acquiring factual knowledge of various cultures; it also entails developing empathy and the capacity to see the world from multiple perspectives (also known as ethnorelativism). It requires active listening, open-mindedness, and the willingness to suspend judgement while engaging with people from other cultural backgrounds. In an increasingly globalised and interconnected world, intercultural literacy is essential for successful cross-cultural collaboration in a variety of contexts—both formal and informal—especially in education, healthcare, business, and diplomacy. To this end, the development of intercultural competencies is widely promoted by educational institutions as part of their internationalisation strategies. Furthermore, actively promoting intercultural literacy can lead to more equitable and harmonious communities. This entry shows that computer-assisted language learning, in addition to face-to-face encounters, offers unprecedented opportunities for learners to build sustainable dialogues and achieve a sense of belonging to a larger global community.

Keywords

Intercultural literacy · Cultural awareness · Intercultural competencies · Intercultural sensitivity · Ethnorelativism · Intercultural citizenship · Global citizenship · Digital literacies

Definition

Intercultural literacy is the ability to understand, respect, and effectively communicate with individuals from diverse cultural backgrounds. It encompasses knowledge of different cultural practices, languages, and worldviews, as well as the skills to navigate and bridge cultural differences. Intercultural literacy involves a deep appreciation for cultural diversity and an awareness of one's own cultural lens and biases.

Intercultural literacy is understood to be a lifelong process. It is widely accepted that 'intercultural competence' comprises five key components (Byram, 1997): (1) *Attitudes* involving having an open-minded and a respectful approach towards other cultures, valuing diversity, and being willing to learn; (2) *Knowledge* such as understanding cultural differences, norms, values, and historical

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_254-1

contexts; (3) *Skills of Interpreting and Relating* which enable effective communication across cultures, e.g. interpreting verbal and non-verbal cues and building relationships; (4) *Skills of Discovery and Interaction* which involve actively seeking out cultural information and engaging in meaningful interactions; (5) *Critical Cultural Awareness or Political Education* which highlights the need to understand power dynamics, social structures, and the impact of culture on behaviour.

The journey of individuals acquiring intercultural literacy implies both internal and external outcomes. The internal outcome is an intimate state which is not easily observable; it represents the internal changes or the shift in frame of reference (Deardorff, 2009) occurring within the individuals who become interculturally competent. It includes qualities of adaptability, flexibility, an ethnorelative view, and empathy. The external outcome concerns the capacity to behave and communicate appropriately in an intercultural situation.

Evolution of Intercultural Literacy Through the Years

In the field of foreign language education, the notion of intercultural literacy has evolved since the 1960s from 'cultural awareness' (Brooks, 1968; Hall, 1976) or 'cultural sensitivity' (Bennett, 1993) to 'cross-cultural communication' (Byram, 1997; Kramsch, 1993) or 'intercultural mediation' (Byram, 1997). Early on, it was emphasised that teaching a foreign language should take place concurrently with teaching culture as the two are intertwined. The notion of culture, in this case, is differentiated into high culture (e.g. literature and the fine arts) on the one hand and deep culture (Brooks, 1968) or the 'silent language' (Hall, 1973) on the other hand. These concepts are also referred to as 'big C' (i.e. high culture such as history, literary, and visual arts) and 'little c' (i.e. anthropological culture such as values, body language, and social practice). In brief, it was recommended that less visible components of a culture such as norms, values, and beliefs were studied in parallel with language acquisition. This cultural awareness focus enabled learners to better understand their own culture, and those of others, and communicate across cultures more effectively.

The concept of cultural competence gained further prominence in the 1980s with the emergence of communicative pedagogy. At the time, language teaching methods shifted from a focus on linguistic competence (grammar, vocabulary, syntax, pronunciation) to communicative competence, i.e. communicating appropriately according to various social situations. This trend continued in the following decades as can be seen in the Council of Europe publications, notably the Common European Framework of Reference for Languages, which provides comprehensive descriptions of knowledge and skills to be acquired by language learners. The original publication (2001) mentioned levels of 'sociocultural competence' required. This was reiterated further in the 2020 Companion Volume where communicative skills of 'mediation', 'sociolinguistic appropriateness', and 'cultural repertoire' were added. In this context, building a pluricultural repertoire (i.e. a unique set of resources drawing on multiple cultures and experiences; ▶ Pluriliteracy) includes but is not limited to the capacity to deal with otherness and ambiguity, the willingness to show sensitivity to differences, and act as an intercultural mediator (2020, p. 124).

One element worthy of note is that while developing 'intercultural competence' is a key aspect of language learning and teaching, it can also be acquired by individuals who do not speak the language of a specific culture. It describes general abilities, as explained earlier in this entry, for interacting with individuals from various cultures regardless of the language used. In contrast, the term 'intercultural *communicative* competence' indicates that a specific individual can speak the language from the target culture (Byram, 1997).

More recently, intercultural literacy has been focusing on 'intercultural citizenship' or 'global citizenship'. These terms signify that as individuals we are all part of a broader global community and that our responsibilities and rights extend beyond national borders. Being interculturally literate entails going further than just changing one's

own attitudes and perspectives; civic engagement and action towards social justice should follow. Such commitments can apply to advocacy towards educational, social, environmental, and economic issues (Council of Europe, 2024).

Challenges in the Modern Digital Age

The recent digital transformation and the globalisation phenomena can facilitate but also hinder linguistic and cultural diversity. While they can be seen as providing opportunities to connect individuals from varied cultures wherever they are located geographically, these developments can also pose significant challenges to intercultural understanding. For instance, they may accelerate the dominance of certain languages and cultures, and as such can cause identity loss, exclusion, and possibly conflict (UNESCO, 2022). Moreover, the overreliance on social media as a source of information or means of communication can create the risk of echo chambers (Olivetti de França et al., 2023) and narrowing intercultural dialogue (▶ Information Literacy). This can be exacerbated by low attention span (Hari, 2022), soundbite journalism (Rinke, 2016), and the advent of monetised digital spaces where polarised opinions prosper (Boler et al., 2024) (▶ Attentional Literacy). In addition, social media sites' algorithms curate feeds based on past behaviours and preferences and impact the content users are exposed to (▶ Media Literacy). By aligning with existing tastes, they limit exposure to diverse points of view (Törnberg et al., 2022) and can reinforce ethnocentrism. In face of these affordances and challenges, UNESCO (2024) calls upon educators and institutions to promote digital and intercultural competencies and foster user agency (▶ Critical Digital Literacy).

The Role of Education and CALL

When used with research-based pedagogy, Internet technologies can transform users' language learning experience through community building, information exchange, co-construction of knowledge, and negotiation of meaning (Godwin-Jones, 2019), all of which are requisite in attaining an intercultural stance. From the early 2000s, practitioners have favoured human computer-mediated communication rather than human-to-computer interaction. Innovative projects that were initially developed with technology-savvy practitioners, e.g. the *Cultura* model (Furstenberg et al., 2001), surged with the democratisation of social media sites. This was a catalyst for telecollaborative initiatives (Guth & Helm, 2010) on multiple platforms, e.g. discussion fora, blogs, Skype and wikis, 3D life-simulation networks such as Second Life, and massively multiplayer online games.

Online intercultural exchanges (Lewis & O'Dowd, 2016), also known as collaborative online international learning (Haug & Jacobs, 2023) or virtual exchange (Helm & van der Velden, 2019), are learner-led interactions and tailored activities facilitated by educators, and enabled by digital or mobile technologies (▶ Mobile Literacy). These (a)synchronous exchanges unfold over time to allow for deep and meaningful intercultural dialogue. In bridging differences and giving a sense of belonging (Evolve, 2024), VE practice supports the fluid concept of intercultural literacy. Real-world tasks (Ollivier, 2018) carried out in virtual exchange can be concerned with contemporary issues of sustainability, equality, and peace as set out in the UN Sustainable Development Goals. VE participants can embody global citizenship by engaging in activities that promote these goals.

There are three types of virtual exchanges; they can be (1) used as blended mobility, i.e. a component of physical mobility or a preparatory/follow-up activity to physical mobility, e.g. *eTandem* (Griggio & Pittarello, 2020); (2) a stand-alone learning project, e.g. *Soliya Connect* which concentrates on conflict resolution (Soliya, 2024); or (3) be integrated in the curriculum like the *Erasmus Speaks* project (O'Regan et al., 2023). Their design enables the development of critical thinking (Duffy et al., 2020), transferable skills (Le Baron-Earle & O'Reilly, 2023), and sought-after graduate attributes that are characteristics of intercultural literacy.

Guidelines on VE Implementation for Educators

For virtual exchanges to be successful in developing intercultural literacy, a number of steps are recommended before embarking on an online partnership:

Prior to Implementation

- Understanding what a virtual exchange is and what it is not. See EVOLVE definition: https://evolve-erasmus.eu/about-evolve/what-is-virtual-exchange/
- Familiarising yourself with the types of VE available. Some are fully integrated in a specific curriculum and others are stand-alone. See various scenarios: https://frames-project.eu/outputs/scenarios/
- Signing up to a webinar or online training. See options offered by UNICollaboration: https://www.unicollaboration.org/index.php/virtual-exchange-training-2/
- Sourcing solid partners. Ensure that your objectives are aligned and schedules are compatible. Be satisfied that your potential partner's flexibility and communicative reactivity are adequate as slow responsiveness can be a red flag. You can meet them at partnering fairs, online VE training sessions or use the virtual exchange partner finding tool: https://www.unicollaboration.org/index.php/finding-a-ve-partner/
- Identifying learning outcomes and designing tasks accordingly. To develop participants' intercultural literacy, tasks should be scaffolded to elicit curiosity, empathetic and dialogic interaction, and fruitful collaboration.
- Deciding on timeline and modes of communication. Agree with your partner(s) on deadlines and where participants should submit their assignments.

During Implementation

- Outlining objectives to participants. For instance, is the main aim of the VE to develop intercultural competencies or language skills? Or both?
- Explaining tasks and marking criteria to participants. Allow time for questions and clarifications.
- Providing a clear time frame and place for delivery of assignments. Make sure that the same information is given to all participants. If there are institutional differences, they need to be made transparent. For example, students from one institution may receive a grade for their work while students of the partner (s) institution take part in the VE on a voluntary basis.
- Monitoring and supporting students to maintain engagement. This can be achieved by having regular online or in-class contact with participants, checking progress, and offering video tutorials on using digital tools. Also, strongly encourage students to have synchronous as well as asynchronous communication with their partner(s).
- Agreeing on deadline extensions with partners if required.

After Implementation

- Receiving accreditation. Completing a virtual exchange can help develop a wide range of skills. Open badges are a means for participants—students and facilitators—to showcase their newly acquired transversal skills (e.g. collaboration and intercultural competencies for the former, designing and implementing for the latter) on their social media profiles or curricula vitae. Learn more about open badges and their potential to boost employability from UNICollaboration: https://www.unicollaboration.org/index.php/virtual-exchange-training/recognition-of-virtual-exchange/
- Organising a debrief with students and partner (s). This enables you to learn what worked well and what needs improvement for future iterations. Feedback can be collected anecdotally in class or more formally via student questionnaires. It is also essential to share your experience with partners and decide whether and how you wish to pursue the exchange.

Guidelines on VE Participation for Students

For participants to fully benefit from the virtual exchange, here are some practical recommendations:

- Understanding the tasks. Make sure that you correctly interpret the work to be carried out. Ask your teacher(s) or facilitator(s) for clarification to alleviate any ambiguity. Check with your partner(s), once contact has been established, that you both have the same understanding of what the assignment entails.
- Contacting your partner(s). Get in touch with your partner(s) as soon as you have their contact details. Note that asynchronous modes of communication (e.g. email, text message) can be helpful when there are gaps in your schedules or a time difference between your geographical locations. However, do not forget to include synchronous modes of communication where you can see and hear each other, such as a video call. This allows for better communication and greater bonding.
- Maintaining engagement. Agree on regular contact times. Let your partner(s) know if there are times when you are not reachable. This could be due to out-of-hours responsibilities or planned periods of leave. For instance, scheduled holidays or examination times may vary from one institution to another. Make arrangements to alleviate periods of busyness or absence.
- Facing breakdown in communication. Do not delay in informing your teacher(s) or facilitator(s) if there is failure of communication with your partner(s).
- Completing tasks and meeting deadlines. Plan ahead and ensure that you finish the assignment on time, especially if substantial input from your partner(s) is needed.
- Showcasing accreditation. If your VE involves the award of an open badge, download it at the end of the project, and display it where appropriate (social media profile, CV or in your open badge passport—https://openbadgepassport.com/). Promoting your newly acquired skills can support job applications and/or promotions.

Intercultural literacy is a cornerstone of modern language education. Educators play an important role in enabling learners to engage with individuals from diverse cultures. In the era of internationalisation at home (Beelen & Jones, 2015), fostering global citizenship can be achieved notably by integrating virtual exchanges in teaching practice. Sustained intercultural interaction promotes cooperation and empathy and can help bridge cultural divides.

Cross-References

- ▶ Attentional Literacy
- ▶ Critical Digital Literacy
- ▶ Information Literacy
- ▶ Media Literacy
- ▶ Mobile Literacy
- ▶ Pluriliteracy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Beelen, J., & Jones, E. (2015). Redefining internationalization at home. In A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott (Eds.), *The European higher education area*. Springer. https://doi.org/10.1007/978-3-319-20877-0_5

Bennett, M. J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In M. R. Paige (Ed.), *Education for the intercultural experience* (pp. 21–71). Intercultural Press.

Boler, M., Yoon, J. K., & Michilin, T. (2024). *Rethinking polarization through the social media dispositif: Affect, melodrama, and digital governmentality in online cross-partisan debate*. Social Media and Society.

Brooks, N. (1968). *Teaching culture in the foreign language classroom*. American Council on the Teaching of Foreign Languages.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Council of Europe. (2001). *Common European framework of reference for languages: Companion volume*. Council of Europe. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Council of Europe. (2024) Intercultural citizenship. https://www.coe.int/en/web/interculturalcities/icc-test#:~:text=What%20is%20intercultural%20citizenship%3F,

cultures%20can%20be%20an%20advantage. Accessed 22 July 2024.

Deardorff, D. K. (2009). Implementing intercultural competence assessment. In D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (pp. 477–491). Sage.

Duffy, L. N., Stone, G. A., Townsend, J., & Cathey, J. (2020). Rethinking curriculum internationalization: Virtual exchange as a means to attaining global competencies, developing critical thinking, and experiencing transformative learning. *SCHOLE: A Journal of Leisure Studies and Recreation Education, 37*(1–2), 11–25. https://doi.org/10.1080/1937156X.2020.1760749

Evolve. (2024) What is virtual exchange? https://evolve-erasmus.eu/about-evolve/what-is-virtual-exchange/. Accessed 22 July 2024.

Furstenberg, G., Levet, S., English, K., & Maillet, K. (2001). Giving a virtual voice to the silent language of culture: The cultura project. *Language Learning & Technology, 5*(1), 55–102.

Godwin-Jones, R. (2019). Telecollaboration as an approach to developing intercultural communication competence. *Language Learning & Technology, 23*(3), 8–28.

Griggio, L., & Pittarello, S. (2020). How a multilingual project can foster and enhance international mobility. In F. Helm & A. Beaven (Eds.), *Designing and implementing virtual exchange – A collection of case studies* (pp. 127–139). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.45.1121

Guth, S., & Helm, F. (Eds.). (2010). *Telecollaboration 2.0: Language, literacies and intercultural learning in the 21st century.* Peter Lang.

Hall, E. T. (1973). *The silent language*. Doubleday.

Hall, E. T. (1976). *Beyond culture*. Anchor Books.

Hari, J. (2022). *Stolen focus: Why you can't pay attention*. Bloomsbury.

Haug and Jacobs. (2023). The design of collaborative online internationalised learning (COIL). In F. Hunter, R. de Ammigan, H. Wit, J. Gregersen-Hermans, E. Jones, & A. C. Murphy (Eds.), *Internationalisation in education: Responding to new opportunities and challenges. Ten years of research by the Centre for Higher Education Internationalisation (CHEI)* (pp. 143–156). EduCATT.

Helm, F., & van der Velden, B. (2019). Erasmus+ virtual exchange: 2018 impact report. https://doi.org/10.2797/668291

Kramsch, C. (1993). *Context and culture in language and teaching*. Oxford University Press.

Le Baron-Earle, F., & O'Reilly, C. (2023). Special issue on virtual exchange and the development of transferable skills: A review of practices across disciplines. *Journal of Virtual Exchange*. https://doi.org/10.21827/jve.6.41364

Lewis, T., & O'Dowd, R. (2016). Online intercultural exchange and foreign language learning: A systematic review. In R. O'Dowd & T. Lewis (Eds.), *Online intercultural exchange: Policy, pedagogy, practice* (pp. 21–66). Routledge.

O'Regan, V., Le Baron-Earle, F., & Batardière, M.-T. (2023). Erasmus speaks: Bringing internationalisation online. *Journal of Virtual Exchange*, *6*-(SI-TransferableSkills), 26–36. https://doi.org/10.21827/jve.6.39207

Olivetti de França, F., Beraldo di Genova, D. V., Camargo Penteado, C. L., & Kamienski, A. A. (2023). Understanding conflict origin and dynamics on Twitter: A real-time detection system. *Expert Systems with Applications, 212*, 118748.

Ollivier, C. (2018). *Digital literacy and a socio-interactional approach for the teaching and learning of languages*. European Centre for Modern Languages. https://www.ecml.at/Portals/1/5MTP/Ollivier/e-lang%20EN.pdf

Rinke, E. M. (2016). The impact of sound-bite journalism on public argument. *Journal of Communication, 66*(4), 625–645.

Soliya. (2024) Soliya Connect Programme. https://soliya.net/connect-program. Accessed 22 July 2024.

Törnberg, P., et al. (2022). How digital media drive affective polarization through partisan sorting. *PNAS, 119*(42), e2207159119.

UNESCO. (2022). We need to talk: Measuring intercultural dialogue for peace and inclusion. https://unesdoc.unesco.org/ark:/48223/pf0000382874?posInSet=1&queryId=6aed109e-3f75-4a78-8f06-b73ed3b59bb1 Accessed 3 March 2025.

UNESCO. (2024). Global citizenship and peace education. https://www.unesco.org/en/global-citizenship-peace-education/recommendation. Accessed 22 July 2024.

C

Code Literacy

Gary Motteram[1] and Felix Kwihangana[2]
[1]University of Manchester, Manchester, UK
[2]King's Academy, King's College London, London, UK

Abstract

This entry describes the value of developing the ability for teachers and students to understand, read and write code. It argues that it is important for everyone to know something about coding in a world that is dominated by digital technologies and where many jobs require people to be able to code at various levels. It describes the role that coding has played in the field of computer-assisted language learning (CALL), e.g. developing software and editing or producing web pages, as well as the changing nature of attitudes to coding in education. Further, it traces some of the history of language teachers learning to be coders and shows how the emphasis appears to be shifting back to learners beginning to code as a part of their regular schooling and how language work can come out of this. It briefly considers the role of artificial intelligence (AI) and the increasing need to understand how best to interrogate large language models (LLMs), making the related skill of AI literacy indispensable for teachers and learners. Finally, it suggests that while teachers may do less traditional coding, they still need coding literacy to adapt, integrate and evaluate digital tools in language teaching.

Keywords

Code literacy · Coding · Software · Authoring · AI literacy · Digital literacies

Definition

Code literacy is being able to understand, edit and/or produce the code that is used to enable computers to run and is usually associated with software.

Introduction

In the early days of personal computing, from the mid-1980s, it was a given that learning to program was an important skill for a better understanding of how the new technologies worked in general but also made it possible to create software (Amiri, 2000) that would be relevant to particular classroom practice (Stevens & Verschoor, 2017). In addition to this came learning how the hardware worked and how to look after it. Before starting any programming, it was necessary to learn how to format a disk, for example. Things have changed considerably from these early days, and we are now flooded with a vast variety of

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_257-1

software types, including apps (→ app literacy) that can support many different aspects of language development. These tools are increasingly described as "smart", and the average user can buy a device that works "out of the box".

However, while "out of the box" hardware and software exist, understanding how these tools do what they do and being able to interact with software and hardware in various ways remains an increasingly important skill.

Learning to Code and Other Literacies

There are many benefits of gaining familiarity with code, for both students and teachers, especially at a time when we are surrounded by coded devices. Amiri (2000) advocated for a move towards language teachers becoming programmers themselves, what he refers to as "end-user programmers" (p. 82). These are people who are not professional programmers but are still capable of producing software, quite often using authoring languages. He argues that their knowledge of the rules and workings of language rules predisposes them to becoming good programmers.

On platforms such as Scratch, where millions of young people learn to code, engagement with code has been likened to "programming-as-writing" (Hagge, 2018, p. 247). This comparison between code literacy and writing represents code literacy as an equally important literacy as traditional literacies, especially in today's digital age, and invites educators and learners to become more code-literate by building on their existing literacies as they would when developing writing or other literacies. For example, Baker-Doyle (2018) notes a link between code literacy and → critical literacy, arguing that code literacy can lead to greater awareness of societal issues and empathy towards people these issues affect. She makes parallels between writing in a language and code writing, highlighting how writing style and code style are equally important in their respective areas, as well as how critical thinking and building on previous texts (for writers) and codes (for coders) are an integral part of language and code literacy.

When applied to texts, criticality entails, among other things, the reader's ability to understand the author's explicit and implicit intentions that may draw from dominant narratives, biases and viewpoints and to appreciate and form a position after examining them. Similarly, code literacy allows users of digital sources and materials to evaluate these critically, enabling them to remain perceptive to the creators' explicit or implicit intentions. In other words, code literacy allows readers to be wary of what advocates for awareness of bias in the code running modern technologies call the "coded gaze" (Buolamwini, 2017, p. 17). Given that by learning to code, students (and their teachers) also learn to express themselves in a different medium (Hagge, 2018), code literacy enhances their ability to analyse, critique and evaluate coded resources and tools in ways that parallel the application of critical literacy to traditional texts.

History of Coding

When the first personal computers started to become generally available, some teachers started to learn programming languages like BASIC and wrote their own software. This often ended up being different conventional exercise types like multiple choice, vocabulary or cloze exercises (which in some cases hid all of the words). Some followed Seymour Papert and experimented with using higher-level languages like Logo to interact directly with objects to make them do things (Papert, 1980). Students could control a "Turtle", a connected robot that would move according to the programming instructions it was given; it would only respond to correct instructions. The underlying educational philosophy for the use of Logo was constructivism, which had links with the ideas of Jean Piaget, with whom Papert had studied.

This is an example taken from the Logo Foundation (n.d.) website, but with the language changed to make it easier to follow. Words in italics are the input from the user, those in bold are the system's responses, and there is commentary in brackets:

fowad

I don't know how to fowad
[The word **fowad** is not a primitive - one of Logo's built in words - nor a procedure that you've defined.]
forward
Not enough inputs to forward
[Now that you've spelled it correctly, Logo knows the word **forward**, but cannot run your instruction because **forward** requires additional information.]
forward 100
[Logo accepts the command. There is no error message. The turtle moves forward 100 steps.]

The early explorations into programming led to exercise creation tools like *Hot Potatoes* (1998), which allow people to create, for example, gap-filling, matching or crossword puzzles activities. (The dates in brackets are when the tools were first introduced.) This tool is still available but is not being updated. There was still some need to code to make these exercises work effectively, and to mount them on the internet, so that students could access them. A more recent manifestation of this type of tool is *H5P* (2013).

From the 1980s, more generic authoring tools like *Authorware* (1984) and *Toolbook* (1990) became available, and teachers were able to produce their software more quickly. Teachers also began to design websites using HTML (1993) and made use of *JavaScript* (1995) to create interactivity.

As blogging software like *WordPress*, first introduced in 2003, matured, it became easier to manipulate the code that produced the webpages and exercises created in *Javascript* could also be embedded. Exercises created in Hot Potatoes and H5P were later included. WordPress became a content management system (CMS) and could be used as an alternative to commercial CMSs such as *Blackboard.* Other tools like *Adobe Captivate* (2001), *iSpring* (2005) and *Articulate 360* (2016) also started to appear, and these allowed task creation of various types to be made much easier and were more readily adopted, as a result.

The Massachusetts Institute of Technology (MIT) also went on to create the more sophisticated version of *Logo* called *Scratch*, which allows students to work on their own mini-projects, where they can also control external devices and engage in different kinds of activities, including learning languages while using these tools. This is often a part of activities that might be referred to as content and language integrated learning (CLIL), where languages are learned alongside other subjects, in this case, programming.

Current Practices

Most tools that are used in language teaching, including Virtual Learning Environments (VLEs) like Canvas, graphics tools like Canva or interactive gaming software like Minecraft, are made more sophisticated if teachers and students can "get under the hood".

Simple activities like embedding a video in a VLE are helped if teachers can customise the code being imported. Here is an example of using an *iframe*, a piece of code that contains some instructions, in this case, to bring a *YouTube* video into the Canvas VLE:

```
<table style="border-collapse:
collapse; width: 98.0322%;"border="0"><
    <tbody>
        <tr>
            <td style="width:
33.3029%;"></td>
            <td style="width:
33.3029%;"><iframe title="YouTube
video player" src="https://www.
youtube.com/embed/qRN6HdbzPPU?
si=NNniWdp4GjUMsNl0" width="560"
height="315"
allowfullscreen="allowfullscreen"
allow="accelerometer; autoplay;
clipboard-write; encrypted-media;
gyroscope; picture-in-picture;
web-share"></iframe></td>
            <td style="width:
33.3029%;"></td>
        </tr>
    </tbody>
</table>
```

If we know a little HTML code, then we can quickly adjust the size of the viewing screen, being careful to maintain the ratio between the

width and the height (see code in bold in the image above). We can also adjust layout, or colour, or in this case the border, if we know that borders are measured in pixels, and there are 72 pixels to the inch, or approximately 38 in a centimetre. In this way, coding gives the user "more control and creativity" in how they choose to present information, which is a feature of the "individual empowerment" argument for learning to code (Vee, 2017, p. 77).

Students can interact with gaming platforms like Minecraft where they work on projects to build houses, national monuments or whole worlds. To do this, the students need to be able to program the world. Research shows that through coding, young people can develop their confidence, connect with others, develop and discover new interests and identify "possible selves" (Roque & Rusk, 2019). Given the importance of teachers in fostering students' interests in learning to code, code-literate teachers are more likely to support students in envisioning coding as a new literacy to develop.

Having given these examples, it cannot be denied that there are now many off-the-shelf tools that do not require any programming, but which allow you to create apps of various types, e.g., Thunkable (https://thunkable.com/) or Jotform (https://www.jotform.com/app-templates/). In addition, with the advent of generative artificial intelligence (GenAI), it is now the case that, if there is a need for code, a GenAI agent can create it (▶ "AI Generation Literacy").

At the same time, GenAI has its own form of coding, which has been called "prompt engineering". Prompt engineering involves interacting with an AI chatbot through natural language, for example, asking questions, then reformulating them based on responses, asking for clarifications and asking for different forms of output. While this is not strictly code, it is the way that good quality and more reliable answers are obtained. If teachers are able to get good answers, they can show their students how best to make use of these tools. In the age of GenAI, an understanding of how the code behind AI tools influences outputs is essential, given the link between code literacy and ▶ information literacy, which is "the use of critical thinking skills to decipher false, biased, and/or irrelevant information" (Kazakoff, 2015, p. 47). Teachers' and students' understandings of how AI tools are programmed (coded) to respond to prompts they receive affects their ability to make the best out of them and appreciate the limitations and shortcomings of GenAI outputs, both in language learning and general contexts.

To return to the work of Vee (2017), the main reasons for learning to code include "individual empowerment", "learning new ways to think", "citizenship and collective progress" and "employability and economic concerns". While it might be argued that the code literacy of the early days of computing which enabled teachers to create software are not as important now, it is still necessary for teachers and students to have an understanding of what coding is and how it impacts what we do with technology.

Implications for educators:

- A knowledge of how code can help us interact with technology is still valuable for teaching, as there is an increasing reliance on various technologies to support teaching. This might be an understanding of how to use an embed code to add a video or a language exercise to a VLE, or being able to support learners in coding on a platform like Minecraft.
- Teachers might go beyond copying and pasting ready-made HTML to altering/customising the code to make it better suit the environment that the code is being included in. Some platforms such as Khan Academy (https://www.khanacademy.org/) or W3Schools (https://www.w3schools.com/) have programming courses that interested educators can use to learn to create and run different types of code.
- Teachers might also want to develop visual coding skills both for themselves and their learners using Scratch (https://scratch.mit.edu/) or Node Red (https://nodered.org/), which can be used in conjunction with the Raspberry Pi (https://www.raspberrypi.org/). The website here includes materials on various aspects of coding, including how AI is coded.

Implications for learners:

- Learners might begin to code as a part of an integrated language lesson, using a tool like Scratch or Minecraft to engage in interactive projects to both develop language skills and programming skills.
- Learners can be encouraged to develop a critical stance towards technology as they interrogate an LLM, so that they are not simply taking as given what is served up to them.
- Learners may develop a range of skills with coding that might support them through life and in employment, e.g. understanding basic programming like loops, or conditionals; debugging code; writing simple algorithms; and knowing basic principles of cybersecurity. This can be done through a tool like Scratch or Raspberry Pi.
- Learning to code may help students identify or develop new interests, connect with peers and like-minded groups as well as empathise with different groups.

While the codes that underlie the tools that are used on a day-to-day basis are often hidden, being able to interact with them can help teachers and learners feel more in control of the digital world we inhabit. Coding can encourage a critical perspective on technology and move people away from simply being consumers. As a result, they may well be better able to utilise software of various kinds to support language development.

Cross-References

► AI Generation Literacy
► App Literacy
► Critical Literacy
► Information Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Amiri, F. (2000). IT-literacy for language teachers: Should it include computer programming? *System, 28*, 77–84.

Baker-Doyle, K. J. (2018). I, pseudocoder: Reflections of a literacy teacher-educator on teaching coding as critical literacy. *Contemporary Issues in Technology and Teacher Education, 18*(2), 255–270.

Buolamwini, J. A. (2017). *Gender shades: Intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*. Doctoral dissertation, Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/114068

Hagge, J. (2018). Coding to create: A subtext of decisions as early adolescents design digital media. *Technology, Knowledge and Learning, 23*, 247–271. https://doi.org/10.1007/s10758-018-9359-y

Kazakoff, E. (2015). Technology-based literacies for young children: Digital literacy through learning to code. In K. Heider & M. Renck Jalongo (Eds.), *Young children and families in the information age. Educating the young child, 10* (pp. 43–60). Springer. https://doi.org/10.1007/978-94-017-9184-7_3

Logo Foundation. (n.d.). https://el.media.mit.edu/logo-foundation/what_is_logo/logo_programming.html

Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. Harvester.

Roque, R., & Rusk, N. (2019). Youth perspectives on their development in a coding community. *Information and Learning Sciences, 120*(5–6), 327–348. https://doi.org/10.1108/ILS-05-2018-0038

Stevens, V., & Verschoor, J. (2017). Coding and English language teaching. *TESL-EJ, 21*(2), 1–15.

Vee, A. (2017). *Coding literacy: How computer programming is changing writing*. The MIT Press. https://doi.org/10.7551/mitpress/10655.001.0001

M

Machine Translation Literacy

Lynne Bowker
Languages, Linguistics and Translation, Université Laval, Quebec, QC, Canada

Abstract

Machine translation tools are technically straightforward to use, but language learners may not instinctively know how to apply them effectively or responsibly to support their learning. Machine translation literacy is a new type of digital literacy that emphasizes critical thinking rather than focusing solely on techno-procedural skills. For language learners, machine translation literacy involves understanding how data is used by and influences data-driven translation tools (e.g. Google Translate, ChatGPT). It also means developing good judgement and being transparent and responsible with regard to tool use, as well as learning how to interact with tools in order to improve the translation quality. Instructors can support students by developing clear policies for machine translation use in their courses and offering structured guidance to help language learners experiment with and compare different tools and use cases.

Keywords

Machine translation · Data-driven machine translation · Transparency · Responsible use · Digital literacies

Introduction

Machine translation literacy is the ability to comprehend key aspects of how machine translation tools work, recognize how they can be used effectively in a particular context, and understand the implications of using them.

Machine translation (or automatic translation) is a tool that can translate a text from one language (e.g. English) to another (e.g. French). Machine translation literacy is a new type of → digital literacy that involves comprehending key aspects of how these tools work, recognizing how they can be used effectively in a particular context, and understanding the implications of using them (Bowker, 2020). The goal of machine translation literacy is to develop informed and responsible tool users. Machine translation literacy is flexible since these tools can be used for different purposes (e.g. understanding vs producing a text in another language) and in different contexts (e.g. language learning vs translating health information). Therefore, machine translation literacy can be adapted for different user groups. This entry focuses specifically on aspects of machine translation literacy that are relevant for

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_258-1

computer-assisted language learning (CALL) (Vinall & Hellmich, 2022).

Machine translation tools date back to the mid-twentieth century, although their underlying architecture has evolved. Early systems were known as rule-based machine translation tools because they used pre-programmed grammatical rules and large bilingual dictionaries. These tools had limited success because they produced awkward word-for-word translations. The low quality meant that rule-based machine translation tools were not often used for CALL, except to show students poor examples or ask them to improve on the tools' efforts (Poibeau, 2017).

By the end of the century, the digital landscape had changed. Computers were smaller, faster, and cheaper and had more storage. People created digital texts using word processors and used the Internet and World Wide Web to share and access texts, inspiring researchers to approach machine translation in a completely different way. Rather than using linguistics (e.g. grammar rules and vocabulary), researchers identified tasks that computers are good at, such as calculations and pattern matching, and used these as the basis for data-driven machine translation. Early data-driven machine translation tools were mainly statistical, but current approaches use artificial intelligence (AI) techniques such as artificial neural networks and machine learning. AI-based approaches have led to neural machine translation tools (e.g. Google Translate), as well as tools based on Large Language Models (LLMs) (e.g. ChatGPT) that can translate as well as performing other language processing tasks. Translations produced by data-driven tools are of a much higher quality and sound more fluent than translations produced by rule-based machine translation tools. As a result, more people are using machine translation, including for language learning, although this has sometimes been controversial (Jolley & Maimone, 2022).

Some instructors initially feared that students might use data-driven machine translation tools to circumvent learning by simply using the tools to do their homework (e.g. translating texts). However, bans are difficult to enforce, and instructors soon saw the value of helping students to improve their machine translation literacy and use the tools wisely (Alm & Watanabe, 2022; Cotelli Kureth & Summers, 2023; Loock et al., 2022).

Implications for Educators or Educational Institutions:

- **Provide clear policies**. Bans are ineffective and deprive students of opportunities to learn about these tools, which are actively used by various types of language professionals (e.g. teachers, translators) and therefore essential to the curriculum. Be clear about when machine translation is encouraged, and explain the nature and scope of any restriction, clarifying why it is necessary. Specify whether and how students should report tool use.
- **Provide guidance and a supportive environment**. Beyond policies, provide suggestions to get students started (e.g. see Bowker and Buitrago Ciro's (2019) machine translation literacy framework for ideas). On the surface, machine translation is easy to use—just copy, paste, click! But the simple interface can mislead students into thinking that translating a whole text is the only or best way to use the tool. Encourage students to reflect more deeply and to experiment with and compare various tools (e.g. Google Translate vs ChatGPT); different use cases (e.g. reading in L1 vs writing in L2); different inputs (e.g. translating words vs sentences or paragraphs, translating formulaic vs creative texts); different languages (e.g. translating out of a widely used language vs a less widely used language); or different prompts (e.g. more vs less detailed).

Implications for Learners:

- **Understand the basic functioning and limits of data-driven tools**. It is not necessary to comprehend the technical details of how neural networks or LLMs work, but it is helpful to understand how data is used by and influences data-driven technologies. Computers can *process* texts using techniques such as pattern matching and counting, but they do not *understand* what the text means. They imitate examples, but to perform well, they need millions of examples in their training corpus. Moreover,

the examples need to be *good* ones; otherwise, the tool will reproduce poor-quality examples. For instance, machine translation tools can reproduce biased language if this is present in the training data. For some language combinations, it is easy to find many good examples, but for other language combinations where there is less translation activity, it can be challenging to obtain enough high-quality training data, leading to poorer performance. Likewise, some subjects or text types are more common than others, and this can also affect translation quality. Keep in mind that each tool (e.g. Google Translate, ChatGPT) has been trained using a different training corpus, and therefore each tool will learn different things and produce somewhat different translations.

