Corpus-based Translation Pedagogy: A Preliminary Case Study

Mengqi LI[0009−0004−3588−4480] and Dongxia PAN[1111−2222−3333−4444]

School of Foreign Studies, Anhui Xinhua University, Hefei, AnHui, 230088, China
mengqili2022@gmail.com, pandongxia@axhu.edu.cn

Abstract. This research is grounded in the practical experiences of teaching translation at a specific local university and aims to introduce an innovative corpus-based pedagogical framework tailored for translation courses. Through a comparative study involving undergraduate students' training experiments, the research seeks to provide valuable insights into the effective integration of corpus technology in translation education, meticulously examining aspects such as course design, corpus resources, teaching content, and implementation strategies. The results of the teaching practice reveal that the newly proposed pedagogical model significantly enhances the expression of translations, elevates students' English proficiency as translators, diversifies learning methods, and fosters increased interest and self-learning capabilities among students. Building upon these findings, future translation education should prioritize the enrichment of corpus teaching resources and dynamic tracking of students' daily corpus usage. Translation instructors must continually refine their corpus usage skills and enhance their digital integrated capabilities to effectively adapt to the evolving needs of translation course reforms in the current context. This research contributes to the ongoing discourse on the integration of corpus technology and Data-Driven Learning (DDL) in language teaching, offering insights into how educators can harness the potential of corpus tools and DDL pedagogy in their classrooms.

Keywords: Corpus-based Translation Pedagogy · Corpus Technology Integration · Data-Driven Learning (DDL) .

1 Introduction

In the ever-evolving landscape of language education, the fusion of technology and pedagogy has redefined how we approach teaching and learning. One of the forefronts of this transformation is Corpus-Based Translation Pedagogy, which offers a promising avenue to enhance the quality and efficacy of translation education.

Our research is rooted in practical experiences gained from the realm of translation teaching at a specific local university, and it embarks on a journey to reshape the landscape of language education. This exploration unlocks invaluable

insights into the seamless integration of corpus technology into the domain of translation education.

In the pages that follow, we will present the findings of a comparative study involving undergraduate students' training experiments. These findings offer valuable insights into the effective integration of corpus technology into translation education. The empirical results of our teaching practice reveal a compelling narrative: the adoption of our newly proposed teaching model significantly enhances translation expression, elevates students' English proficiency as translators, introduces diversity into learning methods, and nurtures heightened interest and self-learning capabilities among students.

The objective of this paper is clear: to craft a corpus-based pedagogical framework that is finely tuned for translation courses. This framework, meticulously designed through a process of inquiry and experimentation, holds the potential to revolutionize translation instruction.

In conclusion, this study significantly contributes to the ongoing discourse surrounding the fusion of corpus technology and Data-Driven Learning (DDL) in language teaching. It serves as a guiding torch for educators, illuminating the path toward acquiring the necessary tools to harness the potential of corpus tools and DDL pedagogy within their classrooms.

2 Literature Review

2.1 Corpus-based Translation Pedagogy Research

Since 1993, scholars such as Mona Baker, Stewart, Zanettin, Lopez-Rodrigues, and Tercedor-Sanchez have explored the application of monolingual corpora, parallel corpora, and other corpus-based resources in translation research and teaching. Their studies have highlighted the positive impact of corpora on translation pedagogy while also addressing the potential and limitations of integrating corpora into translation instruction. Stewart [12] recognized the role of monolingual corpora in translation teaching. Lopez-Rodrigues and Tercedor-Sanchez [5] developed learner corpora and reference corpora to assist students in mastering translation strategies and improving their evaluation skills. Zanettin [15] emphasized the importance of teachers investing sufficient time in guiding students to query corpus data and analyze the information obtained.

In addition, Shih and Shen [11] pointed out that the use of corpora allows for a comparison between the works of professional translators and student translators, suggesting its usefulness in teaching. Such corpora enable students to observe the differences between themselves and professional translators, rather than merely focusing on their own translation errors.

