Check for updates

DOI: https://doi.org/10.34069/AI/2025.86.02.9

How to Cite:
Chen, Y., & Shen, L. (2025). The implementation of a corpus-construction project on student translators: competence development and challenges. *Amazonia Investiga*, *14*(86), 103-113. https://doi.org/10.34069/AI/2025.86.02.9

The implementation of a corpus-construction project on student translators: competence development and challenges

一项学生译者参与的语料库项目：能力发展与挑战

Received: April 1, 2025 Accepted: May 8, 2025

Written by:
Yang Chen[1]
https://orcid.org/0000-0001-9024-3900
Li Shen[2]
https://orcid.org/0009-0009-0694-0056

Abstract

This study investigates a corpus-construction project implemented at a foreign language university over two years, aiming to explore its impacts on student translators and the challenges encountered by teachers in project management. Results show that students have developed multidimensional competence, including enhanced translation skills and technological competence, development of cognitive skills and critical thinking disposition, improvement of collaboration skills and teamwork, as well as historical responsibility and ethical awareness. However, teachers encountered challenges such as corpus difficulty and ambiguity, technological barriers in corpus construction, and difficulties in team management. This research not only enriches the theory of translation education, but also provides practical implications for curriculum design and teaching practices, as well as the implementation of large-scale authentic projects across different domains.

Keywords: Project-based learning, translation education, corpus, project management, competence development.

摘要：

本文阐述了在某外国语学院开展的为期两年的语料库建库项目研究，旨在探究其对学生译者的影响以及教师在项目管理中面临的挑战。结果表明学生在多方面受益，包括翻译技能与技术能力提升、认知技能和思辨特质发展、协作技能与团队合作增强，以及历史责任感和伦理意识的培养。然而，教师也面临了语料复杂和文本模糊、技术障碍以及团队管理困难等挑战。本研究不仅丰富了翻译教育理论，还为课程设计和教学实践提供了建议，并对不同领域开展大规模真实项目提供经验。

关键词： 项目式学习，翻译教育，语料库，项目管理，能力发展.

Introduction

With information technology posing both opportunities and challenges to the translation industry, institutions are seeking for innovative ways to cultivate student translators with comprehensive abilities. In translation education, traditional didactic approaches often fall short in equipping students with the multifaceted skills required in real-world translation scenarios, where technical proficiency, cognitive skills and collaborative competence are paramount. In such context, project-based learning (PjBL) has emerged as a pedagogical strategy to cultivate translators' essential competencies through authentic tasks (Putra et al., 2022; Ribeiro et al., 2023).

[1] PhD of Philosophy, Zhejiang Yuexiu University, China. WoS Researcher ID: LCD-5418-2024 - Email: roy0901@163.com
[2] MD, Zhejiang Yuexiu University, China. WoS Researcher ID: JZT-1625-2024 - Email: 37145035@qq.com

In the realm of translation education, corpus-based projects could enhance learners' translation quality, productivity and familiarity with computer-aided translation (CAT) technologies (Alhassan et al., 2021). Corpus construction projects create opportunities for translator trainees to acquire language patterns, collocations, specialized terminology and translation strategies, thus bridging the gap between theoretical learning and practical translation tasks. However, existing studies predominantly focus on small-scale, classroom-bound corpus activities (Mohammed, 2022), leaving a lacuna in understanding the dynamics of institutionally coordinated, large-scale projects. Meanwhile, PjBL highlights collaborative, experiential learning rooted in social constructivist theories (Vygotsky, 1978), yet its application in specialized translation domains remains under-researched. For example, PjBL integrated into the legal translation domain, presents unique challenges including terminological complexity and ethical sensitivity. This study addresses this gap by implementing a corpus-construction project involving military trial texts at a foreign language university, over an extended period of two years. Situated at the intersection of translator education, corpus linguistics, and project-based pedagogy, the research aims to investigate how such a large-scale authentic project influence students' competence development. Additionally, it probes instructors' challenges in balancing pedagogical goals with project implementation in authentic situations. Such insights are critical for optimizing translator-training models in specialized domains and broader contexts.

Research questions:

RQ 1: How did participation in the corpus-construction project influence the competence development among student translators?
RQ 2: What challenges did instructors encounter in managing such a project?