- **Recognize the relationship between input and output**. The quality of the input text can have a big influence on the quality of the translated text. Sometimes, improving the quality of the input text can be more effective than only fixing up the translation, so experiment with pre- and post-editing and back-translation. Likewise, for generative AI tools (e.g. ChatGPT), the quality of the prompt affects the quality of the results, so experiment with prompt formulation and providing more detailed requests.
- **Be transparent and accountable**. Respect instructors' policies and be clear about when and how you use machine translation to support your learning or future professional translations. Regardless of which tools you use, take responsibility for the final version of the text and recognize that you can often improve the proposed translation.
- **Consider wider implications of tool use**. Tools are not used in a vacuum, and part of being a responsible digital citizen is understanding the broader context of use, including ethical considerations. For instance, tool owners (e.g. Google) can retain information entered into a free online machine translation tool, so do not enter sensitive content. As a language student, you may one day be a translator or writer, so be aware that these tools only work because the data previously produced by language professionals has been repurposed (often without their permission) for training data. Processing such large corpora is computationally intensive, and so these tools have environmental impacts too.

Machine translation tools have straightforward interfaces, but this does not mean that language learners instinctively know how to use them responsibly or effectively. Learners need opportunities and support to develop their machine translation literacy, which places an emphasis on critical thinking, creativity, and developing judgement, rather than solely on techno-procedural skills.

Cross-References

- ▶ AI Practices Literacy
- ▶ AI-Generation Literacy
- ▶ Critical Digital Literacy
- ▶ Digital Literacies

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Alm, A., & Watanabe, Y. (2022). Online machine translation for L2 writing across languages and proficiency levels. *Australian Journal of Applied Linguistics, 5*(3), 135–157.

Bowker, L. (2020). Machine translation literacy instruction for international business students and business English instructors. *Journal of Business & Finance Librarianship, 25*(1–2), 25–43.

Bowker, L., & Buitrago Ciro, J. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community.* Emerald.

Cotelli Kureth, S., & Summers, E. (2023). Tackling the elephant in the language classroom: Introducing machine translation literacy in a Swiss language Centre. *Language Learning in Higher Education, 13*(1), 213–230.

Jolley, J. R., & Maimone, L. (2022). Thirty years of machine translation in language teaching and learning: A review of the literature. *L2 Journal, 14*(1), 22–44.

Loock, R., Léchauguette, S., & Holt, B. (2022). Dealing with the "elephant in the classroom": Developing language students' machine translation literacy. *Australian Journal of Applied Linguistics, 5*(3), 118–134.

Poibeau, T. (2017). *Machine translation*. The MIT Press.

Vinall, K., & Hellmich, E. (Eds.). (2022). Special issue "Machine translation & language education: Implications for theory, research & practice". *L2 Journal, 14*(1), 1–257.

C

Conversational AI Literacy

Louise Ohashi [1] and Antonie Alm [2]
[1]Gakushuin University, Tokyo, Japan
[2]Otago University, Dunedin, New Zealand

Abstract

This entry explores the concept of conversational AI literacy and its implications for language education. Developments in natural language processing technology have led to the emergence of sophisticated AI tools that have the capacity to replicate and respond to human language, presenting language learners with new opportunities for personalised learning and interactive language practice. Conversational AI literacy is a competency that enables language learners to engage effectively with these tools across different modalities, including text, voice, and video. It involves understanding the operational mechanisms, affordances and limitations of tools such as chatbots, virtual assistants, text-to-speech and speech recognition software, as well as new forms of conversational AI that are yet to be created. Language learners should experiment with different conversational AI to explore the affordances of different models and modalities, learn to provide effective input, and develop the skills required to critically evaluate output. Language educators should provide guidance to help learners use this technology appropriately and be able to select or develop their own tools, such as personalised chatbots, for specific learning contexts. As conversational AI tools rapidly evolve to become more multimodal and immersive, it will be increasingly important for both language educators and learners to seize the potential of this technology by developing conversational AI literacy.

Keywords

Conversational AI · Generative AI (GenAI) · Large language models · AI chatbots · AI education · Human-AI collaboration · Natural language processing · Digital literacies

Definition

In the context of language education, conversational AI literacy refers to the ability of language learners and educators to effectively interact with conversational AI systems across different modalities for second language learning and instruction.

Introduction

Conversational AI technology relies on machine learning and natural language processing to process and mimic human language production, making it highly relevant to the field of language

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_259-1

education. Chatbots are a widely known form of this technology, but it also extends to tools such as virtual assistants, text-to-speech software, and speech recognition software (Google Cloud, n.d.). In addition to standalone tools like chatbots (e.g. ChatGPT), conversational AI technology may be built into devices (e.g. the speech-to-text function found in smartphones and laptops), or integrated into other tools, such as the voice recognition function that is part of translation apps (e.g. DeepL and Google Translate). It is increasingly common for conversational AI tools to be multimodal, and each modality and tool presents unique affordances and challenges for language learning.

Defining what conversational AI encompasses is challenging due to the evolving technological landscape. Tools and capabilities are appearing at an unprecedented rate, creating the need for new terms and reconceptualisations of existing meanings. For instance, the concept of conversational AI has changed significantly since the 1960s, when ELIZA, the first chatbot, made "natural language conversation with a computer possible" (Weizenbaum, 1966, p. 36). Today's conversational AI tools are powered by advanced natural language processing and can engage in sophisticated, contextualised interactions across multiple modalities. This development has led to a reconceptualisation of chatbots as "novel interlocutors" (Jeon & Lee, 2024), recognising their position between human-like interaction and machine-based capabilities.

Learners can communicate with conversational AI through both written (text AI) and spoken (voice AI) modalities, with the potential for this to expand further in immersive virtual environments such as the metaverse, which is still under construction (Godwin-Jones, 2023; ▶ "Immersive Literacy"). Conversational AI tools can be used to assist learners at all levels of proficiency, but the skill sets required vary. Conversational AI literacy builds upon the broader framework of ▶ digital literacies (Pegrum et al., 2022) and emerges as a key competency for language learners among a wider set of ▶ AI literacy practices.

Contributions of Conversational AI to Language Learning

Conversational AI offers numerous potential benefits for language learners. Jeon (2024) found that chatbots can provide learners with opportunities for speaking practice, immediate feedback, and a low-anxiety environment for language use, which positively influenced learners' motivation. These findings are consistent with Alm and Nkomo's (2020) research on the use of chatbots by informal language learners, which also highlighted benefits such as personalised learning experiences, opportunities for repetitive practice, and reduced anxiety compared to human interactions. AI-powered chatbots can also serve as virtual tutors, providing students with practice opportunities and personalised feedback and can adapt to learners' proficiency levels (Bibauw et al., 2019; Wan & Moorhouse, 2024), creating a customised learning experience. Personalised learning can offer learners access to materials that address their individual learning needs and interests, which can heighten engagement and foster skill development. For example, an investigation into students' engagement when independently studying vocabulary with a chatbot found their behavioural, cognitive, and emotional engagement improved over time (Pham et al., 2024). Personalisation stands out as a key benefit, but it should be noted that use of the term "personalised" in relation to chatbots may decline in the years ahead, with "individualised" used instead to remove the "person" label from non-human tutors.

Research has furthermore shown that dialogue-based systems using conversational AI can support L2 proficiency development (Bibauw et al., 2022). This includes improvements in various language skills, including pronunciation. In a meta-analysis of automated speech recognition studies, Thi-Nhu Ngo et al. (2024) found that automated speech recognition is largely effective for improving pronunciation when explicit corrective feedback is provided to the user.

There are likely to be even more affordances as this technology continues to expand. On September 11, 2024, Google released a function

on NotebookLM that allows users to upload written files that are used by the software to create human-like discussions about the content between two AI speakers (Wang, 2024). How long will it be before this technology evolves into a system that allows language learners to join authentic-sounding, engaging group discussions with AI interlocutors? The foreseeable benefits of such an invention and others that are yet to come are numerous. However, these tools are not without their limitations, so it is essential to develop conversational AI literacy skills in order to evaluate and use them properly.

Developing Language Learners' Conversational AI Literacy

Conversational AI literacy involves a set of skills that language learners need to engage with AI-powered tools. These competencies can be sorted into the functional, linguistic, strategic, critical, and ethical categories.

On a functional level, learners have to understand the basic principles of conversational AI systems. This includes recognising that these tools generate responses based on patterns in their training data, rather than having true comprehension (Bender et al., 2021). Learners should familiarise themselves with various AI interfaces, learning how to access and use them across different devices.

Linguistic and prompting skills form a core component of conversational AI literacy. Learners need to develop the ability to craft clear and purposeful prompts—the input they provide to initiate and guide conversations with AI. Wan and Moorhouse (2024) emphasise the importance of formulating prompts that elicit targeted language practice and feedback. This involves effectively commanding AI to adjust the complexity and style of language to match the learner's proficiency level and learning goals. Some AI tools are limited to a single language, but others allow prompting in multiple languages, offering even greater flexibility for learners at different stages of language acquisition.

Strategic use of conversational AI involves selecting appropriate tools and tasks that correspond to one's language proficiency and goals (Bibauw et al., 2022). Learners should view AI tools as supplements to other language learning methods, using them for tasks such as vocabulary practice, grammar checking, pronunciation improvement, and, of course, conversational practice.

Critical evaluation skills enable learners to assess the quality, relevance, and accuracy of AI-generated content for their specific learning needs (Kohnke et al., 2023). This involves cross-referencing with reliable sources and discerning when to rely on AI feedback versus seeking human input. Learners should also recognise potential biases or limitations in AI-generated language, including its tendency to reinforce dominant language variety norms, and the role AI can play in perpetuating stereotypes (Bender et al., 2021).

Additional ethical considerations in conversational AI literacy include understanding institutional policies regarding AI use in academic work and discussing uncertainties with instructors (Jeon & Lee, 2024). Furthermore, learners should be aware of privacy concerns related to data input into AI systems in order to make informed decisions about what to share. They should also understand that teachers have a duty of care that extends to the protection of personal data, such as voice recordings. While students may agree to their data being input, teachers need to be as transparent about their use of AI as students are expected to be.

Finally, institutions, teachers, and students should also consider the environmental impact of AI. Large-scale language models consume significant resources and produce high carbon emissions, with impacts unevenly distributed around the world (Ren & Wierman, 2024). While this does not negate the benefits of AI in education, it does highlight the need for responsible use. UNESCO (2024) recommends a balanced approach, with the potential to integrate smaller, more environmentally friendly language models where appropriate.

Considerations for Language Learners When Using Voice AI and Text AI Modes

Modality-specific considerations are also important for language learners. Voice AI tools currently have limitations in understanding the full range of human speech, often performing best with accents and pronunciations well-represented in their training data (Dizon, 2023). Learners may need to adjust their pronunciation, experiment with different input methods, or switch to text-based modalities when voice input is challenging. However, learners and educators should be aware that adjusting speech to imitate "standard" accents may inadvertently reinforce biases in AI systems, potentially leading to less diverse language representation in future AI development.

While text-based AI can often handle a wider range of language input, learners should be aware of potential limitations in understanding context, sarcasm, or culturally specific references. Developing skills in clear and precise written communication can improve interactions with text-based AI tools.

Managing anxiety and building confidence are also important aspects of conversational AI literacy. Although AI chatbots can provide a low-stress environment for speaking (Hayashi & Sato, 2024) and writing practice (Kwon et al., 2023), some learners may still experience anxiety when interacting with these tools. Learners can build confidence through gradual exposure and practice (Dizon, 2023), and by choosing tools with interfaces that make them feel more comfortable. For example, if viewing a humanlike avatar in a voice AI tool like Call Annie induces a sense of time pressure, learners can change the avatar to a non-human character or have real-time conversations on a faceless tool like ChatGPT's voice mode.

Supporting Language Teachers in Developing Conversational AI Literacy

Language educators need opportunities to learn about and engage with conversational AI in order to guide their students towards effective use for language development. Institutions should support their staff in developing the knowledge and skills necessary by providing resources, training sessions, and opportunities for teachers to engage in discussions and share best practices. For students to use conversational AI appropriately, they need to experiment and be taught about ▶ digital literacies (Poole & Polio, 2023), which extends to conversational AI literacy, so they need knowledgeable instructors. However, they also have the capacity to share skills and knowledge, so educators should be open to learning about conversational AI from their students and introduce activities that will enable learning in both directions.

As well as guiding usage, language teachers can create their own chatbots using platforms designed for consumption by the general public (e.g. Poe, http://poe.com, or ChatGPT, https://chatgpt.com/gpts), or others designed specifically for educators (e.g. Mizou, https://mizou.com). Those made for educators often require paid subscriptions, but in return they tend to provide more support for teachers and place a greater emphasis on data privacy. These custom chatbots can be designed to meet specific learning objectives, languages, and learner proficiency levels. By creating chatbots, teachers can gain hands-on experience with conversational AI, deepening their understanding of the technology and its potential applications in language education. Guidance on how to do this would be highly beneficial to instructors.

Institutions should support language teachers in developing conversational AI literacy and gaining the skills needed to create customised tools by:

- Providing training and resources on conversational AI, covering topics such as chatbot design, ▶ prompt engineering, and best practices for integrating chatbots into language curricula. Training could be provided by knowledgeable staff within the institution or outsourced to specialists.
- Encouraging collaboration and knowledge sharing among teachers (including language

teachers and teachers of other subjects), facilitating the exchange of ideas, experiences, and successful strategies for using various forms of conversational AI in the classroom.
- Offering technical support and guidance for teachers creating chatbots, ensuring that they have access to the necessary tools and expertise to develop and deploy their chatbots effectively.
- Establishing guidelines and policies for the use of conversational AI in language learning, addressing issues such as data privacy, security, and ethical considerations.

To guide students effectively, teachers need exposure, as greater awareness of generative AI can lead to increased use by teachers, which can in turn promote positive perceptions of its role in language education (Kaplan-Rakowski et al., 2023). By empowering language teachers with the skills and knowledge required to use conversational AI effectively, institutions can foster a culture of innovation and continuous improvement in language education, ultimately benefiting students and enhancing their conversational AI literacy.

Conclusion

Developing conversational AI literacy is essential for language learners to fully benefit from the growing array of AI-driven language learning tools. By understanding the capabilities and limitations of these technologies, crafting effective prompts, and critically evaluating AI-generated content, learners can make the most of conversational AI to enhance their language skills while developing essential digital literacies for the twenty-first century. Educational institutions should take a proactive stance in fostering conversational AI literacy by providing guidance, support, and opportunities for critical engagement with these technologies. This support should not only be for learners, but also for teachers, who can integrate this technology into their teaching practice in innovative ways.

Cross-References

▶ AI Literacy Practices
▶ Critical Digital Literacy
▶ Digital Literacies
▶ Immersive Literacy
▶ Prompt Engineering

Competing Interest Declaration The authors have no competing interests to declare that are relevant to the content of this manuscript.

References

Alm, A., & Nkomo, L. M. (2020). Chatbot experiences of informal language learners: A sentiment analysis. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 10*(4), 51–65. https://doi.org/10.4018/IJCALLT.2020100104

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Virtual Event, Canada. https://doi.org/10.1145/3442188.3445922

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practise a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning, 32*(8), 827–877. https://doi.org/10.1080/09588221.2018.1535508

Bibauw, S., François, T., Van den Noortgate, W., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology, 26*(1), 1–24. https://hdl.handle.net/10125/73488

Dizon, G. (2023). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal, 54*(3), 848–855. https://doi.org/10.1177/00336882211020548

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology, 27*(2), 6–27. https://hdl.handle.net/10125/73501

Google Cloud. (n.d.). *Conversational AI for richer, more intuitive experiences*. Retrieved September 21, 2024, from https://cloud.google.com/conversational-ai?

Hayashi, K., & Sato, T. (2024). The effectiveness of ChatGPT in enhancing English language proficiency and reducing second language anxiety (L2). In *ISSN: 2759-1182 – WorldCALL2023: Conference proceedings* (pp. 201–208). https://doi.org/10.22492/issn.2759-1182.2023.23

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1–2), 1–26. https://doi.org/10.1080/09588221.2021.2021241

Jeon, J., & Lee, S. (2024). Can learners benefit from chatbots instead of humans? A systematic review of human-chatbot comparison research in language education. *Education and Information Technologies, 17*, 23329–23360. https://doi.org/10.1007/s10639-024-12725-9

Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2023). Generative AI and teachers' perspectives on its implementation in education. *Journal of Interactive Learning Research, 34*(2), 313–338. https://www.learntechlib.org/primary/p/222363/

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Kwon, S. K., Shin, D., & Lee, Y. (2023). The application of chatbot as an L2 writing practice tool. *Language Learning & Technology, 27*(1), 1–19. https://doi.org/10125/73541

Pegrum, M., Dudeney, G., & Hockly, N. (2022). *Digital literacies* (2nd ed.). Routledge.

Pham, T. T., Nguyen, L. A. D., Dang, H. M., & Le, T. T. P. (2024). Exploring tertiary Vietnamese EFL students' engagement in vocabulary learning through the use of an AI tool. In *Proceedings of the AsiaCALL international conference*, *4* (pp. 129–149). https://doi.org/10.54855/paic.23410

Poole, F. J., & Polio, C. (2023). From sci-fi to the classroom: Implications of AI in task-based writing. *Journal on Task-based Language Teaching and Learning, 3*(2), 243–272. https://doi.org/10.1075/task.00022.poo

Ren, S., & Wierman, A. (2024, July 15). The uneven distribution of AI's environmental impacts. *Harvard Business Review*. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL, 36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

UNESCO. (2024, March 18). Small language models (SLMs): A cheaper, greener route into AI. https://www.unesco.org/en/articles/small-language-models-slms-cheaper-greener-route-ai

Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal*. https://doi.org/10.1177/00336882231224813

Wang, B. (2024, September 11). NotebookLM now lets you listen to a conversation about your sources. Google. https://blog.google/technology/ai/notebooklm-audio-overviews/

Weizenbaum, J. (1966). ELIZA–A Computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45. https://web.stanford.edu/class/cs124/p36-weizenabaum.pdf

V

Visual Literacy

Elena Domínguez Romero[1] and Jelena Bobkina[2]
[1]Department of English Language and Linguistics, Universidad Complutense de Madrid, Madrid, Spain
[2]Universidad Politécnica de Madrid, Madrid, Spain

Abstract

Visual literacy, once considered secondary to written literacy, has now emerged as a crucial educational competency, involving the interpretation, evaluation, and creation of visual content. It encompasses the ability to engage with multimodal elements—such as text, imagery, and design—addressing the demands of diverse social and cultural contexts. Initially perceived as a purely cognitive skill, visual literacy is increasingly viewed as a social practice, where meanings are influenced and shaped by cultural and situational factors. This conceptual shift has become progressively relevant in second language (L2) acquisition, particularly with advancements in computer-assisted language learning (CALL). While progress has been made, research continues to emphasise the importance of formally developing visual literacy in education, advocating for its integration into structured curricula.

Incorporating visual literacy into L2 instruction through CALL tools enhances comprehension and retention of language, and cultural awareness, while fostering critical thinking and creative expression. The visual and interactive, engaging nature of CALL can provide an enriched language learning experience, making it both more stimulating and effective. However, achieving effective visual and digital literacy education requires targeted teacher professional development to equip educators with the skills to interpret, evaluate, and produce visual content effectively. Furthermore, assessment methods must evolve to evaluate students' visual, digital, and multimodal competencies, ensuring they can critically and creatively engage with visual media. These strategies not only prepare learners to navigate an progressively visually oriented world but also position them as capable communicators who can thrive in digital and globalised environments.

Keywords

Visual literacy · Digital literacies · Media literacy · Visual critical thinking · Teacher professional development · L2/CALL · Curriculum integration

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_261-1

Definition

Visual literacy refers to the ability to interpret, evaluate, and create visual content. This involves engaging with multimodal elements like text, imagery, and design, all shaped by social contexts and practices (Elkins, 2009; Serafini, 2010). Metros (2008) defines it as "the ability to decode and interpret visual images and to encode and compose meaningful visual communication" (p. 103), emphasising the importance of both analysis and creation in helping L2/CALL learners become more proficient, reflective, and empowered language users.

Introduction

The concept of visual literacy was first introduced by John Debes in 1969, who described it as "a group of vision-competencies a human being can develop by seeing and at the same time having and integrating other sensory experiences" and as necessary "to discriminate and interpret the visible actions, objects, symbols, natural or man-made" (Debes, 1969, p. 27). In the 1990s, definitions of visual literacy evolved to emphasise image interpretation—the ability to understand and derive meaning from images (Wileman, 1993) and the "ability to find meaning in imagery" (Yenawine, 2005, p. 1). Later, visual literacy was reconceptualised as not only a cognitive ability but also a social practice. Sturken and Cartwright (2001) highlighted that meanings are "produced not in the heads of the viewers so much as through a process of negotiation among individuals within a particular culture, and between individuals and the artefacts, images, and texts created by themselves and others" (p. 4). Metros (2008) further formalised visual literacy as the ability to decode and interpret visual messages, as well as to "encode and compose meaningful visual communications" (p. 103), emphasising the importance of visualisation, communication, and the interpretation of visual content.

Further developing this concept, Metros (2008, pp. 13–14) points to visual literacy as a dual-faceted skill set, encompassing both critical and creative dimensions. By developing visual literacy, individuals enhance their ability to critically analyse, interpret, and create visual content, thereby increasing their comprehension, retention, critical thinking, and creativity skills. This approach to visual literacy not only fosters a deeper appreciation of visual media but also stimulates innovative thinking and problem-solving. As individuals become more adept at discerning subtle visual cues and contexts, they are better equipped to navigate and contribute to a world rich in visual and digital content (Kędra & Žakevičiūtė, 2019), which involves browsing the Internet, using apps, or interpreting complex visual data like infographics (Kalaf-Huges, 2023).

Visual Literacy in Language Education

Visual literacy, as an integral aspect of → multimodal literacy, demands focused attention due to its distinct cognitive processes and specialised skill set (Kress, 2010; Kress & van Leeuwen, 2006). It plays a pivotal role in improving language learning outcomes within the contexts of L2 acquisition and computer-assisted language learning (CALL).

Firstly, visual literacy enhances language comprehension, retention, and memory (Guo et al., 2020). CALL tools, including digital storytelling platforms (e.g. Storybird, Book Creator) and interactive applications (e.g. Nearpod, Kahoot), introduce dynamic visual dimensions to L2 instruction, enriching the learning experience and making it more engaging and effective. By critically analysing visual media, such as images, infographics, and videos, L2 learners explore elements like colour, composition, and symbolism. This process deepens their understanding of cultural specificities and sharpens their interpretative skills, thereby enhancing their overall language acquisition process.

Moreover, visual literacy functions as a crucial component of digital literacies (→ history of new literacies), fostering multimodal literacy, critical thinking, and creativity (Serafini, 2014, 2017). The development of visual projects, such as videos, infographics, or digital books, reinforces

these skills by offering practical, hands-on applications. CALL tools like Adobe Express, Canva, and Piktochart empower L2 learners to design visual content and critically evaluate visual messages, identifying biases—skills that are indispensable in today's media-saturated digital landscape.

Thus, the integration of visual literacy into CALL not only strengthens language proficiency but also equips learners with the ability to navigate, interpret, and produce visual content thoughtfully. This preparation enables them to become informed, visually literate communicators capable of thriving in an increasingly digital world (Thompson, 2019).

Challenges and Implementation

Despite the pervasive presence of visual imagery in students' lives, many remain fundamentally visually illiterate (Guo et al., 2020). Research suggests that developing visual literacy skills effectively requires formal instruction (Serafini, 2014; Avgerinou & Pettersson, 2020). Therefore, it is essential to educate students on identifying and using various modes and semiotic resources to convey meaning effectively. To achieve this, the structured integration of visual literacy into the curriculum is essential (Kędra & Žakevičiūtė, 2019), especially as recent research on the use of visual materials in education—particularly in L2 learning—has only just begun to highlight the pressing need for visual literacy development. Various approaches have emerged, with some studies advocating for the integration of visual literacy into curricula through digital tools and multimedia, particularly within the framework of CALL, while others highlight project-based strategies or the fusion of visual comprehension with critical thinking skills (Domínguez Romero & Bobkina, 2021; Unsworth & Mills, 2020). However, several researchers note that the integration of multimodal practices and visual literacy into curricula often remains informal and largely dependent on individual teachers' interests. Metros (2008) observes that in K-12 education, "teachers typically lack the resources to integrate visual literacy into the everyday classroom experience [. . .] they must include in their lessons visual literacy opportunities for students to express ideas and think in multimodal ways" (p. 106).

Practical Implications

As visual literacy becomes an indispensable skill, several key strategies must be considered for its successful incorporation into the curriculum, ensuring students are better prepared for a visually oriented world:

Curriculum Development and Integration:
The pedagogical implications of visual literacy underscore its crucial role in fostering a balanced curriculum where visual and textual literacies are given equal weight (Bamford, 2003). As contemporary education increasingly integrates multimodal approaches, it becomes imperative for educators to design learning objectives and tasks that incorporate visual media in a deliberate and meaningful way. This integration not only supports the development of core language acquisition skills but also nurtures higher-order competencies such as critical analysis and creative problem-solving. By embedding visual literacy into teaching practices, educators can cultivate a more dynamic, digital, and inclusive learning environment that addresses diverse cognitive styles and promotes more nuanced, deeper comprehension of content.

- Graphic novels, film and video analysis, and interactive multimedia exercises (e.g. Duolingo, Rosetta Stone) seamlessly blend visuals, text, and audio to cater to diverse preferences. Some of these formats emphasise the importance of visual literacy by encouraging students to not only interpret visual content but also to create it, fostering a more comprehensive engagement with visual communication alongside diverse modes of learning. Incorporating visual content creation and analysis into L2/CALL instruction offers a sophisticated and engaging approach to language learning.
- Assignments such as designing memes or comic strips inspired by current events not

only foster creativity but also deepen students' linguistic and cultural understanding.

- Advanced AI image generation tools, including Microsoft Designer and DALL-E, require learners to craft detailed prompts, generate visually compelling content, and critically analyse visuals. This process integrates key principles of visual and multimodal literacy and semiotics, helping learners understand how design, symbols, and text–visual interactions create and convey meaning (Domínguez Romero & Bobkina, 2017).
- Complementing these tools, platforms like Storybird and Adobe Express enhance narrative skills and visual comprehension by enabling students to seamlessly integrate images and text in multimodal storytelling.
- Similarly, infographic tools such as Canva and Piktochart support knowledge retention by synthesising textual and visual elements, rendering complex ideas more accessible.
- Interactive whiteboards, such as SMART Boards, further enrich the learning environment by facilitating immersive, multimedia-driven lessons that simultaneously advance language proficiency and stimulate creative thinking.

Teacher Training and Resource Availability

Teachers need professional development opportunities to acquire the skills necessary to effectively integrate visual literacy into their teaching practices. This includes training in interpreting, evaluating, and creating visual communication. Formal teacher education that focuses on the structure, techniques, and modes of multimodal texts is equally crucial for second language teachers to effectively develop their students' critical visual literacy skills (Eutsler, 2021). In the context of L2/CALL, teachers should also develop proficiency in digital tools such as Canva, Genially, Padlet, Piktochart, and Thinglink, among others, which can be used for creating engaging visual content. Collaborative learning with colleagues and leveraging resources from educational organisations can provide valuable support while inspiring innovative approaches to training L2 teachers in visual literacy and CALL. Actively participating in professional networks focused on visual literacy helps teachers stay updated on emerging trends and strategies for effectively integrating digital tools into language instruction. Regular practice in creating and analysing visual content, alongside assigning visual projects such as infographics or digital books, enables teachers to apply these skills in practical, language-centred contexts. Moreover, cultivating critical analysis skills to evaluate visual media empowers teachers to enhance students' comprehension and critical thinking, enriching the L2 learning experience through thoughtfully integrated CALL practices.

Assessment and Evaluation

Traditional assessment methods must be adapted to evaluate students' multimodal abilities, with a particular emphasis on their visual and digital literacy skills. This involves evaluating their ability to interpret and create visual content, as well as their capacity to use visual and digital elements to express ideas and think critically. Assessment tools might include rubrics for creativity, peer review platforms, and digital portfolios (→ e-portfolio literacy). This is particularly important in L2/CALL teaching, where visual and digital aids can significantly enhance language acquisition, comprehension, critical thinking, and creativity. Rubrics specifically designed for multimodal projects can assist in evaluating aspects such as visual design, coherence, and the integration of various modes of communication. The framework proposed by Ross et al. (2020) offers guidance for developing assessment criteria for multimodal work, emphasising four key dimensions—criticality, creativity, holism, and multimodality—while acknowledging the distinct role of visual literacy within these complex projects.

Conclusion

Visual literacy is a vital component of contemporary education, particularly in L2/CALL contexts,

where learners must navigate an increasingly multimodal environment. It enhances comprehension, retention, and engagement by enabling interaction with visually and linguistically rich content. Additionally, in CALL, visual literacy serves as a form of digital literacy, equipping learners with critical and creative skills to navigate today's media-rich landscape. Integrating visual literacy into L2/CALL curricula empowers learners to meet the demands of multimodal education while becoming confident, culturally aware communicators. This comprehensive approach addresses both linguistic proficiency and twenty-first-century skills.

Cross-References

- ▶ e-Portfolio Literacy
- ▶ History of New Literacies
- ▶ Multimodal Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Avgerinou, M. D., & Pettersson, R. (2020). Toward a cohesive theory of visual literacy. *Journal of Visual Literacy, 39*(2), 85–114.

Bamford, A. (2003). *The visual literacy white paper.* Retrieved from https://www.aperture.org/wpcontent/uploads/2013/05/visual-literacy-wp.pdf

Debes, J. (1969). The loom of visual literacy: An overview. *Audiovisual Instruction, 14*(8), 25–27.

Domínguez Romero, E., & Bobkina, J. (2017). Teaching visual literacy through memes. In K. Donaghy & D. Xerri (Eds.), *The image in English language teaching* (pp. 59–71). ELT Council.

Domínguez Romero, E., & Bobkina, J. (2021). Exploring critical and visual literacy needs in digital learning environments: The use of memes in the EFL/ESL University classroom. *Thinking Skills and Creativity, 40*, 100783. https://doi.org/10.1016/j.tsc.2020.100783

Elkins, J. (2009). *Visual literacy*. Routledge.

Eutsler, L. (2021). Making space for visual literacy in literacy teacher preparation: Preservice teachers coding to design digital books. *Tech Trends, 65*, 833–846. https://doi.org/10.1007/s11528-021-00629-1

Guo, D., McTigue, E. M., Matthews, S. D., & Zimmer, W. (2020). The impact of visual displays on learning across the disciplines: A systematic review. *Educational Psychology Review, 32*, 627–656.

Kalaf-Huges, N. (2023). *Promoting information literacy and visual literacy skills in undergraduate students using infographics*. Cambridge University Press.

Kędra, K., & Žakevičiūtė, R. (2019). Visual literacy practices in higher education: What, why and how? *Journal of Visual Literacy, 38*(1–2), 1–7. https://doi.org/10.1080/1051144X.2019.1580438

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.

Metros, S. E. (2008). The educator's role in preparing visually literate learners. *Theory Into Practice, 47*, 102–109. https://doi.org/10.1080/00405840801992264

Ross, J., Curwood, J., & Bell, A. (2020). A multimodal assessment framework for higher education. *E-learning and Digital Media, 17*(4), 290–306.

Serafini, F. (2010). Reading multimodal texts: Perceptual, structural, and ideological perspectives. *Children's Literature in Education, 41*(2), 85–104. https://doi.org/10.1007/s10583-010-9100-5

Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.

Serafini, F. (2017). Visual literacy. *Oxford Research Encyclopaedia of Education*. Retrieved December 2, 2024 from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-19

Sturken, M., & Cartwright, L. (2001). *Practices of looking: An introduction to visual culture*. Oxford University Press.

Thompson, D. S. (2019). Teaching students to critically read digital images: A visual literacy approach using the DIG method. *Journal of Visual Literacy, 38*(1–2), 110–119. https://doi.org/10.1080/1051144X.2018.1564604

Unsworth, L., & Mills, K. (2020). English language teaching of attitude and emotion in digital multimodal composition. *Journal of Second Language Writing, 47*, 100712. https://doi.org/10.1016/j.jslw.2020.100712

Wileman, E. (1993). *Visual communicating*. Educational Technology Publications.

Yenawine, P. (2005). Thoughts on visual literacy. In J. Flood, D. Lapp, & S. B. Heath (Eds.), *Handbook of research on teaching literacy through the communicative and visual arts: Sponsored by the International Reading Association* (1st ed.). Routledge. https://doi.org/10.4324/9781410611161

C

Caption Literacy

Antonie Alm [1] and Yuki Watanabe [2]
[1]Languages & Cultures, University of Otago, Dunedin, New Zealand
[2]Media, Film & Communication, University of Otago, Dunedin, New Zealand

Abstract

Caption literacy is the ability to use on-screen text for language learning purposes. The term captions refers to both same-language transcripts of audio and translations of audio into another language (often referred to as subtitles) added to a video consisting of sound and moving images. Captions appear across different digital platforms, including traditional media, streaming services, social media, and educational content. The availability of captions has opened up new opportunities for language learners to engage with authentic input. Research has shown the benefits of captions for language learning, such as improved comprehension, vocabulary acquisition, and listening skills. Artificial intelligence (AI) technology has enabled the automatic generation of captions, increasing access to authentic content but still requiring human oversight to review and correct any errors. Factors that influence caption literacy development include learner proficiency level, familiarity with captions, and awareness of learning needs when engaging with captions. Educators can foster caption literacy through learner training. Future research should investigate frameworks for developing caption literacy and explore the potential of personalised and adaptive learning strategies integrating AI-based technologies.

Keywords

Caption literacy · Captions · Subtitles · Artificial intelligence · Audiovisual input · Accessibility · Digital literacies

Definition

Caption literacy is the ability to make effective use of video captions (on-screen transcriptions or translations of audio) for language learning.

Introduction

As captions become increasingly common in digital environments, caption literacy is emerging as a skill that enables language learners to productively engage with authentic video content in their target language. Alm and Watanabe (2021) define caption literacy as "the ability to manage captions effectively" (p. 638) in the context of second language (L2) viewing, distinguishing between mechanical caption literacy (the ability to locate and activate caption options) and functional caption literacy

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_263-1

(the ability to make caption choices that support personal preferences and learning needs). Building on this foundational work, the present entry proposes an extended definition of caption literacy that includes a broader range of skills and contexts, as outlined in the definition above.

Captions and subtitles are two forms of on-screen text that accompany video content but serve different purposes. Captions offer a verbatim (or slightly condensed) transcript of the audio in the same language, whereas subtitles provide a translation of the audio content into another language. In this entry, we use the term "captions" as an umbrella term for both forms of on-screen text.

Originally developed to enhance accessibility of audio content for the deaf and hard of hearing, closed captions (often abbreviated as CC, a remnant of the old system that required decoders to display the captions) were first introduced in the 1970s in the United Kingdom and the United States (Vanderplank, 2016). Over time, these same-language transcriptions, typically referred to simply as captions, have become useful tools for a wide range of users, including language learners (Gernsbacher, 2015).

Viewers now have access to a wider range of captioning options across different platforms, from traditional broadcast media to streaming services and social media. These developments include automatically generated captions and translations, enhancing accessibility and providing additional on-screen text options for viewers.

In this context, caption literacy is emerging as a valuable skill in the field of computer-assisted language learning (CALL), highlighting the potential of captions to support L2 development.

Availability and Contexts of Captions

- The widespread use of captions can be observed in various digital contexts. Streaming platforms like Netflix offer professionally created captions, while user-generated content on sites like YouTube often provide automatically generated ones, which can vary in accuracy.
- Social media platforms have also seen a significant increase in captioned video content. Platforms like Facebook, Instagram, and TikTok offer automatic captioning tools, and many videos are now even viewed without audio (Tirumala & Youngblood, 2021). Additionally, these platforms often allow users to add their own captions, further expanding the availability and variety of captioned content.
- Online educational content, such as video lectures, often includes captions (e.g. Rask.ai).
- Virtual meetings on platforms such as Zoom are frequently accompanied by real-time captions to accommodate participants who are deaf or hard of hearing, as well as those who may face language barriers (Jolly et al., 2024).
- Many video games are also offering caption options (Garcia, 2020).

The increasing presence of captions in digital contexts reflects a broader trend towards inclusivity and accessibility. As captions become an integral part of content production and distribution, they are changing the way we consume media. Thompson (2021) highlights that captions allow viewers to engage with video content while engaging in other activities. This trend towards caption-enabled multitasking adds another layer of complexity to the attentional demands on language learners (➔ attentional literacy). As a result, caption literacy is emerging as an important skill in the digital environment, encompassing not only the ability to use captions effectively for various purposes, including language learning, but also the capacity to manage attention in increasingly complex digital environments.

