International Journal of Pedagogics

Integrating Corpus Linguistics for Teaching Pre-Service English Teachers by Using Corpus Tools

Sultonova Munojat Yunusovna

PhD researcher at Fergana State University, Foreign Languages Department, Uzbekistan

Abstract: This study explores the integration of corpus linguistics tools in the training of pre-service English teachers, focusing on enhancing their linguistic competencies and pedagogical effectiveness. By employing tools such as LancsBox X, AntConc, and SKELL, the research examines how these resources can facilitate the development of professional communication skills in English within higher education contexts. The findings indicate that corpus-based approaches offer significant advantages in understanding language use, improving teaching methodologies, and addressing regional linguistic characteristics. The study proposes practical recommendations for incorporating corpus linguistics into teacher training programs to foster a more interactive and data-driven learning environment.

Introduction: The rapid advancement of technology and the increasing availability of linguistic data have transformed language teaching methodologies. Corpus linguistics, which involves the study of language through large collections of texts (corpora), has emerged as a valuable tool in this transformation. For pre-service English teachers, integrating corpus tools into their training can enhance their understanding of language patterns, usage, and nuances, thereby improving their teaching effectiveness. This article investigates the application of corpus linguistics tools—specifically LancsBox X, AntConc, and SKELL—in the development of linguistic competencies among philology students in higher educational institutions.

METHOD

Theoretical Framework

Corpus linguistics provides an empirical basis for language instruction by offering authentic language data. In the context of teaching English to pre-service teachers, corpus tools can support the development of various linguistic competencies, including vocabulary acquisition, grammatical understanding, and discourse analysis. These tools enable students to engage with real-world language use, fostering a deeper comprehension of language structures and functions.

Heather and Helt (2012) conducted a study to evaluate the effectiveness of corpus literacy training for pre-service language teachers. The research involved six case studies of pre-service language teachers enrolled in a semester-long introductory grammar course that integrated corpus linguistics. The study aimed to assess the development of various components of corpus literacy among the participants. The findings indicated that while the corpus literacy training was generally effective, its impact varied among the participants. The authors identified several factors contributing to this variation, including individual differences in prior knowledge, learning styles, and the extent to which the corpus tools were incorporated into teaching practices. Based on these observations, Heather and Helt suggested several practices for teacher educators planning or modifying instruction in corpus literacy, emphasizing the need for tailored approaches that consider the diverse needs of pre-service language teachers.

Farr (2008) grounds her study in the framework of language awareness and data-driven learning (DDL), where learners develop deeper insights into language structure and use through direct engagement with

authentic language data. Drawing on the work of Johns (1991) and others in corpus linguistics, Farr emphasizes that corpus-based instruction empowers both teachers and learners to explore real-life language patterns independently. In teacher education, this approach supports the development of critical pedagogical thinking, allowing future educators to make informed, evidence-based decisions about language teaching. Farr highlights that for corpus-informed methods to be effective, they must be aligned with teacher training goals and supported by institutional and technical resources.

The integration of corpus linguistics into language teacher education is underpinned by several key theoretical perspectives that inform both pedagogical practice and linguistic development. At the core of this study is the concept of Data-Driven Learning (DDL), first introduced by Johns (1991), which advocates for learner-centered discovery through direct interaction with authentic language data. DDL posits that by engaging with real language examples, learners can inductively identify patterns, structures, and usage, fostering a more profound and autonomous understanding of language.

In the context of teacher education, corpus literacy encompasses a multifaceted skill set. Callies (2019) identifies several components of corpus literacy:

-recognizing what a corpus is, the types of corpora available, and the capabilities and limitations of corpus analysis;

-utilizing corpus software tools, such as concordancers, to search corpora and analyze the retrieved data;

-using corpus-derived insights to generate teaching materials and activities that reflect authentic language use.

Callies emphasizes the importance of exposing pre-service teachers to corpora and corpus methods from multiple perspectives: researcher, teacher, and learner. This approach not only enhances their understanding of language but also equips them with practical tools to incorporate corpus-based methods into their teaching practices.

Furthermore, the theoretical framework draws on constructivist pedagogical principles, which advocate for active, learner-centered approaches to teaching and learning. By engaging pre-service teachers in hands-on corpus analysis, they can construct their own understanding of language patterns and usage, leading to deeper learning and more effective teaching strategies. Finally, integrating corpus linguistics into teacher education, as outlined by Callies, provides a comprehensive framework that enhances linguistic competencies, fosters critical thinking, and prepares pre-service teachers for the demands of modern language instruction.

Methodology

The study employed a qualitative approach, incorporating the use of corpus tools in classroom settings. Pre-service English teachers were introduced to LancsBox X, AntConc, and SKELL through guided tutorials and practical exercises. Data were collected through pre- and post-intervention assessments, student feedback, and classroom observations to evaluate the impact of these tools on their linguistic competencies and teaching practices. A case study approach was used to capture rich, contextualized insights into the participants' experiences and development over the course of the intervention.