The structure of the article presents as follows. The literature review section examines existing research on corpus-based pedagogy and PjBL in translation education. The methodology section details our qualitative research design, including data collection and analysis approaches. Next, the results and discussion section present key findings of the study, discussing the impact of PjBL on student outcomes, as well as challenges encountered in implementing the PjBL from the perspectives of instructors. Finally, the conclusion summarizes the main findings, addresses the limitation of the current study and proposes directions for future research.

Literature Review

Corpus-Based Translation Pedagogy

Since the 1990s, the application of monolingual corpora, parallel corpora, and other corpus-based resources has transformed translation research and pedagogy (Li & Pan, 2023). Featured by the data-driven learning mode, the corpus-based pedagogy enhances learners' ability to identify collocations, resolve syntactic ambiguities, and improve terminological precision, which applies widely in translation classrooms. Xiao and Hu (2015) underscored the value of parallel corpora in addressing structural divergences between English and Chinese, which could foster learners' metalinguistic awareness. Mohammed (2022)'s study found that student translators not only used online corpora but also created their own corpora when engaging in the translation projects, which improved their translation efficiency and quality. Zhao and Mu (2020) reported that the corpus-based translation teaching enhanced learners' research thinking and practical translation skills. These findings align with the social constructivist principles (Vygotsky, 1978), where collaborative engagement with authentic tasks promotes cognitive scaffolding and procedural knowledge acquisition. In such case, the integration of corpus tools into PjBL frameworks within translation education receives popularity, in which corpora provide empirical linguistic data to guide translation decisions, while PjBL structures the application of this data in authentic, goal-oriented tasks.

Project-Based Learning and its Application in translation education

PjBL can trace back to the earlier 20st century, when scholars like John Dewey emphasized the importance of experiential learning in guiding learners' engagement and higher-order thinking. According to Krajcik and Shin (2014), there are six crucial elements of PjBL, namely, a driving question, the focus on learning goals, participation in learning practices, engagement in collaborative learning, scaffolded learning technologies, and the creation of tangible artifacts. With these features, PjBL has the prospects of exposing learners to both self-directed learning and collaborative problem-solving processes. From the lens of social

constructivism theory, PjBL engages learners in real-world projects, where meaningful learning occurs through social interaction and scaffolded guidance. Researchers have endorsed the efficacy of PjBL on students' learning such as *affective* outcomes (e.g. perceptions of the benefits of PjBL), *cognitive* outcomes (e.g. knowledge acquisition and higher-order thinking) and *behavioral* outcomes (e.g. skills and engagement) (Guo et al., 2020). Over the years, PjBL has been applied widely across different disciplines, such as engineering, business and economics, STEM, and language education (Dias-Oliveira et al., 2024; Greenier, 2020; Lavado-Anguera et al., 2024; Seo et al., 2024).

With the transformation of translation industry, the translation is no longer an individual work but requires translators to work collaboratively in large-scale projects. PjBL, through engaging translators in the authentic or simulated projects to create the real work environment, receives popularity in translation education. Li, Zhang and He (2015) conducted a project incorporated into business translation; reporting students not only enhanced their translation-specific competencies but other skills including teamwork and collaboration, critical thinking, presentation skills. Another study conducted in an Islamic university echoed that exposing learners to PjBL enhanced their translation performances and soft skills including problem-solving, communication and creativity (Putra et al., 2022). With the advent of digital technology, translation projects can now incorporate elements such as corpus-based translation technology, CAT tools, and online collaboration platforms. For instance, engaging student translators in an online crowdsourcing translation project reported students' significant gains in subtitling knowledge and skills (Tzou, 2024). The online PjBL approach integrated into the technical translation courses was found to enhance learners' self-regulation, and self-confidence in autonomous learning (Ribeiro et al., 2023). These technological advancements have further expanded the scope of PjBL in translation education.

Despite of the benefits of corpus-based pedagogy and PjBL in translation education, critical limitations persist as follows. First, research on corpus-aided translation training has predominantly focused on short-term interventions or isolated skill acquisition, with limited attention to how extended, collaborative projects influence translator trainees' holistic development. Second, existing studies have examined learners' perceptions on PjBL in translation classrooms (Mohammed, 2022; Ribeiro et al., 2023), however, teachers' authentic experiences in the implementation of PjBL are scarcely documented. These gaps collectively underscore the necessity of the present study, which not only advances theoretical discourse on technology-mediated translation education, but also provides practical references for designing future sustainable projects across different domains.

