Advances in Curriculum Design&Education

WISDOM ACADEMIC ISSN: 3080-7352 | E-ISSN: 3080-7360 *Volume 1 , Issue 3*

Article

The Implementation of Corpus-based Language Pedagogy by Teachers in Chinese University

Jinyi Song[1], Meiying Wu[1,*]

[1]Monash University, Clayton 3800, Australia.

[*]**Corresponding author:** Meiying Wu, 18799642131@163.com.

CITATION

Song JY & Wu MY. The Implementation of Corpus-based Language Pedagogy by Teachers in Chinese University. Advances in Curriculum Design&Education. 2025; 1(3): 128.

https://doi.org/10.63808/acde.v1i3.128

ARTICLE INFO

Received: 24 June 2025

Accepted: 20 November 2025

Available online: 21 November 2025

Abstract: This study investigates the implementation of corpus-based language pedagogy (CBLP) in Chinese university EFL contexts, addressing the gap between corpus linguistics research and classroom practice. Despite extensive evidence supporting the effectiveness of corpus-based instruction for vocabulary acquisition through authentic data exposure and data-driven learning, Chinese EFL teachers demonstrate limited adoption of corpus tools due to insufficient training, technical challenges, and contextual constraints. The research examines four strategic approaches for effective CBLP implementation: integration of corpus approaches with existing language curricula to enhance vocabulary learning through concordance analysis and collocation studies; diversification of instructional formats including printed materials, visual aids, and collaborative activities to maintain student engagement; provision of adequate practice opportunities through contextualized exercises and student-centered activities; and systematic guidance addressing technical difficulties and proficiency disparities among learners. The analysis reveals that successful CBLP implementation requires coordinated support from policy-makers for curriculum integration and funding, institutional investment in infrastructure and professional development, and ongoing research to develop context-specific evaluation frameworks and pedagogical materials tailored to Chinese educational settings.

Keywords: corpus-based language pedagogy; vocabulary acquisition; data-driven learning; EFL instruction; Chinese higher education

1. Introduction

People, particularly those learning English as a foreign language (EFL), are having to meet greater language requirements and pay more attention to English proficiency as a result of the world's fast globalization. The importance of improving students' language skills in Chinese colleges is therefore growing. In particular, lexicon learning—which includes vocabulary and collocations—is essential to EFL learners' acquisition of the English language (Yang & Dai, 2012; Shabani & Rahimy, 2020).

Corpus-based language pedagogy (CBLP) is a technique used in education that allows teachers to use corpora as a preferred tool in their language instruction (Ma et al., 2021). As a result, students in higher education can benefit from the development of their corpus literacy while also being able to use direct corpus data for language study and data-driven learning. In corpus learning, for instance, the concordance line is "a line of text taken from a corpus" (Ma & Mei, 2021, p. 178), which includes the word that was searched for, the pertinent keyword, and its context. Students can identify and learn new words with the help of concordance lines.

The effectiveness of corpus-based instruction in improving vocabulary acquisition has been thoroughly examined in prior research over the last 20 years (Daskalovska, 2015; Li, 2017). According to earlier research and case studies in corpus-based language learning, learners' vocabulary acquisition advances when authentic data is incorporated (Hadley, 2002), used for error correction (Gilmore, 2008), and used to increase learners' awareness of language patterns (Gilquin & Granger, 2010). Nevertheless, teachers are still not implementing corpus linguistics in their language classes (Li & Xu, 2022). Furthermore, because they are unsure of the technical features of this method and the knowledge of corpora, few teachers are ready to increase their corpus literacy in corpus-based language education (Zareva, 2017; Chen et al., 2019).

"How can teachers implement corpus-based language pedagogy with EFL students in Chinese universities to improve English language learning?" is the research topic that will be used to demonstrate this essay based on the previously mentioned arguments. In order to give educators, researchers, and other institutional faculty members further insights, this study attempts to make it easier to develop four

tactics for using CBLP in the classroom. Additionally, throughout the language learning process, it might assist instructors in creating corpus-based language courses for university EFL students. which can indicate potential directions for further study.