The participants were a cohort of pre-service English language teachers enrolled in a teacher training program at a higher educational institution. These students were in the final year of their philology degree and had prior exposure to foundational linguistic and pedagogical concepts. The study was embedded within a course on professional communication and English for Academic Purposes (EAP), making it relevant and directly applicable to their future teaching contexts. Students were introduced to these tools through guided tutorials, video demonstrations, and in-class hands-on tasks. The instructional design followed a scaffolded approach—beginning with tool familiarization, progressing to data interpretation, and culminating in the application of corpus data in microteaching and lesson planning activities.

Sample instructions:

It is suggested to open LancsBox and load the provided corpus of academic texts (e.g., journal articles, EAP sample essays, or research reports) and to use the KWIC (Key Word in Context) tool to search the following reporting verbs:

argue   suggest   claim   indicate

For each verb: Examine 20–30 concordance lines.

Note patterns of usage (e.g., Smith (2019) argues that…, or The results suggest that…). Pay attention to:

-the subject (Who is reporting?)

-the grammatical structure (Verb + that-clause, Verb + to + infinitive)

-the level of certainty or tentativeness

Verb	Common Structures	Function in Text	Degree of Certainty
argue	X argues that…	Presenting a strong opinion	High
suggest	X suggests that…	Offering a tentative idea	Medium
claim	X claims that…	Stating a belief (possibly biased)	Medium–High
indicate	Results indicate that…	Showing evidence from data	High (evidence-based)

Table 1. Sample exercise for using corpus in teaching English.

There are some corpus tools given as the following during the research: a) LancsBox X: A powerful corpus analysis tool that supports the analysis of large corpora, such as the British National Corpus 2014. It offers functionalities like concordancing, word frequency analysis, and collocation studies, which are essential for understanding language patterns and structures; b) AntConc: A user-friendly concordance tool that allows for the analysis of text corpora to identify word frequencies, concordances, and collocations. It is particularly useful for examining specific language features and their usage in context; c) SKELL: An online tool that provides example sentences and collocations for words, aiding in the understanding of word usage and meaning in context. It is designed to be accessible to both teachers and students, facilitating the integration of corpus data into language teaching.

These assessments evaluated the students' baseline and post-training levels of linguistic awareness, particularly in areas such as vocabulary use, collocations, and error correction strategies. Student Feedback and Reflective Journals: Participants were asked to maintain reflective journals and complete feedback forms to document their learning experiences, challenges, and perceptions of corpus tools. The instructor-researcher conducted non-intrusive observations throughout the sessions to monitor engagement, tool usage, and collaborative learning dynamics. These artifacts provided evidence of how participants applied corpus insights into teaching content, particularly in grammar instruction, vocabulary development, and EAP-focused activities.

The qualitative data were analyzed using thematic analysis. Patterns were identified across student reflections, teaching artifacts, and observational notes. The analysis focused on changes in corpus literacy, perceived value of the tools, challenges faced, and shifts in pedagogical thinking. Attention was also paid to the psycholinguistic aspects of language processing observed in students' evolving ability to identify patterns and usage through corpus evidence.

RESULTS AND DISCUSSION

The integration of corpus tools into the training of pre-service English teachers yielded several positive outcomes. Students demonstrated a better understanding of language patterns, collocations, and grammatical structures, leading to more accurate language use. Teachers incorporated corpus-based insights into their lesson plans, providing students with authentic language examples and fostering a more interactive learning environment. The use of diverse corpora allowed for the inclusion of regional language variations, promoting cultural awareness and sensitivity in language instruction.

However, challenges such as the initial learning curve associated with these tools and the need for adequate training resources were noted. Addressing these challenges through comprehensive training and support can further enhance the effectiveness of corpus tools in teacher education.

CONCLUSION

Integrating corpus linguistics tools into the training of pre-service English teachers offers significant benefits in enhancing linguistic competencies and improving teaching methodologies. Tools like LancsBox X, AntConc, and SKELL provide valuable insights into authentic language use, supporting the development of effective and culturally sensitive teaching practices. By incorporating these tools into teacher training programs, educational institutions can better prepare future educators to meet the demands of contemporary language instruction.

REFERENCES

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). #LancsBox v. 5.x. [software]. Available at: http://corpora.lancs.ac.uk/lancsbox.

Anthony, L. (2006). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. Proceedings of the 2006 JALT CALL Conference.

Baisa, V., & Suchomel, V. (2014). SkELL: Web Interface for English Language Learning. Proceedings of the Eighth Workshop on Recent Advances in Slavonic Natural Language Processing.

Callies, M. (2019) Integrating corpus literacy into language teacher education. In S. Götz & J. Mukherjee (Eds.), Learner Corpora and Language Teaching (pp. 245–263). Amsterdam, Philadephia: John Benjamins.

Farr, F. (2008). Evaluating the Use of Corpus-based Instruction in a Language Teacher Education Context: Perspectives from the Users. Language Awareness, 17(1), 25–43.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. Language Learning, 67(S1), 155–179.

Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. Journal of Technology and Teacher Education, 20(4), 415–440.