These studies provide compelling evidence of the application value of corpora in teaching and learning. They highlight the advantages of using corpora as a tool for enhancing students' translation competence, language skills, and language awareness. By incorporating corpora into translation pedagogy, instructors can guide students to analyze language data, observe professional

translation practices, and bridge the gap between students' own work and that of experienced translators. Such pedagogical approaches foster critical thinking, promote accuracy, and cultivate a deeper understanding of translation processes and strategies.

While scholars have made significant progress in corpus-based translation pedagogy, there is still room for further exploration and innovation. The integration of monolingual and parallel corpora, along with the incorporation of student translation corpora, holds great potential for creating a dynamic and engaging learning environment that fosters students' translation skills, critical thinking, and self-reflection.

Overall, corpus-based translation pedagogy research has consistently demonstrated the positive impact of corpora on translation learning. It is crucial for educators to harness the potential of corpora in order to enhance the effectiveness of translation instruction and foster the development of competent and reflective translators.

2.2 DDL

The use of language corpora for language teaching and learning, commonly referred to as "data-driven learning" (DDL) as proposed by Johns (1991), has emerged as a well-established and vibrant field of research. According to Boulton & Vyatkina's research [1] in 2021, over the years, nearly 500 empirical studies have contributed to the understanding and application of DDL in diverse educational contexts.

DDL typically encompasses two key approaches: the first involves teachers providing printed concordances of pre-selected corpus data for language learners, while the second empowers learners to directly engage with corpus query software. These approaches serve as valuable tools in aiding the acquisition of various target language constructs, including vocabulary, grammar, collocation, and addressing L2 errors. What makes DDL particularly versatile is its applicability across a wide spectrum of learning and teaching contexts, including English as a Foreign Language (EFL) education.

Recent efforts have aimed to harmonize DDL methodology with established language learning theories. Scholars such as L. Flowerdew [2], Pérez-Paredes [8], and O'Keeffe [7] have highlighted that DDL enhances language learning through a usage-based paradigm. This emphasizes the significance of frequency, salience, and contingency, as articulated by Ellis [3] in 2006. The implementation of DDL involves a blend of constructivist, student-led, and focus-on-form pedagogical activities at the individual level, complemented by teacher- and peer-mediated/scaffolded focus-on-form activities at the sociocultural level. These principles form the core theoretical underpinnings of DDL, contributing to its effectiveness in language education.

The research on DDL has taken various dimensions, shedding light on its impact from multiple perspectives. Studies have investigated classroom corpus use through lenses such as learning gains and reduced L2 errors in pre-/post-test experimental studies. Moreover, research has delved into the perceptions

of both learners and teachers regarding DDL (e.g., Yoon & Hirvela[14], 2004), corpus usage patterns (e.g., Pérez-Paredes et al.[9], 2011), and increasingly, the outcomes of teacher education workshops (e.g., Chen et al. [2], 2019; Schaeffer-Lacroix, [10] 2019).

To address teachers' concerns and enhance the integration of DDL into language education, scholars have called for increased alignment between DDL and language learning didactics. In 2019, Meunier [6] advocates for "constructive alignment," emphasizing that curriculum, learning outcomes, teaching methods, and assessment practices should consistently and coherently align with DDL. The Technology, Pedagogy & Content Knowledge (TPACK) model proposed by Koehler and Mishra [4] in 2009 serves as a promising framework for this alignment, as it encompasses pedagogical content knowledge. This knowledge extends beyond understanding how students learn language to integrating DDL within the curriculum and lesson planning, which has been relatively underexplored in existing DDL-focused teacher education studies. Therefore, this study aims to explore how DDL may be seamlessly integrated into Computer-Assisted Language Learning (CALL) lesson planning in the context of translation pedagogy at the university level, contributing to the broader adoption of DDL in mainstream education.