Methodology

Research Design

Due to the exploratory nature of the study, we adopted a qualitative research design to investigate the benefits and challenges from the perspectives of student translators and teachers. The study employed multiple data sources—including students' reflective journals, semi-structured interviews on both students and teachers—to triangulate findings and ensure the methodological rigor (Creswell & Poth, 2023) .

Participants and sampling

As participants were required to engage in translation and corpus construction tasks, we adopted purposeful sampling to recruit participants mainly from the *Translation and Interpreting Department*. Due to the long-term nature of the project, we recruited participants on a voluntary basis, involving 97 student translators (78% female, 22% male), 15 EFL instructors, together with three project supervisors and two technical experts to oversee quality control. Student participants were randomly assigned into 15 groups (5-8 students each), with one guiding instructor in each group. Ethical approval was secured prior to data collection, and informed consent was obtained from all participants, with assurances of anonymity and confidentiality.

Research procedure

Unlike most studies conducting PjBL activities integrated into a specific course, our study implemented this longitudinal large-scale project due to practical needs. In 2021, the target university received governmental funding support on the translation work of the *Transcripts of the proceedings of XXX Military Tribunals* (volume 11-40). Subsequently, the university decided to launch a corpus construction project, aiming to complete six different corpus covering 8300000 words (volume 11-25). The project aimed to

engage student translators in real-world translation work, help teachers acquire the authentic legal corpus for research work and improve teaching effectiveness, and promote historical responsibility. Prior to the implementation of the project, all participants attended a training session that covered:

a) The historical and legal significance of military tribunal transcripts.
b) The project objectives and expected outcomes.
c) Ethical considerations in handling sensitive legal documents.
d) Copyright and data security protocols.

To start the project, students were required to conduct the initial text proofreading for English PDF-to-WORD conversion. This work was very significant as the source texts were digitally scanned, which contained misspellings, omissions and ambiguous words. To strengthen the accuracy of the subsequent parallel corpus construction work, all participants constantly engaged in generating clean files with both original and target texts. Given to the complexity of the project, the procedure consisted of several phases starting from corpus preparation to finalized corpus (see Table 1).

Table 1.
The schedules and key activities of the project

Timeline	Key Activities & Deliverables
Phase 1 Text proofreading	Conduct initial text proofreading for English PDF-to-WORD conversion Identify misspellings, omissions, and ambiguous texts Complete evidence/witness lists and documented revisions in errata sheets Complete the proofreading and generate clean files in plain texts
Phase 2 Corpus development	Parallel text alignment using CAT tools Terminology extraction and glossary development Quality assurance through triple-checking system
Phase 3 Corpus enhancement	Text annotation for linguistic features (e.g., legal terminology, named entities) Metadata tagging for document properties Conduct inspection and quality assurance
Phase 4 Quality verification and corpus standardization	**Delivered artifacts:(six corpus covering 8300000 words)** ✓ Bilingual sentence-aligned parallel corpus (plain text) ✓ Bilingual sentence-aligned parallel corpus (with word segmentation) ✓ English and Chinese monolingual plain-text corpora ✓ English and Chinese monolingual plain-text corpora (with word segmentation) ✓ Bilingual TMX files Bilingual Trados translation memory

To provide the technical support for this large-scale project, the project organized more than 10 training workshops (online and offline) at regular intervals, covering initial text cleaning and formatting standardization, corpus alignment techniques, terminology management and quality assurance protocols. The project assigned one instructor to each group, mentoring them to complete assigned tasks and conduct quality assurance. During the process, student groups worked collaboratively to negotiate tasks, resolve problems such as language misunderstanding or technical barriers, and consult their supervisor in case of any confusion.

Materials and instruments

This study utilized the qualitative methods to collect data from three instruments to capture multidimensional perspectives. The raw data consisted of students' reflective journals and semi-structured interviews on both teachers and students.

Students' Reflective Journals

Students' reflective journals were collected at strategic intervals throughout the two-year project, and 62 students delivered the journals in the written form on voluntary basis, yielding 43896 words in total.

Student Interviews

The present study also conducted the semi-structured interview on five groups of 30 students via purposive sampling to represent diverse roles and proficiency levels. The interviews were audio-recorded and

transcribed verbatim, and the interviewees were informed of the anonymity and confidentiality. Sample questions were as follows:

- Why do you choose to participate in such a large-scale translation project?
- Have you developed any skill from participating in this project?
- Have you gained any insight in shaping your professional aspirations?
- Have you encountered any challenges during the process? If yes, how did you resolve such problems?

