Exploring the Potential of Mixed Reality in Enhancing Language Acquisition for EFL Children

Chia Hsu*
d11010002@gapps.ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

Yu-Chieh Sun
m11210304@mail.ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

Hsin-Yi Wang
b10915033@gapps.ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

Yu-Jung Liu
m11110302@ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

Yi-ju (Ariel) Wu
yjarielwu@gmail.com
University of Taipei
Taipei, Taiwan

Tse-Yu Pan
typan@ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

Neng-Hao Yu
jonesyu@ntust.edu.tw
National Taiwan University of Science and Technology
Taipei, Taiwan

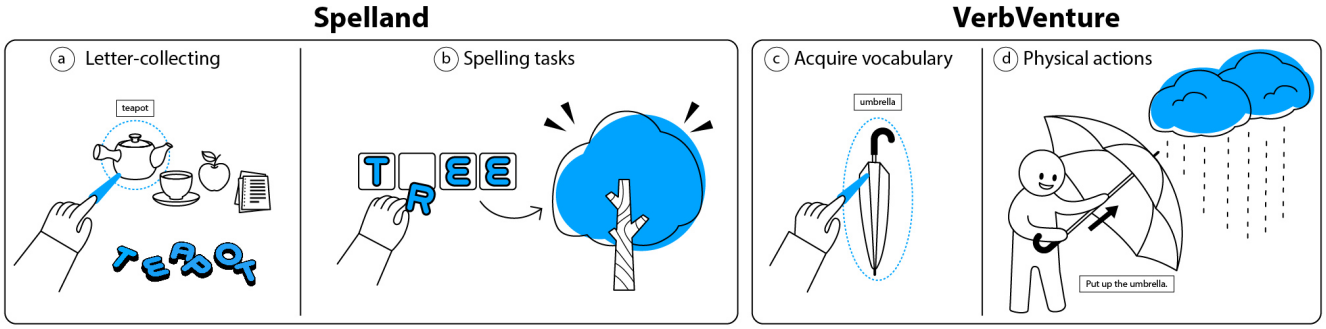

Figure 1: Learning Activities for Vocabulary(Spelland) and Collocations(VerbVenture) through Situated and Embodied Learning. (a) Situated Vocabulary Exploration: Children can freely explore everyday objects and select words to practice pronunciation and spelling. Successful pronunciation earns them letters for later use. (b) Creative Word Building: Children can use their collected letters to spell new words. If spelled correctly, a virtual object representing the word appears in the environment. (c) Vocabulary Practice with Collocations: Children explore everyday objects and practice pronunciation (same as (a)). Successful pronunciation unlocks the collocation interaction mode. (d) Action-based Collocation Practice: A virtual scenario animation appears, prompting children to use real objects to perform corresponding actions (Embodied Learning). This reinforces the connection between words, objects, and actions, strengthening their understanding of collocations in a practical way.

UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3678388

ABSTRACT

Mixed Reality (MR) technology has demonstrated its effectiveness in enhancing education, particularly in task-based training and the visualization of spatial concepts. This study examines the feasibility of using MR to assist EFL (English as a Foreign Language) children in learning English vocabulary and collocations. By analyzing different teaching methodologies, we identify two suitable strategies for MR-based instruction and develop two prototypes: *Spelland* and *VerbVenture*. *Spelland* uses situated learning to help children explore everyday objects and learn vocabulary through context, practicing pronunciation and spelling to earn letters for building

new words and creating virtual objects. *VerbVenture* employs embodied learning to help children understand verb-noun collocations by interacting with virtual scenarios and performing actions, reinforcing word-object-action connections. We conduct a pilot study with EFL children to evaluate these prototypes, providing insights for further refinement and development of MR-based EFL learning applications.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**.

KEYWORDS

Mixed Reality, Tangible Interaction, Embodied Learning, Situated Learning, Language Learning

ACM Reference Format:

Chia Hsu, Yu-Chieh Sun, Hsin-Yi Wang, Yu-Jung Liu, Yi-ju (Ariel) Wu, Tse-Yu Pan, and Neng-Hao Yu. 2024. Exploring the Potential of Mixed Reality in Enhancing Language Acquisition for EFL Children. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3675094.3678388

1 INTRODUCTION

In the context of global linguistic diversity, many countries have implemented language policies to enhance their citizens' language skills and competitiveness. Due to the widespread influence of English, English education is very common in many countries. In countries where English is not the native language, it is often taught as a second or foreign language. For non-native English speakers, differences between their native language and English often result in language transfer issues [7]. These difficulties are particularly pronounced in Asia. For instance, Chinese-speaking *learners* tend to use their native pronunciation, grammatical structures, semantics, and cultural expressions when learning English, leading to errors in expression and difficulties in understanding instructional content.