Captions in Language Learning

Research has shown multiple benefits of captioned videos for language learning, with effects varying based on the type of on-screen text used:

- Same-language captions (L2 captions for L2 audio):
  - Improve comprehension of spoken language (Danan, 2004; Montero Perez et al., 2013)

  - Enhance vocabulary acquisition (Pattemore et al., 2024; Peters et al., 2016)
  - Develop listening skills through improved speech segmentation (Vanderplank, 2016)
  - Support language processing by integrating multiple input channels (Mitterer & McQueen, 2009)
- Subtitles (L1 translation of L2 audio):
  - Reduce anxiety for beginning-level learners (Pattemore et al., 2024)
  - Support comprehension, especially for lower-proficiency learners (Pujadas & Muñoz, 2019)
  - May occasionally interfere with L2 lexical processing (Guichon & McLornan, 2008)
- Reversed subtitles (L1 audio with L2 subtitles)
  - Limited research exists on this approach, but a few studies suggest potential benefits for vocabulary learning (Kanellopoulou, 2019).
  - May provide a unique approach to L2 text comprehension while maintaining L1 audio support.
- Dual subtitles (L2 audio with L1 subtitles and L2 captions):
  - Enhance vocabulary learning and comprehension more than single caption/subtitle options (Dizon & Thanyawatpokin, 2021)
- Glossed captions (L2 audio with L2 captions consisting of clickable words that provide instant L1 translation):
  - Available through tools like Language Reactor, which offers interactive captions for platforms such as Netflix and YouTube, providing instant translations and additional language learning features
  - Support learners across various proficiency levels (Fievez et al., 2023)
- Keyword captions (L2 audio with L2 captions of keywords):
  - Develop listening skills by encouraging less reliance on text (Montero Perez et al., 2014)
  - Benefit higher proficiency learners in particular (Teng, 2022)

The effectiveness of these caption types differs based on learner proficiency, content complexity, and learning goals. For example, while intermediate to advanced learners often benefit more from captions, beginners may find subtitles more accessible (Peters & Webb, 2018).

Research also indicates that the frequency and spacing of vocabulary encounters in captioned content affect learning outcomes. In extensive viewing contexts, massed exposure appears more beneficial for incidental vocabulary acquisition (Fievez et al., 2023).

Familiarity with captions can positively influence language awareness. By exploring captioned L2 content across various social media and digital platforms based on their interests, learners encounter numerous opportunities for incidental language exposure and learning (Alm & Watanabe, 2021). The benefits of captions extend beyond language learners. Gernsbacher (2015) highlights that captions also aid children and adults learning to read, as well as highly literate adults, serving as a useful tool for diverse viewing contexts.

These findings illustrate both the potential and the complexity of learning with captioned content. They highlight the importance of developing caption literacy skills to use different caption types effectively.

Automatic Caption Generation and AI

Artificial intelligence (AI) and automatic speech recognition (ASR) have significantly improved the accessibility and quality of captions. AI-powered tools can now generate captions in real time with increasing accuracy. This has expanded the opportunities for language learners to engage with captioned content across different platforms (Soe et al., 2021). However, despite the progress in ASR technology, automatic captions still require human oversight and correction to ensure reliability, especially in educational contexts (Graham & Choo, 2022).

Strategies for Fostering Caption Literacy in Language Learning

The development of caption literacy is influenced by several factors, including learner proficiency and motivation level, caption quality and synchronisation, multimedia content complexity, and

learner familiarity with caption use (Gass et al., 2019; Pattemore et al., 2024). Educators can support the development of caption literacy by providing training, incorporating caption-based activities, and guiding learners in effective use of captions (Alm & Watanabe, 2021).

To effectively integrate captions into language learning, educators can use several strategies:

- *Training:*
  - Demonstrate how to access and use caption features on different platforms.
  - When using advanced captioning tools like Language Reactor, which offers a range of customisation options, familiarise learners with these features including how and when best to use subtitles and/or captions (➔ app literacy).
  - Guide learners to regularly assess and adjust their caption settings as their proficiency and comfort level with captions improve over time. This fosters the development of functional caption literacy (Alm & Watanabe, 2021).
- *Authentic content selection:*
  - Curate a collection of captioned materials that cover various genres and topics to expose learners to a wide range of language use. Incorporate content with different levels of complexity to accommodate learners' proficiency levels.
  - Encourage learners to explore captioned content independently based on their interests. This intra-formal learning approach (Alm & Watanabe, 2021) fosters autonomy and motivation by bridging learners' informal viewing experiences with formal language learning.
- *Caption activities:*
  - Get learners to create captions for their own videos or those of their peers using caption creation apps (➔AI generation literacy). This activity, which can be shared on social platforms (➔ social media literacy), promotes language production and collaboration while deepening learners' understanding of caption conventions and challenges.
  - Have learners compare and contrast caption practices in specific languages, platforms, or cultural settings. This activity draws on the concept of "cultures-of-use" (Thorne, 2003), which highlights how the use and perception of communication tools can vary across different communities. By analysing these differences, learners develop a more comprehensive understanding of caption usage while simultaneously fostering their ➔ intercultural literacy.
- *Reflective practice:*
  - Provide structured reflection prompts or assign reflective journal entries to guide learners in analysing their experiences with on-screen texts, identifying areas for improvement, and setting goals for future caption-based learning. This metacognitive practice supports the development of functional caption literacy (Alm & Watanabe, 2021).

By incorporating a range of strategies that address both the technical and sociocultural aspects of caption literacy, educators can help learners develop the skills necessary to effectively utilise captions for language learning. The intra-formal learning approach is particularly effective in bridging learners' informal experiences with formal language learning contexts, while the concept of "cultures-of-use" highlights the importance of considering how caption practices may vary across different languages, platforms, and cultural communities. Through a combination of training, authentic content selection, on-screen text creation activities, and reflective practice, learners can become proficient and critical users of captions in diverse language learning contexts.

Conclusion and Future Directions

The increasing presence of captions has created new opportunities for language learners to engage with authentic and diverse language input. However, to use these resources effectively, learners must develop the ability to engage with and process multimodal information and use captions strategically and critically for language learning purposes.

Caption technology continues to evolve, particularly with automatic caption generation

through AI. Future research must further investigate effective strategies for developing caption literacy skills, explore the potential of personalised and adaptive caption-based learning approaches, and address the accuracy issues surrounding AI-generated captions. By including caption literacy in CALL research and practice, we can ensure that language learners are equipped with the skills to benefit from the range of captioned content they encounter in their language learning journeys.

Cross-References

- ▶ App Generation Literacy
- ▶ App Literacy
- ▶ Attentional Literacy
- ▶ Intercultural Literacy
- ▶ Social Media Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Alm, A., & Watanabe, Y. (2021). Functional caption literacy development through intra-formal L2 viewing. *Aula Abierta, 50*(2), 635–642. https://doi.org/10.17811/rifie.50.2.2021.635-642

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67–77. https://doi.org/10.7202/009021ar

Dizon, G., & Thanyawatpokin, B. (2021). Language learning with Netflix: Exploring the effects of dual subtitles on vocabulary learning and listening comprehension. *Computer-Assisted Language Learning Electronic Journal, 22*(3), 52–65.

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2023). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning, 36*(1–2), 26–51. https://doi.org/10.1080/09588221.2021.1899244

Garcia, E. (2020, May 13). Top five reasons games need subtitles. *Can I Play That.* https://caniplaythat.com/2020/05/13/top-five-reasons-games-need-subtitles

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology, 23*(2), 84–104. https://doi.org/10125/44684.

Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights From the Behavioral and Brain Sciences, 2*(1), 195–202. https://doi.org/10.1177/2372732215602130

Graham, R., & Choo, J. (2022). Preliminary research on AI-generated caption accuracy rate by platforms and variables. *Journal on Technology and Persons with Disabilities, 10*, 33–53. https://scholarworks.calstate.edu/concern/publications/z316q8150

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85–93.

Jolly, A. J. J., Macfarlane, C. E., & Barker, B. A. (2024). Deaf/hard of hearing students' experiences with higher education's real-time captioning services. *The Journal of Deaf Studies and Deaf Education, 29*(3), 424–432. https://doi.org/10.1093/jdsade/enae019

Kanellopoulou, C. (2019). Film subtitles as a successful vocabulary learning tool. *Open Journal of Modern Linguistics, 9*(02), 145. https://www.scirp.org/html/10-1640737_92220.htm

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One, 4*(11), e7785. https://doi.org/10.1371/journal.pone.0007785

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141. https://doi.org/10125/44357

Pattemore, A., Suárez, M.-D.-M., & Muñoz, C. (2024). Perceptions of learning from audiovisual input and changes in L2 viewing preferences: The roles of on-screen text and proficiency. *ReCALL, 36*(2), 135–151. https://doi.org/10.1017/S0958344024000065

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal, 47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Soe, T. H., Guribye, F., & Slavkovik, M. (2021). Evaluating AI-assisted subtitling. In *Proceedings of the ACM international conference on interactive media experiences (IMX '21)* (pp. 1–12). https://doi.org/10.1145/3452918.3458792

Teng, F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System, 105*, 102736. https://doi.org/10.1016/j.system.2022.102736

Thompson, C. (2021). *Sorry, but 'I missed the meeting' is no longer an excuse*. Wired. https://www.wired.com/story/ai-means-missing-the-meeting-is-no-longer-an-excuse/

Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology, 7*(2), 38–67. https://doi.org/10125/25200

Tirumala, L. N., & Youngblood, E. (2021). Captioning social media video. *Public Relations Education, 7*(1), 169–187.

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Springer.

Immersive Literacy

Junjie Gavin Wu[1] and Sangmin-Michelle Lee[2]
[1]Faculty of Applied Sciences, Macao Polytechnic University, Macau, China
[2]Department of Metaverse, Kyung Hee University, Yongin, South Korea

Abstract

Immersive technologies, represented notably by extended reality (XR), have sparked groundbreaking transformations in the educational paradigm, necessitating the development of a novel form of literacy—immersive literacy. As an emerging field, immersive literacy encompasses the ability to appropriately select, creatively use, agentively interact with/within, and critically evaluate immersive technologies within multisensory contexts. This entry introduces a definition of immersive literacy, followed by examples and theories related to immersive language learning. The last section offers pedagogical implications that language teachers should consider when integrating immersive technologies into their teaching.

Keywords

Immersive technologies · Extended reality · Literacy · Immersive language learning · Digital literacies

Definition

Immersive literacy is "the ability to effectively navigate, interact with, and communicate and create within multimodal (and increasingly multisensory) immersive environments" (Pegrum et al., 2022, p. 20). It encompasses the skills and competencies required to effectively understand, engage with and manipulate different immersive environments, such as virtual reality (VR) and augmented reality (AR). Beyond technical skills, immersive literacy involves the user's critical awareness and reflective knowledge of immersive environments. In language learning, immersive literacy has emerged as an important yet underexplored concept for language teachers and learners.

Introduction: Overview of Immersive Literacy

Immersive literacy is realised via various immersive technologies, especially VR and AR (Pegrum et al., 2022). VR constructs an artificial, synthetic environment, isolating (or at least partially isolating, in the case of non-immersive, desktop VR) users from the physical world. According to Burdea and Coiffet (2003), the distinctive features of VR include interactivity, immersion, and imagination. In recent years, immersive VR (iVR), which typically employs a head-mounted device and hand controllers, has

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_264-1

offered representational fidelity and a high degree of immersion and interactivity (control factors), which in turn emphasise the first-person perspective, increasing the user's sense of immersion and agency and enhancing affective and cognitive factors such as interest, motivation, and self-regulation, thereby ultimately facilitating effective learning (see Makransky & Petersen, 2021).

While VR involves fully simulated contexts, AR often provides digital support in an actual authentic, everyday context. AR blends physical reality with artificial enhancements by overlaying a fictional world over the physical environment. In AR, users can remain aware of and engaged with the real world while interacting with virtual objects. For instance, in a study by Lee and Park (2020), students created digital stories and superimposed them onto real-world settings—specifically, their university campus—using context-aware, location-based AR technology. Zhang et al. (2020) have proposed a model of AR-enhanced language education, highlighting immersive experience as learners interact across formal and informal learning contexts.

Although VR and AR are different technologies, both have the affordance of creating immersive experiences. Nilsson et al. (2016) offer useful insights into the nature of immersion. By reviewing diverse perspectives on immersion, such as sensory, imaginative, ludic, and engagement immersion, they proposed three types of immersion: (1) *system immersion* that focuses on technology-enabled absorption, (2) *narrative immersion* where narrations are so appealing that learners are fully engaged, and (3) *challenge-based immersion,* in which learners become mentally and physically invested through encountering challenges. An ecological integration of these three types of immersion can maximise the effect of immersive learning.

In the realm of language acquisition, immersion facilitates the development of both oral and written language skills through learners' interactions with the simulated or real environment, other participants, and embedded objects. To maximise the benefits of these environments, learners must cultivate immersive literacy skills. However, proficiency in additional literacies is also essential, given that immersive environments typically incorporate an array of technologies. For instance, considering that immersive environments comprise various modalities (e.g., sound, text, and image) and modes (auditory and written), ▶ multimodal literacy skills are crucial for effective engagement. Furthermore, as immersive environments often present as open worlds where learners autonomously seek and locate information for self-directed learning and interact with peers, competencies in ▶ information literacy and ▶ digital literacies in general become indispensable.

Immersive Literacy in Language Teaching/Learning

An immersive environment can contribute to foreign language learning in several ways. First, it provides authentic, meaningful contexts—or simulations of them—that are often lacking in traditional language learning classrooms, which struggle to offer the situated and contextualised learning, given that language is intrinsically bound to its context (Kaplan-Rakowski & Gruber, 2024; Lee & Park, 2020; Wu et al., 2023). Second, it can promote active participation of learners by engaging them in exploring the environment and interacting with the embedded content, such as non-player characters and virtual objects (Lee et al., 2024). Third, it can enhance the social aspects of language learning, such as interaction and collaboration among learners, which facilitates communication. Fourth, the immersive environment integrates multiple modes of communication (visual, auditory, and textual), engaging learners through different senses (see discussion of embodied cognition in Shapiro, 2011). Finally, language learning in the immersive environment can increase positive affective factors such as motivation, engagement, enjoyment, and satisfaction, and decrease negative affective factors such as anxiety (Lan, 2021; Lee et al., 2024).

Lan (2021) reviewed the theoretical foundations of VR-enhanced language education and proposed that both sociocultural theory and embodied cognition play important parts, stressing the role of social interactions in shaping

cognition via "bodily sensation, perceptions, and actions" (p. 1). In their earlier study on AR, Dunleavy and Dede (2014) posited that AR technology has the capacity to facilitate situated learning, to enable learners to interact with virtual objects within real-world contexts, and to foster enhanced motivation, engagement, and a more profound comprehension of abstract concepts. From this perspective, both VR and AR technologies demonstrate significant potential for advancing language learning by immersing learners in authentic communicative situations.

Numerous studies have confirmed the educational potential of immersive technologies for language education in the cognitive, affective, and linguistic domains (Lee et al., 2024; Topu et al., 2024; Wu et al., 2023, 2024). Lee et al. (2024), for example, utilised iVR in creating immersive learning experiences for elementary language learners in South Korea. Their study highlighted the positive impact on students' behavioural, cognitive, and affective engagement, leading to enhanced language learning performance. From the perspective of educational psychology, Kaplan-Rakowski and Gruber (2024) reported that iVR can reduce cognitive load, boost motivation, and create a stronger sense of presence in language learning.

AR can positively impact several dimensions of language learning, including vocabulary acquisition, speaking, and learner motivation and engagement (for a detailed explanation and examples, see Pegrum, 2019). AR is particularly effective in cultivating intercultural competence by incorporating the real, physical context into learning and allowing students to learn about the target culture within the physical environment (Lan, 2021). In addition, AR has the potential to enhance metacognitive skills such as self-efficacy, attitudes, and learners' ability to transfer their knowledge and skills from one task to another (Lee & Park, 2020).

The development of immersive literacy is of course also important for educators. It is important to acknowledge that teachers do not necessarily possess the knowledge and skills to effectively utilise immersive technologies in language education. Two recent studies by Lee and Wu (2023) and Lee and Wu (2024) drew upon the classic technological, pedagogical, and content knowledge (TPACK) framework (Mishra & Koehler, 2006) and substitution-augmentation-modification-redefinition (SAMR) model (Puentedura, 2013) to train pre-service and in-service teachers. These studies underscore that teacher education in immersive literacy is of prime importance for transforming pedagogical practices. Despite its importance, teachers' immersive literacy remains largely under-taught and under-researched. Researchers have emphasised the importance of basing learning designs on a solid pedagogical framework (Zhang et al., 2020) and, more broadly, the role of language teachers as designers (see also Pegrum, 2019). To maximise learning through the use of immersive technologies, Zhang et al. (2020) suggested that a design thinking model, including five components of empathise, define, ideate, prototype, and test, can be useful in creating immersive experiences for language learning.

Immersive technologies are not without flaws and potential risks. The immersion created by the technologies can lead to a reduced understanding of the problems associated with dangerous situations, such as training firefighters only in virtual environments (Wu et al., 2024). Another scenario involves language learners with special needs, who may benefit from the stress-free virtual environment, but also need to be exposed to the complexities of the real world. Finally, using a head-mounted device may cause physical discomfort and dizziness.

Suggestions for Educators

This entry concludes with a few suggestions for educators:

1. Set distinct learning objectives based on diverse levels of immersive literacy.

   To effectively use immersive technologies, educators need to set the goals for language learning first. Teachers should help students to develop immersive literacy by designing tasks which range from the lower to the higher levels of Bloom's Revised Taxonomy (Anderson et al., 2001), for example, to support learners

"to learn about the benefits of VR in vocabulary learning", "to critically select an appropriate AR platform to practice speaking", and "to create an immersive platform based on students' own learning needs". This can serve as a springboard for educators to agentively update their existing teaching practices to improve learners' immersive literacy.

2. Identify appropriate immersive technological tools for learners' needs.

   Technology selection should be emphasised to align with learners' needs based on the affordances of diverse immersive technologies. For example, iVR may be appropriate for virtual exchange projects, facilitating interaction among learners from diverse geographical locations. Conversely, non-immersive VR could be more suitable for language learners with neurodivergent learners, including autistic learners (Bradley & Newbutt, 2018), prioritising their physical safety and comfort. In addition, to maximise the effectiveness of immersive technologies, educators should balance their use with traditional teaching methods. For example, AR can be a useful tool to complement textbooks by showing 3D avatars and animations to make learning more vivid and engaging (e.g., Hung & Yeh, 2023).

3. Promote metacognitive, cognitive, and affective strategies in using immersive technologies in education.

   Metacognitive, cognitive, and affective strategies play a crucial role in learning, especially in immersive environments, where students need to actively engage with the learning material. According to Makransky and Petersen (2021), learning outcomes in VR environments are shaped by the complex interaction between several learner factors: their interest levels, motivation, belief in their own abilities (self-efficacy), capacity to manage their learning (self-regulation), and mental workload (cognitive load). These factors work together, each influencing and being influenced by the others, to determine how effectively students learn in virtual reality settings. Therefore, instructors should select the VR contents which can enhance students' motivation and interest. While (i)VR contents can attract students, visual details and multimodal materials can increase extraneous cognitive load. Hence, unnecessary environmental details should be avoided. This is the same with AR (see Şimşek et al., 2024).

Unlike classrooms where learning typically follows a predetermined, sequential path, VR and AR learning environments may allow students to freely explore and interact with content in multiple ways and sequences. In these virtual spaces, students can choose their own learning paths, move between different activities at their own pace, and engage with materials in the sequence that best suits their learning needs. This flexibility and freedom of movement make students' metacognitive strategies, such as self-efficacy and self-regulated learning, essential for successful learning. To promote these strategies, instructors need to make the learning goals clear, provide scaffolding and feedback, peer collaboration, and opportunities for reflection, and by doing this, students can enhance their immersive literacy (Wu et al., 2024).

This entry has given an overview of immersive literacy from both technological and pedagogical perspectives. It can serve as a springboard for language educators to utilise immersive environments to support their students' language learning and to help them develop immersive literacy.

Cross-References

► Digital Literacies
► Information Literacy
► Multimodal Literacy

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., & Wittrock, M. (2001). *A taxonomy for learning, teaching and*

assessing: A revision of Bloom's taxonomy. Longman Publishing.

Bradley, R., & Newbutt, N. (2018). Autism and virtual reality head-mounted displays: A state of the art systematic review. *Journal of Enabling Technologies, 12*(3), 101–113.

Burdea, G. C., & Coiffet, P. (2003). Virtual reality technology. John Wiley & Sons.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 735–745). Springer.

Hung, H. T., & Yeh, H. C. (2023). Augmented-reality-enhanced game-based learning in flipped English classrooms: Effects on students' creative thinking and vocabulary acquisition. *Journal of Computer Assisted Learning, 39*(6), 1786–1800.

Kaplan-Rakowski, R., & Gruber, A. (2024). An experimental study on reading in high-immersion virtual reality. *British Journal of Educational Technology, 55*(2), 541–559.

Lan, Y. J. (2021). Language learning in virtual reality: Theoretical foundations and empirical practices. In Y. J. Lan & S. Grant (Eds.), *Contextual language learning* (pp. 1–22). Springer.

Lee, S.-M., & Park, M. (2020). Reconceptualization of the context in language learning with a location-based AR app. *Computer Assisted Language Learning, 33*(8), 936–959.

Lee, S. M., & Wu, J. G. (2023). Teaching with immersive virtual reality: Perceptions of Korean trainee teachers. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 13*(1), 1–14.

Lee, S. M., & Wu, J. G. (2024). Preparing teachers for the future: Microteaching in the immersive VR environment. *ReCALL, 36*(3), 251–269.

Lee, S. M., Yang, Z., & Wu, J. G. (2024). Live, play, and learn: Language learner engagement in the immersive VR environment. *Education and Information Technologies, 29*(9), 10529–10550.

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology, 12*(2), 108–134.

Pegrum, M. (2019). *Mobile lenses on learning*. Springer.

Pegrum, M., Hockly, N., & Dudeney, G. (2022). *Digital literacies*. Routledge.

Puentedura, R. R. (2013). SAMR: Moving from enhancement to transformation. Retrieved from [19]. http://www.hippasus.com/rrpweblog/archives/2013/05/29/SAMREnhancementToTransformatio n.pdf

Shapiro, L. (2011). *Embodied cognition*. Routledge.

Şimşek, B., Direkci, B., Koparan, B., Canbulat, M., Gülmez, M., & Nalçacıgil, E. (2024). Examining the effect of augmented reality experience duration on reading comprehension and cognitive load. *Education and Information Technologies*, 1–20.

Topu, F. B., Yilmaz, R. M., & Tulgar, A. T. (2024). The effects of using augmented reality on vocabulary learning and attitude of pre-school children in English education. *Education and Information Technologies, 29*(10), 11733–11764.

Wu, J. G., Miller, L., Huang, Q., & Wang, M. (2023). Learning with immersive virtual reality: An exploratory study of Chinese college nursing students. *RELC Journal, 54*(3), 697–713.

Wu, J. G., Yang, Z., Wu, S., & Zou, D. (2024). Unveiling the synergy of peer feedback and the Metaverse. *Computers & Education: X Reality, 4*, 100056.

Zhang, D., Wang, M., & Wu, J. G. (2020). Design and implementation of augmented reality for English language education. In V. Geroimenko (Ed.), *Augmented reality in education: A new technology for teaching and learning* (pp. 217–234). Springer.

E

Educating Teachers for Mobile Learning

Agnes Kukulska-Hulme
The Open University, Milton Keynes, UK

Abstract

The use of digital technologies, especially mobile devices, is having significant impacts on education. Educators have diverse responses, ranging from enthusiasm to caution, as they recognise both the potential and challenges of these tools. Mobile learning, characterised by portability and context-awareness, can support more personalised, seamless learning experiences that extend beyond classroom boundaries. This can align with, or disrupt, traditional methods. Teachers may need to rethink their practices to address the individualised nature of mobile learning, while also considering factors like device access and safety concerns, especially with younger students. Community-based professional learning approaches offer valuable support for educators to share strategies and solutions, while guidelines and frameworks provide structured ways to help educators integrate mobile learning meaningfully and responsibly into their educational contexts.

Keywords

Teacher professional development · Community of practice in education · Mobile pedagogy · Mobile assisted language learning · Situated learning · Contextual learning · Activity design frameworks

Introduction

Effective use of digital technologies, including mobile devices such as smartphones and tablets, has become an important part of education both inside and outside the classroom. As our understanding of the impact of these technologies has grown, educators and institutions have adopted different attitudes and practices. Some are embracing the opportunities, while others may be cautious or even antagonistic, concerned about undesirable consequences of the use of technology-mediated learning.

Teacher education and professional development programmes supporting teachers with technology adoption acknowledge that some of the contemporary tools and associated pedagogies may fit in easily with current teaching environments and practices, but others will be much more disruptive. Mobile learning, which incorporates mobile technologies in teaching, learning and assessment, can be in harmony with existing practices, or it may call into question common assumptions about what, where, when, and how students learn and how their learning is evaluated.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_273-1

These possibilities arise because there are diverse conceptualisations of mobile learning (Chen & Tsai, 2021; El-Hussein & Cronje, 2010). Some may view mobile technology as providing easy access to content or communication, while others direct attention to how educational and social experiences are fundamentally altered when mobile technology is used.

From a fundamental change perspective, mobile learning has been described as "a distinctive approach that focuses on the merits of situated, contextual, and seamless learning with portable technology" (Kukulska-Hulme, 2024, p. 302). Following this description, teachers might want to look at their assumptions and current habits around places and times devoted to teaching, learning, and student support and remain open to a variety of new approaches, benefits, and challenges.

What can be achieved with mobile learning varies according to factors such as student age group, student needs, access to suitable devices, application features, flexibility of the curriculum, and social acceptance of a particular technology—to name just a few. Teachers know that allowing children to use smartphones is problematic, so benefits have to be balanced against drawbacks, and safeguards have to be put in place. There are tensions between guidance for schools on prohibiting the use of mobile phones (e.g. Department for Education, 2024) and perspectives that are in favour of constructive use to support teaching and learning (e.g. DePaul, n.d.; Sinaga et al., 2023). It can be easier to use mobile devices among adults, young people who take their education seriously, and in situations where there is limited access to quality teaching resources. Scholars have also noted the positive role of mobile learning in teacher professional development as a means of continually improving the quality of teaching and learning (Gao et al., 2021; Holden, 2021).

Mobile learning can be used in safe ways; for instance, sometimes this can be achieved by lending out devices that have limited software installed and are geared towards specific educational activities. Higher levels of safety may also be achieved by teaching children how to use smartphones responsibly and productively in a learning context and additionally by educating parents and guardians. In the case of adult learners, there could be both benefits and barriers arising from adults' multiple roles in their families, communities, and at work.

How Mobile Learning May Change Language Learning

The question "Will mobile learning change language learning?" remains as relevant today as it was when mobile learning was in its early days of adoption (Kukulska-Hulme, 2009). Classroom-based language learning is characterised by constrained potential to satisfy learners' individual needs in relation to the social situations they will encounter when using or understanding a target language, especially in their daily life or work. Mobile learning can enable learners to capture their difficulties and successes in various media formats on the go, to reflect on their needs and to access immediate help. However, that mode of operation changes their learning, moving it in a more individualised direction than is common in class. Teachers then need to rethink how to teach and support those learners.

Contextual mobile learning may use mobile technologies that automatically recognise a learners' context (e.g. where they are, what they could learn there, what they could create). The "context" could be within a classroom or campus, nearby or far away. Thus, the focus is on the context, and while this can be incorporated into teaching approaches and materials, it could require a change from usual practices. In that case teachers might want to consider the range of contexts that could enhance their students' learning.

Mobile learning is also associated with seamless learning (Wong, 2015), meaning that usual boundaries or "seams" between settings are disrupted or dissolved. For example, students may be encouraged to undertake some activities on the way to class, during and/or after class, at home, and anywhere else that may help them or may be convenient. Sometimes everything is recorded on their mobile device. Seamless

learning should have a strong rationale, for example, in terms of better student engagement in language learning or meeting their needs more effectively.

Some Forms of Education for Teachers Moving into Mobile Learning

Since mobile technologies such as smartphones can be both educational and private life tools, it is important for teachers to understand and to educate themselves on a regular basis as to how these technologies may support learning or deliver educational experiences, as well as to appreciate any negative issues and how to address them. A "community of practice" or "learning community" approach can be an excellent way for teachers to share their knowledge and challenges and to help one another progress (Wang & Ma, 2017). Students may be part of such communities if time, spaces and cultural norms allow it. Community members can propose topics, share experiences, and work together to come up with solutions, materials, methods, or policies. They also need to be supported to try things out in practice, reflect on what worked and what did not work or could be done better, and keep up to date with the latest developments.

In the context of higher education, authors France et al. (2021) ask: "Should you be using mobile technologies in teaching?" and they suggest a "pathway" approach for instructors, with a set of guiding questions addressing pedagogical and technological considerations when incorporating mobile learning into course design and assessment. This is in the context of field-based experiential learning, which is closely aligned to authentic, task-based, and project-based language learning that focuses on real-world settings. By asking a series of questions about aspects of pedagogy (e.g. Is the use of mobile technology expanding or encouraging effective collaboration or communication for the student?) and technology (e.g. Will students have equal access to the mobile technology you are proposing?), instructors can come to a deeper understanding of mobile learning and how it may be implemented, thereby developing themselves professionally while they are also designing experiences for, and with, their students.

In the Mobile Pedagogy guide for teachers of English, informed by research conducted with teachers and learners in ESOL (English for Speakers of Other Languages) and EAP (English for Academic Purposes) contexts, authors Kukulska-Hulme et al. (2015) set out a philosophy and proposed a frame of reference to aid teachers when designing mobile language learning in and beyond the classroom. The framework highlights the use of activities which exploit a dynamic language and technology environment while drawing on the distinctive capabilities of teachers and learners. It supports teachers to design language learning activities that connect the world of language education with the multimodal text processing and creation that learners engage in beyond the classroom. As in the previous example, questions provide a framework for language activity design, for instance, "How does the activity make the most of circumstances and resources to enable more practice?". The framework encompasses people, technologies, the dynamic nature of language, and learner mobilities, meaning "the places and times when people can learn, the range of contexts and cultural settings they occupy and the personal goals that motivate learners to keep on learning beyond the confines of the classroom" (p. 8).

Conclusion

Mobile devices can be used in limited and focused ways, or more extensively as the basis of a radical change in pedagogy. In either case, educators will welcome active support from their institutions and colleagues, so as to be able to try things out and discuss with others. Mobile devices have become sophisticated tools, so conversations around safety and security are as important as discussions around aspects such as learning aims or logistics. Structured frameworks are helpful in guiding thinking and learning experience design. Mobile learning is an effective approach and can be transformative in many contexts; therefore educators

should have opportunities to expand their understanding and practical experience in this domain.

Cross-References

- ▶ Advantages and Disadvantages of Using MALL
- ▶ Call Teacher Communities of Practice
- ▶ Integration of MALL into Curriculum Design and Language Teaching Methodologies
- ▶ Mobile GBLL
- ▶ Preparation and Professional Development of Teacher Educators
- ▶ Professional Learning
- ▶ Self-directed Professional Development
- ▶ Teacher's Digital Competencies
- ▶ Teacher's Digital Literacy
- ▶ Teacher's Technology-Enhanced Learning Design

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Chen, C. H., & Tsai, C. C. (2021). In-service teachers' conceptions of mobile technology-integrated instruction: Tendency towards student-centered learning. *Computers & Education, 170*, 104224. https://doi.org/10.1016/j.compedu.2021.104224

Department for Education. (2024). Mobile phones in schools: Guidance for schools on prohibiting the use of mobile phones throughout the school day. UK Government publication. https://assets.publishing.service.gov.uk/media/65cf5f2a4239310011b7b916/Mobile_phones_in_schools_guidance.pdf

DePaul. (n.d.). Mobile learning. DePaul University Teaching Commons. https://resources.depaul.edu/teaching-commons/teaching-guides/technology/Pages/mobile-learning.aspx

El-Hussein, M. O. M., & Cronje, J. C. (2010). Defining mobile learning in the higher education landscape. *Journal of Educational Technology & Society, 13*(3), 12–21.

France, D., Lee, R., Maclachlan, J., & McPhee, S. R. (2021). Should you be using mobile technologies in teaching? Applying a pedagogical framework. *Journal of Geography in Higher Education, 45*(2), 221–237. https://doi.org/10.1080/03098265.2020.1773417

Gao, M., Tu, X., & Li, J. (2021). A review of the research on pre-service English teachers' professional development based on mobile technology. *Journal of Language Teaching and Research, 12*(3), 404–410. https://doi.org/10.17507/jltr.1203.10

Holden, A. C. (2021). *Mobile learning and teacher quality: Assessing the impact of a mobile application on the positive psychology and professional practice of Panamanian teachers* [EdD Dissertation, University of Cambridge]. https://doi.org/10.17863/CAM.79383

Kukulska-Hulme, A. (2009). Will mobile learning change language learning? *ReCALL, 21*(2), 157–165. https://doi.org/10.1017/S0958344009000202

Kukulska-Hulme, A. (2024). A hopeful future for mobile language learning. In R. Hampel & U. Stickler (Eds.), *The Bloomsbury handbook of language learning and technology* (pp. 302–314). Bloomsbury.

Kukulska-Hulme, A., Norris, L., & Donohue, J. (2015). Mobile pedagogy for English language teaching: A guide for teachers. ELT Research Papers 14.07. The British Council. https://www.teachingenglish.org.uk/publications/case-studies-insights-and-research/mobile-pedagogy-english-language-teaching-guide

Sinaga, R. A., Febrianti, K. V., & Candra, D. (2023). The benefits and drawbacks of adopting mobile learning as perceived by junior high school teachers in Taiwan. *Interactive Learning Environments, 31*(10), 6356–6365. https://doi.org/10.1080/10494820.2022.2036200

Wang, L., & Ma, Q. (2017). Community of practice: Building a mobile learning community in a higher education institution to promote effective teaching and learning. In S. C. Kong, T. L. Wong, M. Yang, C. F. Chow, & K. H. Tse (Eds.), *Emerging practices in scholarship of learning and teaching in a digital era* (pp. 19–38). Springer. https://doi.org/10.1007/978-981-10-3344-5_2

Wong, L. H. (2015). A brief history of mobile seamless learning. In L.-H. Wong, M. Milrad, & M. Specht (Eds.), *Seamless learning in the age of mobile connectivity* (pp. 3–40). Springer.

T

Teacher Role in Online Course and Materials Design

Kate Borthwick
University of Southampton, Southampton, UK

Abstract

This article examines the role of the teacher in online course and materials design for language learning. It considers the skills, competencies and knowledge required by language teachers to be effective in creating and delivering online education and reflects upon how these skills have evolved as technology has developed. Teacher competencies and knowledge have altered in response to technical developments that enable greater networked sharing, collaboration and communication across the internet. This progress has facilitated a change in understanding of how education is delivered, with a move from teacher-centred to learner-centred approaches. As a result, the teacher role has become multi-faceted and includes aspects of designer, expert curator and facilitator of learning. The article reflects upon how new technologies, such as generative artificial intelligence, might further impact the role of the teacher in creating and fostering an effective learning experience. It suggests that the Community of Inquiry framework (COI) and its notion of 'teaching presence' is a suitable way of contextualising and framing how the role of the teacher has adapted and can continue to adapt in a rapidly changing world of technology.

Keywords

Teacher role · Online teacher · Online design · Online materials · Teacher skills · Teacher competencies · Language teaching online

Introduction

Rapid advances in technology and the internet over the last half century have enabled learning and teaching to take place in increasingly blended, or entirely online contexts, and the emergency remote teaching engendered by the Covid-19 pandemic saw an acceleration of this process (Rüschoff, 2023). This evolving situation has required language teachers to develop new skills, competencies and knowledge to both understand the affordances of technical tools and software as well as their application in pedagogically sound ways. At the same time, the development of the internet has enabled extensive, networked creation and sharing of individual content offering the potential for individuals to drive and control their own learning in personalised ways. As a result, traditional teacher-centred approaches, in which the role of the teacher is one of sole leadership and direction of learning, have given way to more learner-centred approaches that are characterised

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_274-1

by collaboration, co-creation of content, flexible modes of engagement and the possibility of personalised and individual learning pathways. As the digital world has altered how we discover and learn information, so the role of the teacher has developed to encompass and demand a range of skills: an understanding of how technology impacts and influences learning, and the design of online learning content, activities and environments that facilitate and enable effective online language learning for diverse audiences. The role of the teacher in online and course materials design has become multi-faceted and requires a teacher, at different times, to be technician, instructional designer, content creator, expert curator of resources, guide and facilitator and role model for digital literacy. All of these aspects have become part of delivering an effective online language learning experience.