3 Current Issues and Reflections on Translation Courses in Higher Education

Translation courses are compulsory subjects for senior students majoring in foreign languages in universities. They are comprehensive courses that combine theory and practice, aiming to apply students' English abilities effectively. However, through teaching practice and research, it has been observed that the practical English application abilities of many students are weak, particularly in terms of English language awareness and discourse knowledge.

In terms of teaching methods and content, traditional teaching methods with a teacher-centered approach have been followed in translation courses in universities for many years. Teachers often determine textbooks and distribute translation materials based on personal preferences. In the classroom, teachers commonly adopt an immediate exercise and on-the-spot feedback model, lacking language awareness training and appreciation of stylistic features. Translations of words and phrases often lack context, and semantic coherence is lacking. The main basis for translation evaluation by teachers is the answer key from reference books. Some teachers also have unclear understanding and grasp of the translated texts, making it difficult for students to be convinced.

In response to the various issues in translation courses in higher education, in recent years, many scholars have proposed various solutions, one of which is the open corpus-based translation teaching model using computer corpora. In 2007, Luo Xuanmin et al. [13] put forward four advantages of the open corpus-based translation teaching method:

1) openness in *teaching ideas and theories* - students should be exposed to various translation phenomena and become familiar with different texts and translations;
2) openness in *teaching platforms* - corpus construction is a dynamic and improving process, and the use and development of corpora can better serve translation teaching;
3) openness in *data-driven learning* - learners actively use word index software to discover, summarize, and induce relevant patterns, thereby stimulating learning interest and forming an autonomous and open learning mode;
4) openness in *translation text evaluation mechanism* - drawing on machine translation software to test students' and teachers' translations in real-time and summarize common issues.

It is recognized that the fourth point involves complex computer technology and software costs, making it challenging to implement. However, the other three advantages are evident and worth exploring by teachers in their teaching practices.

In conclusion, the current issues in translation courses in higher education in China require attention and appropriate solutions. The adoption of an open corpus-based translation teaching model, as proposed by scholars, can address some of the challenges and offer benefits such as exposing students to diverse translation phenomena, utilizing corpus resources, promoting DDL, and implementing effective evaluation mechanisms. Teachers are encouraged to embrace innovative teaching methods and explore the potential of corpus-based translation pedagogy to enhance students' translation competence and overall learning experience.

4 Methods and Objectives

The study examines the students' perspective on the implementation of corpus in English translation classes. The participants received training, followed by guided and independent practice using a freely available corpus. At the conclusion of the semester, an anonymous questionnaire on their experience with DDL was distributed to the students. The aim of this research is to explore the potential of corpus-based approaches in reforming translation courses for English major students.

4.1 Participants

The participants in this study were third-year undergraduate students majoring in English at Anhui Xinhua College. One class consisted of 35 students and served as the control group, while another class consisted of 36 students and served as the experimental group. A total of 71 participants were involved. From March to June 2023, activities such as corpus-assisted translation teaching, testing, and questionnaire surveys were conducted with the students in the experimental

group. During the same period, traditional translation classroom teaching was conducted with the students in the control group. Finally, the same materials were selected for testing in order to compare the results between the experimental and control groups.

4.2 Training Design

The experiment lasted for four weeks. Both the experimental group and the control group received three 90-minute translation class hours every two weeks (with each class hour lasting 45 minutes), resulting in a total of 12 experimental class hours for each group. In these 12 class hours, the teaching content was the same for both groups, but the teaching methods were completely different.