Traditional teaching methods have tried to mitigate language transfer through various media and group activities, but they are often limited by the need for teachers to convey knowledge through explanation. Furthermore, existing teaching methods are predominantly group-based, overlooking the individuality, personal experiences, and knowledge of *learners*, making it difficult to meet their specific needs. Traditional teaching models are teacher-centered, placing *learners* in a passive learning state, which leads to a lack of interactive diversity. Although modern textbooks can be presented through various multimedia formats, they often lack overall coherence, making it difficult for *learners* to resonate with the material and apply what they have learned in daily life.

The most natural way to learn a language is to allow *learners* to fully immerse themselves in a language-rich environment, learning organically as they grow and develop. Therefore, providing more learning opportunities is more effective than forcing *learners* to study in a rigid and monotonous manner. To create meaningful learning environments and contexts, enabling *learners* to engage with diverse and rich language inputs in interesting and stimulating settings, we investigate different teaching strategies that can be adopted into MR technologies to enhance learning activities. We develop two prototypes based on situated and embodied learning to provide young *learners* with immersive learning experiences (Fig.1). This study shares preliminary insights to bring a new approach and progress to the field of language education.

2 RELATED WORK

This section explores the potential of situated and embodied learning within a Mixed Reality (MR) environment to enhance EFL vocabulary and collocation learning for children.

2.1 Situated Learning

Situated learning emphasizes the importance of context in knowledge acquisition [10]. Learners retain information better when encountered in relevant and meaningful contexts [12]. AR/MR technology excels at providing situated learning opportunities by overlaying information on real-world objects [18] [3] [19] [9] [17]. Studies have shown that AR/MR applications effectively label real-world objects with corresponding vocabulary, leading to improved learning outcomes and user engagement compared to traditional methods like flashcards [19]. Additionally, situated learning with AR/MR has demonstrated benefits for children's education, fostering superior learning progress and increased enjoyment [15] [2]. These findings suggest that MR's ability to dynamically integrate information into real-world environments can enhance EFL vocabulary learning by anchoring new words in familiar contexts.

2.2 Embodied Learning

Embodied learning posits that physical actions and interactions play a crucial role in knowledge acquisition [20]. Learners develop a stronger understanding of concepts through physical movement and manipulation [1]. MR offers unique capabilities for embodied learning by enabling physical interaction with virtual objects. Research exploring embodied learning principles in immersive environments demonstrates positive impacts on knowledge retention, engagement, and learning outcomes across various subjects [11]. Studies investigating AR for vocabulary learning show improved results compared to traditional methods, with learners exhibiting greater motivation and satisfaction, particularly when 3D graphics are incorporated [5]. These studies highlight the benefits of AR in English vocabulary learning, yet they lack learners' understanding and application of word contexts.

Our study addresses the gap in current approaches by leveraging situated and embodied learning principles within an engaging MR environment. We propose a system that facilitates EFL vocabulary and collocation learning by associating them with everyday objects. Learners can perform actions corresponding to verb-noun collocations, allowing them to grasp contextual usage through physical interactions with virtual or physical objects. This approach aims to bridge the gap in traditional instruction where collocation teaching is often overlooked, fostering a deeper understanding of word meaning and usage within specific contexts.

3 SYSTEM DESIGN

In addition to understanding semantics, language *learners* must also grasp context and pragmatics. To effectively achieve these learning objectives and create the appropriate learning environment, we

integrated MR technologies into situated and embodied learning methods and developed two language learning prototypes: *Spelland* and *VerbVenture*. This section describes the system design of the two prototypes and analyzes their impact on children's learning motivation, interest, and effectiveness.

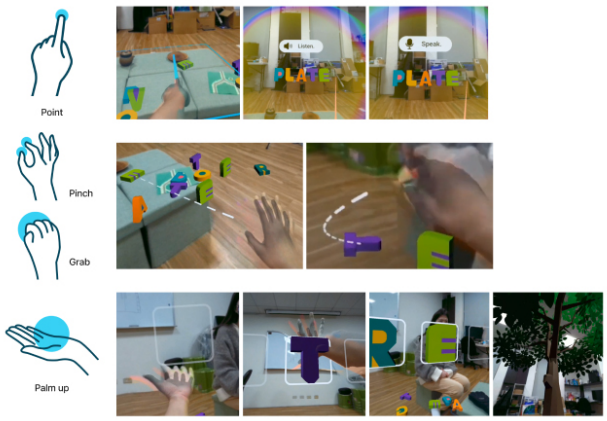

Figure 2: Gestural UI design in Spelland. When the learner points a finger at an object, a word bubble appears above it. Once the learner pronounces the word correctly, the bubble pops and drops the alphabet bricks. They can pinch or grab Alphabet Bricks either up close or from a distance, with a dotted line indicating which one is being selected. Turning the palm up opens the Spelling Board, allowing the learner to create a new word.