Skill and Competencies of the Online Language Teacher

In the early 2000s, as the internet developed and more teaching began to take place through online activities and in online spaces, a need for "*technology-proficient language instructors*" (Hubbard, 2008, p. 177) who could "*effectively and critically engage in technology-related issues*" (Chapelle & Hegelheimer, 2004, p. 300) was identified. It began to be clear that new skills would be required of teachers to ensure effective engagement and learning took place online and there was a growing realisation that it was a myth to assume that good face-to-face teachers could easily teach effectively online (Davis & Rose, 2007). It was even suggested that there needed to be a fundamental change in how language educators perceived their role: "*[there is a]…need for the online instructor to have a paradigm shift in perceptions of instructional time and space, virtual management techniques and ways of engaging students through virtual communications in addition to the communication skills already required for general effective classroom teaching*" (Compton, 2009, Current state of CALL and language teacher education section, para. 5).

The new key skills that online language teachers required were identified by Hampel and Stickler (2005) as:

- Basic Information and Communications Technology (ICT) competence
- Specific technical competence in software or hardware used for teaching
- An understanding of how to deal with the constraints and possibilities of the online medium
- Skills in encouraging socialising online
- The facilitation of communication
- Creativity and choice in the design and implementation of tasks
- The development of an individual, personal style of teaching

Hampel and Stickler presented their idea as a skills pyramid, with each skill enabling the development of the next, with basic ICT competence as the foundation and the top of the pyramid being an ability to demonstrate personal, individual style in online teaching. They revised their ideas in 2015, to indicate that basic ICT competence could now be assumed, to merge other areas together and to add an additional element:

- Basic ICT competence (assumed)
- Specific technical competence and understanding of constraints/affordances of the medium
- Skills in encourage socialising online and facilitating communication
- Creativity, own personal style
- Negotiation of online teaching spaces (Hampel & Stickler, 2015)

These changes reflect the technical development of the internet that saw a proliferation of easy-to-use online content creation tools that integrated social and communicative aspects, and the development of online spaces for the easy sharing of online content. Teachers could engage with online tools and develop a range of audio-visual or textual content more readily than ever before. The addition of 'negotiating online teaching spaces' (Hampel & Stickler, 2015) suggests collaboration with learners on how teaching will be

carried out and indicates a shift in the role of the teacher, to a more learner-centred approach that is flexible and co-developed with learners themselves.

Online Teacher as Learning Designer and Facilitator of Effective Learning

At the same time, it was evident that developing new skills to operate as an educator in the digital twenty-first century was only part of what was needed for effective teaching and learning in online spaces, and that new reflection was required on the pedagogical role of technology in language teaching and learning and the teacher's role within that context: "*Alongside the process of skills development, and enabling it, there needs to be substantial work on developing pedagogical understanding of the affordances of the online medium and acceptance of the transformation required in how teachers perform their role*" (Comas-Quinn, 2011, section 1.1, para. 4). The suggestion of 'transformation' moves consideration of the role of the teacher from a discussion of essential technical skills and competencies to one that is connected to the pedagogy of teaching and learning itself in an online space.

In addition to blended approaches, the variety of courses offered within the landscape of online education, from online degree programmes to short courses to Massive Open Online Courses (MOOCs), demands new awareness and knowledge from teachers in course design (Beetham & Sharpe, 2020). An online language teacher is presented with a context for teaching and learning that may be very different from a traditional, in-person classroom. There may be a more diverse range of learners, of varying ages, from diverse educational or cultural backgrounds, perhaps operating in different time zones, or learning part-time alongside other commitments. Learners are likely to be based in varying geographical locations and studying on their own, rather than in a classroom. In order to foster learning, active consideration needs to be given to the affordances and constraints of the medium, how online content is presented and realised, how meaningful engagement and communication will be fostered, and how learners can collaborate and learn from each other. The role of teacher becomes 'designer of learning' and the task and course design becomes significant for successful learning outcomes (Morgana, 2024). Of equal importance is the learning environment established by the teacher: teachers need to create "*structured, non-threatening, inclusive online learning spaces…[to enable learners] to explore, exchange information, connect and apply ideas and successfully to develop their language skills*" (Murphy, 2015, p. 61). Therefore, the role of the teacher takes on aspects of learning designer and facilitator of engagement, collaboration and communication, so that learners can learn from each other in an environment designed to foster their full, engaged participation.

The development and delivery of particular kinds of online course, for example, MOOCs, requires further skills in designing engaging and useful experiences at scale, for a mass cohort of students, who are unknown to the course designer and have varied motivations for learning. Consideration of open course design and engagement in open courses are essential for engaged learning to take place (Liyanagunawardena et al., 2015), especially in fostering the development of communicative skills in language (Appel & Pujolà, 2021).

The Online Teacher as Expert Curator and Creator of Content

The ability of learners to learn from peers (or from others in general) has been transformed by the expansion of content on the internet in what has been called "*an age of abundance in terms of resources, opportunities and networks*" (Beetham & Sharpe, 2020, p. 3). As content has proliferated, the teacher's role has evolved to include expertise in the curation of digital third-party content. This kind of content is not created by the teachers personally but is part of the vast knowledge database of the internet.

As part of this vast database, there are resources specifically created for language education that are openly licensed for use and adaptation by others (open educational resources—

OER) and offered through websites or repositories (Comas-Quinn & Borthwick, 2015), such as the Center for Open Educational Resources and Language Learning (COERLL), University of Texas, USA (COERLL, 2024). There are also increasing numbers of MOOCs in languages available globally that are designed for anyone to take, are free and are self-access (Martín-Monje & Borthwick, 2021) and can also be integrated into teaching, as courses or individual open resources. The benefits for teachers of drawing upon this global database of open content have been identified as offering additional content to learners, providing content outside of teachers' knowledge or expertise or supporting language teaching with a variety of authentic, multimodal resources (Comas-Quinn & Borthwick, 2015). Any language educator can also create and share their own work online for use by others. The meaningful, expert curation and presentation of this 'abundance' of open content for the specific benefit of their learners has become another element of the role of the effective online teacher.

The Online Teacher/Designer in an Era of Generative Artificial Intelligence

The rapid advances in development of generative Artificial Intelligence (GenAI) over 2022–23 and its widespread adoption (Reuters, 2023) present the need for teachers to reflect once more on how a particular technology might support the design and delivery of effective online educational experiences, and how a teacher's role might need to adapt.

Generative Artificial Intelligence tools, such as ChatGPT (OpenAI, 2024), have the capacity to generate audio-visual and text content in response to questions and instructions given to them, including translated content. They are often dialogic in nature (learners use chatbots to engage with them), can adapt their responses to further instruction and have the capacity to do this at great speed (Belda-Medina & Kokošková, 2023).

Benefits for learning have been identified with the tools taking teaching support roles such as a personal tutor giving corrective feedback, a study buddy offering definitions and summaries, an idea generator to help in getting started on work, a collaborator to assist in creation and editing, or a translator (UNESCO, 2023). These abilities offer a tremendous opportunity to support language teachers in the delivery of key activities, such as feedback and marking or translation of content, while giving a personalised learning experience for each learner. GenAI tools suggest the potential to free up teachers from certain tasks in order to focus on more creative or productive activities with their learners. Similarly, learners may be able to demonstrate more control and autonomy over their learning by making use of GenAI to support their own individual learning aims.

The ability to rapidly create authentic, personalised content is useful for language teachers seeking to create targeted, relevant linguistic content for their own learners. A 2023 study for the UK's British Council indicated that language educators are already using GenAI to support their teaching in key ways, with 57% of respondents using AI to create materials; 53% using AI to help learners practice English language; 43% using AI to create lesson plans, 23% for grading/assessment and 33% using it to correct learners' English (Edmett et al., 2023). As GenAI develops, the challenge will be understanding how far its role in teaching support alters the role of the human teacher in the learning process.

One area where teachers can play an important role is in developing their learners' digital- and AI- literacy skills in understanding how to evaluate and use GenAI content effectively, as the flaws and limitations of GenAI content are widely reported (Edmett et al., 2023). For example, the tools are trained on data models that reflect bias (Gross, 2023), are dominated by Western perspectives so often lack the cultural and idiomatic nuance necessary in language education (CIOL, 2024), and give information that is often inaccurate or out-dated (OpenAI, 2024). Teachers are well-placed to guide learners on how to evaluate the quality of GenAI-output (Bearman et al., 2024) and role model appropriate, effective use of the tools.

From Teachers to Teaching Presence: A Conceptual Lens Through Which to View the Role of Teacher

One way to make sense of the evolving role of the teacher in online course and material design, in a never-ending world of innovation and technical development, is to use the Community of Inquiry (COI) framework (Garrison et al., 1999, 2010) as a lens through which to view the learning experience as a whole.

The COI is a well-known theoretical framework that enables learning designers to consider how the interplay of collaboration, communication, engagement, interaction and directed content works to create the most effective educational experience. It "*identifies the education experience as occurring at the convergence of three presences: cognitive, teaching and social presence. Presence is defined as a state of alert awareness, receptivity, and connectedness to the social, cognitive, emotional and physical workings of both the individual and the group in the context of their learning environments*" (Cleveland-Innes, 2020, p. 87, adapted from a definition by Rodgers & Raider-Roth, 2006). The framework suggests that the presences influence each other and should be approached in an inter-connected way.

We will focus on the 'teaching presence' element of the framework here. The idea of 'teaching presence' is defined as "*the design, facilitation and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes*" (Anderson et al., 2001, p. 5). The term specifically focuses on 'teaching' not 'the teacher', in order to account for the varied ways that teaching can be manifested: through interaction with peers and fellow classmates, online via networks, engaging with open educational content, GenAI chatbots, as well as direct content transmission from a teacher. Conceptualising the role of the teacher within a 'teaching presence' that is realised in multi-faceted ways, helps to give context to how the role of the teacher—and the nature of the skills that online teachers need—has evolved through advances in technology. It is important to note that it does not minimise the role of the embodied, human teacher in the educational process, but acknowledges the importance of the teacher as a designer within the holistic educational experience establishing content, activities and the learning environment. It has been noted that the role of "*. . .teaching presence in creating and sustaining social and cognitive presence, encouraging an open and constructive learner-teacher relationship and helping to boost learner satisfaction [is significant]*" (Murphy, 2015, pp. 60–61). Viewing the role of the teacher in the light of 'teaching presence' also allows for flexibility and adaptation when faced with the advent of new technologies (such as GenAI).

Conclusion

The role of the teacher in online course and materials design is multi-faceted and is realised in ways that align to the development of the internet that encourages networked, social, collaborative, personalised learning. In this context, it is perhaps more appropriate to think of the teacher role as operating within a notion of 'teaching presence' that requires a teacher to be both a creator and expert curator of content and a designer of rich, collaborative, communicative, inclusive and personalised learning experiences.

Cross-References

- ▶ CALL Teacher Communities of Practice
- ▶ Language Teacher Education for CALL
- ▶ Preparation and Professional Development of Teacher Educators
- ▶ Teacher's Digital Competencies
- ▶ Teacher's Digital Literacy
- ▶ Teacher's Technology-Enhanced Learning Design

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anderson, T., Rourke, L., Garrison, D., & Archer, W. (2001). Assessing teaching presence in a computer conferencing context. *Online Learning, 5*(2), 1–17. https://doi.org/10.24059/olj.v5i2.1875

Appel, C., & Pujolà, J. T. (2021). Designing speaking interaction in LMOOCs: An eTandem approach. *ReCALL, 33*(2), 161–176. https://doi.org/10.1017/S0958344021000045

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment and Evaluation in Higher Education, 49*, 893–905. https://doi.org/10.1080/02602938.2024.2335321

Beetham, H., & Sharpe, R. (Eds.). (2020). *Rethinking pedagogy for a digital age* (3rd ed.). Routledge. https://doi.org/10.4324/9781351252805

Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education, 20*, 62. https://doi.org/10.1186/s41239-023-00432-3

Chapelle, C. A., & Hegelheimer, V. (2004). The language teacher in the 21st century. In S. Fotos & C. M. Browne (Eds.), *New perspectives on CALL for second language classrooms* (pp. 299–316). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410610775

CIOL Voices. (2024). *AI and translation: A CIOL white paper on the future for languages and linguists*. Chartered Institute of Linguists. Retrieved August 12, 2024, from https://www.ciol.org.uk/ciol-ai-voices-white-paper

Cleveland-Innes, M. (2020). Community of inquiry theoretical framework: Designing collaborative online and blended learning. In H. Beetham & R. Sharpe (Eds.), *Rethinking pedagogy for a digital age* (pp. 85–102). Routledge. https://doi.org/10.4324/9781351252805

COERLL. (2024). *Center for Open Educational Resources and Language Learning*. University of Texas. Retrieved August 11, 2024, from https://coerll.utexas.edu/coerll/

Comas-Quinn, A. (2011). Learning to teach online or learning to become an online teacher: An exploration of teachers' experiences in a blended learning course. *ReCALL, 23*(3), 218–232. https://doi.org/10.1017/S0958344011000152

Comas-Quinn, A., & Borthwick, K. (2015). Sharing: Open educational resources for language teachers. In R. Hampel & U. Stickler (Eds.), *Developing online language teaching* (pp. 96–112). https://doi-org.soton.idm.oclc.org/10.1057/9781137412263_7

Compton, L. (2009). Preparing language teachers to teach language online: A look at skills, roles and responsibilities. *Computer Assisted Language Learning, 22*(1), 73–99. http://www.tandfonline.com/doi/abs/10.1080/09588220802613831.

Davis, D., & Rose, R. (2007). *Professional developments for virtual schooling and online learning*. Report for the North American Council for Online Learning. http://files.eric.ed.gov/fulltext/ED509632.pdf

Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2023). *Artificial intelligence and English language teaching: Preparing for the future*. British Council. https://doi.org/10.57884/78EA-3C69

Garrison, D., Andersen, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D., Andersen, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education, 13*(1–2), 31–36. https://doi.org/10.1016/j.iheduc.2009.10.003

Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences, 12*(8), 435. https://doi.org/10.3390/socsci12080435

Hampel, R., & Stickler, U. (2005). New skills for new classrooms: Training tutors to teach languages online. *Computer Assisted Language Learning, 18*(4), 311–326. https://doi.org/10.1080/09588220500335455

Hampel, R., & Stickler, U. (2015). *Developing online language teaching: Research-based pedagogies and reflective practices*. Palgrave Macmillan. https://doi.org/10.1057/9781137412263

Hubbard, P. (2008). CALL and the future of language teacher education. *CALICO Journal, 25*(2), 175–188. https://www.learntechlib.org/p/74242/.

Liyanagunawardena, T., Kennedy, E., & Cuffe, P. (2015). Design patterns for promoting peer interaction in discussion forums in MOOCs. *elearning Papers, 42*, 1–16. https://centaur.reading.ac.uk/40453/

Martín-Monje, E., & Borthwick, K. (2021). Researching massive open online courses for language teaching and learning. *ReCALL, 33*(2), 107–110. https://doi.org/10.1017/S0958344021000094

Morgana, V. (2024). Fostering English speaking and writing subskills for the Cambridge B2 First through technology-mediated tasks. *ReCALL, 36*(2), 119–134. https://doi.org/10.1017/S0958344023000277

Murphy, L. (2015). Online language teaching: The learner's perspective. In R. Hampel & U. Stickler (Eds.), *Developing online language teaching*. Palgrave Macmillan. https://doi.org/10.1057/9781137412263

OpenAi. (2024). *Introducing ChatGPT*. https://openai.com/blog/chatgpt

Reuters. (2023, February 2). ChatGPT sets record for fastest growing user base – Analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Rodgers, C. R., & Raider-Roth, M. B. (2006). Presence in teaching. *Teachers and Teaching, 12*(3), 265–287. https://doi.org/10.1080/13450600500467548

Rüschoff, B. (2023). The impact of the Covid pandemic on the future of digitally enhanced language education. In R. Rossner & F. Hayworth (Eds.), *Rethinking language education after the experience of COVID: Final report* (pp. 41–49). European Council of Modern Languages [ECML]. https://www.ecml.at/Portals/1/documents/ECML-resources/Rethinking-language-education-after-the-experience-of-Covid-EN.pdf

UNESCO. (2023). ChatGPT and Artificial Intelligence in higher education: A quick start guide. Retrieved August 18, 2024, from https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf

P

Professional Learning

Chun Lai
Faculty of Education, The University of Hong Kong, Hong Kong, China

Abstract

This entry provides an overview of the different dimensions CALL teacher professional learning entails and the characteristics of CALL teacher professional learning as well as its influencing factors. It further synthesizes key principles of effective CALL teacher professional learning experience and discusses how CALL teacher professional learning can be supported. The entry highlights the complexity of CALL teacher professional learning and the necessity of adopting a longitudinal and personalized perspective in supporting professional learning.

Keywords

Professional learning · Technology integration · Informal learning · Formal learning · Professional identity · TPACK · Agency

Introduction

Professional learning refers to the emotional and cognitive processes that teachers go through individually or collaboratively in inducing changes in knowledge, skills, attitudes, beliefs and orientations, as well as professional self-images and agency pertaining to technology in language teaching and learning and related instructional actions. By highlighting the inner cognitive and emotional process that teachers go through, professional learning differs from the commonly used term, professional development, which often carries the connotation of externally directed activities with are often associated with 'finite episodes with a beginning and end' (Webster-Wright, 2009, p. 2). Scholars like Webster-Wright (2009) argue for moving beyond the current focus on providing professional development activities to understanding the fundamental process of how professionals learn. This entry adopts the term 'professional learning' to capture the ongoing, active nature of teachers' learning journey, which can occur across spatio-temporal boundaries.

Professional learning is a process of acquiring and developing knowledge, competencies and dispositions that contribute to professional practice and professional growth. The knowledge and competency components and structure as well as key dispositions manifest differently in different professional practice areas. Teacher professional learning does not take place in a vacuum, but rather is nested within complex systems, emerging and evolving through the complex interplay between activities of teachers as autonomous entities and layers of systems teachers are situated in,

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_275-1

such as school, policy, sociopolitical educational contexts (Opfer & Pedder, 2013). This entry will examine the nuances of teacher professional learning in relation to the use of technology for language education.

Dimensions of CALL Teacher Professional Learning

The integration of technology in language education is a complex social activity, which makes CALL teacher professional learning a multifarious concept beyond technical skills alone (Tafazoli, 2022). A major component of CALL teacher professional learning is the knowledge and skills teachers need to develop to maximize the potential of technology for language teaching and learning. TPACK, i.e., the competency of selecting and using technology appropriately to empower optimal pedagogical practices in supporting the mastery of a given content, has been proposed as a key competence that draws attention to context-specific strategies in technology integration (Tafazoli, 2021). At the same time, scholars like Kukulska-Hulme et al. (2017) caution that TPACK treats language as content, failing to capture that language is a medium. Thus, scholars incorporated aspects like digital literacies to set the necessary skills and knowledge in the context of CALL (Gönen & Aşık, 2019). Thus, the knowledge and skill base CALL teachers need to grasp is rather multifaceted.

Whether and how teachers are going to utilize relevant knowledge and skill for technology integration depends on a few additional dimensions: (1) teachers' creativity; (2) teachers' teaching philosophy and pedagogical orientations as well as broader belief profiles, such as epistemological beliefs, pedagogical beliefs and conceptions of technology in teaching and learning; (3) teachers' pedagogical repertoire; and (4) teachers' abilities to redesign and reappropriate technology for individual purposes as well as associated cyber-ethics concerning responsible digital behaviours (Hanson-Smith, 2018; Karatay & Hegelheimer, 2021). The interplay of these dimensions determines the intensity and nature of technology integration. Moreover, the implementation of and sustained engagement in technology integration depend on relevant instructional and assessment competencies, such as classroom managerial skills and assessment literacy (e.g., Crosthwaite et al., 2023; Park & Son, 2022; Tafazoli, 2022), as well as attitudinal, emotional and dispositional dimensions, such as attitudes towards the technological behaviour, openness to new technologies, techno-stress and cognitive and emotional dispositions towards novel and uncertain situations (Lai, 2023; Levin & Wadmany, 2008). Thus, CALL teacher professional learning involves not only the acquisition of the knowledge and skills but also the development and (re)construction of these various dimensions.

Moreover, CALL teacher professional learning is also a process of the (re)construction of professional identity and the emergence of agency (Kitade, 2015; Meihami & Esfandiari, 2021). CALL teacher professional learning involves developing affiliation to communities of practice, attachment to relevant ideologies about teaching process and teaching approaches, methods and techniques, and autonomy in accepting and assuming responsibility of teaching and relevant decision making (Meihami & Esfandiari, 2021; Werbińska, 2016). Teachers' professional self-image and role identification in turn shape how teachers integrate technology in teaching practices (e.g., Lai & Jin, 2021, and Lai et al., 2025). CALL teacher professional learning also involves the internalization of the professional agency as mediators/boundary brokers across various communities (e.g., the teaching community and the ICT community, the teacher community and student community) (Kitade, 2015) and as agentive actors in negotiating, evaluating and transforming instructional practices and power relationships in the teaching contexts (Meihami & Malmir, 2024; Meihami, 2025).

These different dimensions and processes are interrelated in CALL teacher professional learning. Thus, a holistic and relational view is needed when understanding and supporting CALL teacher professional learning.

Characteristics of CALL Teacher Professional Learning and Influencing Factors

CALL teacher professional learning has the following characteristics: (1) It is a long-term recursive progression. Scholars use terms such as 'spiral process' (Kozlova & Priven, 2015) and 'multigenerational process' (Koehler et al., 2007) to capture this non-linear learning trajectory. The recursive learning trajectory invalidates one-shot interventions to support teacher professional learning. (2) It has stage-like properties but does not follow a one-stage-at-a-time evolution pattern (Levin & Wadmany, 2008). According to Levin and Wadmany (2008), teacher professional learning in relation to information technology evolves along two developmental continuums: a human learning dimension (from external authoritative influences, through dialogical processes, towards personal needs-driven self-directed learning) and an ICT adoption knowledge dimension (from technical and organizational knowledge, through knowledge concerning the process of change, to knowledge restructuring process critical to conceptual transformation). They further pointed out that seemingly discrepant views and practices may co-exist and interplay at each stage. (3) Teachers show varying learning trajectories, which are influenced by a myriad of factors.

Some factors are related to the technological behaviours per se, such as goals of technological behaviours, context-specific technological experiences and related interactions (colleagues, students, time pressure, etc.), as well as teachers' subjective interpretation of the experience and interactions (Morsink et al., 2011). Some are contextual factors, including differential challenges at the technical, resource, strategic (e.g., policy, institutional culture; collegial support; subject-specific culture; professional communities) and socio-economic (e.g., cultural resistance, infrastructure, financial issues) level. Others are teacher-level factors, such as prior knowledge, attitudes, capital and general pedagogical knowledge, which may also shape differential reactions to challenges (Arnold & Ducate, 2015; Masoumi & Noroozi, 2023). In addition, factors like gender, school level and prior teaching and technological experience may moderate teachers' professional learning through technological experience (Lai, 2023). For instance, teachers' career stage may shape differential learning trajectories, as beginning teachers may attune to different aspects of teaching and face different challenges from experienced teachers (Louws et al., 2017).

Thus, CALL teacher professional learning trajectories vary across individuals, subject to the influence of factors at various levels.

Principles of Effective CALL Teacher Professional Learning Experience

Decades of research has generated some key principles of effective CALL teacher professional learning experience. These principles include (1) authenticity and contextualization of the learning experience, which highlights the importance of situating and embedding professional learning experience in teachers' working conditions and classroom practices; (2) sustained support for interaction around technology, which underscores easy access to long-term, on-demand and flexible support from diversified sources for independent learning or collective participation; (3) hands-on experience with technology that treats teachers as agentive and competent players and engages them in active exploration and inquiry as well as provides them with mastery experience; (4) opportunities to engage in reflections on individual experimentation and collective participation and (5) attending to the multifaceted nature of CALL teacher professional learning so that the learning experience not only strengthens relevant knowledge, competence and dispositions but also boosts teachers' self-efficacy and strategic competence in dealing with contextual challenges as well as strengthens emotional resilience against failed experience (Gönen & Aşık, 2019; El Shaban & Egbert, 2018).

El Shaban and Egbert's (2018) model for CALL teacher professional learning is an epitome of how these key principles could be integrated together to provide effective professional learning

experience for CALL teachers. According to them, effective CALL teacher professional learning should involve experiences that address why (i.e., seeing the relevance and importance of the technological behaviour and having relevant self-efficacy beliefs), what (i.e., explore attributes of technology) and how (i.e., contextualized implementation of specific technologies) of CALL experience. They further argue that such experience needs to be reinforced with an enabling environment and regular opportunities for supported and contextualized communication and collaboration.

Sources of and Approaches to CALL Teacher Professional Learning

There are two major sources of teacher professional learning: formal and informal professional learning. Formal professional learning is often structured in time and space, follows a pre-specified curriculum, is agenda-based and content-driven and is planned by others. Examples of formal professional learning include workshops, courses, seminars and conferences. Informal learning is self-initiated and self-directed and often follows an unplanned process through which individuals deliberately, reactively or implicitly acquire knowledge and skills outside of formal educational settings (Kyndt et al., 2016; Lecat et al., 2019). There are social sources of informal learning (e.g., interacting with colleagues), non-interpersonal sources of informal learning (e.g., reading professional documents) and inner sources of informal learning (e.g., reflection) (Jones & Dexter, 2014; Lecat et al., 2019). Different sources of professional learning contribute to different aspects of teacher learning. Formal learning is often associated with changes in instructional beliefs, sources of knowledge and reflections on professional practices. In contrast, informal learning is often associated with changes in instructional practices and attitudes towards technology, sources of new ideas and innovations and a sense of belonging (Kessler, 2007; Lantz-Andersson et al., 2018). Scholars hence advocate combining formal and informal sources of professional learning (Arnold & Ducate, 2015; Ngo & Nguyen, 2023). Some scholars suggest following up formal training at the earlier stages with supportive environments and informal learning opportunities at later stages to support CALL teachers' professional learning (Levin & Wadmany, 2008; El Shaban & Egbert, 2018).

Researchers have experimented with various approaches to supporting CALL teachers' professional learning. Torsani (2016) summarized three major categories of teacher professional learning models in the CALL field: models that focus on the development of technological knowledge and skill development, models that focus on teacher role reconstruction and models that focus on collaborative learning.

The first category of models emphasizes developing knowledge and skills. Various approaches have been adopted to achieve this goal, including breadth-first approach (i.e., systematic instruction on knowledge and skills related to a comprehensive range of CALL alternatives), depth-first approach (i.e., focusing on intensive experience in a single area), integrated approach (i.e., integrating technology in multiple places in coursework rather than a single, stand-alone class) and online approach (i.e., learning about technology through doing and participating in the process). These approaches have been realized through various means such as lecture/demonstration, microlearning lessons, inquiry-oriented projects, situated learning, reflective learning and portfolio-based process (Kohnke & Foung, 2023; Meihami & Esfandiari, 2021). Egbert and Shahrokni (2019) further suggested going beyond knowledge and skills around technological applications to focus on competence-based learning. They advocate building a resources portal to support teachers' self-paced learning towards key competences laid out in CALL teacher education standards (e.g., Healey et al., 2011).

The second category of models focuses more on teacher role reconstruction. This has normally been achieved through role reversal and role experimentation. For instance, Yang and Kuo (2020) engaged student teachers as CALL system co-developers to work with system engineers in developing CALL systems, implement the

systems in practice teaching and provide feedback to inform system refinement. They discovered that shifting from a passive consuming role to an active development role resulted in a change in pedagogical concepts. Similarly, Meihami and Malmir (2024) adopted the ADDIE (analysis, design, development, implementation, evaluation) model to turn L2 teachers into agentive actors in designing and redesigning online classes. They found that assigning teachers the role of CALL designers enhanced their professional agency.

The third category of models focuses on constructing formal and informal collaborative experiences, such as critical friend groups, university-school collaboration and informal communities on social networking sites such as twitters (Ngo & Nguyen, 2023). For instance, Morsink et al. (2011) documented teacher professional learning trajectories in a teacher development initiative where school teachers were paired with university researchers, receiving sustained individualized and intensive support from university researchers after training seminars. Nami, Marandi and Sotoudehnama (2016) traced CALL teacher professional learning through lesson study where teachers co-designed lessons, observed each other's lessons and provided peer feedback to improve the quality of lesson plans. Teachers reported that their knowledge of technological pedagogical knowledge and knowledge of technological glitches increased and found their confidence in teaching with technology got boosted through the collaborative learning.

Existing literature further points out that there are no-size-fit-all approaches to supporting teacher professional learning as individual situations and contextual circumstances vary. Teachers at different career stages are driven by different motives for professional learning and prefer different professional learning activities (e.g., Louws et al., 2017). Moreover, there are great variations in professional learning trajectories amongst individuals (e.g., Levin & Wadmany, 2008). Scholars, like Levin and Wadmany (2008), hence argue that support mechanisms need to strike a right balance between supporting meaningful groups of teachers and meeting the idiosyncratic developmental needs of individual teachers. Providing a resource repertoire to support teachers' self-directed learning is a promising direction to pursue (e.g., Egbert & Shahrokni, 2019, and Kohnke & Foung, 2023). Recent development of generative artificial intelligence (GenAI), with its capacity of generating human-like multilingual textual and visual output powered by advanced natural language processing and machine learning techniques, may further empower teachers' self-directed personalized learning to meet idiosyncratic developmental needs.

Summary

CALL teacher professional learning is a socially mediated, multifaceted phenomenon that is characterized by long-term recursive progression. Its complexity requires an ecological, longitudinal and personalized perspective that supports the active integration of both formal and informal sources of professional learning to address the needs of individual teachers and counter multilevel challenges at various developmental stages.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Arnold, N., & Ducate, L. (2015). Contextualized views of practices and competencies in CALL teacher education research. *Language Learning & Technology, 19*(1), 1–9.

Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392–1420.

Egbert, J. L., & Shahrokni, S. A. (2019). Balancing old and new: Integrating competency-based learning into CALL teacher education. *The JALTCALL Journal, 15*(1), 3–18.

El Shaban, A., & Egbert, J. (2018). Diffusing education technology: A model for language teacher professional development in CALL. *System, 78*, 234–244.

Gönen, S. İ. K., & Aşık, A. (2019). A review of studies on CALL teacher education: Trends and issues. In L. Guerra, G. Eksi, D. Werbinska, & Y. Bayyurt

(Eds.), *Research trends in English language teacher education and English language teaching* (pp. 343–362). Universidade de Évora.

Hanson-Smith, E. (2018). CALL (computer-assisted language learning) materials development. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching*. Wiley.

Healey, D., Hanson-Smith, E., Hubbard, P., Ioannou-Georgiou, S., Kessler, G., & Ware, P. (2011). *TESOL technology standards: Description, implementation, integration*. Alexandria: TESOL.

Jones, W. M., & Dexter, S. (2014). How teachers learn: The roles of formal, informal, and independent learning. *Educational Technology Research and Development, 62*, 367–384.

Karatay, Y., & Hegelheimer, V. (2021). CALL teacher training – Considerations for low-resource environments: Overview of CALL teacher training. *CALICO Journal, 38*(3), 271.

Kessler, G. (2007). Formal and informal CALL preparation and teacher attitude toward technology. *Computer Assisted Language Learning, 20*(2), 173–188.

Kitade, K. (2015). Second language teacher development through CALL practice: The emergence of teachers' agency. *CALICO Journal, 32*(3), 396–425.

Koehler, M. J., Mishra, P., & Yahya, K. (2007). Tracing the development of teacher knowledge in a design seminar: Integrating content, pedagogy and technology. *Computers & Education, 49*(3), 740–762.

Kohnke, L., & Foung, D. (2023). Exploring microlearning for teacher professional development: voices from Hong Kong. In D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development: Voices from under-represented contexts* (pp. 279–292). Springer.

Kozlova, I., & Priven, D. (2015). ESL teacher training in 3D virtual worlds. *Language Learning & Technology, 19*(1), 83–101.

Kukulska-Hulme, A., Lee, H., & Norris, L. (2017). Mobile learning revolution: Implications for language pedagogy. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 217–233). Wiley.

Kyndt, E., Gijbels, D., Grosemans, I., & Donche, V. (2016). Teachers' everyday professional development: Mapping informal learning activities, antecedents, and learning outcomes. *Review of Educational Research, 86*(4), 1111–1150.

Lai, C. (2023). *Insights into autonomy and technology in language teaching*. Castledown Publishers.

Lai, C., & Jin, T. (2021). Teacher professional identity and the nature of technology integration. *Computers & Education, 175*, 104314.

Lai, C., Lyu, B. N., Jiang, L. J., & Gong, Y. (2025). Language teachers' professional identities and classroom technology integration. *Language Learning & Technology, 29*(1), 1–19.

Lantz-Andersson, A., Lundin, M., & Selwyn, N. (2018). Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teaching and Teacher Education, 75*, 302–315.

Lecat, A., Raemdonck, I., Beausaert, S., & März, V. (2019). The what and why of primary and secondary school teachers' informal learning activities. *International Journal of Educational Research, 96*, 100–110.

Levin, T., & Wadmany, R. (2008). Teachers' views on factors affecting effective integration of information technology in the classroom: Developmental scenery. *Journal of Technology and Teacher Education, 16*(2), 233–263.

Louws, M. L., Meirink, J. A., van Veen, K., & van Driel, J. H. (2017). Teachers' self-directed learning and teaching experience: What, how, and why teachers want to learn. *Teaching and Teacher Education, 66*, 171–183.

Masoumi, D., & Noroozi, O. (2023). Developing early career teachers' professional digital competence: A systematic literature review. *European Journal of Teacher Education*, 1–23. https://doi.org/10.1080/02619768.2023.2229006

Meihami, H. (2025). Situated learning in CALL teacher preparation programs: an ecological perspective to student-teachers' agency. *Computer Assisted Language Learning, 38*(1–2), 235–261.

Meihami, H., & Esfandiari, R. (2021). Exploring EFL teachers' professional identity development in a CALL teacher preparation program. *JALT CALL Journal, 17*(2), 135–157.

Meihami, H., & Malmir, A. (2024). Student-teachers' professional agency development through ADDIE model of CALL teacher preparation. *Language Teaching Research*, 13621688241254907. https://doi.org/10.1177/13621688241254907

Morsink, P. M., Hagerman, M. S., Heintz, A., Boyer, D. M., Harris, R., Kereluik, K., … & Withey, K. (2011). Professional development to support TPACK technology integration: The initial learning trajectories of thirteen fifth-and sixth-grade educators. *Journal of Education, 191*(2), 3–16.

Nami, F., Marandi, S. S., & Sotoudehnama, E. (2016). CALL teacher professional growth through lesson study practice: an investigation into EFL teachers' perceptions. *Computer Assisted Language Learning, 29*(4), 658–682.

Ngo, P. L. H., & Nguyen, C. N. B. (2023). Online community of practice on facebook: A case study of EFL teachers in Vietnam. In D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development: Voices from under-represented contexts* (pp. 531–548). Springer.

Opfer, V. D., & Pedder, D. (2013). Teacher change and changing teachers via professional development. In C. McLaughlin (Ed.), *Teachers learning: Professional development and education* (pp. 93–117). Cambridge University Press.

Park, M., & Son, J. B. (2022). Pre-service EFL teachers' readiness in computer-assisted language learning and

teaching. *Asia Pacific Journal of Education, 42*(2), 320–334.

Tafazoli, D. (2021). CALL teachers' professional development amid the COVID-19 outbreak: A qualitative study. *CALL-EJ, 22*(2), 4–13.

Tafazoli, D. (2022). CALL teacher education and professional development: A retrospective synthesis of the two decades of teaching English with technology. *Teaching English with Technology, 22*(3/4), 1–19.

Torsani, S. (2016). *CALL teacher education: Language teachers and technology integration*. Springer.

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research, 79*(2), 702–739.

Werbińska, D. (2016). Language-teacher professional identity: Focus on discontinuities from the perspective of teacher affiliation, attachment and autonomy. In C. Gkonou, D. Tatzl, & S. Mercer (Eds.), *New directions in language learning psychology* (pp. 135–157). Springer.

Yang, Y. F., & Kuo, N. C. (2020). New teaching strategies from student teachers' pedagogical conceptual change in CALL. *System, 90*, 102218.

ICALL and Individual Differences

Majid Elahi Shirvan[1] and Tahereh Taherian[2]
[1]Department of Foreign Languages, University of Bojnord, Bojnord, Iran
[2]English Language Department, Yazd University, Yazd, Iran

Abstract

In the evolving field of second/foreign language acquisition (SLA), individual differences (IDs) significantly shape learner experiences, behaviours and outcomes. Whilst considerable progress has been made in understanding how IDs influence SLA and computer-assisted language learning (CALL), there remains a gap in systematic reviews focusing on their intersection with intelligent computer-assisted language learning (ICALL). Exploring these differences within the realm of ICALL is an emerging area of study that requires thorough examination. This chapter aims to address this gap by providing an overview of how various IDs impact ICALL. It delves into the importance and extent of these differences in ICALL, reviews existing research on their significance and explores methods for evaluating differences within this context. By exploring these aspects, this chapter aims to improve our understanding of how ID factors can be incorporated into ICALL practices to enhance language learning outcomes.