For the students in the experimental group, the teacher organized autonomous learning by utilizing basic information about corpora and establishing their own corpora on SketchEngine. The distribution of class hours was shown in Figure 1 as follows:

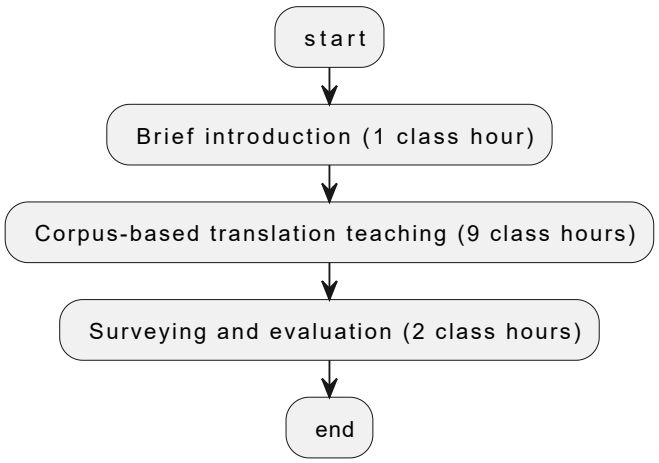

Fig. 1. Illustration of the distribution of class hours.

The specific operations for corpus-based translation teaching were as follows: The students were organized to collect Chinese political materials from the internet and create their own parallel corpus of Chinese political texts on the SketchEngine platform.

Additionally, the teacher distributed approximately 800-word Chinese political texts to the students at the beginning of the class, and the students used the parallel corpus to translate the materials. During this process, students were allowed to discuss with each other. The teacher guided students

in summarizing the terminology, collocations, and sentence patterns in the materials. After completing the translation task, the teacher provided the students with the standard translation of the Chinese materials, allowing the students to autonomously check and supplement their own translations.

For the students in the control group, the teacher adopted the traditional translation teaching mode. In other words, the teacher explained the materials in detail in class, enabling the students to acquire translation skills.

After four weeks, both the experimental group and the control group of students underwent a closed-book translation test using the same test papers, which were related to the teaching content within the four weeks, in order to compare the teaching effects of the two groups.

4.3 Analysis of Pre- and Post-test Scores for Experimental and Control Groups

The final test involved translating a Chinese text of approximately 300 characters within one hour. Initially, the instructor scored the translations based on faithfulness and fluency. To ensure the accuracy of the scores, additional translation instructors were invited to provide their own evaluations. The final score for each student was determined by averaging the scores from two teachers.

Taking advantage of SPSS, the author conducted the Independent Samples T Test on the experimental group and the control group, and calculated the average scores of the two groups. The results are as follow.

As is shown in the Table 1, the average score of the experimental group is 71.278 and the average score of the control group is 77.143. Since the P value (significance) is 0.005 which is less than 0.05, we can assume that there is a significant difference between the experimental group and control in terms of overall translation level. It can be concluded that there is a significant difference in overall translation proficiency between the experimental group and the control group, with the experimental group demonstrating superior translation quality. This suggests that corpus-based translation pedagogy contributes to enhancing the effectiveness of translation instruction.

Table 1. Results of Independent Samples T Test for the Experimental Group and Control Group in Terms of Overall Language Level. The mean difference is significant at the 0.05 level.

Group	N	Mean Scores	Std. Deviation	t	Sig.
Experimental Group	36	71.278	7.648	-2.921	0.005***
Control Group	35	77.143	9.217		

4.4 Questionnaire Survey Results

After the test, the students of the experimental group filled out a questionnaire about corpus-based translation pedagogy. Based on the questionnaire, the author

could explore the students' attitude towards the new teaching mode as well as whether corpus-based translation pedagogy can help them to acquire translation skills better and improve their translation level.

The survey results showed that: over 60 percent of the participants believed that corpus-based teaching helped them learn translation better. The details can be listed as Figure 2.