Teacher Interviews

To triangulate the data analysis, the present study also conducted a post-project interview on five teacher participants, including one project supervisor and one technical expert. The interview questions on teachers were mainly to collect their responses to:

- The pedagogical design of the corpus project;
- Their perceived alignment between project outcomes and curriculum goals;
- Challenges encountered in mentoring student translators during the project;
- Recommendations for implementing future scaling corpus-based projects.

Data Analysis

This study employed a qualitative methodology to examine the impacts of the large-scale project on student translators' learning outcomes. Following Braun and Clarke's (2006) thematic analysis framework, we systematically analyzed qualitative data from the above three data sources. Two researchers independently read and interpreted the data to generate categories and themes. To ensure methodological rigor, we implemented multiple validation strategies, including data triangulation across sources and consultation with external qualitative research experts in case of any disagreements. Due to the exploratory nature of the present study, our comprehensive analytical approach was suitable to develop nuanced understandings of both the transformative potential and implementation challenges of longitudinal, corpus-based translation projects in EFL contexts.

Results

RQ 1: How did participation in the corpus-construction project influence the competence development among student translators?

Through our thematic analysis, four core themes emerged from the qualitative data (see Table 2), which were presented below with supporting evidence from coded excerpts.

Table 2.
Students' perceived competence development

Competence development	Descriptions and examples
Translation skills and technological competence	▸ Mastery of legal English conventions, interdisciplinary knowledge and terminological precision ▸ Application of "faithfulness, expressiveness, elegance" translation principles and flexibility in selecting translation strategies ▸ Enhanced information literacy (grammar checks, tool utilization such as CAT)
Cognitive skills and critical thinking dispositions	▸ Analytical rigor, reasoning skills (seeking for linguistic equivalence and identifying translation errors) ▸ Open-mindedness and cognitive maturity
Collaboration skills and teamwork	▸ Peer support in encountering problems ▸ Effective knowledge sharing and collaboration (e.g., online collaborative platforms for real-time resolution) ▸ Negotiation in solving disagreements
Historical responsibility and ethical awareness	▸ Moral values as professional practitioners ▸ Ethical awareness in adhering to historical facts ▸ Unbiased decision-making in translating authentic events

Translation skills and technological competence

During the corpus cleaning and proofreading work, students had to compare and analyze whether the target texts aligned with the source texts, and they did a lot of revision work. The majority of students reported significant improvements in the translation skills and mastery of specialized terminology. Some excerpts were as follows:

There were many legal terminologies, long and complex sentences in the original texts. Translation strategies such as omission/addition and passive voice conversion were effective, which are critical for legal text fidelity and offer recommendations for my future translation work. (S23- student journal).

Through comparative analysis of Chinese and English texts, I improved my translation skills by initially identifying problematic areas in translations, consulting instructors with questions, and making revisions based on experts' feedback. (S7-interview-ref).

After phase one was completed (proofreading and text cleaning), students were trained and prepared for the corpus construction work. During this process, they improved their digital literacy and technical proficiency, including the mastery of CAT tools (e.g., Trados, AntConc, MemoQ) and AI applications.

I never used the CAT tools in my translation practices before. This project opens up new insights for me to acquire translation technology, which improves the translation efficiency and facilitates the large-scale project management. After completing the project, I tried to create my own bilingual corpus in political texts translation, which aid my future translation work. (S25- student journal).

I am glad that those students have not only improved their translation skills but also enhanced the digital literacy. The project engages learners in authentic tasks, requiring them to create specialized terminology, produce TMX files with certain tools, as well as master the CAT software. (T4-interview-ref)

Cognitive skills and critical thinking dispositions

When exposing to authentic translation tasks, student participants also developed their cognitive skills, including analytical rigor, independent reasoning and decision-making skills, as S6 stated in the interview: When addressing challenges like mistranslation, or conceptual gaps, we need to exercise preliminary judgment and strategic decision-making, which strengthened our logical reasoning capabilities and analytical precision in maintaining linguistic equivalence.