2. Four main strategies for teachers' implementation of corpus-based language pedagogy facilitating students' vocabulary acquisition

2.1. Combination of corpus approach and language courses

Teachers can use the corpus in conjunction with linguistic expertise in the classroom to enhance students' vocabulary acquisition from a variety of corpus-based perspectives. Instead of being a stand-alone activity, the examination of corpus data would be a fundamental component of language education, according to Chambers (2019). It is feasible to investigate a word's part of speech based on its literal meanings through corpus learning. When working with terms that have different meanings or applications, this is especially helpful (Xu, 2014). For example, the corpus, such as COCA, will give specific usages of a single word in various situations. Students can determine the grammatical category of a single word by looking at the words that surround it in a single sentence that the corpus provides (Xu, 2014). To improve learners' vocabulary and reduce repetition of a single word, the synonym function in a corpus can be used in addition to part of speech in vocabulary education to explore words with comparable meanings (Ma & Mei, 2022). Additionally, two writers emphasize that teachers might use corpora to help students identify often collocated terms and assess if they are utilizing them in acceptable ways. It is then possible to raise students' understanding of proper word usage. Students can also be taught to concurrently recognize the subtle differences between synonyms.

Despite the unique constraints of Chinese education, there are ways to deal with these issues. On the one hand, teachers can use learner corpora in the classroom to help students satisfy their needs (Granger, 2009). The learner corpus is made up of language created by EFL students and is compiled by publishers or scholars. It helps teachers create more effective language courses by providing them with information about typical errors made by pupils during the lexical learning process. Teachers can

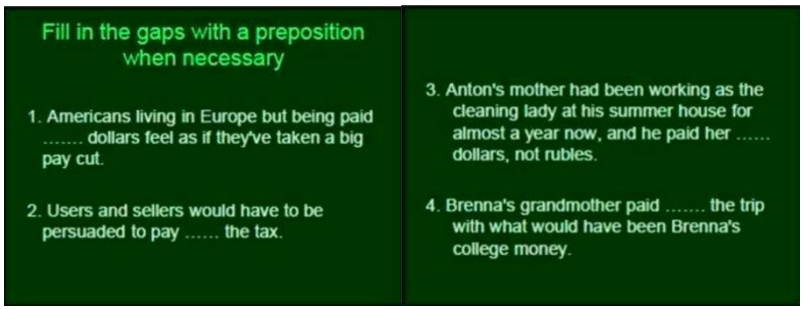
modify their practical education to meet the unique test needs of their pupils by compiling a list of vocabulary usage faults. According to students' potential English proficiency, corpus problems and previous language resources can also be explored in class (Zareva, 2017). It is possible to effectively teach language resources such prepositions, synonyms, and collocations (Çalışkan & Kuru Gönen, 2018). Schools can conduct additional supportive workshops to help instructors improve their corpus literacy and language pedagogy in order to address the absence of direction in their implementation.

2.2. Variety of forms in corpus-based language learning

When it comes to presenting corpora in the classroom, giving pupils interesting material can be quite important for vocabulary teaching. In addition to several kinds of computer-based digital corpora, there are several ways to assist pupils in directly understanding the meaning of every word. The significance of striking a balance between corpus technology and other teaching resources is emphasized by Ma and Mei (2021). They draw attention to the fact that an over-reliance on concordance lines or other corpus functions might cause boredom, which can impair students' motivation to learn new words. More than 92.6% of students in English classes taken as a second language reported feeling bored, according to a survey done at Chinese institutions (Li et al., 2021). Teachers can therefore use a variety of vocabulary teaching strategies to draw in pupils and increase their involvement in class. According to Boulton (2010) and Liu (2013), printed corpus data, as opposed to other available materials, may be more beneficial for inexperienced students who have little background knowledge or corpus usage expertise. Specifically, students can receive explicit instruction in data-driven learning and practice through printed concordance lines of keywords in hands-on corpus practice (Çalışkan & Kuru Gönen, 2018). As a result, even inexperienced students can participate in corpus-based learning successfully and easily. Along with paper materials, students can also benefit from the corpus-based approach's use of clear slides or specific graphics to demonstrate word usage, especially when dealing with abstract terminology (Gao, 2013). According to Frankenberg-Garcia (2014), teachers can use the concordance lines they choose and display to teach students about collocation studies，as shown in **Figure 1**.