Keywords

Individual differences · Intelligent computer-assisted language learning · Second language acquisition · Language learning outcomes · Assessment approaches

Introduction

In the field of SLA, IDs have an important effect on learners' experiences, behaviours and achievements. Although the importance of IDs in SLA is extensively documented, encompassing studies on language aptitude (Li & DeKeyser, 2021), working memory (Wen & Li, 2019), motivation (Al-Hoorie et al., 2023), learning strategies (Pawlak & Oxford, 2018), language engagement (Hiver et al., 2024) and emotions (Dewaele, 2015; Elahi Shirvan et al., 2021; Kruk et al., 2023; Li & Li, 2024; Pawlak et al., 2023; Shao et al., 2024), investigations into their function in CALL and ICALL remain sparse (Pawlak & Kruk, 2022). ICALL, where AI interacts with language learners, offers novel prospects for improving language acquisition. This chapter provides an overview of how IDs influence ICALL, highlighting their importance, reviewing current research and suggesting further research in this developing domain.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_289-1

Individual Differences (IDs): Conceptualizations and Functions

For years, ID research in SLA was based on a traditional view. In this view, IDs were perceived as distinct, consistent psychological traits that were largely internal to the learner and independent of one another. This viewpoint resulted in the classification of IDs into dimensions such as cognitive and affective factors, aiming to pinpoint those that significantly impact SLA outcomes (Dörnyei, 2009; Dörnyei & Ryan, 2015). This method facilitated extensive research; however, it has been criticized for neglecting the dynamic aspects of learners' characteristics. Critics contend that the conventional perspective overlooks the variability of learner traits influenced by external factors, including instructional strategies, social interactions and individual experiences. It also neglects to consider the interrelationships between various IDs (Dörnyei & Ryan, 2015).

In light of the emergence of complex dynamic systems theory (CDST; Larsen-Freeman & Cameron, 2008) and sociocultural theory (Storch, 2017), the current perspective on IDs is indicative of their context-dependent, dynamic and interconnected structure. CDST posits that IDs interact with one another and their environment, experiencing phases of transformation (i.e. change) and equilibrium (i.e. stability). Sociocultural theory highlights the impact of social context on these characteristics. Consequently, IDs are regarded as dynamic entities influenced by external factors, such as instructional approaches and educational settings.

IDs are generally regarded as predictors of learning outcomes rather than outcomes themselves (Dewaele & Li, 2022). However, there is a possibility that learning outcomes could influence IDs. For example, motivation is recognized as a factor that predicts achievement, yet previous research has indicated that high achievement can also predict higher motivation (Li et al., 2024). Furthermore, IDs may function as mediators or moderators. For example, Elahi Shirvan et al. (2024) show that emotions can predict L2 willingness to communicate (L2WTC) and coping strategies, both of which are necessary for language achievement (e.g. coping strategies and L2WTC act as mediators in the relationship between emotions and language achievement).

Research on ID Factors Within ICALL Settings

In this chapter, based on the framework established by Li et al. (2022), we classified IDs into four clusters: cognitive/metacognitive, conative, affective and sociocultural/demographic differences. Cognitive IDs include language aptitude and metacognition; conative factors involve motivation and willingness to communicate in a second language (L2); affective factors pertain to emotions and beliefs; and sociocultural/demographic aspects encompass age, proficiency level and literacy. This taxonomy offers a structured approach to understanding the complexities of ID factors in SLA research (Li et al., 2022).

It should be mentioned that we utilized the classification framework in this review to arrange and present IDs, aiming to facilitate a more comprehensive understanding. Nevertheless, it is crucial to recognize that we do not disregard the interconnectedness and dynamic nature of these factors. Although IDs frequently emerge from intricate interactions amongst cognitive, emotional and social factors, our objective was to present a systematic framework to assist readers in understanding the variety of variables involved. Furthermore, due to the substantial number of empirical studies in certain domains (e.g. motivation), the focus is on the most representative examples.

Research Into Cognitive and Metacognitive Factors

Cognitive factors. Cognitive IDs refer to variations in mental capacities and processes amongst individuals that impact their perception, learning and task performance (Li et al., 2022). These factors include abilities such as working memory, attention, processing speed and declarative

memory. However, research on cognitive IDs in ICALL remains limited. Ruiz et al. (2019) found that, in web-based ICALL, form-focused methods (i.e. reading texts and completing multiple-choice gap exercises) may be more effective for learners with stronger working memory, whilst meaning-focused approaches (i.e. reading texts without additional tasks) may better support learners with declarative memory strengths. This highlights the need for adaptive learning environments in ICALL platforms that can tailor instruction to the cognitive profiles of learners.

Metacognitive factors. Metacognition refers to the awareness and understanding of one's own thought processes, often summed up as 'knowledge about knowledge' (Weinert, 1987). Dinsmore et al. (2008) describe it as a process of reflecting on and abstracting cognitive structures, whether new or existing. Roebers (2017) associates metacognition with executive functions, including intentional, purposeful and self-regulated cognitive processes. Research suggests that metacognitive factors in ICALL significantly influence learning outcomes. AI-powered tools that promote self-regulation and enhance learning efficiency have shown substantial impact (Qiao & Zhao, 2023). For example, Wei (2023) compared AI-mediated instruction with traditional methods amongst university students and found that the AI group exhibited significantly better outcomes in terms of English motivation and self-regulation strategies.

Research Into Conative Factors

Conative ID factors in language learning are the personal characteristics and motivational dimensions that drive a learner's efforts to achieve their objectives (Li et al., 2022). These factors refer to the actions and behaviours that students engage in to achieve their goals, as opposed to cognitive factors, which include analysis, interpretation, reasoning, memory and symbol use.

Motivation. Motivation plays a crucial role in SLA. Dörnyei and Ryan (2015) emphasize that individuals, regardless of their exceptional abilities, cannot attain long-term goals without adequate motivation and that the absence of motivation cannot be offset by high-quality curricula or teaching methods. Various theoretical perspectives highlight the complex nature of motivation. Gardner's (1985) instrumental and integrative orientations emphasize learning outcomes, whereas Dörnyei's (2009) L2 motivational self system integrates future selves with the dynamic nature of the learning process. Self-determination theory (SDT; Deci & Ryan, 1985) emphasizes the needs of learners. This range of theories underscores the complexity and diversity of motivation in SLA (Al-Hoorie et al., 2023). In the contemporary socio-dynamic phase, which draws from CDST (Larsen-Freeman & Cameron, 2008), motivation is viewed as both a changing and stable phase that interacts with other ID factors.

The significance of motivation in ICALL has received growing focus because of its crucial influence on learning outcomes. Research has revealed ambiguous results regarding the impact of ICALL on motivation. Fryer et al. (2019) found that learners' motivation to use chatbots for language learning decreased significantly after the early stages of a course. Ericsson and Johansson (2023) conducted a study in Sweden with beginner-level language learners, utilizing the Enskill English platform to provide AI-driven speaking scenarios. Although a number of students found the experience to be motivating, others expressed frustration with the system's limited automatic speech recognition abilities. This highlights the significance of technological reliability in maintaining learner motivation. Alm and Nkomo (2020) identified technological constraints, including difficulties in comprehending multiple utterances and restricted conversational paths, as major challenges affecting motivation.

On the other hand, positive effects of ICALL on motivation and interest are reported by other studies. In their study on writing motivation in ICALL, Song and Song (2023) identified significant enhancements in writing proficiency and motivation amongst EFL students who received AI-assisted instruction, in contrast to those who experienced traditional instruction. Guo et al. (2023) demonstrated that chatbot-assisted in-class debates improved students' argumentation skills

and task motivation, indicating that AI can facilitate challenging and enjoyable learning experiences.

L2WTC. L2WTC refers to an individual's readiness to engage in conversation using L2. Initially conceptualized as a stable personality trait (MacIntyre et al., 1998), the concept of WTC has evolved from being seen as a stable personality trait (McCroskey & Richmond, 1987) to a more complex construct involving psychological, linguistic and contextual factors in SLA (Dörnyei & Ryan, 2015). MacIntyre et al.'s (1998) pyramid model of L2WTC reflects this complexity, covering long-term influences like intergroup environment and short-term motivations for specific interactions. In the field of ICALL, research has shifted to examine how technology enhances learners' L2WTC. Studies highlight AI's positive impact on L2WTC and language learning outcomes. For instance, Kim and Su (2024) found that AI chatbots improved students' L2WTC by providing engaging speaking practice. Similarly, Fathi et al. (2024) compared AI-assisted learning with traditional peer interaction, finding that the AI platform, Andy English Chatbot, was more effective in improving speaking skills and L2WTC. Students expressed positive attitudes towards AI-enhanced instruction, reinforcing AI's potential to transform language learning through interactive and motivating exercises.

Engagement. Language engagement is a multifaceted complex concept involving the interplay between the learners' environment, the individual and their actions (Hiver et al., 2024). It includes cognitive, emotional, and behavioural, social and agentic dimensions, all of which influence the depth and quality of learner involvement. Recent studies indicate that AI-driven applications and tools significantly enhance language engagement in diverse educational contexts (Huang & Zou, 2024; Wu et al., 2024). Huang and Zou (2024) examined the potential of AI-powered chatbots to facilitate adaptive and personalized learning experiences, thereby enhancing engagement within the educational context. This is consistent with findings from other studies indicating that AI technologies, through the provision of individualized feedback and contextual responses, can sustain an interactive learning environment that engages students.

Research Into Affective Factors

Affective variables include emotions like anxiety, enjoyment, boredom and playfulness. They are also related to beliefs, attitudes and perceptions regarding particular learning tools and methods, including AI-based language instruction.

Emotions. Emotions are complex psychological states that involve subjective experience, a physiological response and a behavioural or expressive response. Emotions significantly impact language learning, affecting motivation, engagement and overall success. Foreign language anxiety (FLA) and foreign language enjoyment (FLE) are two significant emotions that have been extensively researched across diverse SLA contexts. FLA is a distinct form of anxiety defined by self-perceptions, beliefs, emotions and behaviours associated with language learning in a classroom context that arise from the particular characteristics of the SLA process (Horwitz et al., 1986). It can substantially impede students' performance and L2WTC (Dörnyei & Ryan, 2015; MacIntyre, 2002). High FLA is associated with diminished academic performance and decreased language proficiency (MacIntyre, 2002). Research indicates that AI tools in ICALL settings can effectively diminish FLA. For example, Hsu et al. (2021) demonstrated that a task-oriented chatbot system effectively reduced speech anxiety amongst Taiwanese students, thereby enhancing their spoken English proficiency.

The significance of addressing FLE in language learning has been increasingly acknowledged. FLE denotes a positive emotional condition that integrates aspects of challenge, happiness, interest, enjoyment, pride and a sense of belonging (Dewaele & MacIntyre, 2014). Research in ICALL settings indicates that AI applications in FLE can improve enjoyment by creating a less intimidating environment and facilitating authentic communication experiences. Wang et al. (2022) demonstrated that interaction

with an AI coach enhanced L2 enjoyment amongst language learners, which subsequently predicted improved learning outcomes. The cognitive presence of the AI coach and the learners' positive feelings towards the AI significantly enhanced their enjoyment and performance.

Although there have been significant advancements in the understanding of FLA and FLE within ICALL contexts, the phenomenon of foreign language boredom (FLB) has not been extensively investigated. FLB, characterized as a state of disengagement resulting from insufficient interest and involvement (Pawlak et al., 2020), represents a critical emotion that can adversely affect language acquisition. There exists a significant gap in research examining boredom within ICALL contexts. The potential of AI to develop dynamic and interactive learning environments presents a valuable avenue for future research into its capacity to mitigate boredom. Other unexplored emotions that significantly influence ICALL include curiosity (Mahmoodzadeh & Khajavy, 2018), hope (Oxford & Curtis, 2021) and playfulness (Shao et al., 2024).

Beliefs. Conceptually, beliefs can be described as assumptions about L2 learning that individuals hold, which have the potential to impact how they approach learning tasks (Gregersen & Macintyre, 2014). These beliefs encompass learners' and teachers' perceptions, attitudes and assumptions about various aspects of language learning and teaching. Several studies have indicated that learners' beliefs have been linked to metacognition (Wenden, 1999), epistemological beliefs (Mori, 1999) and emotions (Elahi Shirvan et al., 2024), emphasizing their complexity and dynamism. These perspectives underscore that beliefs are crucial contributors to success or failure in L2 learning.

A critical review of the studies on beliefs in ICALL settings reveals several key points. First, learners' and teachers' beliefs are dynamic and significantly shaped by their direct experiences with technology. For learners, the perceived usefulness of AI tools like ChatGPT and Alexa is crucial for their acceptance and sustained use, though technical issues can hinder engagement (Liu et al., 2024; Liu & Ma, 2023). For example, Liu et al. (2024) investigated L2 learners' acceptance and use of large language model platforms for informal digital learning purposes. They revealed that hands-on experimentation with chatbots increased perceived usefulness, which in turn enhanced the intention to use and actual use of these tools. This study underscores the dynamic negotiation between learners and the affordances and constraints of AI technologies, highlighting how learners' beliefs evolve through interaction with these tools.

For teachers, beliefs are influenced by their familiarity and experience with LLMs rather than external infrastructural support. Gao et al. (2024) explored EFL university teachers' beliefs regarding the integration of AI in education. They identified concerns such as neglecting traditional resources, academic integrity issues and over-reliance on technology. Previous experiences with AI and frequency of use significantly shaped teachers' beliefs, whereas IT personnel and infrastructure availability had little impact. This indicates that personal experience and direct interaction with technology are more influential in shaping teachers' beliefs than external support factors.

Research Into Sociocultural and Demographic Factors

Research on sociocultural and demographic factors in ICALL highlights the importance of language proficiency and technology literacy in the effective use of AI tools in EFL classes. Language proficiency significantly affects how well AI chatbots work for EFL students. Novice learners, as Divekar et al. (2021) note, often struggle to initiate conversations due to their limited skills. This makes it challenging for them to use general-purpose chatbots, which may not meet their needs. Beginner learners benefit more from structured, goal-orientated chatbots designed within an educational framework (Li & DeKeyser, 2021). Technology literacy also influences AI integration in language education. As De la Vall and Araya (2023) suggest, both teachers and students often lack the technological know-how to use AI

systems effectively. This gap limits the benefits AI can offer in classrooms. Financial and technical barriers further complicate the situation, as Hazaymeh et al. (2024) point out, with high costs and complexity making AI tools inaccessible, particularly in resource-limited regions.

Conclusion

This chapter explores the interplay between ICALL and IDs, highlighting how AI-based personalized learning experiences can enhance language acquisition. The findings underscore the importance of tailoring ICALL systems to accommodate learners' unique cognitive, emotional and motivational traits. However, challenges remain, particularly regarding the quality of underlying algorithms and the systems' ability to adapt in real-time to shifts in learners' needs or emotional states. To address this, more sophisticated AI-driven models are needed that can integrate cognitive, affective and behavioural data for truly individualized learning pathways.

Whilst ICALL has significant potential for personalized language education, educators must be cautious not to rely solely on technology to address learner diversity. Human mediation, particularly emotional and social support, remains essential. A blended model combining the adaptability of ICALL with human elements of teaching may be the most effective approach.

A key limitation in assessing IDs in ICALL is the difficulty in accurately measuring such traits. Treating IDs as static, isolated variables overlooks their dynamic nature. IDs are shaped by learners' contexts and interactions, which traditional assessment methods fail to capture. Future research should develop dynamic and comprehensive approaches to assessing IDs, such as idiodynamic and process tracing methods. Integrating both micro- and macro perspectives using techniques like latent change score modelling (LCSM) and dynamic structural equation modelling (DSEM) can provide a more comprehensive view of how IDs evolve over time, offering insights into their role in language acquisition and informing the design of more effective ICALL systems.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Al-Hoorie, A. H., Hiver, P., & In'nami, Y. (2023). The validation crisis in the L2 motivational self-system tradition. *Studies in Second Language Acquisition, 46*, 307. https://doi.org/10.1017/s0272263123000487

Alm, A., & Nkomo, L. M. (2020). Chatbot experiences of informal language learners. *International Journal of Computer-Assisted Language Learning and Teaching, 10*(4), 51–65. https://doi.org/10.4018/ijcallt.2020100104

De La Vall, R. R. F., & Araya, F. G. (2023). Exploring the benefits and challenges of AI-language learning tools. *The International Journal of Social Sciences and Humanities Invention, 10*(1), 7569–7576. https://doi.org/10.18535/ijsshi/v10i01.02

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. In E. L. Deci & R. M. Ryan (Eds.), *Perspectives in social psychology.* Springer. https://doi.org/10.1007/978-1-4899-2271-7

Dewaele, J. M. (2015). On emotions in foreign language learning and use. *The Language Teacher, 39*(3), 13–15. https://eprints.bbk.ac.uk/id/eprint/12944/

Dewaele, J. M., & Li, C. (2022). Foreign language enjoyment and anxiety: Associations with general and domain-specific English achievement. *Chinese Journal of Applied Linguistics, 45*(1), 32–48. https://doi.org/10.1515/cjal-2022-0104

Dewaele, J. M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2021). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning, 35*(9), 2332–2360. https://doi.org/10.1080/09588221.2021.1879162

Dörnyei, Z. (2009). The L2 motivational self-system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters. https://doi.org/10.21832/9781847691293-003

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge eBooks. https://doi.org/10.4324/9781315779553

Elahi Shirvan, M. E., Taherian, T., & Yazdanmehr, E. (2021). Foreign language enjoyment: A longitudinal confirmatory factor analysis–curve of factors model. *Journal of Multilingual and Multicultural Development, 45*(2), 224–242. https://doi.org/10.1080/01434632.2021.1874392

Elahi Shirvan, M., Taherian, T., Pawlak, M., & Kruk, M. (2024). The bright side and the dark side of foreign language boredom in learner engagement: A moderated mediation model of foreign language playfulness and boredom-coping strategies. *Language Learning Journal*, 1–17. https://doi.org/10.1080/09571736.2024.2384484

Ericsson, E., & Johansson, S. (2023). Sustained speaking experiences with virtual humans over time. *INTED Proceedings*. https://doi.org/10.21125/inted.2023.1589

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior, 93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023

Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: A study in China. *Asia Pacific Journal of Education, 44*(1), 29–44. https://doi.org/10.1080/02188791.2024.2305173

Gardner, R. C. (1985). Social psychology and second language learning: The role of attitudes and motivation. https://ci.nii.ac.jp/ncid/BA00118070

Gregersen, T., & MacIntyre, P. D. (2014). *Capitalizing on language learners' individuality*. Multilingual Matters eBooks. https://doi.org/10.21832/9781783091218

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education, 203*, 104862. https://doi.org/10.1016/j.compedu.2023.104862

Hazaymeh, W. A., Bouzenoun, A., & Remache, A. (2024). EFL instructors' perspective on using AI applications in English as a foreign language teaching and learning. *Emerging Science Journal, 8*, 73–87. https://doi.org/10.28991/esj-2024-sied1-05

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research, 28*(1), 201–230. https://doi.org/10.1177/13621688211001289

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Hsu, C. C., Chen, Y. L., Lin, C. Y., & Lien, W. C. (2021). Cognitive development, self-efficacy, and wearable technology use in a virtual reality language learning environment: A structural equation modeling analysis. *Current Psychology, 41*(3), 1618–1632. https://doi.org/10.1007/s12144-021-02252-y

Huang, F., & Zou, B. (2024). English speaking with artificial intelligence (AI): The roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior, 159*, 108355. https://doi.org/10.1016/j.chb.2024.108355

Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System, 122*, 103256. https://doi.org/10.1016/j.system.2024.103256

Kruk, M., Pawlak, M., Taherian, T., Yüce, E., Elahi Shirvan, M., & Barabadi, E. (2023). When time matters: Mechanisms of change in a mediational model of foreign language playfulness and L2 learners' emotions using latent change score mediation model. *Studies in Second Language Learning and Teaching, 13*(1), 39–69. https://doi.org/10.14746/ssllt.37174

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.

Li, S., & DeKeyser, R. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue. *Studies in Second Language Acquisition, 43*(3), 473–497. https://doi.org/10.1017/s0272263121000024

Li, B., & Li, C. (2024). Achievement goals and emotions of Chinese EFL students: A control-value theory approach. *System, 123*, 103335. https://doi.org/10.1016/j.system.2024.103335

Li, S., Hiver, P., & Papi, M. (2022). *The Routledge handbook of second language acquisition and individual differences*. Routledge.

Li, J., Wang, C., & King, R. B. (2024). Which comes first? Modeling longitudinal associations among self-efficacy, motivation, and academic achievement. *System, 121*, 103268. https://doi.org/10.1016/j.system.2024.103268

Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching, 18*, 125. https://doi.org/10.1080/17501229.2023.2240316

Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2024.2310288

MacIntyre, P. (2002). Motivation, anxiety, and emotion in second language acquisition. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 45–68). John Benjamins.

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal, 82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Mahmoodzadeh, M., & Khajavy, G. H. (2018). Towards conceptualizing language learning curiosity in SLA: An empirical study. *Journal of Psycholinguistic Research, 48*(2), 333–351. https://doi.org/10.1007/s10936-018-9606-3

McCroskey, J. C., & Richmond, V. P. (1987). Willingness to communicate. In J. C. McCroskey & J. A. Daly (Eds.), *Personality and interpersonal communication* (pp. 129–156). SAGE.

Mori, Y. (1999). Epistemological beliefs and language learning beliefs: What do language learners believe about their learning? *Language Learning, 49*(3), 377–415. https://doi.org/10.1111/0023-8333.00094

Oxford, R., & Curtis, A. (2021). Exploring peace language: Hope, help and harmony. In R. Oxford, M. Olivero, & M. Harrison (Eds.), *Peacebuilding in language education: Innovations in theory and practice* (pp. 228–241). Multilingual Matters. https://doi.org/10.21832/9781788929806-019

Pawlak, M., & Kruk, M. (2022). *Individual differences in computer assisted language learning research*. Routledge. https://doi.org/10.4324/9781003240051

Pawlak, M., & Oxford, R. L. (2018). Conclusion: The future of research into language learning strategies. *Studies in Second Language Learning and Teaching, 8*(2), 525–535. https://doi.org/10.14746/ssllt.2018.8.2.15

Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System, 91*, 102259. https://doi.org/10.1016/j.system.2020.102259

Pawlak, M., Solhi, M., Elahi Shirvan, M., Kruk, M., & Taherian, T. (2023). Revisiting after-class boredom via exploratory structural equation modeling. *IRAL – International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2022-0151

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology, 14*, 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review, 45*, 31–51. https://doi.org/10.1016/j.dr.2017.04.001

Ruiz, S., Rebuschat, P., & Meurers, D. (2019). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research, 25*(4), 510–539. https://doi.org/10.1177/1362168819872859

Shao, K., Barabadi, E., Elahi Shirvan, M., Taherian, T., & Tabar, M. R. (2024). Conceptualization and measurement of foreign language playfulness via exploratory structural equation modeling. *System, 123*, 103321. https://doi.org/10.1016/j.system.2024.103321

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Storch, N. (2017). Sociocultural theory in the L2 classroom. In S. Loewen & M. Sato (Eds.), *The handbook of instructed second language acquisition* (pp. 69–83). Routledge.

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning, 4*, 814. https://doi.org/10.1080/09588221.2022.2056203

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology, 14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Weinert, F. E. (1987). Introduction and overview: Metacognition and motivation as determinants of effective learning and understanding. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 1–16). Lawrence Erlbaum.

Wen, Z., & Li, S. (2019). Working memory in L2 learning and processing. In J. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 365–389). Cambridge University Press. https://doi.org/10.1017/9781108333603.016

Wenden, A. L. (1999). An introduction to metacognitive knowledge and beliefs in language learning: Beyond the basics. *System, 27*(4), 435–441. https://doi.org/10.1016/s0346-251x(99)00043-3

Wu, J., Li, Y., Zhou, J., & Chen, S. (2024). The impact of intelligent personal assistants on Mandarin second language learners: Interaction process, acquisition of listening and speaking ability. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2024.2317849

C

CALL and Vocabulary Learning

Shu-Li Lai[1], Hsiao-Ling Hsu[2] and Hsiao-Hui Yang[3]
[1]Center for General Education, National Taipei University of Business, Taipei, Taiwan
[2]Language Center, National Chung Hsing University, Taichung, Taiwan
[3]Applied English Department, Hungkuo Delin University of Technology, New Taipei City, Taiwan

Abstract

Vocabulary learning is the cornerstone of language acquisition, which has evolved greatly over time. This chapter will introduce the nature of vocabulary learning, its current trends in computer-assisted language learning, and the emerging role of artificial intelligence (AI) in this field. As vocabulary proficiency is crucial for effective communication and overall language competence, the primary goal of this chapter is to present a variety of approaches to enhance vocabulary acquisition. This chapter will examine four key trends in vocabulary learning: (1) multimodal input, which integrates multimedia resources to provide rich, contextualized input that aids in the retention and application of new vocabulary; (2) game-based learning to create engaging and motivating learning environments, offering experiential opportunities and encouraging active participation in vocabulary learning; (3) mobile-assisted vocabulary learning, which offers flexibility and convenience through the use of mobile devices, allowing learners to practice vocabulary anytime and anywhere; and (4) corpus-based, data-driven learning, which enables language learners to build vocabulary knowledge independently through authentic, large language databases, enhancing their understanding of words, patterns, and contextual appropriateness. The chapter will conclude by introducing AI tools for vocabulary learning. These technologies have revolutionized the field by facilitating more efficient and targeted vocabulary learning, opening new avenues for personalized learning experiences.

Keywords

Vocabulary learning · AI-generated materials · Multimodal inputs · Digital games · Mobile-assisted learning · Corpus-based learning · Data-driven learning · Chatbot · Learning motivation · Incidental learning

Introduction

The landscape of vocabulary learning has significantly advanced due to modern technology and artificial intelligence (AI). This chapter will

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_293-1

introduce the crucial role of vocabulary in language acquisition, focusing on the impact of computer-assisted language learning (CALL) on vocabulary development. The chapter will begin with an overview of the basics of vocabulary learning, setting the stage for understanding the subject. The chapter will then introduce current trends transforming vocabulary learning, including game-based learning, mobile-assisted vocabulary learning, and corpus-based learning. The final section will discuss AI's potential to enhance vocabulary learning and influence teaching methodologies. By explaining these technological approaches, this chapter will provide educators and learners with insights into the evolving field of vocabulary acquisition in the digital age.

Nature of Vocabulary Learning

Vocabulary learning is considered a crucial aspect of language learning, forming the foundation for effective communication and comprehension (Schmitt, 2010). A robust vocabulary enables learners to express complex ideas and understand nuanced texts. To facilitate effective vocabulary learning, some critical issues must be addressed.

Form, Meaning, and Use

According to Nation (2001, 2020), knowing a word involves understanding multiple aspects of the word: its form, meaning, and use. To know a word, one must know the spoken form, the written form, and the word's parts and be able to connect its form and meaning. Without a successful connection between form and meaning, one cannot claim to learn a word. In addition, one must know how the word is used in context, including its lexical grammar, collocation, and any constraints on its use.

Receptive and Productive Vocabulary

Receptive vocabulary refers to words that learners recognize and understand when they encounter those words in listening or reading. Productive vocabulary, on the other hand, is the ability to use words to produce language for speaking and writing purposes (Nation, 2001). It is generally believed that productive vocabulary use require more time and effort to acquire than receptive vocabulary use (Griffin & Harley, 1996).

High- and Low-frequency Words

Word frequency plays a pivotal role in vocabulary learning. Nation (2001) distinguished four kinds of vocabulary in texts: high-frequency words (i.e., 2,000 most frequent word families), academic words, technical words, and low-frequency words. Knowledge of high-frequency words is crucial as it allows learners to recognize a substantial proportion of words in both spoken and written texts (Dang et al., 2022). Schmitt and Schmitt (2014) proposed a new classification of vocabulary, namely, high-, mid-, and low-frequency words, expanding high-frequency vocabulary from 2,000 word families to 3,000 word families; additionally, they introduced the concept of mid-frequency vocabulary to describe words that fell between high and low frequency (Vilkaitė-Lozdienė & Schmitt, 2020). Recognizing the importance of high-frequency words, language educators can prioritize teaching words that are most commonly used in the language.

Intentional and Incidental Learning

Vocabulary acquisition can occur through intentional learning (i.e., the deliberate study of words) and incidental learning (i.e., the by-product of a meaning-focused task) (Webb, 2020). Researchers have agreed that intentional learning leads to efficient gains in vocabulary knowledge that may provide a crucial foundation for incidental learning to occur (Na & Nation, 1985). A balanced approach to vocabulary learning incorporates both intentional and incidental methods, which is more likely to lead to richer vocabulary (Teng, 2023a).

Current Trends in Vocabulary Learning

Multimedia and Vocabulary Learning

In the digital age, there are abundant free, high-quality multimedia resources for language learning, including YouTube channels, vocabulary learning apps, interactive e-books, and online

dictionaries with multimedia features. These materials provide multimodal inputs, including text, image, audio, video, subtitles or captions, and interactive features, creating rich learning environments that facilitate cognitive processes.

Clark and Paivio's (1991) dual coding theory of cognition suggests that vocabulary learning involves both verbal and nonverbal representations that are activated during the initial study of word pairs and subsequent attempts to recall translations. Therefore, by engaging learners with multiple inputs, they can strengthen their word retention and deepen their understanding of new words. This multimodal approach aligns with cognitive theories suggesting that information processed through multiple channels is more likely to be encoded in long-term memory.

The effects of multimodal inputs on language learning have gained increasing attention (Peters & Muñoz, 2020). Recent studies have shown the positive impacts of engaging students with audio and visual content, including subtitled television programmes, multimedia annotations, and interactive multimedia materials (e.g., Peters & Webb, 2018; Rodgers & Webb, 2020). These inputs enhance vocabulary learning by providing dynamic, contextual, and interactive experiences that traditional methods cannot match (Al Seghayer, 2001; Teng, 2023b).

Multimedia aids vocabulary retention by helping learners construct connections between verbal and visual representation, thus reducing learners' cognitive load while maintaining engagement. Multimedia enhances attention, facilitating the transformation of input into intake (Schmidt, 1993) and promoting incidental learning (Fievez et al., 2021; Puimège & Peters, 2019).

Furthermore, multimedia increases students' motivation, engagement, and enjoyment in vocabulary learning (Teng, 2023b). Variety prevents boredom, keeps learners curious, and makes the process more interesting than traditional methods. Incorporating audiovisual elements into vocabulary instruction is crucial for improving learning outcomes.

Game-Based Vocabulary Learning

Global players have grown significantly, reflecting the substantial expansion of the gaming industry. Digital games have become a popular form of entertainment for people of all ages. This widespread appeal has led to increased interest in the educational potential of digital games, particularly their effect on vocabulary learning. Studies have shown that digital games can greatly enhance learners' motivation and vocabulary learning (Prensky, 2001; Raffone, 2022; Tsai & Tsai, 2018; Zou et al., 2021). Enhanced learning motivation is a crucial factor for success in learning a new language (Dörnyei, 2014; Dörnyei & Ushioda, 2021). Game-based learning fosters this motivation by incorporating elements such as competition, achievement, progress, and enjoyment, making the learning experience more engaging and effective.

The importance of enjoyment in game-based learning cannot be overstated. Studies have shown that games transform learning into an interesting and entertaining experience (Pivec et al., 2003; Rasti-Behbahani & Shahbazi, 2022). Adventure games, for example, engage players in rich narratives and interactive challenges that encourage exploration and problem solving (Chen et al., 2016). These games focus on puzzle solving and narrative progression, typically involving intricate storylines that require players to actively engage with dialogue and environmental clues. Specifically, adventure games can enhance vocabulary by presenting contextualized language within a meaningful narrative, encouraging players to infer meaning and use words based on context and storylines (Chen et al., 2016; Chen & Yang, 2012; Vahdat & Behbahani, 2013).

In addition to adventure games, a variety of game types are available to teachers and learners, catering to different interests, language learning goals, and task types. Popular game types include massively multiplayer online role-playing games (MMORPGs), simulation games, and serious games. MMORPGs provide immersive virtual environments where players interact with one another and their surroundings, often assuming distinct roles and collaborating on tasks. These games can significantly enhance vocabulary

learning through extensive communication and teamwork, engaging players with a diverse range of vocabulary relevant to their in-game activities and interactions (Rama et al., 2012; Suh et al., 2010; Zheng et al., 2015).

Simulation games are valuable tools in language learning contexts because they can replicate real-world contexts where players use language to navigate tasks, solve problems, and interact with other characters or players. These games, including business simulations, city-building games, and historical simulations, provide opportunities to practice vocabulary related to specific scenarios (Miller & Hegelheimer, 2006; Ranalli, 2008). With advancements in technology, simulation games now offer immersive experiences through virtual reality (VR) headsets, such as Meta Quest, enhancing engagement and interactivity in the language learning process. This immersive context fosters meaningful interactions through cooperative and task-based learning. By creating a fully interactive environment, VR makes vocabulary acquisition more engaging and memorable while allowing learners to use language in context, reinforcing understanding and enhancing retention (Alfadil, 2020; Cho, 2018).

Adventure games, MMORPGs, and simulation games are commonly referred to as commercial off-the-shelf (COTS) games. While these games have been shown to enhance learners' motivation and vocabulary knowledge, some researchers have raised concerns about the presence of violent content in COTS games, which may be inappropriate for educational purposes (Carnagey & Anderson, 2005; Kirsh, 1997). In response to these concerns, serious games have been specifically designed with educational content and objectives, blending game elements with targeted learning goals.

In conclusion, digital games significantly enhance learners' motivation and vocabulary learning by incorporating elements of competition, achievement, progress tracking, and enjoyment. Different types of games, including adventure games, MMORPGs, and simulation games, cater to various interests and learning objectives, offering interesting and appealing environments that support vocabulary learning. Therefore, digital games present a dynamic and engaging approach to improving vocabulary learning.

Mobile-Assisted Vocabulary Learning

The growth of mobile technology has influenced billions of people worldwide and plays an important role in our daily lives. Due to the widespread ownership and frequent use of these devices, researchers and educators have identified a great opportunity to transform traditional language learning environments. Therefore, mobile-assisted language learning (MALL) has gained importance, allowing learners to incorporate language practice into their everyday activities (Hsu, 2017; Wong, 2013).

Specifically, the portability and availability of mobile devices provide unique advantages for language learning, making them valuable tools (Kukulska-Hulme, 2013; Sharples et al., 2009). This flexibility overcomes the limitations of traditional classroom settings, fostering a mobile learning environment that extends language learning beyond the classroom context. It also allows learners to engage in language learning activities tailored to their changing locations and situations, enhancing contextual and situational learning (Chen & Li, 2010; Kukulska-Hulme, 2009; Liu, 2009).

Furthermore, mobile devices empower language learners by providing access to a wide range of multimedia resources, interactive applications, and communication tools (Hsu, 2017; Wong, 2013). These resources include apps, online dictionaries, flashcards, podcasts, video lessons, and social media platforms that facilitate interaction with authentic language use. With mobile devices, learners can interact with diverse learning content anytime and anywhere, seamlessly integrating their studies into daily activities (Chinnery, 2006; Kukulska-Hulme & Shield, 2008).

Given the potential of mobile-assisted language learning, most MALL studies have focused on vocabulary learning (Burston, 2014; Burston & Giannakou, 2022). Those studies found that mobile devices allowed learners to practice and

learn new vocabulary effectively and engagingly (Kukulska-Hulme & Shield, 2008). Integrating mobile devices into daily routines can support both incidental and intentional vocabulary learning. Incidental vocabulary learning occurs when learners encounter new words in context, such as while reading news articles or listening to audiobooks and music. This form of learning is often spontaneous and meets learners' immediate interests and needs. Intentional learning, on the other hand, can be facilitated through dedicated language learning apps, interactive quizzes, digital flashcards, and virtual language exchange platforms. These tools, such as Kahoot!, allow learners to practice vocabulary through structured exercises and real-time interactions.

A particularly advantageous aspect of mobile devices for vocabulary learning is their implementation of spaced repetition. This technique involves presenting vocabulary items at increasing intervals to enhance long-term retention, as opposed to massed repetition, where items are reviewed in quick succession. Studies have shown that spaced repetition is effective for vocabulary retention (Kennedy & Levy, 2008; Lu, 2008; Song & Fox, 2008; Thornton & Houser, 2005). The mobile apps Anki and Quizlet can easily implement spaced repetition algorithms, providing personalized and adaptive learning experiences that help learners improve their vocabulary knowledge over time.

With their accessibility and flexibility, mobile devices offer valuable learning content, making them crucial tools for enhancing vocabulary. By using these devices, teachers can create interactive learning experiences that extend beyond the confines of the classroom. This approach also encourages learners to engage with language materials anytime and anywhere, fostering continuous learning and promoting active participation (Okumuş Dağdeler, 2023).