We learnt to identify and analyze sentences that violated translation standards such as over-translation or mistranslation and then offered rational revision notes, which significantly enhanced my analytical and reasoning skills. Gaining an in-depth understanding of case details enabled me to approach ambiguous sentences with greater logical reasoning, ultimately enhancing my translation accuracy and explanatory precision. (S32- student journal).

Additionally, the student participants not only improved their cognitive skills, but their critical thinking dispositions as well. As indicated by Teacher 2, those students had developed cognitive maturity in translation practices, through searching for multidimensional resources, and were open-minded in absorbing different perspectives in the peer review process.

Collaboration skills and teamwork

As this is a long-term project implemented on student translators, they perceived the significance of teamwork in improving project efficiency. Knowledge sharing on effectively detecting errors in the proofreading phase is significant, which saves them a lot of time. Additionally, they perceived that it significant to offer help to their peers and promptly respond to others. Some excerpts were selected from students' reflective journals:

We often encounter difficulties in the first phase; for example, how to detect the mismatch of original texts and translation texts, identify the punctuation errors, confirm the legal terminology, etc. At this time, posing the questions in the online chatting boards and seeking for help is a good choice. (S12)

Translation is not individual work, which requires teamwork when engaging in large-scale projects. The instructor asked me to act as the leader in my group, so I need to coordinate with my team members effectively. (S41).

The instructors also reported the enhanced collaboration skills among students. A case in point was resolving discrepancies in translating "*command responsibility*", students collaboratively analyzed existing parallel corpora, consulted the specialized databases, and negotiated solutions, thereby transforming disagreements or conflicts into consensus-building opportunities.

Historical Responsibility and Ethical Awareness

Through dealing with authentic military tribunal documents, student participants demonstrated a heightened sense of historical responsibility throughout the project. In their reflective journals, many emphasized the importance of accurately representing historical events. For example, one stated:

I realized that every word we translate and every correction we make in the corpus is a step towards preserving the true history of the military tribunals. A single error could potentially distort the original historical events, so we must be extremely serious regarding this challenging yet meaningful task (S16-student journal).

Numerous students, indicating a collective understanding that their work was not only involving linguistic transfer, but also related to moral values as professional practitioners, echoed this sentiment. During the process, many students have developed the sense of historical responsibility in handling the authentic corpus.

When translating sensitive contents related to war crimes, Student 16 noted, "It was emotionally challenging to translate the testimonies of those criminals, but I knew I had to set aside my personal feelings and translate accurately". This reflects an ethical approach to translation, where students recognized the significance of providing an unbiased account of history. In cases where historical terms had controversial translations, student participants also engaged in ethical decision-making processes. One student said, "Language equivalence is not difficult, but we had to consider the historical and cultural implications. We referred to multiple academic sources and consulted with teachers to ensure our translation was both accurate and respectful of historical facts" (S29-interview-ref). This shows how students took responsibility for their translation choices, going beyond surface-level translation to engage with the complex ethical and historical dimensions of the task.

Moreover, teachers themselves felt a sense of historical responsibility in guiding the project. Teachers' awareness of their role in shaping students' ethical and historical understanding was a significant factor in promoting these values within the project. One instructor responded,

> As mentoring instructors, we have the responsibility of guiding students to understand the importance of adhering to historical accuracy. Enhancing students' translation skills is not the only purpose, it is equally crucial to develop the sense of moral ethics towards facts (T5-interview-ref).

RQ 2: What challenges did instructors encounter in managing such a project?

While the corpus project yielded significant educational benefits, teacher interviews revealed three major challenges encountered in the implementation of such a longitudinal project. These obstacles including *the corpus difficulty and text ambiguity, technological barriers, and difficulties in team management*.

Regarding corpus difficulty, teachers pointed out that the historical archives as primary resources presented a multitude of difficulties. A wide variety of specialized terms made it challenging for both teachers and students to understand the source material accurately, which might affect the quality of the translation without specialized training. Just Teacher 3 stated.

The legal and military jargon in these documents was highly specialized. There were terms that were unique to the military tribunals of that era, and their meanings were not always clear even after consulting multiple dictionaries.

Additionally, some teachers responded that the original military tribunal transcripts were in a poor state. There were missing pages, faded texts, and unclear printings, which made it extremely difficult to piece together the complete story. As teacher 5 said: "taking texts in Volume 23 as example, there were several keywords where the printings were unclear, we had to rely on guesswork and cross-referencing with other sources to fill in the gaps". Therefore, this lack of completeness not only hindered the translation process but also raised concerns about the accuracy of the final corpus. Both student and teacher participants had to do a lot of work in the proofreading work, which required a lot of time and devotion to the project.

