

CORPUS-BASED LANGUAGE TEACHING PEDAGOGY

Toshpulatova Mexriniso Kilichevna

Associated professor of Termez State University

Abstract

Despite the increasing prevalence of corpus linguistics in scholarly circles in recent years, its application in teaching remains largely unexplored in primary and secondary education. Building on Shulman's notion of pedagogical content knowledge, this research delineates the distinction between two fundamental concepts — corpus literacy and corpus-driven language pedagogy — and investigates how a cohort of teacher-training students in TESOL (Teaching English to Speakers of Other Languages) developed their corpus literacy skills and acquired proficiency in corpus-based instructional methods through a two-stage training program.

Keywords: Corpus Linguistics, Data-Driven Learning (DDL), Corpus-based Approach, Authentic Language Usage, Concordance Lines, Learner Corpus, Pedagogical Corpus, Reference Corpus.

The findings indicate that the majority of participants achieved a satisfactory level of corpus literacy, as assessed through a self-designed questionnaire. Additionally, they demonstrated a commendable level of proficiency in implementing corpus-informed teaching strategies, as evidenced by the evaluation of their lesson plans and the analysis of lesson content and interview data. The findings support a clear distinction between corpus literacy and corpus-driven language pedagogy, providing evidence for the efficacy of a two-stage corpus-based teacher education program. The research offers valuable insights into how teachers can be guided through corpus-based instruction and how students can be taught to navigate corpus resources in order to address their lexical needs and challenges. Furthermore, several considerations are raised regarding the implementation of effective corpus-based teaching strategies in educational settings.

Key word: *Corpus-based language pedagogy, corpus literacy; teacher training; vocabulary learning; online collaborative learning*

Introduction

Research has demonstrated that corpus linguistics represents a groundbreaking and effective method for studying and analyzing language. According R. Reppen within the context of English language teaching, corpus linguistics holds immense potential for assisting educators in designing instructional activities.

Nonetheless, the use of a corpus-driven linguistic approach remains relatively unknown to the majority of the professional teaching community due to various factors, including the lack of exposure to corpus learning during teacher training, teachers' perception of corpus linguistics as primarily associated with research rather than pedagogical applications, and teachers' challenges in mastering corpus technology.

A crucial concept in the realm of corpus-driven teacher education is corpus literacy (CL), first introduced by Mukherjee (2006) and later defined by Heather and Helt (2012) as the capacity to employ corpus linguistics tools for exploring language and enhancing student language development (p.417). Few studies have explored the development of CL among teachers. Notable exceptions include the works by Heather and Helt (2012), Leńko-Szymańska (2014), and Zareva (2017).

Another concept largely overlooked by researchers is CBLP, which is analogous to Shulman's pedagogical content knowledge. CBLP refers to the integration of corpus linguistics technology into language pedagogy in the classroom, facilitating language teaching. This study introduces CBLP as a means of effective corpus-based teacher training, supported by the current research.

Mukherjee and John stated that Corpus linguistics has firmly established itself as a groundbreaking and highly effective approach to the study of language. The corpus-based method of learning languages is regarded as revolutionary due to

the authenticity and depth of its data, as well as its potential to empower learners as active participants in their own learning process

The use of corpora fosters autonomous discovery and personalized learning experiences (Boulton & Cobb, 2017; McEnery & Wilson, 1997). Boulton and Cobb's meta-analysis (2017) revealed that the impact of corpus use on learners' language proficiency was substantial, with effect sizes ranging from 0.95 to 1.50 (Cohen's *d*). This finding is highly encouraging, indicating the effectiveness of the corpus-based approach.

Boulton, Callies stated that for these reasons, corpus linguistics has emerged as a mainstream tool in language learning and research. However, its application in language teaching remains relatively limited Chambers, Granger, and Mukherjee all highlight the potential discrepancy between the enthusiasm of corpus linguists for the pedagogical application of corpora and the reluctance of average teachers to incorporate them into their classroom practices. Mukherjee specifically notes that there may be a disconnect between the research conducted by corpus linguists and the reality of language teaching, particularly among non-specialist teachers (p. 7). Chambers's perspective echoes this view, stating that there is a gap between the findings of corpus linguistic research and the practices of non-corpus linguist language teachers . Several factors contribute to this discrepancy.