Corpus-Based, Data-Driven Learning

Corpus-assisted language learning has gained prominence in second language acquisition research over the past two decades (e.g., Boulton & Cobb, 2017; Casal & Kessler, 2024; Lee et al., 2019; Pérez-Paredes, 2022; Vyatkina & Boulton, 2017). This approach, often called "data-driven learning" (DDL), encourages learners to construct vocabulary knowledge through authentic language input (Johns, 1994; Sinclair, 2004). This inductive process promotes deeper thinking and language pattern discovery (Flowerdew, 2015). Moreover, keyword in context (KWIC) facilitates inference, generalization, and "noticing" of target items (Schmidt, 2001).

Cognitive factors play a key role in DDL. The lexical processing model by de Bot et al. (1997) outlines the steps to successful vocabulary comprehension, including determining unknown words, decoding input, matching word form with mental lexemes, and connecting these to syntactic and semantic features. Multiple example sentences aid in this process (Lee et al., 2020).

Autonomous learning is another key term in DDL. By observing corpus data, including word frequency, collocation, and lexical grammar, learners become language researchers (Johns, 1994). Learners' language proficiency levels may influence their ability to explore concordance lines, a valuable skill. It is not just about which words or phrases learners learn; it is about guiding students in learning how to learn a word. Studies have shown the benefits of DDL in both short-term and long-term L2 vocabulary learning (Lee et al., 2019).

However, DDL requires proper guidance and scaffolding. Learners need training in corpus consultation strategies and inferencing skills (Hu & Nassaji, 2014; Lee et al., 2020). A common limitation is the difficulty of interpreting authentic language examples (Yoon & Hirvela, 2004). Some researchers have addressed this by providing simplified example sentences (Lai et al., 2022) or user-friendly tools such as SKELL.

Now, generative AI offers new directions for DDL by generating customized corpora (Crosthwaite & Baisa, 2024; Godwin-Jones, 2024). AI can generate multiple native-like example sentences and texts tailored to learners' needs and levels, potentially offering new opportunities for vocabulary learning by combining existing corpus-driven tools with AI-generated data (Godwin-Jones, 2024).

AI and Vocabulary Learning

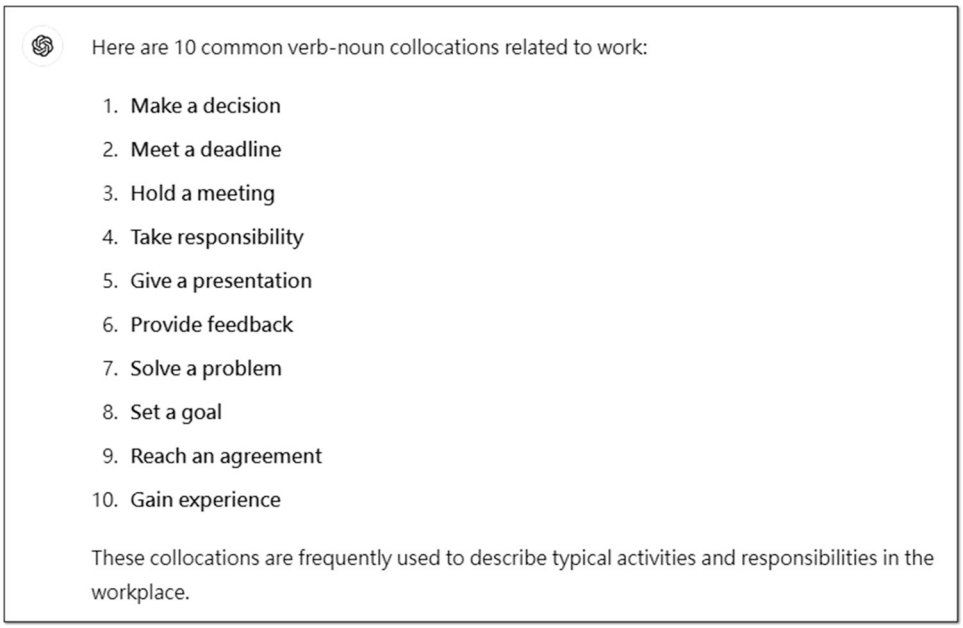
Generative AI offers valuable opportunities for English vocabulary learning (Godwin-Jones, 2024; Kohnke et al., 2023; Yang & Li, 2024). Tools such as ChatGPT, Claude, Perplexity, and Gemini enable the creation of AI-powered, learner-tailored materials in various formats through appropriate prompts (Kohnke et al., 2023; Moorhouse & Kohnke, 2024). AI provides several key applications. At the word level, AI can create personalized word lists, including challenging collocations, based on learners' needs and proficiency. For example, learners can request AI to suggest Verb-Noun (VN) collocations used in professional contexts (see Fig. 1). Collocations often pose challenges for ESL/EFL learners due to negative L1 transfer, so ESL/EFL learners often need to learn them explicitly.

At the sentence level, AI produces quality example sentences and paragraphs, adjusting difficulty levels using frameworks such as CEFR. Vocabulary is best learned in context, and AI tools can easily limit sentence length, keywords, and difficulty level. For example, based on the collocations suggested in Fig. 1, learners can ask ChatGPT to provide example sentences: "Please provide two example sentences at the CEFR B1 level for the following five phrases. Bold the phrase". The results are shown in Fig. 2.

AI can construct paragraphs and short texts in various genres (e.g., narrative, descriptive, argumentative) that incorporate multiple target vocabulary items into cohesive contexts. This capability allows for more extensive vocabulary exposure across different contexts and genres.

In addition to text, AI can generate voices and images for vocabulary learning. These multimodal materials have reinformed learning through diverse associations. According to Peters and Muñoz (2020), such multimodal inputs improve comprehension, retention, and recall, enhancing the effectiveness and engagement of vocabulary learning.

Generative AI systems simulate human-like conversation, offering valuable opportunities for language learners, especially in EFL contexts (Moorhouse & Kohnke, 2024). For example, CoolE Bot, developed by Lai et al. (2024), integrates various features such as text-to-speech, automatic speech recognition, and translation features (see Fig. 3) and offers themed-based conversations,


Here are 10 common verb-noun collocations related to work:

1. Make a decision
2. Meet a deadline
3. Hold a meeting
4. Take responsibility
5. Give a presentation
6. Provide feedback
7. Solve a problem
8. Set a goal
9. Reach an agreement
10. Gain experience

These collocations are frequently used to describe typical activities and responsibilities in the workplace.

CALL and Vocabulary Learning, Fig. 1 10 Common Verb-Noun Collocations Related to Work Suggested by GPT-4o

Sure, here are the example sentences with the phrases in bold:

1. Make a Decision

1. "After discussing all the options, we finally **made a decision** about the new project."
2. "She found it hard to **make a decision**, but she eventually chose the red dress."

2. Meet a Deadline

1. "We need to work extra hours this week to **meet the deadline** for the report."
2. "If we don't **meet the deadline**, the client will be very unhappy with us."

3. Hold a Meeting

1. "The manager will **hold a meeting** tomorrow to discuss the new company policies."
2. "We need to **hold a meeting** with the team to plan the upcoming event."

4. Take Responsibility

1. "He was brave enough to **take responsibility** for the mistake and apologize."
2. "As the team leader, she always **takes responsibility** for the group's work."

5. Give a Presentation

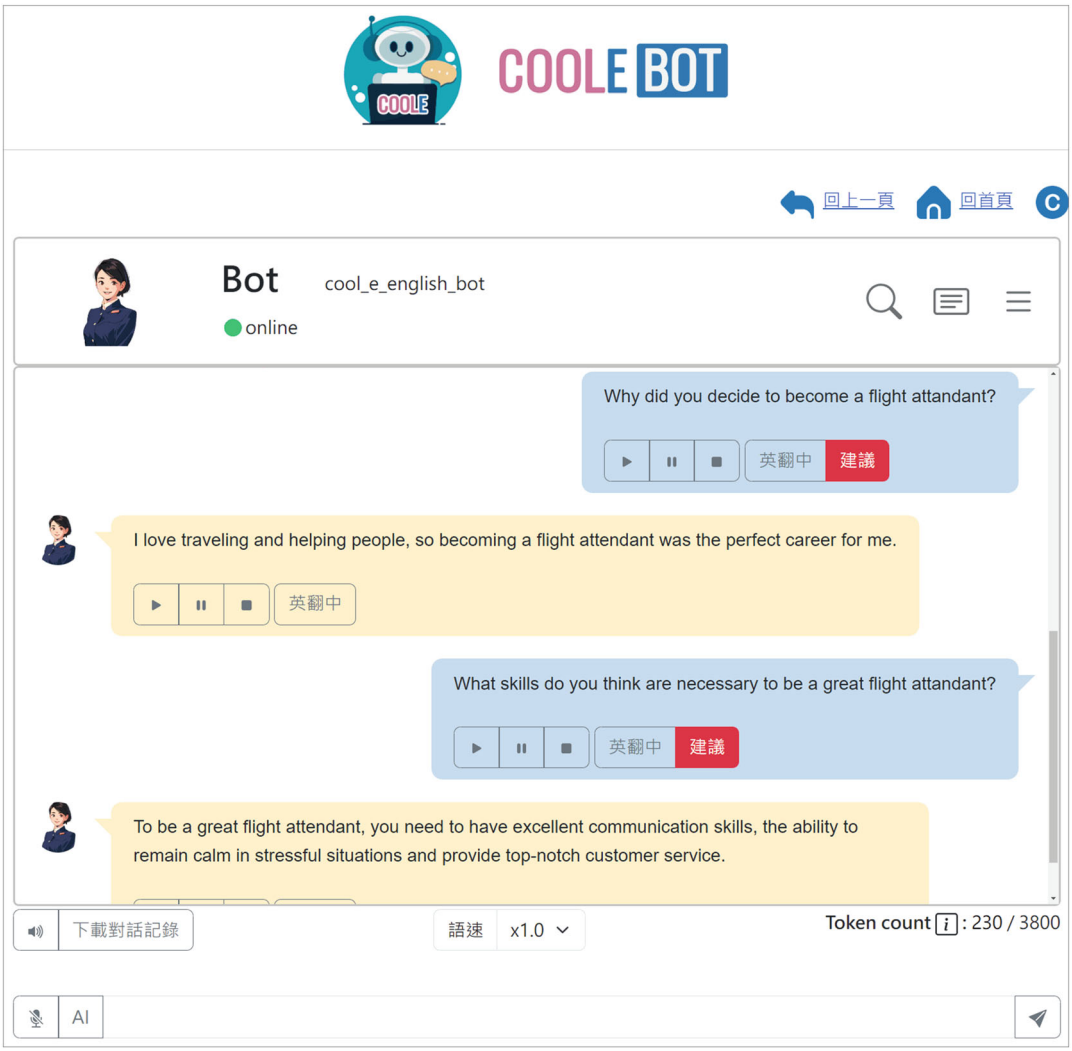
1. "Next week, I have to **give a presentation** about our new product."
2. "He is nervous because it's his first time **giving a presentation** to the board."

CALL and Vocabulary Learning, Fig. 2 Example sentences provided by ChatGPT based on the five most common V+N collocations used at work

featuring incidental vocabulary acquisition. Chatbots provide a range of expressions and vocabulary that human partners might lack, serving as powerful learning tools for EFL learners. This promotes productive vocabulary use.

AI can also facilitate the creation of adaptive self-learning and testing materials (Jeon, 2023; Kostka & Toncelli, 2023) by generating customizable vocabulary tests with various question types and themes, focusing on areas that need improvement. AI also provides immediate feedback on vocabulary exercises, enabling comprehension checks and promoting autonomous learning.

In conclusion, AI-powered vocabulary learning materials offer personalized, context-rich, and



CALL and Vocabulary Learning, Fig. 3 The interface of CoolE Bot and a chat with the flight attendant Chatbot

adaptive experiences, addressing traditional challenges in ESL and EFL vocabulary acquisition. This approach aligns with effective language learning principles, and more profound influences in language teaching are expected as AI tools become more sophisticated and ubiquitous (Godwin-Jones, 2024).

Conclusion

This chapter explored contemporary approaches to vocabulary learning in ESL and EFL contexts, fundamental concepts, and key trends and examined multimedia-enhanced learning, gamification, mobile-assisted language learning, and corpus-based approaches. These methods engage learners, provide flexible access to resources, and promote autonomous learning through authentic input. While challenges exist, recent AI advancements offer promising solutions, creating customized materials and enhancing vocabulary pedagogy. The integration of these technologies and methods signifies a transformative shift in vocabulary instruction, emphasizing personalized, context-rich, and adaptive learning experiences.

Cross-References

- ▶ Data-Driven and Corpus-Based CALL
- ▶ Digital Game-Based Language Learning and Teaching
- ▶ Intelligent CALL (ICALL)
- ▶ Mobile-Assisted Language Learning (MALL)
- ▶ Pedagogical Approaches and Methods of CALL

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Al Seghayer, K. (2001). The effect of multimedia annotation modes on l2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202–232.

Alfadil, M. (2020). Effectiveness of virtual reality game in foreign language vocabulary acquisition. *Computers & Education, 153*, 103893.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393.

Burston, J. (2014). A survey of MALL curriculum integration: What the published research doesn't tell. *CALICO Journal, 31*(3), 303–322.

Burston, J., & Giannakou, K. (2022). MALL language learning outcomes: A comprehensive meta-analysis 1994–2019. *ReCALL, 34*(2), 147–168.

Carnagey, N. L., & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychological Science, 16*(11), 882–889. https://doi.org/10.1111/j.1467-9280.2005.01632.x

Casal, J. E., & Kessler, M. (2024). New theoretical and practical horizons in data-driven learning: Introduction to the special issue. *TESOL Quarterly, 58*. https://doi.org/10.1002/tesq.3331

Chen, C.-M., & Li, Y.-L. (2010). Personalized context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments, 18*(4), 341–364.

Chen, H. J. H., & Yang, T. Y. C. (2012). The impact of adventure video games on foreign language learning and the perceptions of learners. *Interactive Learning Environments, 21*(2), 129–141. https://doi.org/10.1080/10494820.2012.705851

Chen, M.-H., Tseng, W.-T., & Hsiao, T.-Y. (2016). The effectiveness of digital game- based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology, 49*(1), 1–9.

Chinnery, G. M. (2006). Emerging technologies: Going to the MALL: Mobile assisted language learning. *Language Learning & Technology, 10*(1), 9–16.

Cho, Y. (2018). *How spatial presence in VR affects memory retention and motivation on second language learning: A comparison of desktop and immersive VR-based learning*. Doctoral dissertation, Syracuse University.

Clark, J. M., & Paivio, V. (1991). Dual coding theory and education. *Educational Psychology Review, 3*, 149–210.

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*. https://doi.org/10.1075/ijcl.23056.cro. Advance online publication.

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research, 26*(4), 617–641.

De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition, 19*, 309–329.

Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation*. Routledge.

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2021). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning, 36*(1–2), 26–51.

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins.

Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology, 28*(2), 5–30.

Griffin, G. F., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics, 17*, 223–460.

Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter? *Computers & Education, 106*, 137–149.

Hu, H. M., & Nassaji, H. (2014). Lexical inferencing strategies: The case of successful versus less successful inferences. *System, 45*, 27–38.

Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning, 36*(7), 1338–1364.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical*

grammar. Cambridge applied linguistics (pp. 293–313). Cambridge University Press.

Kennedy, C., & Levy, M. (2008). L'italiano al telefonino: Using SMS to support beginners' language learning. *ReCALL, 20*(3), 315–330.

Kirsh, S. (1997). Seeing the world through "Mortal Kombat" colored glasses: Violent video games and hostile attribution bias. *Childhood, 5*(2), 177–184.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550.

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendation. *TESL-EJ, 27*(3), 1–19.

Kukulska-Hulme, A. (2009). Will mobile learning change language learning? *ReCALL, 21*(2), 157–165.

Kukulska-Hulme, A. (2013). Mobile-assisted language learning. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3701–3709). Wiley.

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271–289.

Lai, S. L., Chang, J., Lee, K. L., & Huang, W. C. (2022). Linggle 2.0: A collocation retrieval system with quality example sentences. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds.), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 234–239). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1464

Lai, S. L., Chen, H. J., Suen, J. H., & Huang, E. J. (2024). Developing a chatbot for EFL learners: Process and insights. In *JALTCALL 2024 Conference*, May 17–19, Meijo University, Nagoya.

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753.

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal, 104*(2), 345–362.

Liu, T. Y. (2009). A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning, 25*(6), 515–527.

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning, 24*(6), 515–525.

Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL: Incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education, 3*(4), 311–328.

Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System, 122*. https://doi.org/10.1016/j.system.2024.103290

Na, L., & Nation, I. S. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16*(1), 33–42.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15–29). Routledge.

Okumuş Dağdeler, K. (2023). A systematic review of mobile-assisted vocabulary learning research. *Smart Learning Environments, 10*(1), 19.

Pérez-Paredes, P. (2022). How learners use corpora. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 390–405). Routledge.

Peters, E., & Muñoz, C. (2020). Introduction to the special issue: Language learning from multimodal input. *Studies in Second Language Acquisition, 42*, 489–297.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*, 1–27.

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003). Aspects of game-based learning. In *The 3rd international conference on knowledge management* (Vol. 304, pp. 14–16), Graz.

Prensky, M. (2001). *Digital game-based learning*. McGraw-Hill.

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*, 424–438.

Raffone, A. (2022). Gamifying English language learning through interactive storytelling and MALL technologies. *Language Teaching Research*. https://doi.org/10.1177/13621688221110206

Rama, P., Black, R., van Es, E., & Warschauer, M. (2012). Affordances for second language learning in World of Warcraft. *ReCALL, 24*(3), 322–338.

Ranalli, J. (2008). Learning English with the Sims: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning, 21*(5), 441–455.

Rasti-Behbahani, A., & Shahbazi, M. (2022). Investigating the effectiveness of a digital game-based task on the acquisition of word knowledge. *Computer Assisted Language Learning, 35*(8), 1920–1945.

Rodgers, M. P., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics, 171*(2), 191–220.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206–226.

Schmidt, R. (2001). Attention. In P. J. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 284–503.

Sharples, M., Milrad, M., Arnedillo-Sánchez, I., Milrad, M., & Vavoula, G. (2009). Mobile learning: Small devices, big issues. In N. Balacheff, S. Ludvigsen, T. Jong, A. Lazonder, & S. Barnes (Eds.), *Technology enhanced learning: Principles and products* (pp. 233–249). Springer.

Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 271–299). John Benjamins.

Song, Y., & Fox, R. (2008). Using PDA for undergraduate student incidental vocabulary testing. *ReCALL, 20*(3), 290–314.

Suh, S., Kim, S. W., & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning, 26*(5), 370–378.

Teng, M. F. (2023a). Understanding incidental vocabulary learning in practice. *Asian Journal of English Language Teaching, 32*(1), 7–28.

Teng, M. F. (2023b). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching, 17*(3), 738–754.

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning, 21*(3), 217–228.

Tsai, Y.-L., & Tsai, C.-C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education, 125*, 345–357.

Vahdat, S., & Behbahani, A. R. (2013). The effect of video games on Iranian EFL learners' vocabulary learning. *The Reading Matrix, 13*, 61–71.

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 31–96). Routledge.

Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology, 21*(3), 1–8.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Wong, L. H. (2013). Analysis of students' after-school mobile-assisted artifact creation processes in a seamless language learning environment. *Educational Technology & Society, 16*(2), 198–211.

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 1–14.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257–283.

Zheng, D., Bischoff, M., & Gilliland, B. (2015). Vocabulary learning in massively multi-player online games: Context and action before words. *Educational Technology Research and Development, 63*(5), 771–790.

Zou, D., Huang, Y., & Xie, H. (2021). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning, 34*(5–6), 751–777.

G

Genre-Based Approaches to CALL

Christoph A. Hafner and Gabriel Tetteh
Department of English, City University of Hong Kong, Hong Kong, China

Abstract

Genre-based approaches to language learning are those that focus on regularly occurring communicative events that are realized in the form of particular text types, like argumentative essays, lab reports, or academic presentations, for example. A typical pedagogical process involves teachers and students in stages of: (1) context analysis; (2) model analysis; (3) guided composition with teachers and students working together; (4) extended composition with the student working on their own; and (5) comparison, which lays the foundations for transfer to other genres. Research of relevance to a genre-based approach to CALL falls into three main categories. First, research on digital genres suggests that innovative genres in digital contexts can be 'paired' with traditional genres in order to stimulate discussion and reflection around the communicative affordances and constraints of such pairs of 'new' and 'old' genres. Second, research has also examined the role that digital tools like collaborative writing platforms, interactive blogs, and wikis can play in the genre writing process. Such digital tools can expand the 'learning environments' or 'spaces' available in the genre writing process as well as make authentic audiences available for language learners. Third, another strand of research on 'digital multimodal composing' has examined the way that language learners can be engaged with multimodal forms of communication enabled by digital tools.

Keywords

Digital literacies · Genre-based pedagogy · Digital genres · Collaboration · Digital multimodal composing

Introduction

Genre-based approaches to CALL are those that focus on the development of communicative competence with respect to specific, identifiable genres. Here, the term 'genre' refers to regularly occurring communicative events that are realized in the form of particular text types, like argumentative essays, lab reports, or academic presentations, for example. Genre-based approaches are informed by genre pedagogy developed in English for specific purposes, rhetorical genre studies, and the Sydney school approach to genre analysis. They are particularly well suited to project-based approaches to CALL (Warschauer et al., 2000) and ecological

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_299-1

approaches to CALL (Blin, 2016), which are sensitive to the context of communication and tend to foster authentic environments for the production of target genres. In this entry, we begin by providing an outline of principles of genre pedagogy and then examine how such a pedagogy can be integrated into technological environments with reference to three main strands of research on (1) digital genres, (2) collaborative writing platforms, and (3) digital multimodal composing, respectively.

Genre-Based Pedagogy

The overarching concern for genre theorists, Hyon (1996) notes, relates to "helping students become more successful readers and writers of academic texts and workplace texts" (p. 700). Thus, a genre-based pedagogy is one in which teachers and learners are focused on selected genres, i.e. text types and related activities, with a view to developing the necessary knowledge to use those genres effectively. Underpinning genre-based pedagogy is the notion of *genre*, variously defined by Bhatia (1993), Martin et al. (1987), Miller (1984), and Swales (1990). While no single definition of genre has been agreed upon by all, some common themes are identifiable in how it has been conceptualized. These include genre as:

- A typified rhetorical social action;
- A response to particular recurrent rhetorical situations;
- A communicative event for the accomplishment of common communicative purposes in a discourse community;
- A conventional form with typical organizational structure and language forms;
- A dynamic and not static process and form (Hafner & Miller, 2019).

In other words, genres can be conceived of as conventionalized text types that are used for particular communicative purposes and arise in response to regularly occurring social and rhetorical situations. Furthermore, a genre is also seen as a multimodal semiotic activity which forms part of a wider genre ecology.

A number of genre-based pedagogies have been developed, three of which are presented here (Hafner & Miller, 2019; Hyland, 2018; Hyon, 2018). Hyland (2018, p. 4) proposes a five-stage pedagogy: (1) setting the context (revealing genre purposes and the settings in which it is commonly used); (2) modelling (analysing representative samples of the genre to identify its stages and key features and the variations which are possible); (3) joint construction (guided, teacher-supported practice in the genre through tasks which focus on particular stages or functions of the text); (4) independent construction (independent writing by students monitored by the teacher); and (5) comparing (relating what has been learnt to other genres and contexts to understand how genres are designed to achieve particular social purposes). In contrast, Hyon (2018) presents a three-stage pedagogy: (1) rhetorical consciousness-raising tasks; (2) text production activities; and (3) process activities. Hafner and Miller (2019) also advance a three-stage process-oriented genre approach to attend to multimodality and digital literacies. According to them, teachers should: (1) identify and situate authentic genres and practices; (2) examine the role of multimodality and digital tools; and (3) design projects, tasks and materials, adopting Hyland's (2018) five-stage pedagogy. Researchers taking these genre-based approaches have engaged with technology in a range of ways in the learning process.

Digital Genres

A focus on digital genres aligns with a digital literacies approach to CALL, which aims to foster language learners' ability to 'learn to use technology' to communicate and write for the digital context (Hafner & Pun, 2020; see also Shetzer & Warschauer, 2000). Such a focus is in contrast to approaches that 'use technology to learn', a practice which was the mainstay of previous CALL approaches. With digital literacies then, teachers following a genre-based approach consider how

digital media has shaped generic forms and processes, and set out to understand those changes, with a view to developing students' genre competence in the digital media context. The digital genre represents a new, digital target form that draws on the affordances of digital media in complex digital contexts. Such digital media affordances—including multimodality, hypertextuality and interactivity—facilitate entextualization, recontextualization and semiotic remediation, resulting in interdiscursivity and intertextuality, "a fluidity that complicates both the notion of text and the context of texts" (Belcher, 2023, p. 36). In terms of context, it is worth noting that the audience of online genres frequently goes beyond the normal specialist audience that genre-based pedagogy often focuses on. Furthermore, it includes not only a 'diversified' audience of both specialists and non-specialists but also non-human algorithmic audiences, with writers crafting texts that will appeal to the algorithm in order to gain attention in the online context.

A number of practical strategies have been suggested to engage learners with digital genres, both as target forms and as a stimulus to promote reflection, comparison, and contrast. For example, Darvin (2023) notes that 'traditional' and 'digital' genres exist on a 'genre continuum' and illustrates how university students can benefit from recursive contrastive analyses of online and offline genres with shared purposes and structures. Similarly, Tardy (2023) calls for the teaching of high-stakes academic genres to be paired with digital counterparts like social media posts or infographics. For example, scientific articles written for a scientific community audience could be paired with digital genres like academic blogs and other social media genres like tweets, which are less formal and have diversified audiences. Hafner and Miller (2019) describe a project-based approach in which science students present the findings of a simple scientific field study first as a multimodal digital video uploaded to YouTube and aimed at a public audience, and then as a written scientific report aimed at a specialist audience. Such recontextualization activities serve to promote genre awareness, as students reflect on how they need to adapt rhetorical strategies, making use of digital tools to design their compositions for different kinds of audiences.

Collaborative Learning Environments

Practitioners taking a genre-based approach are also using technology to create an interactive environment that facilitates genre learning in language classrooms. Such an approach aligns well with an ecological approach to CALL (Blin, 2016), allowing for the design of interactive and collaborative learning environments. Such environments allow for plentiful feedback as part of a classroom community as shown in Bloch (2007) and Kuteeva (2011). For example, Bloch (2007) examines the influence of blogging on students' writing competence in an L2 composition course at a university. Blogging was found to foster a sense of community among students of disparate backgrounds, provide a space for students to generate critical views for their writing assignments, and facilitate an interaction between readers and writers in the blogosphere. Similarly, Kuteeva (2011) reports on how the use of wikis in a genre writing process (involving texts such as formal letter, argumentative text, summary or abstract of a formal report and special genre-based project) not only engenders peer review and collaboration but also provides students with a platform for shaping their academic writing competences. In both cases, the affordances of the online tools that allowed students to respond to each other's blogs and post comments on the discussion pages of the wiki fostered collaborative online interactions and promoted a sense of community among learners.

Moreover, such online learning environments can allow students to engage with an authentic audience. In an academic writing course focusing on academic essay writing, King (2015) required university students in Hong Kong to develop and publish Wikipedia articles. Through this process, they participated as legitimate peripheral members of the Wikipedia community and developed identities as competent writers and contributors of Wikipedia articles. By engaging with the wider Wikipedia community and responding to readers'

critiques, the students were motivated "to think deeply about their writing and. . .invest in writer identities online and in their local context" (p. 106).

In the same vein, Parks et al. (2003) reported how teachers who tasked their high school students with creating genres such as PowerPoint presentations and websites recruited a more public audience for their students. To facilitate this authentic interaction, teachers organized a museum event where students presented their projects to their parents, among others. They also contacted a NASA official electronically to provide feedback to students about their space projects.

Recently, generative AI has also been conceived of as a kind of collaborative partner capable of offering similar kinds of feedback. Automated writing evaluation (AWE) systems are being developed to help provide feedback for student writers. Feng and Chukharev-Hudilainen (2022) report on a genre-based AWE system created to provide feedback to Taiwanese engineering graduate students on their written research abstracts. Feedback from the system enhanced students' construction of rhetorical moves in their abstracts.

Digital Multimodal Composing

As mentioned above (Digital genres), a common genre-based pedagogy adopted to address use of digital technologies in language learning is to examine the way that the affordances of digital tools can enable new forms of multimodal meaning-making. In particular, researchers and practitioners have explored the possible use of digital multimodal composing (DMC) activities in language teaching and learning. DMC has been defined as 'a textual practice that involves the use of digital tools to produce texts by combining multiple semiotic modes that include, but are not limited to, image, word, and soundtrack' (Jiang, 2017, p. 413). For example, students could be engaged in learning genres like, among others, websites, PowerPoint presentations, infographics, posters, brochures, and digital video scientific documentaries. A DMC approach focuses specifically on the role that multimodal semiotic resources (including language) play in the construction of the genre.

Research on the use of DMC has highlighted a range of benefits, from improved multimodal communicative competence to increased motivation and investment among students (see Belcher, 2017 for a useful summary). However, critics question the effectiveness of the approach in fostering language development, conceived as linguistic competence (Manchón, 2017). Nevertheless, a recent experimental study comparing students following a traditional essay-based curriculum with students doing DMC, showed that students doing DMC outperformed the essay group on a 'content and language' measure (Kim et al., 2023). Such a finding perhaps demonstrates that students engaged in DMC pay significant attention to language in the course of their multimodal composing process; for example, students creating a video production may devote significant time and energy to producing a high-quality script. According to Jiang and Hafner (2024), future research on DMC should focus on: effectiveness of DMC for writing development, DMC task design, L2 teacher education for DMC, feedback practices, DMC assessment, collaborative DMC as a translanguaging space, and the deployment of DMC for critical digital literacies.

Conclusion

This entry has provided a summary of genre-based approaches to CALL, beginning with a review of genre-based pedagogy and followed by accounts of three main areas: digital genres, collaborative learning environments, and digital multimodal composing. Future research on genre-based approaches to CALL can continue to elaborate on the hybrid, intertextual forms of digital genres and the complex processes of collaborative genre construction. More research is needed to shed light on the way that power is constructed in online spaces, examining: (1) the affordances and constraints of online spaces and how algorithms and ideologies shape our online interactions and behaviour; (2) the use of digital genres

to misinform, disinform, and manipulate. Future pedagogically oriented research can help us to understand how genre-based approaches can be adapted to help students to develop capacities for innovation and creativity in digital contexts, as well as a critical awareness of the way that online platforms shape writing activity and how this tendency might be resisted. Finally, an important avenue of future research will be to consider the impact of generative AI on genre-based approaches to CALL, especially its effect on the writing process and on learning.

Cross-References

- ▶ Digital Literacies
- ▶ Genre Awareness and Data-Driven Learning
- ▶ Multiliteracies
- ▶ Multiliteracies-Based Approaches to CALL
- ▶ Project-Based Approaches to CALL
- ▶ Task-Based Language Learning in CALL Environments

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Belcher, D. D. (2017). On becoming facilitators of multimodal composing and digital design. *Journal of Second Language Writing, 38*, 80–85.

Belcher, D. D. (2023). Digital genres: What they are, what they do, and why we need to better understand them. *English for Specific Purposes, 70*, 33–43.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.

Blin, F. (2016). *Towards an 'ecological' CALL theory: Theoretical perspectives and their instantiation in CALL research and practice*. Routledge.

Bloch, J. (2007). Abdullah's blogging: A generation 1.5 student enters the blogosphere. *Language Learning & Technology, 11*(2), 128–141.

Darvin, R. (2023). Moving across a genre continuum: Pedagogical strategies for integrating online genres in the language classroom. *English for Specific Purposes, 70*, 101–115.

Feng, H., & Chukharev-Hudilainen, E. (2022). Genre based AWE system for engineering graduate writing: Development and evaluation. *Language Learning & Technology, 26*(2), 58–77.

Hafner, C. A., & Miller, L. (2019). *English in the disciplines: A multimodal model for ESP course design*. Routledge.

Hafner, C. A., & Pun, J. (2020). Editorial: Introduction to this special issue: English for academic and professional purposes in the digital era. *RELC Journal, 51*(1), 3–13.

Hyland, K. (2018). Genre and second language writing. In *The TESOL encyclopedia of English language teaching* (pp. 1–6). Wiley.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly, 30*(4), 693–722.

Hyon, S. (2018). *Introducing genre and English for specific purposes*. Routledge.

Jiang, L. J. (2017). The affordances of digital multimodal composing for EFL learning. *ELT Journal, 71*(4), 413–422.

Jiang, L. J., & Hafner, C. A. (2024). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching*, 1–19.

Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *Journal of English for Academic Purposes, 64*, 1–13.

King, B. W. (2015). Wikipedia writing as praxis: Computer-mediated socialisation of second-language writers. *Language Learning & Technology, 19*(3), 106–123.

Kuteeva, M. (2011). Wikis and academic writing: Changing the writer–reader relationship. *English for Specific Purposes, 30*, 44–57.

Manchón, R. M. (2017). Disciplinary dialogues: The potential impact of multimodal composition on language learning. *Journal of Second Language Writing, 38*, 94–95.

Martin, J. R., Christie, F., & Rothery, J. (1987). Social processes in education: A reply to sawyer and Watson (and others). In I. Reid (Ed.), *The place of genre in learning: Current debates* (pp. 46–57). Deakin University Press.

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech, 70*, 151–167.

Parks, S., Huot, D., Hamers, J., & H.-Lemmonier, F. (2003). Crossing boundaries: Multimedia technology and pedagogical innovation in a high school class. *Language Learning & Technology, 7*(1), 28–45.

Shetzer, H., & Warschauer, M. (2000). An electronic literacy approach to network-based language teaching. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 171–185). Cambridge University Press.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Tardy, C. M. (2023). How epidemiologists exploit the emerging genres of twitter for public engagement. *English for Specific Purposes, 70*, 4–16.

Warschauer, M., Shetzer, H., & Meloni, C. (2000). *Internet for English teaching*. TESOL, Inc.

E

E-Portfolio Development Literacy

Ricky Lam
Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

Abstract

E-Portfolios, the digital version of print portfolios, have been adopted for over two decades. Despite their advantages, not all students have adequate competence to understand and achieve specific goals related to building e-Portfolios which are intended to improve their second language learning. E-Portfolio development literacy, an emerging subset of digital literacies, underscores the process-orientated rather than product-orientated nature of e-Portfolio assessment, which can enhance one's language learning development through iterative and feedback-rich learning experiences. For teachers, e-Portfolio development literacy involves the mastery of a knowledge base, skills, and awareness of how e-Portfolios can integrate into new literacies curricula, bridge the digital divide among students, and synergise the formative and summative purposes of assessment pedagogically. For students, it entails the ability, strategies, and dispositions needed when they construct their digital dossiers to monitor, review, and adjust language learning metacognitively. Students also need to acquire understanding and skills to reflect upon their past learning in order to envision their future learning by responding to multimodal feedback, formulating planning strategies, and utilising other Internet resources. To foster e-Portfolio development literacy, teachers can provide students with scaffolded input (portfolio pedagogy promoting self- and co-regulated learning) and guidance on ethics (avoidance of academic dishonesty) to facilitate e-Portfolio compilation tasks in a safe, trustworthy, and ethical manner.

Keywords

Print portfolios · Electronic portfolios · e-Portfolios · Assessment · Reliability · Ethics · Digital literacies

Definition

For teachers, e-Portfolio development literacy is about the knowledge, skills, and awareness required to incorporate e-Portfolio-based pedagogy into new literacies curricula to support student learning. For learners, this literacy refers to their ability, strategies, and dispositions regarding how they take stock of their learning development through active participation in key portfolio compilation processes, including the creation, curation, review, and circulation of digital artefacts.

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_301-1

Introduction

E-Portfolios refer to a digital space wherein learners document and showcase their language learning for a variety of educational purposes. Their characteristics are process-orientated, student-centred, multimedia, and communal, aligning with attributes of ▶ new literacies or ▶ digital literacies curricula where dynamic meaning-making and content creation experiences are prioritised. Owing to their organic and iterative nature, e-Portfolio development processes entail learners' creation, curation, review, and circulation of digital artefacts to generate knowledge and make sense of learning, as well as demonstrate their learning to the target audience, such as peers, teachers, parents, or the public (Yancey, 2019). These developmental processes are usually ongoing, metacognitive, and collaborative as opposed to product-orientated second language instruction which highlights didactic and deductive pedagogies. E-Portfolio development literacy is an emerging form of digital literacies, in which teachers need to acquire specific expertise to design and incorporate e-Portfolio assessment into their instructional approaches to empower students to be independent. Similarly, students need to acquire related strategies and dispositions to partake in these new literacy compilation journeys by learning how to self- and co-regulate their learning. Self- and co-regulation allude to socio-psychological developmental processes of how students independently and collaboratively manage their learning in relation to individual set goals and institutional learning outcomes. This entry emphasises the developmental qualities of e-Portfolio assessment and how these qualities can be further enhanced by improving teachers' and students' assessment literacy.