Technological challenges emerged as another major barrier, particularly during the early phases of the project. As not all teachers or students were familiar with the CAT tools, while it took some time for all participants to master specialized software required in the corpus construction work. Tools like ABBYY FineReader struggled with ambiguous evidentiary notes, exhibiting a large error rate for identifying original texts, while Trados SDL was incompatible with some original documents. Teacher 2 recalled the difficulties in bringing students up to speed: "We had to train students from zero on XML tagging for metadata—some initially marked entire pages as 'unclear' when encountering a single illegible character."

Perhaps the most persistent challenges occurred at the team management, where coordinating over so many participants with varying skill levels required innovative solutions. The two-year duration of this large-scale project led to motivational fluctuations, with task completion rates dropping from 92% in the first six months to 72% in the final six months as student fatigue set in. Nearly 30% of late-stage journals mentioned "*burnout from repetitive quality checks*". This lack of motivation and consistency in performance affected the overall efficiency and continuity of the project. Additionally, students with verifying competence levels and personalities made the team management work more challenging.

Collectively, these challenges extended the project timeline beyond initial projections. However, as several teachers noted, the struggles themselves became valuable learning experiences, which could provide references for future PjBL implemented on students.

Discussion

The implementation of the two-year corpus construction project for student translators provides a multifaceted lens to examine the interplay between project-based learning, translation pedagogy, and interdisciplinary skill development. While PjBL has been widely investigated in translation education (Garcia Gonzalez & Veiga Diaz, 2015; Moghaddas & Khoshsaligheh, 2019), its application in large-scale corpus construction remains underexplored. Unlike traditional PjBL approaches focusing on the design of appropriate teaching activities, our study engaged learners in authentic translation corpus, and interaction with CAT tools, machine translation software and AI applications in completing authentic tasks. Such human-machine interaction not only enhanced students' technology literacy, but improved their metacognitive awareness of translation knowledge and strategies as well. These findings echo the cognitive theory of multimedia learning, which proposes that multimodalities create conditions for learners to process contents in an effective way (Mayer, 2005; Shadiev et al., 2024). By aligning bilingual texts and annotating terminology, learners internalized patterns of language equivalence and divergence, reinforcing the interplay between declarative knowledge (e.g., legal terminology and sentence patterns) and procedural skills (e.g., corpus alignment).

The findings also indicate the development of *social* and *cognitive* skills (e.g. team work, critical thinking) among student participants, which align with the social constructivist approach positing that knowledge is constructed through authentic and socially mediated tasks (Vygotsky, 1978). When working towards the assigned team tasks, individuals need to share knowledge or offer help to their peers, which could strengthen their collaboration skills and teamwork. Further, those student translators need discuss, negotiate and reconcile different perspectives when encountering idea collision and disagreements during the project; which stimulated their critical thinking development (Alwafi, 2023).

Moreover, as our project involves translating military tribunal texts with legal jargon, historical contexts, and ethical dilemmas, students also developed their *historical responsibility and ethical awareness*. Taking authentic historical corpus as situated ethical practice, this study extends Baker's (2006) concept of "*translator activism*" to educational settings. For example, students' discussions on translating culturally sensitive terms (e.g., "justice" in cross-legal systems) highlighted their growing awareness of linguistic hegemony and their social responsibility to preserve historical nuance—a finding that aligns with recent

calls for "incorporating social responsibility in translation projects" (Muftah, 2024). From this lens, our study extends the traditional boundaries of translation education, highlighting its role in cultivating socially and ethically responsible translators.

Despite of the reported benefits, teachers faced challenges including *corpus ambiguity and difficulty*, *technological barriers* and *team management difficulty* during the project. Due to the long-term nature of this project, instructors had to manage quality control, project coordination, and guide learners acquire domain-specific knowledge and CAT tools. These findings highlight areas for improvement in curriculum design, training on both students and teachers, as well as institutional support on managing a large-scale, long-term project. The following section offers practical implications on initiating future corpus-based projects in translation education.