Firstly, corpus linguistics research often focuses on advanced educational settings where learners already possess a high level of proficiency in English. At this level, concerns tend to center around the development of advanced linguistic skills required for academic writing in English, as highlighted by Charles (2014) and others.

Crosthwaite, 2017, 2020; Lee & Swales, 2006; Thurston & Candlin, 1998) Thus, few studies have targeted school settings or lower proficiency learners. Second, corpus-based learning activities intended for university-level students typically focus on the use of concordance lines, which may be considered too difficult for low-level school learners (Caliskan & Gönen, 2018; Poole, 2020).

Third, teachers may encounter various problems, including insufficient hardware and a scarcity of free and user-friendly corpus resources (Chambers, 2019), and they may also have limited computing skills (Römer, 2010; Tribble, 2012). Finally, as revealed by a few studies, teachers are reluctant to exploit corpora because of a lack of corpus-based teacher training (Boulton, 2017; Breyer, 2009; Callies, 2019; Chambers, 2019; Leńko-Szymańska, 2017). Without adequate preparation, educators may become disinclined to engage with corpora, leading them to perceive the use of corpora as a research endeavor limited to advanced academic settings (Boulton, 2017, p. 483).

Corpus-based teacher training

In contrast to the plethora of empirical investigations into the use of corpora for language learning, there is a dearth of empirical studies specifically focused on teacher training in this area. Corpus-based instruction is either integrated into study programmes as newly developed courses (e.g., Breyer, 2009; Callies, 2019; Leńko-Szymańska, 2017), or incorporated into existing courses within the framework of a programme (Farr, 2008; Heather & Helt, 2012; Zareva, 2017).

Nevertheless, there are several critical aspects that must be taken into account when designing effective corpus-based teacher training programmes. A fundamental question arises: what knowledge and skills should be prioritized? Mukherjee (2006) introduced the concept of "corpus literacy" (CL), encompassing four dimensions:

1. Understanding the nature of a corpus.
2. Awareness of what can and cannot be achieved with a corpus.
3. Proficiency in analysing corpus data.
4. Ability to draw meaningful conclusions about language usage based on corpus analysis.

These four dimensions strive to elucidate the intricacies of employing and scrutinizing corpus data as an educational instrument. Recently, Leńko-Szymańska (2017) delineated three essential competencies in teacher education: technical

proficiency, corpus linguistic expertise, and pedagogical acumen, encompassing the art of designing corpus-driven instructional materials and executing pedagogical practice. While the knowledge and abilities proposed by various researchers exhibit a certain degree of overlap, a closer examination reveals that they can be categorized into two broad domains:

1. The ability to effectively employ corpus data as an educational resource.
2. The integration of corpus-based resources into language instructional practices.

To clarify, the first domain aligns with Mukherjee's concept of CL, emphasizing the use of corpora for learning purposes. We refer to the second domain as corpus-based language pedagogy (CBLP), enabling educators to utilize corpora as an instructional tool.

While training in CL has yielded varying degrees of success for teachers, leading them to incorporate corpora into their learning practices, their understanding of the pedagogical implications of corpora remains limited within the classroom context (e.g., Breyer, 2009; Lęko-Szymańska, 2017; Naismith, 2017). It is our hypothesis that the CBLP has been overlooked, or at the very least, has not been as successful in the corpus-based training provided to educators. The pedagogical competence necessary for incorporating corpus resources into classroom instruction is of paramount importance and may be linked to what Shulman proposed in his works from 1986 and 1987. Shulman defined pedagogical content knowledge as a unique amalgamation of content and pedagogical principles that is exclusively the domain of educators and their specialized form of professional comprehension. If CL emphasizes the ability to master the use of corpus tools as part of content knowledge, CBLP underscores the significance of integrating CL within language pedagogy within real-world classroom settings. This requires careful consideration of how to represent, articulate, and integrate the knowledge and skills associated with corpus linguistics, as well as designing appropriate teaching activities that incorporate this knowledge.. Thus, CBLP can be defined as the capacity to integrate corpus linguistic technologies into

classroom language instruction in order to enhance language teaching. In this context, trainees must acquire a certain degree of CL proficiency in order to conduct corpus searches and analyses, as well as to generate practical strategies for conveying CL concepts to their students in order to facilitate their learning through corpora. If pedagogical content knowledge represents the missing link between content and pedagogy, as proposed by Shulman (1986), CBLP acts as a bridge between CL and instruction in language classrooms. It provides a novel theoretical framework for effectively organizing corpus-based training programs for ESL/EFL instructors, incorporating a wide range of cutting-edge English language teaching activities. Secondly, another crucial aspect related to teacher training based on corpus data concerns the specific areas of content that should be prioritized. Ideally, every aspect of language learning, including vocabulary, grammar, discourse analysis, writing skills, reading comprehension, and listening comprehension, should be incorporated.