The predecessor of e-Portfolios is print portfolios, which date back to their original use in higher education in the late 1970s. Print portfolios emerged when university-level and school educators became dissatisfied with product-orientated writing instruction and the validity of timed impromptu essay testing. Unsurprisingly, the adoption of a *workshop* mode of pedagogy highlighting recursive composing processes, reflective thinking mindsets, teacher–student dialogues, and interactive use of formative feedback was regarded as a better alternative to the then-existing instructional and assessment practices (for a complete review of print portfolios, see Lam (2017)). The 1990s marked the heyday of print portfolios and also signalled the point when many K-12 educators began to transition to the cousins of print portfolios, e-Portfolios, despite certain challenges such as low inter-rater reliability and additional costs for assessment training.

After the turn of the century, e-Portfolios became mainstream as a result of the continuing spread of technologies such as the world wide web and mobile devices. The transition from print to e-Portfolios involves some innovative educational practices, such as how students create multimedia artefacts to document their learning, and how teachers evaluate portfolio development, which requires technological skills such as accessing online resources via URLs and providing electronic feedback. The 2000s till now have witnessed the building of e-Portfolios using a plethora of software and platforms, including word processing software (e.g. Microsoft Word), weblogs (e.g. WordPress), learning management systems (e.g. Mahara), e-learning apps (e.g. Seesaw), instant messaging apps (e.g. WeChat Moments), and social media platforms (e.g. Facebook; for a chronicle of e-Portfolio development, see Farrell (2020)).

E-Portfolios remain contentious because of inadequate validation of their usefulness. For instance, most e-Portfolio studies have investigated stakeholders' perceptions rather than their learning outcomes as a result of e-Portfolio use. Thus far, there are no tried-and-tested scales which can accurately measure the effectiveness of any oft-cited portfolio approaches (Lam, 2023). On the other hand, there has been no shortage of evidence in support of effective e-Portfolio implementations, such as increased motivation, improved writing skills, and mastery of self- and co-regulated learning. Nonetheless, this evidence has often come from studies that included a small number of e-Portfolio programmes which were still product-orientated, de-contextualised, teacher-centric, and, worse still, devoid of a proper reflective component (Lam, 2023). Without understanding the idea of

e-Portfolio development trajectories, teachers and researchers may misinterpret how a metacognitive and learner-centred e-Portfolio programme will impact students' second language acquisition and the formation of an active learner identity within a ▶ multimodal literacy environment. Therefore, the following implications are proposed to promulgate e-Portfolio development literacy.

Implications for teachers and educational institutions:

Maximise integration into the curriculum. To sustain e-Portfolio development processes like curation and reflection for a longer time, teachers and programme directors should ensure a smooth e-Portfolio integration into the existing language and academic writing curricula, respectively. E-Portfolios could serve as a tool, facilitating the delivery of new literacies content. After all, e-reading and digital multimodal composing are commonplace and should be seamlessly blended into the twenty-first-century language curriculum. Teachers should also master knowledge, skills, and awareness concerning how to integrate e-Portfolio assessment into school-based and university curricula (Lam, 2024). To effectively integrate e-Portfolios, teachers can consider different approaches based on their curriculum needs and resources. For instance, there are three integration modes, namely (1) the immersive mode which fully integrates into the existing curriculum; (2) the parallel mode where e-Portfolios are adopted as a one-off or short-term programme alongside the existing curriculum; and (3) the segregated mode wherein students use e-Portfolios as a learning companion with minimal intervention. The choice of integration mode may depend on factors such as institutional support, teacher readiness, and student needs.

Narrow the digital divide. To help students compile their digital artefacts across time (say in a 3-year window), schools should provide both financial and infrastructure support, for instance, loans of electronic equipment (e.g. tablets), partial or full sponsorship of software subscription fees, and unlimited use of campus Wi-Fi. This is because the making of some e-Portfolio artefacts involves collaborative writing tasks, off-campus online participation (e.g. forum posts), and mini-research projects requiring literature reviews. Teachers and school leaders should not take for granted that every student can afford the expenses incurred by the use of e-Portfolios, especially those from less privileged families. The use of cross-platform apps via a mobile phone rather than a laptop may help narrow the digital divide and alleviate the financial burdens, as not every student can afford to own a laptop or bring one to class. By addressing the digital divide through various support measures and alternative solutions, schools can ensure that e-Portfolio implementation is inclusive and equitable for all students.

Use e-Portfolios for formative and summative evaluation. Adopting e-Portfolios for formative and summative evaluation concurrently remains a complex undertaking. Teachers are advised to evaluate students' digital artefacts or an entire e-Portfolio near the end of a semester or a teaching unit so that students can focus on the compilation, curation, review, and revision of their e-Portfolio content for formative evaluation, namely, self-assessment, peer reviews with classmates or chatbots, and synchronised teacher e-feedback on early or interim drafts/work (Cheng & Zhang, 2024). Suspension of summative grading in the e-Portfolio development process can encourage better and more frequent text revision and error corrections, which can also bring about improved language learning skills. Synergising the formative and summative evaluation in e-Portfolios appears to be feasible although teachers need to be assessment-literate in judging and supporting student learning at the same time. For example, teachers could use formative assessments throughout the semester to guide students' e-Portfolio development and then use the final e-Portfolio for summative grading, taking into account the progress demonstrated in the previous formative assessments.

Implications for learners:

Develop e-Portfolio literacy. Students should work to acquire pertinent skills and strategies to regularly curate and reflect on their artefacts in line with set goals or course-based assessment criteria. The acts of plan-monitor-review-and-adjust are indispensable in nurturing e-Portfolio development literacy because these reflective practices are learning-orientated. Acquiring e-Portfolio literacy entails interactive uptake among the following major skills, including technological know-how, linguistic competence, learning strategies, emotion regulation, rudimentary assessment principles, and digital communication skills (Lam, 2024). Because of that, students are encouraged to attend related e-Portfolio training comprising language, technology, and communication components before an e-Portfolio programme begins.

Use Internet materials and GenAI sources ethically. Throughout the e-Portfolio development process, students frequently browse online information, cite source texts, borrow digital resources, and utilise AI-powered tools to increase the quality of their multimedia artefacts. This could include editing software (e.g. Grammarly), image creator apps (e.g. Copilot), machine translation tools (e.g. DeepL), or AI research assistants (e.g. Elicit and Consensus). While these tools can increase the quality of student work, their unethical use may be linked to academic dishonesty, including plagiarism. This has been a growing concern since the arrival of ChatGPT in November 2022. To this end, students need to raise their awareness of ▶ AI literacy practices, which are an integral part of e-Portfolio development literacy. Students need training on the ethical use of AI tools and online resources and clear guidelines on how to properly attribute AI-generated content.

E-Portfolios represent an innovative medium, which uniquely supports meaning-making and knowledge-creating processes in new literacies. To effectively fulfil their pedagogical, learning-orientated, and evaluative goals, teachers and students need to attain a moderate level of e-Portfolio literacy. This can be achieved through workshops about e-Portfolio integration, webinars on technology-assisted alternative assessment methods, academic readings, and participation in action research or exploratory practice projects independently or with university researchers.

Cross-References

▶ AI Literacy Practices
▶ Digital Literacies
▶ Multimodal Literacies
▶ New Literacies

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Cheng, X., & Zhang, L. J. (2024). Examining second language (L2) learners' engagement with AWE-teacher integrated feedback in a technology-empowered context. *Asia-Pacific Education Researcher, 33*, 1023–1035. https://doi.org/10.1007/s40299-024-00877-8

Farrell, O. (2020). From portafoglio to e-portfolio: The evolution of portfolio in higher education. *Journal of Interactive Media in Education, 2020*(1), 19. https://doi.org/10.5334/jime.574

Lam, R. (2017). Taking stock of portfolio assessment scholarship: From research to practice. *Assessing Writing, 31*, 84–97. https://doi.org/10.1016/j.asw.2016.08.003

Lam, R. (2023). E-Portfolios: What we know, what we don't, and what we need to know. *RELC Journal, 54*(1), 208–215. https://doi.org/10.1177/0033688220974102

Lam, R. (2024). *Integrating e-Portfolios into L2 classrooms: Education for future*. Multilingual Matters.

Yancey, K. B. (Ed.). (2019). *E-Portfolio as curriculum: Models and practices for developing students' e-Portfolio literacy*. Stylus Publishing.

T

Technology Standards for Language Teachers

Deborah Healey
University of Oregon, Portland, OR, USA

Abstract

As technology and its use continuously evolve, standards to inform practice become increasingly helpful. The International Society for Technology in Education (ISTE) has developed technology standards aimed at teachers, learners, and administrators, with the latest versions in 2016, 2017, and 2018 (International Society for Technology in Education, 2016, 2017, 2018). TESOL International Association developed the *TESOL Technology Standards Framework*, with versions for learners and for teachers, in 2008. This document is now freely available for non-commercial use, with attribution (CC BY SA NC). TESOL expanded the document into a book in 2011 with performance indicators, vignettes, and can-do statements for each standard. The 2011 standards themselves have been validated by recent research (though some of the performance indicators and vignettes have become dated. More recently, separate targeted technology standards were developed for Canadian adult language learners, teachers, and programs. (All three of the standards documents include the standards, performance indicators, reflection questions, and self-assessment checklists. The instructor and program standards also include vignettes. The standards are freely available for non-commercial use, with attribution (CC BY SA NC). Teachers and administrators looking for guidance about incorporating technology appropriately and effectively would be well served by exploring the TESOL International Association and Canadian standards.

Keywords

Standards · Technology self-assessment · Teacher vignette · Open access · Administrative guidance · Adult learners

Introduction

As a field matures, it develops standards for its practitioners. The use of technology in English language teaching is now robust, and standards can set guidelines for practitioners, teacher educators, and researchers. While the technology standards developed by the International Society for Technology in Education (ISTE) are helpful, they are aimed at a general audience of educators. More focused standards are those from TESOL International Association for English language teachers and learners (TESOL, 2008) with expanded information in 2011 (Healey et al., 2011). New Language Solutions (NLS) developed

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_309-1

targeted standards aimed at Canadian adult language learners, teachers, and programs in 2023–24 (NLS, 2024a, b, c).

TESOL's Technology Standards

Goal

The goal of TESOL's technology standards was to produce a document that would help English language teachers and learners globally assess and improve what they did with technology. It would form a basis for best practice using what was locally available, recognizing that high-, mid-, and low-resource contexts existed within and across countries.

Process

A team of researchers from the US and Cyprus met virtually and face-to-face over the course of 2 years to develop first the learner standards, then the teacher standards. They reviewed the ISTE standards then available as well as the UNESCO ICT Guidelines to build the standards. Next, teachers internationally were asked to submit vignettes based on draft standards and following a vignette template. The team created the initial framework document (now freely available) and asked for comment within the TESOL community. The framework document was approved by TESOL International Association in 2008. The additional information in the published volume was informed by research and comments from forums and conferences. The team's work ended in 2011.

Components

The initial, freely available framework document (TESOL, 2008) consists of separate learner and instructor documents. Each has goals, the two to five standards for each goal, a few performance indicators, and a few vignettes from practitioners related to specific standards. The published volume from TESOL Press (Healey et al., 2011) contains the goals and standards from the framework document. It adds a full set of four to six performance indicators for each standard. The performance indicators include both a basic level and an "expert" level. Vignettes for each standard are for low-resource, mid-resource, and high-resource settings. Chapters address research; a comparison with ISTE and UNESCO standards; and recommendations for teacher educators, administrators, and online teaching and learning. Can-do statements for self-assessment by teachers, learners, administrators, and online programs are included as well. The full volume gives more depth and context to the standards.

Comments

TESOL's 2008 standards avoid mention of specific tools, giving them a longer shelf life. Researchers have indicated that the standards themselves are still valid, but the performance indicators and vignettes need updating (Sun, 2022). There is no mention of artificial intelligence, for example. However, the standards remain instructive. The vignettes give real-world examples from educators. The self-assessments via can-do statements are based on the performance indicators, but they can still be helpful with minor adjustment for teachers, administrators, online teachers and administrators, and teacher educators.

New Language Solution's Avenue Technology Standards

Goal

The Avenue technology standards are aimed at the Canadian adult settlement (immigrant and refugee) context. The goal is to provide focused instructor, learner, and program standards for technology-enhanced language learning in this context. The standards are not designed to be used for teacher assessment, but rather as guides for self-assessment and program improvement. Releasing the standards under a Creative Commons license (free for non-commercial use and revision with attribution) allows users elsewhere to adapt the standards for their specific context.

Process

The shift to online instruction during the pandemic made it clear that adult educators and

language programs in the settlement sector needed guidance for effective use of the online resources provided by NLS. NLS initially examined the ISTE and TESOL International Association standards for applicability to their context but felt that their audience would be better served by updated, focused standards. Three of the initial TESOL Technology Standards authors were recruited, plus a long-time language educator and program director in Canada. The combination brought together prior experience with technology standards development and a deep understanding of the Canadian settlement sector. The group created an initial set of standards together, then individuals took specific standards to flesh out with performance indicators and explanatory text. As a group, the team focused on making the standards cohesive, appropriate to the context, teacher-friendly, up-to-date, and clearly usable. This was a very effective process (Hubbard & Healey, 2024). The standards were shared with mentor teachers and others and have been presented in several Canadian and international venues for comments. Development of courses that incorporate the standards and pilot online programs is ongoing.

Components

The Avenue standards consist of three separate documents: Instructor, Program, and Learner Standards for Technology-Enhanced Language Learning. All three documents have a similar format. Each starts with brief notes to users. A guiding philosophy that sets the tone for the standards is next, followed by the standards. Every standard has performance indicators (PIs) that provide concrete suggestions for meeting that standard. An expanded text accompanies each PI to give context and clarity. A short self-assessment and reflection questions follow each PI. An appendix puts the self-assessments from all the PIs together for easy access. The instructor and program standards include vignettes that serve as examples of how practitioners have met some of the PIs and standards. A glossary is included with each of the three documents. The learner standards are currently addressed to instructors and administrators rather than to the learners themselves. The learner standards have reflections for administrators and teachers and suggested tasks for learners. A version of the learner standards is envisioned that would be written in language understandable to intermediate and advanced learners. The learner standards are arguably the most context-specific of the three documents.

Comments

The Avenue standards, like TESOL's standards, refer to ways of using technology and avoid naming specific tools. The three instructor, program, and learner standards documents are closely interrelated. Program administrators should understand how to utilize administrative technology well and enable teachers to employ instructional technology effectively. Instructors work with administrators to establish environments for appropriate use of technology by learners. The standards for learners expect adult learners to have agency in their technology use and, where feasible, be part of instructional and program technology decisions. This interconnectedness is an additional incentive for all technology users to come together to listen, reflect, and collaborate.

Conclusion

The past few years have seen some major shifts in ways that technology is employed in education. The use of technology in English language teaching developed explosively with the forced shift to online instruction during the COVID pandemic in 2020–2022. The advances and issues posed by artificial intelligence are changing the field in new and currently unpredictable ways. Standards can and should assist policy-making as the field continues to evolve. English language teachers and program administrators can take advantage of the freely available material from TESOL International Association and NLS to review their progress and assess their current state of technology use. Teachers, learners, and programs can all benefit from this reflection and follow-up action.

Cross-References

- ▶ Critical Approaches to CALL
- ▶ Digital Literacies
- ▶ Language Teacher Education for CALL
- ▶ Self-Assessment in Technology-Mediated Language Learning
- ▶ Teacher's Digital Competencies

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Healey, D., Hanson Smith, E., Hubbard, P., Ioannou-Georgiou, S., & Kessler, G. (2011). *TESOL technology standards: Description, implementation, integration*. TESOL Press.

Hubbard, P., & Healey, D. (2024). Developing targeted technology standards for teachers: Report on a project for the Canadian settlement sector. In *WorldCALL 2023: Conference proceedings*. IAFOR Research Archive. https://papers.iafor.org/wp-content/uploads/papers/worldcall2023/WorldCALL2023_75515.pdf

International Society for Technology in Education. (2015). *ISTE standards for students*. https://iste.org/standards/students

International Society for Technology in Education. (2016). *ISTE standards for educators.* https://iste.org/standards/educators

International Society for Technology in Education. (2017). *ISTE standards for education leaders.* https://iste.org/standards/education-leaders

New Language Solutions. (2024a). *Avenue instructor standards for technology-enhanced language learning*. https://avenue.ca/avenue-instructor-standards-for-tell/

New Language Solutions. (2024b). *Avenue learner standards for technology-enhanced language learning*. https://avenue.ca/avenue-learner-standards-for-tell/

New Language Solutions. (2024c). *Avenue program standards for technology-enhanced language learning*. https://avenue.ca/avenue-program-standards-for-tell/

Nozawa, K. (2019). Language teachers in the 21st century: Professional qualifications and challenges to implement the latest technologies. In *Proceedings of the 1st international conference on mathematics, science, language, and economics in education 2019* (pp. 1–16). UMMY.

Sun, X. (2022). Ten years later: Reexamining the TESOL technology standards for language teachers. *TESOL Journal, 13*(4), e684. https://onlinelibrary.wiley.com/doi/pdf/10.1002/tesj.684

TESOL. (2008). *TESOL technology standards framework*. TESOL Press. https://www.call-is.org/WP/wp-content/uploads/2023/06/TESOL-Technology-Standards-Framework-Open-2023.pdf

H

History of CALL Teacher Education

Philip Hubbard
Stanford University, Stanford, CA, USA

Abstract

Teacher education in computer-assisted language learning (CALL) goes back more than four decades. By tracing its developmental history, this entry serves as a foundation for others in the CALL Teacher Education section of the *Palgrave Encyclopedia of Computer-Assisted Language Learning*. It begins with reports of examples of pioneering CALL courses and professional organizational support for language teachers. It then shifts to describing how CALL moved from the periphery to being a significant element in language teacher preparation and professional development through the actions of professional organizations, journal publications and edited volumes. It next moves through a discussion of recurrent challenges for teachers and teacher educators alike, in particular inertia, ignorance and lack of time. It then discusses CALL teacher education approaches, frameworks and the role of language teacher technology standards. It concludes by acknowledging the impact of the COVID-19 crisis and the emerging area of generative AI in language teaching.

Keywords

CALL · Language teacher education · Language teacher professional development · Communities of practice · Teacher education frameworks · Teacher education models · Technology standards

Introduction

The Teacher Education section of the *Palgrave Encyclopedia of Computer-Assisted Language Learning* offers numerous entries covering the current range of issues, options and resources related to bringing CALL into the domain of language teacher education and professional development. As a foundation for interpreting all these topics, it is useful to take a look at the history of this sub-field. Formal courses devoted to helping language teachers and by extension their students use computers in all their forms—desktops, laptops, tablets, smartphones and various supporting technologies—have been offered for well over 40 years (Hubbard, 2023). Increasingly, language teacher education as a whole integrates significant elements of CALL, reflecting the move towards normalization predicated by Bax (2003). This entry reports on a number of areas connected to the history and evolution of CALL teacher education. It begins with early examples from the time when formal coursework often included learning programming skills, and only a handful of institutions had such offerings. This time

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_313-1

coincided with the era of the 'teacher-programmer', where individual language teachers chose to leverage a course or two in general computer programming or taught themselves enough independently to produce or adapt software for their students (Hubbard, 2023). It then follows the progress of CALL teacher education as it grew as a professional sub-field through the availability of early online training resources, the formation of special interest groups in CALL organizations and the increasing presence of CALL teacher education research and practice in academic journals and dedicated edited volumes. Subsequent sections address reported challenges and some proposed approaches to CALL teacher education, models and frameworks, the influence of technology standards and professional development beyond formal degree programs. The entry concludes with a note on the impact of COVID-19 and the emerging issues surrounding generative AI.

Early Development

The term *computer-assisted language learning* (CALL) came to prominence in the early 1980s as the preferred descriptor for the nascent field (Davies & Higgins, 1982). However, even before the term appeared, individuals and institutions were engaged in CALL teacher education. For example, Hubbard (2023) notes that a professor at Ohio University, Marmo Soemarmo, offered linguistics graduate students courses in computer programming specifically to support language teaching and learning as early as 1979. Beyond rare dedicated course offerings such as those, early CALL teacher education was more commonly in the form of professional development for practicing teachers. Curtin and Shinall (1985) report on one such initiative, a 2-week workshop for high school and university language teachers. Additionally, several books appeared during this period that provided gateways to CALL for adventurous language teachers and teacher educators: Kenning and Kenning (1983), Wyatt (1984) and Ahmad et al. (1985).

By 1987, the TESOL CALL Interest Section (founded in 1983) was offering informal training and mentoring to teachers who dropped into the TESOL convention's 'CALL Hospitality Room' (see https://web.archive.org/web/20231130073834/https://www.call-is.org/WP/about-call-is/call-is-community-history/community-history-1983-2003/). A few years later, the room was rebranded the TESOL Electronic Village and became one of the highlights of the annual convention for teachers eager to learn about CALL courseware, tools and practices from more informed and experienced peers. To reach a larger group, in 2001 members of the CALL-IS pioneered the Electronic Village Online (EVO). The EVO offered free 5-week online sessions on CALL-focused and other classroom-oriented topics prior to the formal TESOL conference in March: see http://evosessions.pbworks.com/w/page/41929022/EVO%202001%20CfP and more recently http://evosessions.pbworks.com/w/page/10708587/What%20is%20EVO. They continue to offer these annually, and the sessions have moved from being primarily guided discussion forums to more organized hands-on workshop-style presentations. Interested moderators initially propose topics, and if accepted, they then receive training from a team of experienced online instructors. For example, the 2011 training timeline for TESOL 2012 can be found here: http://evo-training.pbworks.com/w/file/40506342/EVO2012Training.jpg.

In terms of more formal coursework, Chapelle (2006) observes that in the 1980s it was more typical for language teacher training programs to point interested students towards general education technology courses rather than to any dedicated specifically to CALL. By the 1990s, a few institutions were offering whole degree programs devoted to CALL. Partridge (2006) describes one connected to applied linguistics at the University of Kent, which ran from 1993 to 2002. He notes that it ended when professors with the expertise to conduct the classes moved on, rather than for a lack of interest. In 1999, Graham Davies led a team that created the Information and Communication Technologies for Language Teaching (ICT4LT) website (www.ict4lt.org), consisting of

modules that language teachers could use for self-study and teacher educators could use to support CALL courses and workshops.

The status of CALL teacher education as a recognized sub-field arguably goes back at least to 2002, when *Language Learning & Technology* released a special issue on the topic. That issue included five articles as well as an introduction by the editors (Zhao & Tella, 2002). That introduction acknowledged four key areas needing further exploration: (1) teachers' cognitive and psychological processes in integrating technology; (2) effective social and organizational arrangements to promote that integration; (3) the relationship between teachers and technology, including the degree to which technology might replace teachers; and (4) the differences and similarities between online and face-to-face teaching. A further development was a volume on language teaching with technology edited by Lomicka and Cooke-Plagwitz (2004) that included four chapters related to language teacher education. Two years later, Hubbard and Levy (2006a) published the first fully dedicated edited volume on the topic, *Teacher Education in CALL*. It included 20 chapters: an introduction and 19 research reports and case studies by CALL teacher educators from around the world. The following year saw another dedicated volume, this one published by CALICO (Kassen et al., 2007), underscoring the growing importance of CALL teacher education. Amongst other contributions, it included three papers on the use of electronic portfolios for documenting teacher reflections and development. A number of books dedicated to CALL teacher education have followed (e.g., Son, 2018, and Torsani, 2016). As a testament to how widely the field has expanded, Tafazoli and Picard (2023) put together an edited volume on CALL teacher education in underrepresented contexts, featuring case studies from around the world.

Additional moves towards professionalizing CALL teacher education came when the Computer Assisted Language Instruction Consortium's (CALICO) Teacher Education Special Interest Group (SIG) formed in 2004, followed in 2009 by the European Association for Computer Assisted Language Learning (EUROCALL) Teacher Education SIG. These two groups, still thriving today, provided important communities of practice for their organizations and beyond, through the many workshops, sponsored panels and group meetings at their respective professional conferences.

Challenges

Despite the widely acknowledged need for language teachers to be competent and confident with technology (Kessler, 2018; TESOL, 2008), a number of significant challenges to achieving that end have been noted. Hubbard (2008) identified seven of these: inertia, ignorance, insufficient time, insufficient standards, insufficient infrastructure, lack of established methodology and lack of experienced CALL teacher educators. The last four points have been dealt with to some degree in the past 15 years. Technology standards have been promulgated by TESOL and other organizations, as noted in the section on standards below. Infrastructure in many settings has greatly improved as computer labs have increasingly been replaced by teachers (and in many cases students) using their own device, and for connecting online, both Wi-Fi and cellular data Internet connections are more reliable, more pervasive and faster. There are more teacher educators with experience in CALL, gained either through training or direct experience, and besides the teacher educator communities of practice like the aforementioned special interest groups of CALICO and EUROCALL, informative conference presentations, webinars and publications are widely available. These avenues provide teacher educators with a means of exchanging ideas and resources to promote their *own* professional development. There are also a number of reports of CALL-focused courses that have shown success (Son & Windeatt, 2017) and from which strategic and methodological insights can be drawn.

Nevertheless, the first three challenges remain widespread (Hubbard, 2023). Inertia is still a major barrier to overcome: established language teachers are too often satisfied teaching the way

they have in the past, and thus both they and their programs may lag far behind the pace of technology integration in the society at large. Amongst other problems, failing to keep up with technology for language teaching and for promoting learner interaction in the L2 ignores the fact that students are likely to engage with real-world language *use* through technology-mediated means: online videos, games, social media, etc. Ignorance of available options similarly remains a problem despite the fact that a skilful web search can readily lead to resources for teacher self-study. Yet of all the challenges, insufficient time (or the perception of it) seems to be the biggest block. For formal teacher education, it is finding time for CALL within an increasingly crowded curriculum. For professional development, it is acknowledging that practicing teachers are often already overburdened by their teaching demands. For this reason, various sources recommend that language programs set aside time for CALL professional development or provide teachers additional compensation to engage in it (e.g., Healey et al., 2011, 2024).

In addition to echoing the preceding challenges to achieving success employing CALL, Son (2018) adds several others, some of which can be overcome with appropriate training and institutional support. These include addressing curricular restrictions, social pressures and an often conservative academic culture. He also emphasizes the critical importance of the attitudes and skills of the teacher educators themselves in this venture.

Approaches

To prepare pre-service and in-service teachers for using CALL, teacher educators have employed a variety of approaches in their courses (Hubbard, 2008). The *breadth-first* approach reflects the traditional survey course common in other disciplines like economics or psychology, where the field as a whole is given an overview but at a broad but relatively shallow level. A number of CALL textbooks such as Arnold and Ducate (2019), Beatty (2010) and Blake and Guillen (2020) are consistent with this approach, as are a number of 'handbooks' catering to a mixed audience of CALL researchers and practitioners (e.g., Chapelle & Sauro, 2017; Farr & Murray, 2016; and Tafazoli & Picard, 2023). The assumption is that such a survey will lead language teachers to an understanding of a wide range of options so that they can make more informed decisions regarding the use of technology in their classes. This contrasts with a *depth-first* approach, where one or two single areas might be focused on deeply. Chao (2006), for example, introduces technology to language teachers through a course devoted to developing an individual WebQuest lesson for their own classes. Eskenazi and Brown (2006) built their CALL course around producing a software application using the then emerging technology of speech recognition. These examples show how depth-first approaches are consistent with project-based learning. For the third option, the *integrated* approach, CALL learning moves from a single, standalone course to being threaded through the teacher education curriculum. One intriguing model described by Hegelheimer (2006), a blend of the breadth first and integrated approaches, has students in a 2-year master's program developing technology knowledge and skills in their first term through a survey course in computers for applied linguistics. Faculty teaching subsequent classes on a variety of topics can then build on those foundational competencies shared by all students in the program. The final approach in Hubbard (2008) was labelled simply *online* but more accurately refers to the concept of learning technology through using technology. Cooke-Plagwitz (2004) sums this approach up nicely in the title of her paper, 'Using the Internet to train language teachers to use the Internet: A special topics course for teachers of German'. In an even more focused example (Kolaitis et al., 2006), a group of ESL teachers at a community college embarked on a professional development project where they took on the role of language students using software available at that time to learn a new language independently. Through an online discussion board and regular in-person meetings, they shared their experiences with one another. They

concluded that the lessons they learned using technology in the language learner role helped them train and advise their own students in strategies for using assigned tutorial programs more effectively.

In line with the final approach, a number of studies with teacher candidates have been done involving virtual exchange (or telecollaboration), where teachers at two or more institutions work together on tasks or projects (Lord & Lomicka, 2004) and/or engage in intercultural communication. Müller-Hartmann (2006) justifies the latter purpose as follows.

Because Internet-mediated intercultural exchange is a highly complex social activity, teachers cannot be adequately prepared for their crucial role in such exchanges by simply reading the relevant academic literature in a typical pre-service teacher education seminar; instead, they must build their knowledge base via experiential learning and model teaching, two practices that will afford them the opportunity to become intercultural speakers (p. 63).

This experience of learning technology through using technology as members of collaborative teams and/or learners of one another's cultures places teacher candidates in a more informed position for trying virtual exchange with their own students. An impressive example of both using virtual exchange and refining the process through 12 years of telecollaborative experiences is Sadler and Dooly (2016), who repeatedly connected students in their teacher education courses in the USA and Spain, respectively.

Teacher Technology Standards

Previous sections have focused on the history of CALL teacher education along with challenges and approaches. This section looks briefly at what the objectives of CALL teacher education should be as reflected in standards that delineate targets for teacher technological knowledge and skills. Early applications of technology standards for language teacher education, when present at all, built primarily on existing general education standards. For example, Murphy-Judy and Youngs (2006) review, 'three sets of standards for teacher education and technology on three different continents: in a US public university; at the national level in Colombia, S.A; and across the transnational Common European Framework' (p. 46). The first two of these involve general technology standards not specific to language teaching. The third builds on the Common European Framework (CEF) standards for languages, which advocates the increasing use of appropriate technology in language teaching, but as the authors note, without providing any specific technology standards.

The first set of technology standards aimed exclusively at foreign and second language teaching was the *TESOL Technology Standards Framework* (TESOL, 2008). This framework covered competency targets for both language teachers and language learners. For teachers, there were 14 standards divided across 4 overarching goals: acquiring and maintaining foundational technology knowledge and skills; effectively integrating pedagogical knowledge and skills with technology to improve language learning, applying technology appropriately in record-keeping, feedback and assessment and using technology to improve communication, collaboration, and efficiency. Each standard was subdivided into several more explicit performance indicators. In the follow-up volume (Healey et al., 2011), there was a separate chapter aimed at teacher educators, providing guidance on how to integrate the standards into CALL courses. Unfortunately, TESOL did little to promote these technology standards and nothing to keep them up to date. Nevertheless, they have informed a number of CALL courses. Tschichold (2016), for example, offers details of how she integrated the TESOL standards into her CALL class.

In line with growing calls for updated and more relevant language teacher technology standards by, e.g., Nozawa (2019) and Sun (2022), the Canadian organization New Language Solutions embarked on a project in early 2023. This was largely in response to the lack of language teacher preparedness for teaching online manifested by the COVID crisis. The goal was to develop a targeted set of technology enhanced language

learning (TELL) standards for teachers, learners and programs in the settlement language sector, especially in online or blended settings (Allan et al., 2024). Heavily based on the original TESOL Technology Standards framework (TESOL, 2008), these nevertheless differ in three significant respects. First, they are updated (TESOL's last publication of its framework was Healey et al., 2011), reflecting the technological and pedagogical shifts that have occurred. In particular, they address fully online or blended teaching situations, supported by learning management systems, and they reference competency targets for using generative AI. Second, unlike the broad, decontextualized standards from TESOL, ISTE, UNESCO and other groups, they are highly targeted, created specifically for teachers working with the adult newcomer population in Canada. Thus, they are not intended mainly for pre-service teachers but rather for those who may have been teaching for some time and need to be brought up to speed quickly. Third, because of being more targeted, they are also more compact: there are only 7 teacher standards compared to TESOL's 14. Finally, the overall set includes TELL program standards, directing administrators and staff to provide sufficient support not only for the mission of teaching and learning but also for teacher professional development.

Frameworks and Models

To bring some structure into CALL teacher education and professional development, a variety of frameworks and models other than technology standards have been proposed. Son (2002) introduced a conceptual model for classroom teachers using CALL based on the range of roles they play. Though couched more in terms of professional development than formal teacher education, Son identified several new roles in addition to the traditional classroom ones of tutor, guide and facilitator: CALL observer, designer, implementer, evaluator and manager.

A framework by Hubbard and Levy (2006b) leveraged sociological role theory to define two types of roles CALL roles: institutional and functional. Institutional roles were pre-service teacher, in-service teacher, CALL specialist and CALL professional. Pre-service teachers rely on formal CALL training; in-service teachers engage in professional development initiatives. CALL specialists are of two types, experts and adjuncts. Taking pronunciation as an example, a CALL *expert* pronunciation specialist would have deep knowledge of pronunciation and pronunciation pedagogy as well as an extensive technical skill set. An adjunct in contrast could be either a general pronunciation expert with limited specialized technical skills or a technical expert with limited knowledge of pronunciation pedagogy. See Hubbard (2009) and Torsani (2016) for additional guidance for educating CALL specialists. Finally, a CALL professional would be someone with broad knowledge of the CALL field and deeper knowledge and/or skills in one or more specialist areas—basically someone whose professional career is closely connected to CALL. Turning to functional roles within the institutional ones, Hubbard and Levy identified practitioner, developer, researcher and trainer. Someone holding an institutional role like in-service language teacher might engage in one or all of the functional roles for different purposes, for example, taking the role of trainer to show students how to effectively engage in an online collaborative task.

Torsani (2016) presents a comprehensive review and analysis of the history and present state of the field of CALL teacher education, culminating in a proposal for a 'theory of CALL curriculum'. The theory includes (1) needs analysis and goals, (2) content and (3) processes and evaluation. She adapts Hubbard and Levy's (2006b) role-based framework for needs analysis and goals. Reinders (2009) and Hegelheimer (2006) amongst others are drawn on for content. For processes, she mentions a number of sources but all revolve strongly around experiential learning. Evaluation of a CALL course or whole curriculum is presented as the main challenge, with self-evaluation using the TESOL Technology Standards (2008) given as one option. Torsani follows the presentation of her theory with eight scenarios, each outlining a syllabus for a CALL course within a specific curriculum. Some serve a

broad objective (e.g., a CALL survey course for pre-service teachers/practitioners), whilst others serve a more targeted one (e.g., a course for CALL specialists in courseware design for vocabulary).

In his monograph on technology and teacher education, Son (2018) proposes a CALL teacher education model called 'ECCR' for its four interacting components: exploration, communication, collaboration and reflection. Of particular importance is the exploration part, where teachers introduced to a relevant tool are placed in the role of learners and experience the technology hands on. Stages involving communication and collaboration strengthen and refine the exploratory phase, and a final reflection completes the ECCR cycle, leading to another round of exploration.

A final influential framework in CALL teacher education is that of technological pedagogical content knowledge (TPACK) (Koehler & Mishra, 2009). The TPACK framework applied to the field of language teaching builds on the intersection of content knowledge (knowledge of the target language), pedagogical knowledge (knowledge of teaching) and technological knowledge (knowledge of hardware, software and systems). These connect as knowledge sets first in pairs (e.g., pedagogical content knowledge is knowledge of the target language coupled with how to teach it) until the final TPACK goal is reached: knowledge of how to teach the target language using technology. The model has been used in a number of articles for CALL teacher education. For instance, in a special issue of *Language Learning & Technology* on teacher education, Liu and Kleinsasser (2015) reported on a TPACK-based professional development in Taiwan. Tai (2015) showed how designing TPACK into a workshop for in-service teachers in Taiwan led them to integrate workshop content into their individual teaching contexts. In a critical review of 51 studies of TPACK in language teacher education, Tseng et al. (2022) discovered that the majority simply surveyed teachers' TPACK knowledge base, and only a handful involved research on TPACK in practice. Nevertheless, TPACK remains a viable option as a model for CALL teacher education because of its premise that technological pedagogical content knowledge for language teaching is discipline specific.