Practical implication

First, the corpus difficulty and technical anxiety experienced by students suggest the need for more systematic training integration within translation curricula. The novice translators could start from acquiring declarative knowledge (e.g. domain-specific terminology and conventions), and basic corpus use in foundational courses, before they move to complex bilingual alignment and terminology management tasks. In this way, students could develop the readiness to use CAT tools in translation, identify the errors generated by machine translation, and improve the quality of translation outputs (Loock & Léchauguette, 2021; Shadiev et al., 2024).

Second, regular motivation-boosting activities are necessary to combat student fatigue and ensure consistent project management. This could include progress-based rewards, such as recognition for meeting milestones or small incentives for outstanding work. For example, offering participants with rewards such as academic credits or certificates might provide tangible motivation for sustained student engagement throughout multi-semester projects. Additionally, establishing clear role specialization such as assigning students to focused tasks like corpus alignment, terminology management, might improve individual accountability while mitigating the coordination challenges encountered in our study.

Furthermore, the time-intensive nature of the project necessitates institutional support to ensure its sustainability. This includes providing access to a wide range of domain-specific corpus resources, such as historical and legal resources for corpus interpretation, investing in the latest corpus software and hardware. As proposed by faculty interviewees, the current CAT tools were not in alignment with the required corpus construction work; institutions could subscribe to specialized historical databases or purchase licenses for advanced CAT tools. Additionally, training on guiding instructors to guarantee the project quality is also necessary. For example, themed workshops, academic conferences, and online courses can help teachers to stay updated on the latest trends in translation technology, domain-specific knowledge, and project management knowledge.

Conclusions

This study provides compelling evidence for the educational values of a large-scale corpus construction project in developing multi-faceted competences among novice translators. Grounded in social constructivism, our study not only validates existing theoretical frameworks but also extends their application to specialized translation contexts. The success of the project highlights the unique capacity of authentic, ethics-grounded PjBL to address the complex demands of contemporary translator education. However, the implementation of such a time-consuming project also encounters some challenges, which underscores the importance of elaborate project design and institutional support. The recommendations proposed in this study suggest pathways for enhancing both the effectiveness and accessibility of corpus-based learning models. If properly designed and supported, the large-scale, collaborative projects have the potential to prepare translators for the handling complexities of their profession in future authentic translation work. By analyzing the perceived benefits and challenges in PjBL environments, this research contributes to both theoretical understanding and practical applications in translation education and beyond.

This study bears some potential limitations. First, the single-institution context, while allowing for in-depth qualitative analysis, limits the generalizability of the findings across different cultural and educational systems. It suggests future longitudinal tracking of participants' retention and practical applications of skills developed through the large-scale project. It also recommends that cross-cultural comparative studies might

offer new insights for future investigation. Replicating similar corpus projects in different contexts, for instance, working with other countries might test the generalizability of our findings. Second, though we use multiple data resources to triangulate our results, the qualitative method for assessing skill development may arise some potential bias. Future studies could adopt mixed methods by adding quantitative assessments, such as pre- and post-test instruments on translation competence or standardized tests of technical proficiency, to complement participants' perceived skills development.

Bibliographic references

Alhassan, A., Muhammad Naguib Sabtan, Y., & Omar, L. (2021). Using Parallel Corpora in the Translation Classroom: Moving towards a Corpus-driven Pedagogy for Omani Translation Major Students. *Arab World English Journal*, *12*(1), 40–58. https://doi.org/10.24093/awej/vol12no1.4

Alwafi, E. M. (2023). The impact of designing an online learning environment based on cognitive apprenticeship on students' critical thinking and interaction in CSCL. *Educational Technology Research and Development*, *71*(2), 2. https://doi.org/10.1007/s11423-022-10180-2

Baker, M. (2006). Translation and Activism: Emerging Patterns of Narrative Community. *The Massachusetts Review*, *47*(3), 462–484.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Creswell, J. W., & Poth, C. N. (2023). *Qualitative Inquiry and Research Design* (5th ed.). SAGE Publications, Inc. https://uk.sagepub.com/en-gb/eur/qualitative-inquiry-and-research-design/book266033

Dias-Oliveira, E., Pasion, R., Vieira da Cunha, R., & Lima Coelho, S. (2024). The development of critical thinking, team working, and communication skills in a business school–A project-based learning approach. *Thinking Skills and Creativity*, *54*, 101680. https://doi.org/10.1016/j.tsc.2024.101680