However, among the limited empirical studies on teacher training, all have focused primarily on grammar, as demonstrated by Breyer (2009), Farr (2008), Heather and Helt (2012), and Zareva (2017). This is understandable, as grammar is often considered less complex compared to other aspects of language learning, particularly in terms of the quantity and diversity of information.

Nevertheless, recent decades have seen significant advancements in corpus linguistics, leading to valuable findings in the fields of vocabulary, collocation, and phraseology, exemplified by works such as those by McCarthy and Carter (2004), Schmitt (2004), and Sinclair (1991). These findings have provided invaluable insights for English language teachers when it comes to developing vocabulary skills. Additionally, vocabulary acquisition is a critical aspect for both learners and educators alike, as emphasized by Ma (2009). Consequently, there is a pressing need for further investigation into corpus-driven vocabulary instruction for educators, as this approach holds the potential to have both practical and motivational implications for both instructors and learners.

Thirdly, there is a paucity of systematic assessments regarding the learning outcomes of studies exploring corpus-based teacher education (Farr, 2008; Heather & Helt, 2012), and what exists relies on surveys or qualitative data such as reflective essays, interviews, and document analysis (e.g., Breyer, 2009; Farr, 2008; Heather & Helt, 2012). These investigations typically measure the participants' perceptions or attitudes. However, what is notably absent is a comprehensive evaluation of the progress made in terms of proficiency in working with corpus data (CL) or the ability to integrate CL into language teaching in the classroom (CBLP). While Leńko-Szymańsка's (2017) study provides a noteworthy example of analyzing corpus-based lessons created by trainee teachers, the analysis was primarily qualitative, focusing on identifying features specific to corpus-based lesson designs created by trainees.. The present study builds upon previous research on corpus-assisted teacher training, such as that conducted by Callies (2019), Heather and Helt (2012), Leńko-Szymańsка (2014 and 2017), and Zareva (2017). It expands upon this body of work in several significant ways.

Firstly, the study establishes a framework that distinguishes between CL (Corpus Linguistics) and CBLP (Corpus-Based Language Pedagogy). This framework also proposes a two-stage training model. This approach provides a theoretical basis for designing effective corpus-assisted training programs for educators.

Secondly, the research delves into how pre-service teachers design corpus-informed vocabulary lessons tailored to meet the specific needs of Chinese EFL learners. The findings may provide valuable insights into the effective utilization of corpus resources for vocabulary instruction in primary and secondary school settings.

Thirdly, this study proposes the implementation of online collaborative learning to enhance participant engagement and interaction.

Finally, the study conducts a comprehensive evaluation of participants' learning outcomes in CL and assesses a subset of their CBLP through quantitative

and qualitative analysis. Given the complexity of measuring pedagogical content knowledge (Hill et al., 2007; Morrison & Luttenegger, 2015; Shulman, 1988), the current research is limited in its ability to fully gauge CBLP. Instead, only the corpus-informed lesson plans developed by student teachers were analyzed to gauge their initial development of CBLP skills.

Discussion and implications

In the context of corpus-based instruction, scaffolding teachers is essential due to the complexity of such training. Participants in this type of training need to be supported in various ways, as research on corpus applications typically focuses on learners rather than teachers.

It is crucial for teachers to have hands-on experience with corpus searching in a step-by-step manner before teaching students how to utilize corpus tools (Breyer, 2009; Johns, 1991; Mukherjee, 2004). Additionally, teachers can be supported through various strategies to assist them in integrating their acquired corpus knowledge into classroom activities. To this end, they should receive access to a variety of pedagogical resources, such as a dedicated website that includes design principles for corpus-informed lessons and samples of such lessons.