3.1 Spelland: Situated Vocabulary Acquisition

This study targets early EFL *learners* aged 6 to 8 and uses mixed reality to provide ample opportunities to practice spelling and learn new vocabulary. We adopted the situated learning strategy [10] into the game [8] to enable everyday objects as collectible words.

In our game design, we employ a teaching method based on Presentation, Practice, and Production (PPP), which cultivates the habit of independent learning through repetition, practice, and demonstration. The *learners* first interact with real-world objects via MR, where the objects' spelling and pronunciation appear (Presentation). The *learners* then repeat the pronunciation (Practice) to collect the letters of these objects, and finally use the collected letters to spell out the target words (Production), which then transform into interactive 3D objects. We designed the learning experience and content tools using gestural UI (Fig.2), voice input, and an object-to-word engine [16]. Therefore, *learners* can naturally point/manipulate the physical/virtual objects and practice pronunciation and spelling of the corresponding words in a fun way.

We conducted a round of pilot testing, which included a semi-structured interview after the experiment. Due to the performance issues of the object-to-word function, we simulated the experience by employing a Wizard of Oz technique for evaluating the prototype. The results showed that our game allows children to develop an interest in understanding new vocabulary and motivates them to learn through the process of collecting and reorganizing words. Initial interview results show positive feedback, with children generally finding the interactive process interesting and quickly discovering letters and spelling words correctly. In terms of interactive design, children were particularly impressed by the special effects of words, such as the appearance of Alphabet Bricks, the bursting of bubbles, and the falling of Alphabet Bricks. In terms of gesture recognition, we observed that when children pinched letters, they sometimes retracted their arms, causing their hands to move out of the headset range and affecting recognition accuracy. In the future, the sensing technique should be improved, or clear guidance can be provided on the scene to remind children not to exceed the operational range. In terms of speech recognition, it effectively recognized children's pronunciation, showing expected results.

In summary, children in the preliminary user study found the game to be immersive, actively exploring everyday objects and trying to create new words. The bubble effect, letter-collection task, and word-to-object transformation made the kids eagerly play with everyday objects. Furthermore, they showed increased interest in learning about other nearby objects after playing the game.

3.2 VerbVenture: Embodied Verb-noun Collocations Acquisition

The vocabulary knowledge of EFL *learners* encompasses phrases such as collocations or idioms. Among these, collocations refer to the natural cooccurrence of a group of words. The emphasis in using collocations lies more on understanding context and pragmatics. While their meanings can be inferred from literal concepts, both beginner and advanced EFL *learners* often struggle with their correct usage due to the influence of their native language. Research by Nesselhauf indicates that the most common error type is the collocation of verbs [13], and collocations involving gerunds are also popular learning targets in teaching. Therefore, we design *VerbVenture*, grounded in embodied learning strategy, enable the *learners* to perform verb-noun and verb-prep-noun collocations with relevant physical activities. This approach mimics the natural language acquisition process, encouraging children to freely explore real-world objects and engage in corresponding actions that exemplify verb-noun collocations (Fig.3).

Utilizing embodiment aids in creating a sense of presence, credibility, and immersion, helping *learners* to better integrate into learning contexts. Based on the aforementioned concept, the system comprises two phases: 1) Vocabulary Acquisition: *Learners* use MR to interact with objects, learning their names and pronunciations. 2) Collocation Practice: Animated guidance helps *learners* use real objects for context-based verb collocations. This approach enhances mental representations and real-world learning. The young participant in the pilot study indicates *VerbVenture* boosts her interest and understanding of collocations. Future plans involve assessing embodied MR's effectiveness in collocation learning, focusing on memory retention and long-term outcomes.