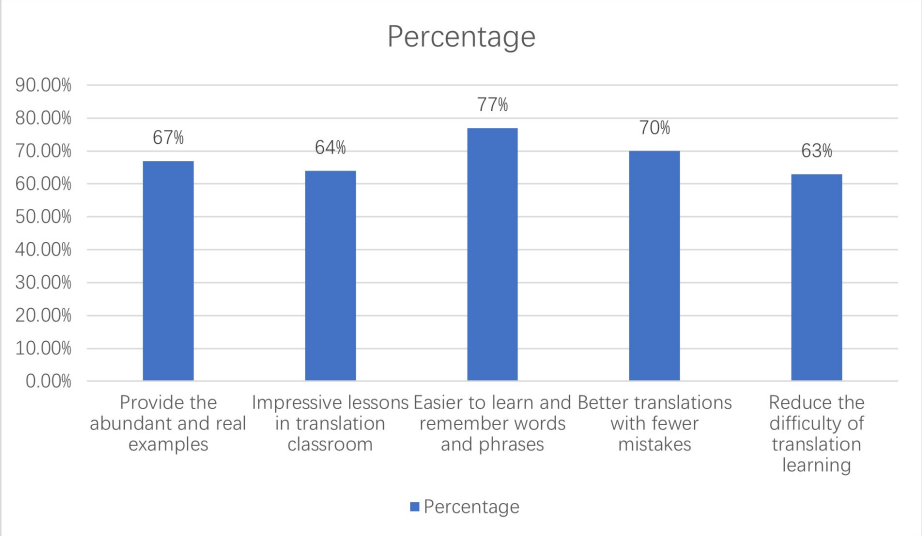

Fig. 2. Advantages of using corpora for translation assistance.

According to Figure 2, we can see that, 67% thought that the abundant real examples provided by the corpus made it easy for them to remember what they needed to search for. 64% found the content learned in the corpus-based translation classroom impressive even after class. 77% stated that using words and phrases retrieved from the corpus made a deeper impression on them. 70% believed that using the corpus for translation made their target language more authentic and helped reduce translation errors. 63% found that using the corpus for translation reduced the difficulty of translation learning.

Additionally, when asked about future translation learning, 79% of the students believed that it is necessary to integrate corpus-based pedagogy into traditional translation teaching, and in the future, when faced with more tasks, students would rely on corpus tools for translation.

However, during the process of translation teaching, the author also identified certain disadvantages of corpus-based translation pedagogy, which include:

One challenge was the limited language variety covered by specific corpora, hindering students' ability to translate texts from diverse dialects and regional variations. Additionally, the static nature of corpora and the lack of real-time

updates may affect the teaching of contemporary language usage and emerging terminology. Moreover, corpora may not always provide sufficient cultural and contextual information, necessitating the supplementation of learning with additional resources.

All in all, the corpus-based translation pedagogy presents both advantages and disadvantages, but the advantages outweigh the disadvantages. This approach enhances translation accuracy, promotes language proficiency, and fosters a systematic understanding of translation principles. However, it is important to address the limitations, such as the lack of context and domain specificity, to ensure a comprehensive and well-rounded translation education. Integrating supplementary resources and providing real-time updates can further enhance the effectiveness of corpus-based translation teaching.

5 Conclusion and Future Work

This paper presents a preliminary case study on the corpus-based translation pedagogy, aiming at investigating its effectiveness in translation teaching practices. By examining various aspects, including course design, corpus resources, teaching content, and implementation strategies, a new pedagogical model was proposed and evaluated at a local university.

The experimental group, which received parallel corpus-based translation instruction guided by constructivism and autonomous learning theory, exhibited higher average scores and improved translation accuracy compared to the control group. These results demonstrate that the integration of corpora in translation teaching positively impacts students' translation abilities and overall learning outcomes.

This study might shed light on translation teaching, which can be briefly summarized but not limited to as follows:

1) *Advocating Autonomous and Cooperative Learning*: The findings underscore the significance of fostering autonomous and cooperative learning environments in translation classrooms. Integrating computer-aided translation software and encouraging students to collaboratively discuss and solve translation challenges enhance their understanding and application of translation principles, leading to more effective learning experiences.
2) *Enhancing Learning Experience with Authentic Resources*: Providing students with authentic learning resources and advanced tools, such as computer-aided translation software, can significantly enhance their engagement and enthusiasm in translation learning. Utilizing a rich parallel corpus and specialized translation software opens up new possibilities for creating more interactive and engaging learning experiences.
3) *Recognizing the Impact of Technology*: The study emphasizes the need to explore the role of technology in corpus-based translation teaching. Understanding how different translation software and emerging technologies, such as machine translation and artificial intelligence, can be effectively

integrated into the pedagogical approach will be crucial for optimizing the learning experience in the future.