COVID, AI, and the Future

The COVID-19 crisis with its forced move to emergency remote teaching represented a wakeup CALL not just for language teachers and programs but also for language teacher education programs and the faculty who implement them. In a review of papers published about the response by language teachers and students, Moorhouse and Kohnke (2021) identify the following themes: (1) the potential and affordances of online teaching with technology as the primary mediator; (2) the challenges, diversity and complexity of teaching online and (3) the continuing presence of the digital divide. They conclude 'The effects of the pandemic have been dramatic and are potentially long-lasting. The widespread use of technology for teaching and learning could fundamentally change the educational landscape' (p. 375).

Not long after the COVID-19 crisis subsided, on November 30, 2022, OpenAI released ChatGPT to the public. It is not yet known *how* this and other generative AI presences will impact CALL teacher education, but the impact will be profound. Already, there are multiple AI competency frameworks from organizations such as UNESCO, ISTE and Educause, and these need to be interpreted for language teaching. Technology standards such as those in Healey et al. (2024) are calling on language programs to develop clear policies and importantly, to support teacher professional development in AI. Professional organizations are also stepping up: the previously mentioned TESOL EVO for 2025 listing shows a typical range of interesting topics, and half of them are focused directly on helping language teachers with AI integration: https://sites.google.com/view/evosessions2025/evo25-sessions.

For teacher educators, whether in formal courses and degree programs or engaged in workshops and webinars for professional development, the world has changed dramatically in the past few years. The future history of language teacher

education over the next few decades will almost certainly include how technology—and especially AI—became fully integrated into both the methodologies of language teaching and learning and the teacher education process itself. Language teacher educators will be at the forefront of these developments.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Ahmad, K., Corbett, G., Rogers, M., & Sussex, R. (1985). *Computers, language learning and language teaching* (Vol. 1). Cambridge University Press.

Allan, J., Healey, D., Hubbard, P., Kesler, G., McBride, R., Rajabi, S., & Sturm, M. (2024). Developing targeted technology standards for avenue language instructors, programs, and learners: An initiative of new language solutions. *TESL Ontario Contact, 50*(Spring), 16–30.

Arnold, N., & Ducate, L. (Eds.). (2019). *Engaging language learners through CALL: From theory and research to informed practice*. Equinox.

Bax, S. (2003). CALL – past, present and future. *System, 31*(1), 13–28.

Beatty, K. (2010). *Teaching & researching: Computer-assisted language learning*. Routledge.

Blake, R., & Guillen, G. (2020). *Brave new digital classroom* (3rd ed.). Georgetown University Press.

Chao, C. C. (2006). How WebQuests send technology to the background: Scaffolding EFL teacher professional development in CALL. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 221–234). John Benjamins.

Chapelle, C. (2006). Foreword. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. vi–vii). John Benjamins.

Chapelle, C. A., & Sauro, S. (Eds.). (2017). *The handbook of technology and second language teaching and learning*. Wiley.

Cooke-Plagwitz, J. (2004). Using the internet to train language teachers to use the internet: A special topics course for teachers of German. In L. Lomicka & J. Cooke-Plagwitz (Eds.), *Teaching with technology* (pp. 65–71). Heinle.

Curtin, C., & Shinall, S. (1985). A CAI workshop from expectation to evaluation. *CALICO Journal, 3*(2), 27.

Davies, G., & Higgins, J. (1982). *Computers, language and language teaching*. Centre for Information on Language Teaching.

Eskenazi, M., & Brown, J. (2006). Teaching the creation of software that uses speech recognition. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 135–151). John Benjamins.

Farr, F., & Murray, L. (Eds.). (2016). *The Routledge handbook of language learning and technology.* Routledge.

Healey, D., Smith, E. H., Hubbard, P., Ioannou-Georgiou, S., Kessler, G., & Ware, P. (2011). *TESOL technology standards: Description, implementation, integration*. TESOL.

Healey, D., Hubbard, P., Kessler, G., & Rajabi, S. (2024). *Avenue program standards for technology-enhanced language learning*. New LanguageSolutions. https://avenue.ca/avenue-program-standards-for-tell/

Hegelheimer, V. (2006). When the technology course is required. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 117–133). John Benjamins.

Hubbard, P. (2008). CALL and the future of language teacher education. *CALICO Journal, 25*(2), 175–188.

Hubbard, P. (2009). Educating the CALL specialist. *International Journal of Innovation in Language Learning and Teaching, 3*(1), 3–15.

Hubbard, P. (2023). Contextualizing and adapting teacher education and professional development. In D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development: Voices from under-represented contexts* (pp. 3–14). Springer Nature Singapore.

Hubbard, P., & Levy, M. (Eds.). (2006a). *Teacher education in CALL*. John Benjamins.

Hubbard, P., & Levy, M. (2006b). The scope of CALL education. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 3–20). John Benjamins.

Kassen, M., Lavine, R., Murphy-Judy, K., & Peters, M. (Eds.). (2007). *Preparing and developing technology-proficient L2 teachers*. CALICO.

Kenning, M. J., & Kenning, M. M. (1983). *Introduction to computer assisted language teaching*. Oxford University Press.

Kessler, G. (2018). Technology and the future of language teaching. *Foreign Language Annals, 51*(1), 205–218.

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.

Kolaitis, M., Mahoney, M. A., Pomann, H., & Hubbard, P. (2006). Training ourselves to train our students for CALL. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 317–332). John Benjamins.

Liu, M. H., & Kleinsasser, R. C. (2015). Exploring EFL teachers' CALL knowledge and competencies: In-service program perspectives. *Language Learning & Technology, 19*(1), 119–138.

Lomicka, L., & Cooke-Plagwitz, J. (Eds.). (2004). *Teaching with technology.* Heinle.

Lord, G., & Lomicka, L. L. (2004). Developing collaborative cyber communities to prepare tomorrow's teachers. *Foreign Language Annals, 37*(3), 401–408.

Moorhouse, B. L., & Kohnke, L. (2021). Responses of the English-language-teaching community to the COVID-19 pandemic. *RELC Journal, 52*(3), 359–378.

Müller-Hartmann, A. (2006). Learning how to teach intercultural communicative competence via telecollaboration: A model for language teacher education. In J. Belz & S. Thorne (Eds.), *Internet-mediated intercultural foreign language education* (pp. 63–84). Heinle.

Murphy-Judy, K., & Youngs, B. L. (2006). Technology standards for teacher education, credentialing, and certification. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 45–60). John Benjamins.

Nozawa, K. (2019). Language teachers in the 21st century: Professional qualifications and challenges to implement the latest technologies. In *International conference on mathematics, science, language, and economics in education* (pp. 1–16). Solok Indonesia.

Partridge, J. (2006). Matching language and IT skills: The life-cycle of an MA Programme. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 63–79). John Benjamins.

Reinders, H. (2009). Technology and second language teacher education. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to second language teacher education* (pp. 230–237). Cambridge University Press.

Sadler, R., & Dooly, M. (2016). Twelve years of telecollaboration: What we have learnt. *ELT Journal, 70*(4), 401–413.

Son, J. B. (2002). Computers, learners and teachers: Teamwork in the CALL classroom. *English Language Teaching, 14*(2), 239–252.

Son, J. B. (2018). *Teacher development in technology-enhanced language teaching*. Springer.

Son, J. B., & Windeatt, S. (Eds.). (2017). *Language teacher education and technology: Approaches and practices*. Bloomsbury Publishing.

Sun, X. (2022). Ten years later: Reexamining the TESOL technology standards for language teachers. *TESOL Journal, 13*(4), e684.

Tafazoli, D., & Picard, M. (2023). *Handbook of CALL teacher education and professional development: Voices from underrepresented contexts*. Springer.

Tai, S.-J. D. (2015). From TPACK-in-action workshops to classrooms: CALL competency developed and integrated. *Language Learning & Technology, 19*(1), 139–164.

TESOL. (2008). *TESOL technology standards framework*. TESOL Press. https://www.call-is.org/WP/wp-content/uploads/2023/06/TESOL-Technology-Standards-Framework-Open-2023.pdf

Torsani, S. (2016). *CALL teacher education: Language teachers and technology integration*. Springer.

Tschichold, C. (2016). Meeting the technology standards for language teachers. In S. Papadima-Sophocleous, L. Bradley, & S. Thouesny (Eds.), *CALL communities and culture – short papers from EUROCALL* (pp. 445–449). Research-publishing.net.

Tseng, J. J., Chai, C. S., Tan, L., & Park, M. (2022). A critical review of research on technological pedagogical and content knowledge (TPACK) in language teaching. *Computer Assisted Language Learning, 35*(4), 948–971.

Wyatt, D. H. (1984). *Computers and ESL*. Harcourt Brace Jovanovich.

Zhao, Y., & Tella, S. (2002). Technology and teacher education [Special issue introduction]. *Language Learning & Technology, 6*(3), 2–5.

O

Online Continuing Professional Development

Gary Motteram
University of Manchester, Manchester, UK

Abstract

This entry explores online continuing professional development for CALL showing this has been a topic of interest in the field for some time. While complete courses that focus on this topic tend still to be taught face-to-face, there have been other more informal developments in the field of online teaching and learning. Many online language teacher education courses do include a specific module in CALL, and these continue to be researched. There is also the growing interest in the use of teacher online communities that make use of various social media platforms to support professional development, which are currently under-researched.

Keywords

Continuing professional development (CPD) · Online · TESOL technology standards · Course design · Online teacher communities

Introduction

Continuing professional development (CPD) is a term that has come to prominence since the 1980s, although the concept of professionals seeking further training or professional growth has been around since the early twentieth century. It is thought to have originated in the UK and adopted as a standard term by the UK General Medical Council in the mid-1990s. Education more broadly has also made use of the term for some time and other terms meaning some similar, like In-Service Training (INSET) are also used to describe teacher development.

In the field of language teaching, particularly English Language Teaching (ELT), teacher development more recently has focused on personal professional development, so not development where teachers are trained, or educated by others, but where the development is led by the individual teacher and much effort has been made to promote this distinction. The International Association of Teachers of English as a Foreign Language (IATEFL) has two separate Special Interest Groups (SIGs): Teacher Training and Education (TTEd) SIG and the Teacher Development (TD) SIG and engage in different forms of activity. This article is not going to make this binary distinction, however, and will explore Training, Education and Development.

The second aspect of this entry is the online element, and this article will explore how access to training and development has changed in recent years with the easier access to a variety of online tools. It will also speculate on the impact of the COVID-19 pandemic in accelerating the growth of online professional development. It will look at

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_320-1

how CALL teacher education has developed and in what ways this has also moved online.

Definitions and Developments in Continuing Professional Development and Online Learning

Although many teachers in the world receive initial training, a considerable number do not, and even teachers who are trained can struggle with the realities of the day-to-day teaching world. They quickly discover they need to revisit this initial training to further develop skills learned during pre-service courses, or work on developing new ones that were not covered.

Professional development has been defined by Day (1999) as follows:

> Professional development consists of all natural learning experiences and those conscious and planned activities which are intended to be of direct or indirect benefit to the individual, group or school, which contribute through these, to the quality of education in the classroom. It is the process by which, alone and with others, teachers review, renew and extend their commitment as change agents to the moral purposes of teaching; and by which they acquire and develop critically the knowledge, skills and emotional intelligence essential to good professional thinking, planning and practice with children, young people and colleagues throughout each phase of their teaching lives. (p. 4).

This definition makes clear that CPD can consist of both regular, planned activities, as well as the process of teaching itself, which when viewed critically, supports the development process throughout their careers as teachers, whether they work with children or adults. Teachers are encouraged more and more to explore their own classrooms, engaging in different forms of action research and much recent research has focused on this.

When it comes to professional development for CALL itself, we are talking about adult learners and while a few pre-service courses will introduce the skills necessary to work effectively with technology to support language development, most will not, and so for many this is a part of training that occurs after the initial courses are completed. This may take the form of short training courses, or may be a part of further professional courses like an MA TESOL, for example. It may be that many pre-service courses offer basic, generic digital literacy, but they may be less likely to focus directly on CALL specifically.

The literature on courses that are suitable for teachers has continued to be mostly focused on face-to-face opportunities and most focus on postgraduate courses. The topic of teacher education for CALL has been written about for some time, with, for example, the edited collection by Hubbard and Levy in 2006 being a useful starting point. That volume did include one article concerned with online training, Bauer-Ramazani (2006). A more recent publication, Son and Windeatt (2017) also contained an article on teacher training with CALL online by the same author.

If we search for 'online' 'teacher education/training' and 'CALL' or 'computer-assisted language learning', the search results mostly focus on teaching languages online, not on teacher development. One recent article that does explore this topic is Karatay and Hegelheimer (2021) who focus on CALL training in low-resource environments. There are also articles that explore the interaction between members of online communities, but the topics of discussion in these communities do not focus on CALL. A typical example of this would be Satar and Akcan (2018) whose article analyses the interaction that the students who took part in the course engaged in. While the course they were studying was not specifically focused on CALL, small elements of the course focused on the topic.

Bauer-Ramazani (2006) discusses the initial creation of her online course, essentially moving an existing face-to-face course online to meet the needs of students who were working abroad, taking course materials that were already available online and adding interactivity to try to match the face-to-face and online learning experiences.

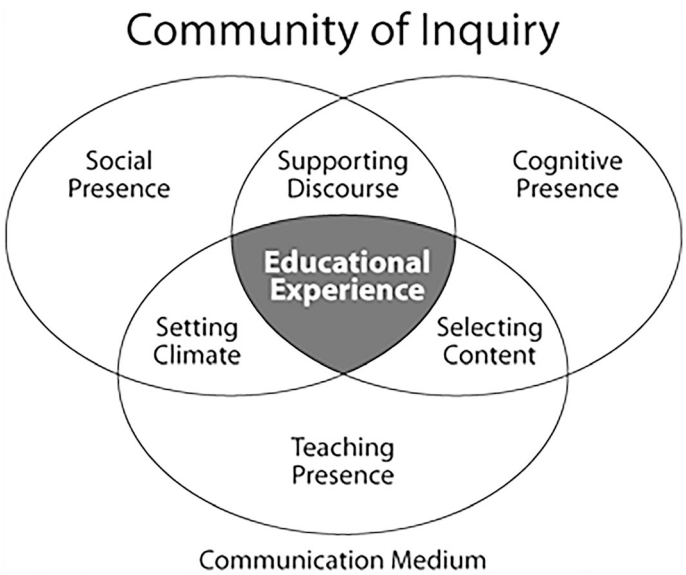
- "providing a teacher training experience through online CALL that results in active learning as demonstrated by high-quality outcomes

- finding ways to create an effective, involved, enthusiastic learning community that transcends the boundaries and limitations of the online environment while also taking advantage of the tools and resources available through online delivery
- modelling effective incorporation of technology into the classroom through hands on, task- and project-based experiences that follow a constructivist, inquiry-based approach (Butler-Pascoe and Wiburg, 2003)." Bauer-Ramazani (2006, p. 4).

The presentation of the course is framed by the TESOL technology standards (TESOL, 2008) that were being implemented throughout the US-focused CALL training community at the time, but which essentially match with the more general model of online training that was created by Garrison et al. (2000) and also found here: https://coi.athabascau.ca/coi-model/ (Fig. 1).

In this model, a course consists of content, interaction and engagement with both tutors and the learning community. The engagement with the content—cognitive presence—the supporting discourse surrounding the material—social presence—and the role members of the community play in managing the process and in selecting content—teaching presence—are also vital. Even in 2006, Bauer-Ramazani recognised that synchronous communication was an important part of the required ecology, supporting the development of the community. In a chapter focusing on an updated version of the course in 2017, Bauer-Ramazani suggests that the overall aims of the course had not changed, although technologies for delivery and the tools being explored had changed. Even though synchronous technologies had become more sophisticated by 2017, and *Skype* (rather than the currently ubiquitous Zoom) is mentioned as a tool for online meetings, it was still not easy to deploy them because of the distributed nature of the student community. Keeping the materials up-to-date with the constant changes in the field and the large quantity of new materials produced by the CALL community was also challenging. This article, of course, pre-dates the pandemic when so many more people who had never been involved in online teaching discovered the value of such online practices for education. Although the pandemic had a significant impact at the time, it does not seem to have moved the dial in terms of the further development of the formal, online training of CALL.

In their article Karatay and Hegelheimer (2021) explore teacher development in low-resource contexts, but refer back to Hubbard and



Online Continuing Professional Development, Fig. 1 Garrison, Anderson & Archer's community of inquiry model

Online Continuing Professional Development, Table 1 Characteristics of settings (Karatay & Hegelheimer, 2021, p. 273)

Characteristics	Settings
Low resource, low access setting	Computer, no projector or internet Internet outside of the classroom
Mid-resource setting, mid-access setting	Computer lab with low bandwidth internet Computer and projector in the classroom
High resource, high access	Lab available, high-speed internet, computers in the classroom

Levy (2008) for the foundations of the ideas that generated the course that they created:

> . . .both language teachers in training and practicing teachers will find themselves at a disadvantage if they are not adequately proficient in computer assisted language learning. . . (p. 272).

They also refer to the technology standards (TESOL, 2008) when they describe the different low-resource contexts they are referring to (Table 1).

It could be argued that by 2021, these standards needed some re-thinking with many contexts, particularly low-resource ones making using of mobile devices, principally mobile phones, which generally have good internet connectivity (Motteram, 2024a, b). This has led to the development of many online communities which support CPD. Articles that discuss CALL in these contexts are not readily available, but we see many recordings of online sessions that cover CALL-related topics, offered for example by organisations like the British Council, e.g. Technology supported language learning (https://africa.teachingenglish.org.uk/practical-introduction-technology-supported-language-learning), or the Flipped Classroom (https://africa.teachingenglish.org.uk/skills/webinars/flipped-learning-benefits-and-effects-teaching-and-learning-efl-classroom). The presentations of these topics are usually supported by a teacher professional development community. Publishers' websites also include large quantities of such material.

Conclusions

While online courses for language teachers have grown, this has not been the case for courses that specifically focus on CALL. Many teacher education programmes around the world would include a module that focuses on technology-supported language learning and Son and Windeatt (2017) includes several examples. We also see the growth of more self-initiated teacher development online making use of social media. Most of the more formal courses follow standard guidelines for creation and delivery contained in such guidance from, e.g. Garrison et al. (2000). Clearly this is an area for future research, particularly the exploration of more informal groups.

Cross-References

- ▶ CALL teacher communities of practice
- ▶ History of CALL teacher education
- ▶ Preparation and Professional Development of Teacher Educators

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Bauer-Ramazani, C. (2006). Training CALL teachers online. In P. Hubbard & M. Levy (Eds.), *Teacher education in CALL* (pp. 183–200). https://doi.org/10.1075/lllt.14.16bau

Bauer-Ramazani, C. (2017). Teacher training with CALL online (distance): A project- and standards-based approach. In J.-B. Son & S. Windeatt (Eds.), *Language teacher education and technology: Approaches and practices*. Bloomsbury. (Chap. 9).

Butler-Pascoe, M. E., & Wiburg, K. (2003). *Technology and teaching English language learners*. Pearson Education.

Day, C. (1999). *Developing teachers: The challenges of lifelong learning*. Falmer.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education model. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Hubbard, P., & Levy, M. (Eds.). (2008). Teacher education in CALL. John Benjamins Publishing Company.

Karatay, Y., & Hegelheimer, V. (2021). CALL teacher training—Considerations for low-resource environments: Overview of CALL teacher training. *CALICO Journal, 38*(3), 271–295. https://files.eric.ed.gov/fulltext/EJ1324755.pdf

Motteram, G. (2024a). *Creating and running successful online groups for teacher development*. British Council. https://doi.org/10.57884/44ZX-3715

Motteram, G. (2024b). *Improving teacher development through the effective use of social media groups*. British Council. https://doi.org/10.57884/GFVF-W338

Satar, H. M., & Akcan, S. (2018). Pre-service EFL teachers' online participation, interaction, and social presence. *Language Learning & Technology, 22*(1), 157–183. https://doi.org/10125/44586

Son, J.-B., & Windeatt, S. (Eds.). (2017). *Language teacher education and technology: Approaches and practices*. Bloomsbury.

TESOL. (2008). *Technology standards framework document*. TESOL.

F

Future Challenges and Opportunities for Data-Driven Learning

Pascual Pérez-Paredes and
Carlos Ordoñana-Guillamón
Universidad de Murcia, Murcia, Spain

Abstract

This entry examines the evolving landscape of data-driven learning (DDL) in language education, addressing both opportunities and challenges. As DDL expands beyond higher education, issues such as the lack of mobile-ready tools, teacher training, and accessibility to diverse corpora are highlighted. Methodological concerns in research are also discussed. Opportunities arise from the integration of artificial intelligence (AI), multimodal tools, and personalized learning, which can enhance DDL's impact across diverse learning contexts. The entry emphasizes the need for collaboration between DDL, computer-assisted language learning (CALL) and AI researchers to optimize future language education practices.

Keywords

Corpora · Corpus linguistics · Corpus-based instruction · DDL · Language teaching

Introduction

Data-driven learning (DDL) can be conceptualized as the use of corpus tools and methods for either learning or using a second language (L2) (Boulton & Pérez-Paredes, 2024). Its future presents both exciting opportunities and significant challenges, particularly as the field evolves beyond highly structured settings like universities and well-known language data sources such as corpora. Expanding DDL beyond higher education environments could contribute to transforming how languages are taught and learned, making language learning more accessible for diverse contexts while addressing different learner needs. However, this broader application to wider learning contexts brings several challenges, especially in terms of technology adaptation and adoption, learner autonomy, and the integration of new types of language data against the backdrop of the unrelenting emergence of artificial intelligence (AI). This entry identifies challenges and opportunities for DDL in language education. In doing so, we set out to suggest areas of both educational application and research that may increase the impact of DDL, providing suggestions for future research and development.

Challenges

DDL and *challenges* have gone hand in hand since Johns (2002) noted that for the practising

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025
L. McCallum, D. Tafazoli (eds.), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*,
https://doi.org/10.1007/978-3-031-51447-0_350-1

teacher, the direct use of concordances in language teaching poses a number of challenges. (A concordance is a line of text from a corpus that shows the occurrence of word or phrase in the middle.) He described them in the abstract of the chapter as technical, linguistic, logistic, pedagogical and philosophical. Two decades later, it is worth exploring whether these challenges still hold on and how they have evolved.

There is a relatively high degree of consensus in recent DDL meta-analyses and systematic reviews about the main challenges for the field. While some of the challenges identified in DDL are essentially linked to the implementation of DDL in classrooms, others offer critical perspectives on how DDL research should drive the field into new directions. In the following paragraphs, we will discuss the two types of challenges in more detail. In our discussion, we largely draw on recent synthetic DDL research that offers a unique opportunity to analyse whether these challenges have changed over the last decade or so (Boulton & Cobb, 2017; Boulton & Vyatkina, 2021; Boulton & Pérez-Paredes, 2024; Crosthwaite & Boulton, 2024; Pérez-Paredes, 2022).

Institutional Challenges

Anthony (2024) has noted that some of these challenges are tool-driven and affect how the data is presented to the learners, how usable the software or application for hands-on DDL is, and how the learners find relevant texts and upload them to the application. (Hands-on DDL involves learners using corpus software for language learning purposes.) Apart from concordance software-specific issues, the most widely voiced challenge is that DDL is predominantly happening in Higher Education (Chambers, 2019). Historically, DDL has been largely confined to formal learning environments, where language learners work with corpus data under the guidance of instructors. This context has proven effective for fostering awareness of language patterns, allowing learners to investigate real examples of usage and patterns directly from corpora. 'Prototypical DDL' (Boulton & Pérez-Paredes, 2024) or 'traditional DDL' (Anthony, 2024) uses different types of corpora (see the entry on ▶ 'Types of Corpora in DDL') and has mainly taken place in university classrooms, where access to resources and teacher guidance is more feasible. The goals of prototypical DDL are often focused on lexicogrammar, utilizing concordances, clusters, and collocates in context. Most DDL experiences with university students involve direct interaction with concordance tools, occasionally starting with paper-based resources, and frequently employing reference corpora, such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) (Boulton & Vyatkina, 2021).

Although there have been significant efforts in widening the scope of DDL for groups of secondary school learners (e.g. Braun, 2007; Crosthwaite, 2019) and secondary school teachers (e.g. Crosthwaite et al., 2023), DDL practice remains well established in university education, more specifically in learning-to-write instruction (Pérez-Paredes, 2022). University language education is just one of the different types of formal scenarios where language learning takes place. Based on recent statistics by the European Commission (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language_learning_statistics and https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20230926-1), most language learning in Europe alone takes place in primary and secondary education settings, with universities playing a comparatively smaller role. In 2021, approximately 86.3% of pupils in primary education, 98.5% of students in lower secondary education, and 91.0% of students in upper secondary education in the European Union were learning at least one foreign language. In contrast, the number of students studying languages in higher education is significantly smaller, and university language education is primarily focused on specialized or advanced language studies. Most of the language acquisition in Europe is mandated at earlier educational stages, making university learners a much smaller segment of the language-learning population (see Pérez-Paredes, 2010).

Methodological Challenges

Although Boulton and Cobb's (2017) metanalysis was not explicitly designed to identify challenges in

DDL research, they suggest the presence of important methodological problems in the field, pointing to the need to improve the validity of DDL research. One of them is the lack of DDL researchers' enhanced criticality when discussing the affordances of the approach, often 'cast from a positive or even enthusiastic perspective' (p. 351). Other challenges include the low number of subjects in the samples analysed—only a third of the studies use >50 individuals; the over-representation of L1 Chinese learners, as well as a lack of research outside classrooms and labs. Their results also point to a lack of both longitudinal studies and studies that, although longitudinal in nature, go over one or two terms. For Boulton and Vyatkina (2021) and Crosthwaite and Boulton (2024), the methodological challenges include the lack of engagement with Second Language Acquisition theoretical frameworks, a more varied sample of language learners representative of the contexts where language learning occurs and an overwhelming focus on short-term learning outcomes, such as vocabulary acquisition and error correction, which works against the development of critical thinking and learner autonomy. This suggests that the field has not made much progress in how research has been carried out since Boulton and Cobb (2017). These challenges will be hopefully widely addressed in future research, allowing for a richer discussion about the impact of DDL and corpora on language learning.

Implementation Challenges

A third group of challenges include the lack of appropriate teaching resources, training and support for language teachers and the integration of DDL in the curriculum (Pérez-Paredes, 2022). Some of these classroom-oriented challenges have been widely identified in the specialized literature (Chambers, 2007, 2019; Pérez-Paredes, 2010, 2022). While technical challenges concerning the logistics of using corpora and computers in classrooms do not seem to be a threat to the normalization of DDL (Pérez-Paredes, 2022), the use of DDL on mobile devices (see entry on ▶ 'Mobile Assisted Language Learning (MALL)' and ▶ 'MALL and Data-Driven Learning') is not a reality yet. Most desktop software for DDL such as AntConc (Anthony, 2024) is not available on tablets or mobile devices, and online services such as Sketch Engine (https://www.sketchengine.eu) are not free and available to all learners. SKELL (https://skell.sketchengine.eu) or CorpusMate (Crosthwaite & Baisa, 2024) are free to use, but it is uncertain how useful these generic corpora and traditional concordancing can be for all learners.

Another challenge is the lack of freely available, downloadable corpora for use in the language classroom. Although there is a multitude of resources available, the collection and use of texts for further use in the classroom presents not only technical but primarily ethical challenges related to data processing and use in a legally complex international context. The development of tools that are flexible, mobile, easy to use and offer guidance and support to both language teachers and learners is possibly one of the most important challenges we face in the field. Another important challenge is to increase the representativeness of languages other than English. Recent contributions to Italian (Forti, 2023; Forti & Spina, 2019), German (Vyatkina, 2024), as well as Spanish and French (Pérez-Paredes & Chambers, in press) are welcome but clearly insufficient. We need a DDL that represents the richness and complexity of human communication more reliably. It is also urgent that DDL applications become more inclusive and begin to address software usability issues for learners with both cognitive and sensory difficulties.

Opportunities

The future of DDL in language education points towards a multimodal, adaptive, individualized, multi-corpus, and mobile-ready approach (Boulton & Pérez-Paredes, 2024). This vision reflects ongoing shifts in education and technology (Godwin-Jones, 2021, 2023), aiming to make language learning more flexible and responsive to learners' needs.

Evolving Data-Driven Learning

A multimodal DDL would involve the integration of various types of media beyond text-only input,

which can enhance students' engagement and accommodate diverse learning aims. The use of YouTube videos for DDL, for example, is a promising avenue for researchers. Lin (in press) has developed an online tool that extracts formulaic language from a database containing information on the frequency and semantic transparency of 40,000 English expressions. The tool allows learners to replay scenes to observe context, gestures, and tone of voice linked with these expressions, thereby enhancing comprehension and retention. This serves as an excellent example of how corpus linguistics and multimodal online resources can be integrated to enrich teaching practices, lowering the complexity in using the tools involved.

While the tools for analysing language data can still be daunting for many learners, particularly those without a background in linguistics or computational skills, simplifying these tools or creating more user-friendly interfaces could help make DDL more accessible to a broader audience. Due to the constant gripes among academics with the limited uptake of DDL practices across all educational levels, some researchers have attempted to revise some of the core beliefs in prototypical DDL. Pérez-Paredes (2010) suggests that DDL should focus on the learning affordances of corpus tools and data and less on the training of language learners to use corpus research methods. Meunier (2022) suggests that DDL should test the boundaries of its own potential by experimenting with other vehicles of implementation that remain largely unexplored, such as taking advantage of accessible, Natural Language Processing (NLP)-based software available on the Internet, exploring multimodal input for DDL practices, or even tapping the sheer potential of using already existing virtual environments in the form of videogames to foster linguistic interaction with authentic language in context.

A promising development in the field is the emergence of 'Broad Data-Driven Learning' (BDDL) (Pérez-Paredes, 2024). BDDL encompasses a wider variety of digital resources, such as online dictionaries, NLP tools, and language-learning applications. Unlike prototypical DDL, BDDL is not strictly tied to formal education contexts. It also includes informal and self-directed learning opportunities, which have become increasingly important with the rise of digital technologies. Learners can now access language data on their mobile phones, use apps for translation, or consult online vocabulary tools, making language learning more personalized and integrated with everyday activities. Personalization implies that the learning environment can adjust dynamically to the user, offering tailored experiences based on individual progress and preferences not necessarily guided by a teacher or in the context of a syllabus. DDL has much to offer to MALL (Pérez-Paredes et al., 2019; entry on ▶ 'MALL and Data-Driven Learning'), but DDL practitioners need to open up to collaboration with CALL researchers to test the feasibility of integrating, among others, frequency information, visual representation of lexical relations, collocation behaviour, genre-specific patterning or salient concordances into language learning apps. So far, the best-selling language learning apps seem to ignore the type of approach fostered in DDL (Li, 2024).

Data-Driven Learning and AI

The widespread impact of Large Language Models (LLMs) on language education (Han, 2024; Kern, 2024; Tolstykh & Oshchepkova, 2024) provides an unprecedented territory for applied linguists to explore and potentially challenge established notions on the naturalness of language use and what constitutes authentic linguistic production. The capacity of dialogic AI to respond to users' questions is certainly astonishing, reducing the language learners' cognitive load involved in interpreting complex language data. However, DDL researchers need to address the implications that generative AI will have on language education, understanding the contributions of corpora and DDL to a new ecology of language learning and L2 use. In a context where writing and learning is seemingly transitioning to a post-plagiarism era where human/AI collaboration will be the norm (Eaton, 2023), corpora and DDL will necessarily need to be able to outline the specific data literacy and corpus skills that can contribute to improve the process. We point two

of those opportunities. The first one is the role of frequency information in a context where AI applications are reluctant to reveal their sources. This deprives language learners of the opportunity to examine how genres and contexts of use affect language use. The second involves taking part in relevant conversations around the notion of attested use and what constitutes real communication evidence. As LLMs extract linguistic patterns that convincingly imitate human communication, it remains ripe for discussion whether AI-generated language constitutes authentic language. Another potential danger of present conversation AI agents is that LLMs end up defining the future standards for communication and language use, transforming an author's writing style into a cliché and flowery prose for the sake of correctness. In such conversations, corpus linguists can bring forward their insights into the many social, cultural, linguistic and contextual complexities involved in communication.

Conclusion

This entry has briefly discussed some challenges and opportunities for the foreseeable future of DDL. Challenges such as the over-reliance on higher education settings, lack of mobile-ready tools, and limited language diversity highlight the need for more accessible and inclusive approaches. Additionally, methodological concerns, such as small sample sizes and short-term outcomes, suggest that DDL research must evolve. However, opportunities abound with advancements in technology and increased understanding of language learners' uptake of DDL. NLP, multimodal resources, and AI-driven tools can expand DDL's reach, reshaping language learning to be more adaptive and personalized. AI also raises critical discussions on what constitutes authentic language, especially as LLMs increasingly influence communication standards. By addressing these challenges and embracing AI's potential, DDL can play a pivotal role in modern language education, offering new avenues for learners to engage with authentic language data in increasingly diverse and flexible ways.

Cross-References

▶ MALL and Data-Driven Learning
▶ Types of Corpora in Data-driven Learning

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Anthony, L. (2024). Addressing the challenges of data-driven learning through corpus tool design—In conversation with Laurence Anthony. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 9–24). Routledge. https://doi.org/10.4324/9781003413301

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Pérez-Paredes, P. (2024). Data-driven language learning. In R. Hampel & U. Stickler (Eds.), *The Bloomsbury handbook of language learning and technology* (pp. 212–225). Bloomsbury. https://doi.org/10.5040/9781350340350

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. http://hdl.handle.net/10125/73450

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*(3), 307–328. https://doi.org/10.1017/s0958344007000535

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 1–16). Rodopi. https://doi.org/10.1163/9789401203906_002

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching, 52*(4), 460–475. https://doi.org/10.1017/s0261444819000089

Crosthwaite, P. (Ed.). (2019). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge. https://doi.org/10.4324/9780429425899

Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*. Advance online publication. https://doi.org/10.1075/ijcl.23056.cro

Crosthwaite, P., & Boulton, A. (2024). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In H. Tyne, M. Bilger, L. Buscail, M. Leray, N. Curry, & C. Pérez-Sabater (Eds.), *Discovering language: Learning and affordance*. Peter Lang.

Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning, 36*(7), 1392–1420. https://doi.org/10.1080/09588221.2021.1995001

Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity, 19*(23). https://doi.org/10.1007/s40979-023-00144-1

Forti, L. (2023). *Corpus use in Italian language pedagogy: Exploring the effects of data-driven learning*. Routledge. https://doi.org/10.4324/9781003137320

Forti, L., & Spina, S. (2019). Corpora for linguists vs. corpora for learners: Bridging the gap in Italian L2 learning and teaching. *Educazione Linguistica Language Education, 8*(2), 349–362. https://doi.org/10.30687/elle/2280-6792/2019/02/005

Godwin-Jones, R. (2021). Evolving technologies for language learning. *Language Learning & Technology, 25*(3), 6–26. http://hdl.handle.net/10125/73443

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology, 27*(2), 6–27. https://hdl.handle.net/10125/73501

Han, Z. (2024). Chatgpt in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition, 46*(2), 301–306. https://doi.org/10.1017/S0272263124000111

Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107–118). Rodopi.

Kern, R. (2024). Twenty-first century technologies and language education: Charting a path forward. *The Modern Language Journal, 108*(2), 513–533. https://doi.org/10.1111/modl.12924

Li, C. (2024). A review of theories, pedagogies and vocabulary learning tasks of English vocabulary learning apps for Chinese EFL learners. *Journal of China Computer-Assisted Language Learning*. Advance online publication. https://doi.org/10.1515/jccall-2023-0026

Lin, P. (in press). Using corpus linguistics in formulaic language learning. In G. Brookes, N. Curry, & R. Love (Eds.), *Applications of corpus linguistics: Established and emergent contexts*. Cambridge University Press.

Meunier, F. (2022). Revamping DDL: Affordances of digital technology. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 344–360). Routledge: Taylor & Francis.

Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: Appropriation and the possibilities scenario. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 53–73). Peter Lang.

Pérez-Paredes, P., et al. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System, 84*, 145–159.

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pérez-Paredes, P. (2024). Data-driven learning in informal contexts? Embracing broad data-driven learning (BDDL) research. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 211–226). Routledge. https://doi.org/10.3726/978-3-0351-0166-9

Pérez-Paredes, P., & Chambers, A. (in press). Corpora and the learning and teaching of romance languages. In H. Tyne & S. Spina (Eds.), *Applying corpora in teaching and learning romance languages*. John Benjamins.

Tolstykh, O. M., & Oshchepkova, T. (2024). Beyond ChatGPT: Roles that artificial intelligence tools can play in an English language classroom. *Discover Artificial Intelligence, 4*(1), 60. https://doi.org/10.1007/s44163-024-00158-9

Vyatkina, N. (2024). *Corpus applications in language teaching and research: The case of data-driven learning of German*. Routledge. https://doi.org/10.4324/9781003223627