4 TECHNICAL LIMITATIONS AND FUTURE WORK

This study explored the potential of Mixed Reality (MR) for language education, focusing on vocabulary and verb-noun collocation

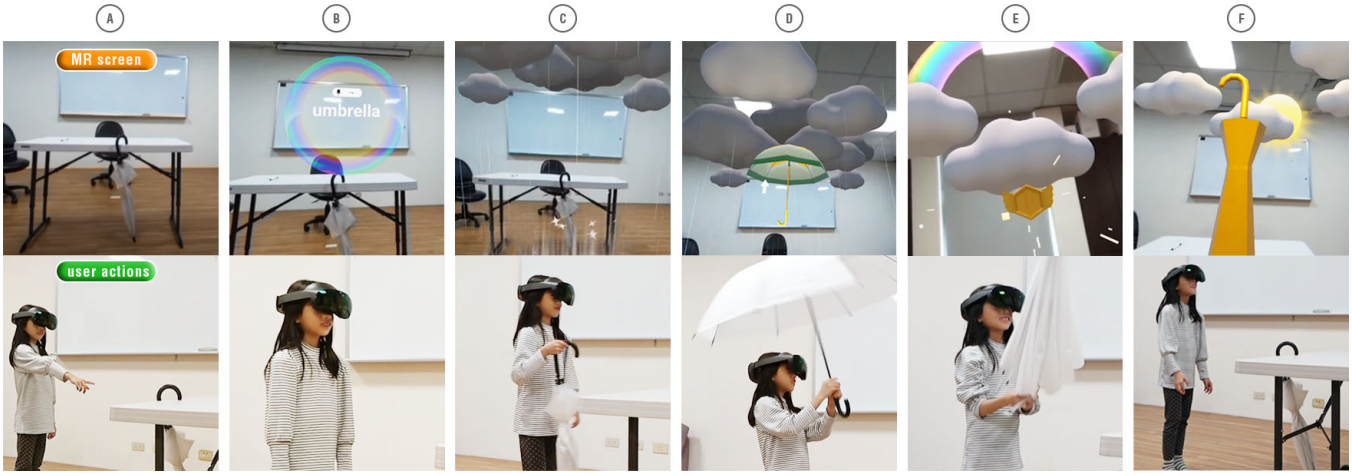

Figure 3: Interactive flow within the *VerbVenture* prototype.

learning for English as a Foreign Language (EFL) learners. We developed two prototypes, *Spelland* and *VerbVenture*, using Unity3D and implemented them on the Meta Quest Pro.

Our prototypes incorporated hand gesture recognition and speech-to-text (STT) to enhance the usability of the learning activities. However, the object-to-text engine requires high accuracy and performance to display the correct word when the user points at an object. Due to privacy restrictions on the Meta platform, we were unable to obtain real-time images for computer vision recognition. With the advanced development of 3D Scene Understanding [14], we envision that every object in the environment could be sensed in situ. The new challenge will be determining what word should be displayed above the object to stimulate the learner to acquire new vocabulary.

In *VerbVenture*, we did not integrate an action recognition function into the prototype. The Wizard of Oz (WoZ) prototype study demonstrated the feasibility of the embodied learning design. Studies like ego4D [4] aim to provide a large-scale dataset for first-person activity recognition. We aim to observe more young learners' behavior while interacting with the embodied learning process and contribute to the dataset or provide potential insights on learner's action sensing in the future.

Furthermore, we will conduct a formal user study on our system to evaluate its effectiveness and long-term impact on EFL learning. This study will help us refine our prototypes and potentially expand the use of MR in language education.

5 CONCLUSION

This study explored the potential of Mixed Reality technology to enhance EFL vocabulary and collocation learning for children. We investigated the theoretical foundations of situated and embodied learning, demonstrating their suitability for language acquisition within an MR environment.

Our findings, based on the development and pilot testing of two prototypes, *Spelland* and *VerbVenture*, suggest that MR can create engaging and interactive learning experiences that promote EFL vocabulary and collocation acquisition in children. *Spelland*, utilizing situated learning principles, encourages children to explore virtual environments and connect real-world objects with corresponding vocabulary. *VerbVenture*, leveraging embodied learning, allows children to interact with objects and perform actions corresponding to verb-noun collocations, fostering a deeper understanding of word meaning and usage within specific contexts.

While further research with a larger population is needed, this study provides valuable insights into the potential of MR for EFL learning. The combination of situated and embodied learning principles within an MR environment creates a promising avenue for developing engaging and effective language learning tools for children. Future research directions include exploring the effectiveness of different MR interaction modalities (e.g., voice, gesture) and evaluating the long-term learning outcomes of children using MR-based language learning applications. By continuing to explore the potential of MR combined with AI technology[6], we can pave the way for innovative and engaging learning experiences that can transform the way children acquire new languages.

6 ACKNOWLEDGMENTS

This study was supported by the National Science and Technology Council, Taiwan, under grant NSTC 112-2221-E-011-101-MY3.