Garcia Gonzalez, M., & Veiga Diaz, M. T. (2015). Guided Inquiry and Project-Based Learning in the field of specialised translation: A description of two learning experiences. *Perspectives-Studies in Translation Theory and Practice*, *23*(1), 107–123. https://doi.org/10.1080/0907676X.2014.948018

Greenier, V. T. (2020). The 10Cs of project-based learning TESOL curriculum. *Innovation in Language Learning and Teaching*, *14*(1), 27–36. https://doi.org/10.1080/17501229.2018.1473405

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research*, *102*, 101586. https://doi.org/10.1016/j.ijer.2020.101586

Krajcik, J. S., & Shin, N. (2014). Project-Based Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 275–297). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.018

Lavado-Anguera, S., Velasco-Quintana, P.-J., & Terrón-López, M.-J. (2024). Project-Based Learning (PBL) as an Experiential Pedagogical Methodology in Engineering Education: A Review of the Literature. *Education Sciences*, *14*(6), 6. https://doi.org/10.3390/educsci14060617

Li, D., Zhang, C., & He, Y. (2015). Project-based learning in teaching translation: Students' perceptions. *The Interpreter and Translator Trainer*, *9*(1), 1-19. https://www.tandfonline.com/doi/abs/10.1080/1750399X.2015.1010357

Li, M., & Pan, D. (2023). Corpus-Based Translation Pedagogy: A Preliminary Case Study. In H. Xie, C.-L. Lai, W. Chen, G. Xu, & E. Popescu (Eds.), *Advances in Web-Based Learning – ICWL 2023* (pp. 28–38). Springer Nature. https://doi.org/10.1007/978-981-99-8385-8_3

Loock, R., & Léchauguette, S. (2021). Machine translation literacy and undergraduate students in applied languages: Report on an exploratory study. *Revista Tradumatica*, (*19*), 19. https://doi.org/10.5565/rev/tradumatica.281

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004

Moghaddas, M., & Khoshsaligheh, M. (2019). Implementing project-based learning in a Persian translation class: A mixed-methods study. *The Interpreter and Translator Trainer*, *13*(2), 190–209. https://doi.org/10.1080/1750399X.2018.1564542

Mohammed, T. A. S. (2022). The Use of Corpora in Translation into the Second Language: A Project-Based Approach. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.849056

Muftah, M. (2024). Incorporating social responsibility into translator training through situated learning in translation projects: Making room for a sustainable learning environment. *Interactive Learning Environments*, *32*(9), 4950–4967. https://doi.org/10.1080/10494820.2023.2207188

Putra, H. R., Retnaningsih, W., & Nugroho, A. (2022). Enhancing Students' Translation Skills using Project Based Learning: A Case of An Islamic University. *Ta'dib: Journal of Islamic Education*, *26*(2), 93–106. https://doi.org/10.19109/td.v26i2.10006

Ribeiro, S., Tavares, C., Lopes, C., & Chorão, G. (2023). Competence Development Strategies after COVID-19: Using PBL in Translation Courses. *Education Sciences*, *13*(3), 3. https://doi.org/10.3390/educsci13030283

Seo, S., Van Orman, D. S. J., Beattie, M., Paxson, L. C., & Murray, J. (2024). Transforming Learning Orientations Through STEM Interdisciplinary Project-Based Learning. *Education Sciences*, *14*(11), 11. https://doi.org/10.3390/educsci14111154

Shadiev, R., Chen, X., & Altinay, F. (2024). A review of research on computer-aided translation technologies and their applications to assist learning and instruction. *Journal of Computer Assisted Learning*, *40*(6), 3290–3323. https://doi.org/10.1111/jcal.13072

Tzou, Y.-Z. (2024). Situated learning in the translation class: Exploring the effects of engaging college students in an online crowdsourcing translation project. *Interpreter and Translator Trainer*, *18*(4), 563–580. https://doi.org/10.1080/1750399X.2024.2423476

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Xiao, R., & Hu, X. (2015). *Corpus-Based Studies of Translational Chinese in English-Chinese Translation*. Springer.

Zhao, W., & Mu, Y. (2020). An Empirical Study of Corpus-Based Translation Teaching in Higher Vocational Colleges in China. In E. Popescu, T. Hao, T.-C. Hsu, H. Xie, M. Temperini, & W. Chen (Eds.), *Emerging Technologies for Education* (pp. 280–284). Springer International Publishing. https://doi.org/10.1007/978-3-030-38778-5_30