Figure 1

An example of lexical collocations with slides (Frankenberg-Garcia, 2014)



Teachers can boost students' engagement in corpus classes by involving them in various patterns of class interpretation in addition to the types of materials (Ma & Mei, 2021). In particular, students' involvement in corpus learning can be enhanced by group and peer collaboration with classmates. Students' potential weariness with repetitive concordance data in solo work can be reduced through group work (Leńko-Szymańska, 2014). Students can participate in collaborative work in the language classroom in a variety of ways, and their success in group projects will boost their confidence in data-driven learning with corpora, according to Liu and Jiang (2009). Additionally, Ma and Mei (2021) stress the critical importance of various substitute activities, such as storytelling and games. For example, to review their corpus study, students can be instructed to either create stories utilizing the terms they need to learn from the corpus in their textbooks or play some fun games.

2.3. Sufficient opportunities for students' practice in corpus-based lessons

Apart from the variety of learning styles in corpus instruction, students also benefit greatly from the practical application of data-driven learning in corpus-based language classes. Successful vocabulary instruction requires adequate time in corpus training programs and ample opportunity for language practice (Gass et al., 2013; Karras, 2016). A particular context or a recurring subject across the entire session can be used as exercise to support students' data-driven learning in vocabulary acquisition if they want additional opportunities for hands-on practice (Ma & Mei, 2021). Teachers can specifically show students movies with a theme in a certain context, like Finding Nemo, which will both pique students' interest and help them practice at the same time. Additionally, some classic exercises, like as multiple-choice questions, gap-filling, and sentence-making tasks, can also be used in corpus practice

(Frankenberg-Garcia, 2014). Both conventional and innovative methods can be used to train students' data-driven learning.

These exercises, however, differ from those found in current Chinese language instruction. It is clear that Chinese teachers are used to adopting teacher-centered traditional pedagogy in the classroom, which primarily emphasizes teachers' interpretations above students' performance and practice (Zhang, 2021). However, CBLP necessitates that students take the lead in practice exercises and that professors function as facilitators, offering encouraging assistance to students as they practice using corpora (Ma & Mei, 2021). Teachers' fear of losing control of the entire class is exacerbated by the fact that students will be organizing the class (Breyer, 2009).

Similarly, one of the corpus properties can be considered an impediment. Instead of teaching language for the first time, corpora are intended for language research (Almutairi, 2016). Students may therefore run into errors in the language usage in the corpus when they study and practice with it in a real-world setting, which is detrimental to the precision of language usage in vocabulary learning. According to Çalışkan and Kuru Gönen (2018), some grammatical faults in corpus data will motivate teachers to reduce opportunities for students to practice and explore the corpus on their own.

One solution to these issues is for teachers to take part in workshops that would help them guide their corpus teaching. In the case study (Schmidt, 2023), instructors receive training on students' DDL and the corpus teaching framework, which boosts their confidence and proficiency in using CBLP in the classroom. Consequently, educators can create more successful corpus-based courses and give students additional corpus exposure possibilities. Students might be given extra homework assignments to reinforce their learning with corpus tools and provide opportunities for corpus practice (Ma & Mei, 2021).