REFERENCES

[1] David Birchfield and Colleen Megowan-Romanowicz. 2009. Earth science learning in SMALLab: A design experiment for mixed reality. *Computer Supported Learning* 4 (2009), 403–421. https://doi.org/10.1007/s11412-009-9074-8

[2] Harvey Thornburg David Birchfield, Sarah Hatton M. Colleen Megowan-Romanowicz, Igor Dolgov Brandon Mechtley, and Winslow Burleson. 2008. Embodiment, multimodality, and composition: Convergent themes across HCI and education for mixed-reality learning environments. *Advances in Human-Computer Interaction* (2008). https://doi.org/10.1155/2008/874563

[3] Fiona Draxler, Audrey Labrie, Albrecht Schmidt, and Lewis L. Chuang. 2020. Augmented Reality to Enable Users in Learning Case Grammar from Their Real-World Interactions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376537

[4] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. 2022. Ego4d: Around the world in 3,000 hours of egocentric video. In *Proceedings*

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 18995–19012. https://doi.org/10.48550/arXiv.2110.07058

[5] Kuang-Chung Hao and Lai-Chung Lee. 2021. The development and evaluation of an educational game integrating augmented reality, ARCS model, and types of games for English experiment learning: an analysis of learning. *Interactive Learning Environments* 29, 7 (2021), 1101–1114. https://doi.org/10.1080/10494820.2019.1619590

[6] Teresa Hirzle, Florian Müller, Fiona Draxler, Martin Schmitz, Pascal Knierim, and Kasper Hornbæk. 2023. When XR and AI Meet - A Scoping Review on Extended Reality and Artificial Intelligence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 730, 45 pages. https://doi.org/10.1145/3544548.3581072

[7] Daniel Horowitz, Charles Wordell, Jane Wieman, and Kyoko Nozaki. 1989. Language Transfer. *Journal of the Japan association of language teachers* 11, 2 (1989).

[8] Chia Hsu, Yu Chen, Yu-Jung Liu, Yu-Cheng Chang, and Min-Jui Lee. 2023. Spelland: Situated Language Learning with a Mixed-Reality Spelling Game through Everyday Objects. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)* (2023), 6 pages. https://doi.org/10.1145/3544549.3583830

[9] Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Höllerer, Dorothy Chun, and John O'Donovan. 2018. ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE transactions on visualization and computer graphics* PP (2018). https://doi.org/10.1109/TVCG.2018.2868568

[10] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, UK.

[11] Robb Lindgren and Michael J. Moshell. 2011. Supporting children's learning with body-based metaphors in a mixed reality environment. *Proceedings of the 10th International Conference on Interaction Design and Children (IDC '11)* (2011), 177–180. https://doi.org/10.1145/1999030.1999055

[12] Christopher R. Madan and Anthony Singhal. 2012. Encoding the world around us: Motor-related processing influences verbal memory. *Consciousness and Cognition* 21, 3 (2012), 1563–1570. https://doi.org/10.1016/j.concog.2012.07.006

[13] Nadja Nesselhauf. 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24, 2 (2003), 223–242. https://doi.org/10.1093/applin/24.2.223

[14] Songyou Peng, Kyle Genova, Chiyu Jiang, Andrea Tagliasacchi, Marc Pollefeys, Thomas Funkhouser, et al. 2023. Openscene: 3d scene understanding with open vocabularies. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 815–824.

[15] Sara Price and Yvonne Rogers. 2004. Let's get physical: The learning benefits of interacting in digitally augmented physical spaces. *Computers & Education* 43, 1 (2004), 137–151. https://doi.org/10.1016/j.compedu.2003.12.009

[16] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *arXiv* (2018).

[17] Marc Ericson Santos, Arno Lübke, Takafumi Taketomi, Goshiro Yamamoto, Ma Rodrigo, Christian Sandor, and Hirokazu Kato. 2016. Augmented reality as multimedia: The case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning* 11 (01 2016), 1–23. https://doi.org/10.1186/s41039-016-0028-2

[18] Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. 2017. Serendipitous Language Learning in Mixed Reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2172–2179. https://doi.org/10.1145/3027063.3053098

[19] Maheshya Weerasinghe, Verena Biener, Jens Grubert, Aaron Quigley, Alice Toniolo, Klen Čopič Pucihar, and Matjaž Kljun. 2022. VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3748–3758. https://doi.org/10.1109/TVCG.2022.3203116

[20] Shaoqun Wu, Margaret Franken, and Ian H. Witten. 2010. Supporting collocation learning with a digital library. *Computer Assisted Language Learnig* 23, 1 (2010), 87–110. https://doi.org/10.1080/09588220903532971