In light of these findings, future research in corpus-based translation pedagogy should explore various directions:

1) Exploring Different Types of Corpora: Researchers should investigate the application of different types of corpora in translation teaching beyond parallel corpora. Exploring the use of comparable corpora or specialized corpora will provide a comprehensive understanding of their potential effectiveness in enhancing translation skills and language proficiency.
2) Investigating the Impact on Different Language Pairs: Extending the study to examine the impact of corpus-based teaching on various language pairs will help determine the generalizability and adaptability of this pedagogical model across diverse linguistic contexts.
3) Emphasizing the Role of Technology: Future research should focus on the integration of cutting-edge technologies in corpus-based translation teaching, exploring their potential to further optimize learning outcomes and enrich the translation process.

In conclusion, the preliminary case study demonstrates the promising potential of corpus-based translation pedagogy in improving translation skills, language proficiency, and overall learning experiences. The findings highlight the importance of creating dynamic and interactive learning environments and embracing technological advancements to enhance translation education. By addressing the implications and future research directions, educators can continuously refine and advance the corpus-based translation pedagogy to benefit students and foster their proficiency in translation and language learning.

References

1. Boulton, A., Vyatkina, N.: Thirty years of data-driven learning: Taking stock and charting new directions over time (2021)
2. Chen, M., Flowerdew, J., Anthony, L.: Introducing in-service english language teachers to data-driven learning for academic writing. System **87**, 102148 (2019)
3. Ellis, N.C.: Language acquisition as rational contingency learning. Applied linguistics **27**(1), 1–24 (2006)
4. Koehler, M., Mishra, P.: What is technological pedagogical content knowledge (tpack)? Contemporary issues in technology and teacher education **9**(1), 60–70 (2009)
5. López-Rodríguez, C.I., Tercedor-Sánchez, M.I.: Corpora and students' autonomy in scientific and technical translation training 1 (2008)
6. Meunier, F.: A case for constructive alignment in ddl: Rethinking outcomes, practices, and assessment in (data-driven) language learning. In: Data-driven learning for the next generation, pp. 13–30. Routledge (2019)
7. O'Keeffe, A.: Data-driven learning–a call for a broader research gaze. Language Teaching **54**(2), 259–272 (2021)

8. Pérez-Paredes, P.: 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 (2022)
9. Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J.M., Jiménez, P.A.: Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation. Computer Assisted Language Learning **24**(3), 233–253 (2011)
10. Schaeffer-Lacroix, E.: Barriers to trainee teachers' corpus use. In: Data-Driven Learning for the Next Generation, pp. 47–64. Routledge (2019)
11. Shih, C.L., Shen, C.A.: A corpus-based investigation into student technical translation. Journal of Translation Studies pp. 1–31 (2005)
12. Stewart, D.: Supplying native speaker intuitions or normalising translation? translating into english as a foreign language with the british national corpus. In: From the abstract of a paper in the provisional programme of the UMIST/UCL Conference: Research Models in Translation Studies, Manchester. pp. 28–30 (2000)
13. Xuanming, L., Bin, L.: On open-style corpus-based translation teaching. Foreign Language Education **30**(6), 45–48 (2009)
14. Yoon, H., Hirvela, A.: Esl student attitudes toward corpus use in l2 writing. Journal of second language writing **13**(4), 257–283 (2004)
15. Zanettin, F.: Corpus-based translation activities for language learners. The Interpreter and Translator Trainer **3**(2), 209–224 (2009)