2.4. Guidance for students' corpus learning

EFL learners will be deterred by the intricacy of corpus learning, hence teachers' assistance is required throughout the process. On the one hand, using the corpus to study vocabulary may present certain technological challenges for pupils (Boulton, 2009; Breyer, 2009). Accordingly, some students find it too challenging to understand the concordance lines during the corpus learning process (Abdel Latif, 2020; Ma & Mei, 2021; Şahin Kızıl & Savran, 2018). Teachers can help students with targeted

corpus questions that address this issue by employing complete concordance lines and minimizing the number of concordance lines (Lin & Lee, 2015). Together with the concordance tools, the corpus's slowness in loading vocabulary results and the restricted access for free users will make it more difficult to use the corpus for language instruction and increase search time (Çalışkan & Kuru Gönen, 2018). Teachers can choose a few common concerns to provide feedback on corpus approaches that students frequently experience during corpus practice in order to solve these technical challenges and increase class efficiency (Zareva, 2017). Crucially, other forms of organizational support, including technical assistance from the school administration, can also be crucial for giving devices access to a more dependable and high-quality wireless network.

However, in a corpus-based language instruction class, the disparity in vocabulary acquisition between Chinese students and English will present a barrier. Liu and Lei (2017) propose that higher-level learners possess potential in the data-driven learning approach and will achieve greater success than those at lower proficiency levels. Regrettably, regional differences in education result in varying admission exam requirements between areas, which in turn affects students' language proficiency at Chinese universities (Tao, 2019). It might also make it more difficult for teachers to practice teaching corpus. According to Frankenberg-Garcia (2014), it may be challenging for students with lower English ability to learn how to use a word in a variety of real-world scenarios if they are unable to even recognize its meaning. As a result, students with low proficiency in the English language may view concordance analysis as a laborious and less efficient method of learning word meanings than consulting dictionaries directly (Frankenberg-Garcia, 2014). This affects their attitudes toward data-driven learning through corpora in the classroom. Additionally, the time constraint will affect how instructors in Chinese colleges support data-driven learning.

Teachers can use methods like general and individualized aid to address the aforementioned specific issues in the Chinese setting. Given the limitations mentioned, topics like defining patterns in corpus research, methods for spotting different patterns in corpus data, and the creation of relevant research questions can all be addressed and discussed in the classroom to improve students' corpus learning skills (Zareva, 2017). Problems that may come up throughout the corpus learning process can be addressed by employing a one-on-one strategy and step-by-step instruction (Ma &

Mei, 2021; Zareva, 2017). This strategy entails teachers giving students individualized advice to assist them get past obstacles they have when utilizing corpus-based learning techniques.

3. Implications

To promote corpus-based language pedagogy (CBLP) in Chinese universities, policy-makers should integrate corpus teaching into the national curriculum and provide funding for teacher training and corpus development (Ma & Mei, 2021). University administrations can allocate resources to support infrastructure, expert consultation, and corpus research aligned with students' needs (Gao, 2013). Teachers play a central role in implementation; they should enhance their corpus literacy through workshops and practical training (Frankenberg-Garcia, 2014). Effective CBLP also requires teachers to consider students' proficiency and adjust classroom activities accordingly. Future research should focus on practical guidance, especially in diverse Chinese contexts, and develop evaluation systems for CBLP effectiveness (Karras, 2016). Collaboration with publishers to create classroom materials is also recommended to support novice teachers.

4. Conclusion

To sum up, corpus-based language pedagogy (CBLP) offers authentic language input and supports data-driven learning for vocabulary acquisition. However, Chinese EFL teachers are often hesitant to apply corpus tools due to limited training and contextual challenges. To address this, teachers can adopt four strategies: integrating corpus into language courses, diversifying learning formats, offering practice opportunities, and providing student guidance. Meanwhile, support from policy-makers, universities, and researchers is essential for enhancing teachers' corpus literacy. Future studies should further explore CBLP's effectiveness and contextual adaptation in China to promote its broader application.

Conflict of interest: The author declares no conflict of interest.