In our study, the majority of groups followed either an exact or slightly modified version of a four-step framework in designing their lessons. Apart from individual work, collaborative efforts promote interaction and ideation among participants, thereby facilitating the exchange of thoughts in the development of teachers' pedagogical knowledge (Shulman, 1987). Furthermore, it helps to alleviate the monotony often associated with working with concordance data .

In essence, providing educators with a diverse range of learning experiences, including online collaboration, can infuse vitality and empowerment into corpus-based training for teachers.

The outcomes of collaborative learning activities have demonstrated that the participants in the teacher training program have developed crucial perspectives that contribute to their professional growth in the field of content-based language

pedagogy (CBLP). Specifically, these perspectives were evident in the stages of "comprehension" and "transformation" within Shulman's model of pedagogical content knowledge development.

Furthermore, the data collected through interviews indicate that online collaborative learning provides a valuable framework for fostering pedagogical consciousness among the trainees. This consciousness encompasses their ability to critically reflect on how to effectively incorporate and adapt corpus data in their teaching practices. These aspects of pedagogical thinking and practical considerations are integral components of their overall development as CBLP professionals.

The current research posits that CL (corpus linguistics) and CBLP (corpus-based language pedagogy) represent distinct concepts. While CL pertains to the acquisition of corpus linguistic skills, encompassing proficiency in search and analysis, CBLP entails the integration of these skills into a practical teaching environment, drawing upon the framework of Shulman's pedagogical content knowledge .

Building on this conceptual distinction, the study proposes a two-stage approach to corpus-based teacher education, which was implemented with a cohort of TESOL students preparing to teach English to Chinese learners in primary and secondary school settings. The findings indicate that the participants self-assessed their proficiency in the various dimensions of CL as being reasonably well-developed. Furthermore, they acknowledged the advantages of utilizing corpora, including access to authentic language samples and the acquisition of collocations.. The assessment of the lessons designed by the participants, as well as their analysis, has revealed that most of the groups were able to create a suitable corpus-based lesson, demonstrating a deep understanding of how to incorporate corpus resources into language teaching in the classroom.

The evidence supporting the participants' development of pedagogical content knowledge (CBLP) was particularly evident during the early stages of the

process, namely, during the phases of "comprehension" and "transformation" (Shulman, 1987). This study provides support for the formulation of two distinct concepts, CL and CBLP, which contribute to the development of corpus-based training for teachers.

Given that the concept of CBLP is relatively new, further research is necessary to explore and define the specific knowledge dimensions that teachers require to effectively integrate corpus resources into their teaching practices.

References

1. Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching*, 50(4), 483–506. doi:10.1017/S0261444817000167
2. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. doi:10.1111/lang.12224
3. Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172. doi:10.1080/09588220902778328
4. Caliskan, G., & Gönen, S. I. K. (2018). Training teachers on corpus-based language pedagogy: Perceptions on vocabulary instructions. *Journal of Language and Linguistic Studies*, 14(4), 190–210.
5. Callies, M. (2019). Integrating corpus literacy into language teacher education. *Learner Corpora and Language Teaching*, 92, 245.
6. Dever, M. T., & Hobbs, D. E. (1998). The learning spiral taking the lead from how young children learn. *Childhood Education*, 75(1), 7–11. doi:10.1080/00094056.1998.10521968
7. 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. doi:10.2167/la414.0
8. Granger, S. (2015). Data-driven learning and language learning theories. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language*

corpora for data-driven learning(pp. 486–510. Amsterdam, The Netherlands: John Benjamins.

9. Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. doi:10.1016/j.esp.2005.02.010

10. Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260–278. doi:10.1017/S095834401400010X

11. Leńko-Szymańska, A. (2017). Training teachers in data driven learning: Tackling the challenge. *Language Learning & Technology*, 21(3), 217–241.

12. Ma, Q. (2009). *Second language vocabulary acquisition*. Bern, Switzerland: Peter Lang. McEnery, T., & Wilson, A. (1997). Teaching and language corpora (TALC). *ReCALL*, 9(1), 5–14. doi:10.1017/S0958344000004572

13. Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective*(pp. 239–250). Amsterdam, The Netherlands: Rodopi.

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

15. Naismith, B. (2017). Integrating corpus tools on intensive CELTA courses. *ELT Journal*, 71(3), 273–283.