References

[1] Abdel Latif, M. M. M. (2020). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL*, 33(1), 34–48. https://doi.org/10.1017/s0958344020000129

[2] Almutairi, N. D. (2016). The effectiveness of corpus-based approach to language description in creating corpus-based exercises to teach writing personal statements. *English Language Teaching*, 9(7), 103–115. https://doi.org/10.5539/elt.v9n7p103

[3] Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37–54. https://doi.org/10.1017/S0958344009000068

[4] Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

[5] Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172. https://doi.org/10.1080/09588220902778328

[6] Çalışkan, G., & Kuru Gönen, S. İ. (2018). Training teachers on corpus-based language pedagogy: Perceptions on using concordance lines in vocabulary instruction. *Journal of Language and Linguistic Studies*, 14(4), 190–210. https://dergipark.org.tr/en/pub/jlls/issue/43366/528108

[7] Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4), 460–475. https://doi.org/10.1017/S0261444819000089

[8] Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130–144. https://doi.org/10.1080/09588221.2013.803982

[9] Frankenberg-Garcia, A. (2014). How language learners can benefit from corpora, or not. *Cahiers de l'ACEDLE*, 11(1). https://doi.org/10.4000/rdlc.1702

[10] Gao, L. (2013). College English Curriculum Requirements in China: Expectations and responses. *Problems of Education in the 21st Century*, 51, 47–62. https://doi.org/10.33225/pec/13.51.47

[11] Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An*

introductory course (4th ed.). Routledge.

[12] Gilmore, A. (2008). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363–372. https://doi.org/10.1093/elt/ccn056

[13] Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359–370). Routledge. https://doi.org/10.4324/9780203856949

[14] Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13–32). John Benjamins Publishing. https://doi.org/10.1075/scl.33.04gra

[15] Hadley, G. (2002). An introduction to data-driven learning. *RELC Journal*, 33(2), 99–124. https://doi.org/10.1177/003368820203300205

[16] Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, 28(2), 166–186. https://doi.org/10.1017/S0958344015000154

[17] Li, C., Dewaele, J.-M., & Hu, Y. (2021). Foreign language learning boredom: Conceptualization and measurement. *Applied Linguistics Review*, 14(2), 223–249. https://doi.org/10.1515/applirev-2020-0124

[18] Li, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology*, 21(3), 153–171. https://doi.org/10125/44625

[19] Li, S., & Xu, M. (2022). Corpus literacy empowerment: Taking stock of research to look forward for practice. *Journal of China Computer-Assisted Language Learning*, 2(1), 126–155. https://doi.org/10.1515/jccall-2022-0002

[20] Liu, D. (2013). Using corpora to help teach difficult-to-distinguish English words. *English Teaching*, 68(3), 27–50. https://doi.org/10.15858/engtea.68.3.201309.27

[21] Liu, D., & Lei, L. (2017). *Using corpora for language learning and teaching*. TESOL Press.

[22] Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177–190. https://doi.org/10.1515/jccall-2021-2008

[23] Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online

collaboration. *Computer Assisted Language Learning*, 35(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

[24] Şahin Kızıl, A., & Savran, Z. (2018). The integration of corpus into EFL speaking instruction: A study of learner perceptions. *International Online Journal of Education and Teaching (IOJET)*, 5(2), 376–389. http://iojet.org/index.php/IOJET/article/view/368/241

[25] Schmidt, N. (2023). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, 35(1), 40–57. https://doi.org/10.1017/S0958344022000106

[26] Yang, W., & Dai, W. (2012). Vocabulary memorizing strategies by Chinese university students. *International Education Studies*, 5(1), 208–214. https://doi.org/10.5539/ies.v5n1p208

[27] Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69–79. https://doi.org/10.1093/elt/ccw045

[28] Zhang, B. (2021). A comparison between pedagogical approaches in UK and China. *Journal of Comparative & International Higher Education*, *13*(5), 232–242. https://doi.org/10.32674/jcihe.v13i5.2629