

Examining The Integration of Corpus Technology into English Language Teaching: A Systematic Review (2020-2024)

Shuang Yang, Wan Nur Asyura Wan Adnan

Faculty Of Social Sciences and Humanities, University Technology Malaysia

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.9020340>

Received: 18 February 2025; Accepted: 22 February 2025; Published: 22 March 2025

ABSTRACT

Corpus linguistics (CL) has been found to greatly influence both the content and practice of English Language Teaching (ELT). With the application of the large practical corpora and technological tools, it has brought promising advantages to greatly enhance the foreign or second language teaching and learning in the current language education. Moreover, this field continues to evolve, driven by ongoing technological advancements and the ever-expanding availability of digital texts. To date, corpus technology has emerged as a pivotal role in foreign or L2 education. Nevertheless, EFL teachers may still face the challenges on how to integrate corpus technology into pedagogical practices in the class. Therefore, this systematic review study thoroughly analyzed the current empirical studies of this domain by reviewing 20 included studies published from 2020 to 2024 through Web of Science and Scopus, and explored the synergy effect between the use of corpus technology and pedagogical practices in the context of English language education. Included studies were reviewed to extract six major research parameters to construct a coding book: publication information, research methods, data collection methods, challenges of using corpus technology, pedagogical practices of developing corpus technology, and the effectiveness of using corpus technology. The findings revealed that the uses of corpus technology have demonstrated an effective strategy following innovative pedagogical practices, like Corpus-Based Language Pedagogy (CBLP), Data-Driven Learning (DDL) and other corpus-based instructions. These identified practices served as the latest evidences and reference to establish novel and valuable corpus-based pedagogical model for the future research direction.

Keywords: Corpus Linguistics, Corpora, Corpus Technology, Corpus-Based Language Pedagogy (CBLP), Data-Driven Learning (DDL), English Language Teaching (ELT)

INTRODUCTION

Corpus linguistics emerged in the technological era as a result of advancements in computing power and data storage. It is a transformative field within the broader study of language education, offering unprecedented insights into the effective implementation of foreign or second language (L2) pedagogy in a real classroom context. An increasing number of relevant studies have shown that corpus linguistics is an effective and creative approach for English teachers to design teaching activities (Ma et al., 2022; McCarthy & Carter, 2001; Yoon, 2005). As is well known, the advent of corpus linguistics coincides with advances in computer technology, demonstrating the potential of empirical data to revolutionize English language teaching (ELT) research. Therefore, as the earliest language approach mediated by technology, computer-assisted language learning (CALL) has made much influence on language education.

To date, this field continues to evolve, driven by ongoing technological advancements and the ever-expanding availability of digital texts. Most importantly, corpus technology, as a strong branch of CALL, has emerged as a pivotal role in foreign or L2 research. According to Ma et al. (2024), corpus technology refers to the use and application of technology associated with corpus linguistics and corpora for language learning and teaching. In the context of teaching English as a Foreign Language (TEFL) in the mainland of China, the use of corpora has expanded beyond the compilation of dictionaries and grammatical description. It is greatly known that corpora has been frequently utilized by many corpus linguists in other

varieties of applications, including teaching design teaching methods, and classroom writing teaching, all of which are

important for English language teaching (ELT) in the authentic classroom setting (Roslim et al., 2020).

The integration of corpus technology into ELT would provide empirical significance to the teaching methodology. However, corpus technology has had challenges in integrating into language teachers' classroom instruction, or been less successful in exploring its path into language teachers' classroom teaching, because most of the language teachers are unfamiliar and reluctant to implement it in classroom teaching (Ma et al., 2024). Thus, many scholars are still sceptical about its broader adoption in the context of education. Over the recent years, the studies on the integration of corpus technology into English language teaching has steadily aroused and introduced new pedagogic approach. In this regard, more and more corpus-based studies in foreign or L2 pedagogy have affirmed that the value of the corpus approach for teaching English constructs inductive language learning process based on effective instruction.

For instance, Barlow (1996) demonstrates that corpora and the corpus tools derived from any of the corpus-based investigation can mostly impact lesson design and the content of teaching. More recently, the increasing number of academic articles and proceedings on related topics has demonstrated the impact of corpora, including corpus tools, and corpus methods on L2 learning and teaching (McEnery & Hardie, 2011; Römer, 2011; Sinclair, 2004). These studies have shown that the corpus approach not only improves learners' knowledge of grammar in context, but also broadens their comprehension of word functions in context. Also, Flowerdew (2012) illustrates how the corpus-based pedagogy can be systematically incorporated into business English writing class through designing a corresponding teaching materials. Tsai (2021) found that the integration of corpus consultation with business English writing instruction has enhanced students' linguistic awareness and brought considerable improvements in aspects of lexical and syntactic application. It is important to note that corpus-based studies in English writing highlights the co-occurrence and typical contexts of words, namely collocation. In addition to collocation, frequency information is also required to take into consideration when selecting teaching materials for business English instruction (Tsai, 2021, p. 2). Furthermore, Boulton and Cobb (2017) conduct a meta-analysis on the effectiveness of corpus use in second language learning, identifying 205 empirical studies up to mid-2014. The found that corpus-based approach plays an effective role in the enhancement of second language learning, while further studies also need to address the more comprehensive understanding of the effectiveness of corpus technology in second language learning, aligning with more evidences conducted in various contexts. In general, corpus technology has the potential to be a groundbreaking tool that facilitates to instruct teaching for Chinese EFL learners. Despite the previous studies have presented a systematic understanding of corpus use in ELT, latest empirical evidences are required to investigate the current trend of research in this domain as corpus technology advances fast.

Building on the current trend of relevant studies, this review study aims to answer the following questions:

Q1. What challenges do language teachers encounter when integrating corpus technology into ELT classes?

Q2. What pedagogical practices do language teachers implement when integrating corpus technology into ELT classes?

Q3. What is the effectiveness of integrating corpus technology into ELT classes?

In answering these questions, this study conducts a comprehensive and systematic synthesis of corpus technology use in ELT scholarship in 2020 to 2024, retrieving articles from high quality databases, including Web of Science (WoS) and Scopus, and a total of 20 articles with full texts were included.

METHODOLOGY

The current study adopted the systematic review method. Bettany-Saltikov (2010) defined that a systematic review is a comprehensive analysis of the scholarly literature pertaining to a certain topic, with

the aim of identifying, selecting, evaluating, and synthesizing all the high-quality research evidence that was relevant to that topic. In general, A systematic review adhered to a well-defined protocol that outlined its aims, principles, and methods in advance. Its purpose was to synthesize the existing research in order to reach reliable findings. Compared with other review designs such as narrative review, the systematic review could provide a more suitable example of taking adequate measures to minimize error and bias (Littell et al., 2008). In this study, the review protocol was established based on the research questions to search for the relevant studies.

Paper Retrieval

This review was undertaken by an intensive comparison of peer-reviewed journals on the corpus linguistics domain. To entail a comprehensive and in-depth study, Green et al. (2006) argued that it is typically advisable to search across a minimum of two databases. In pursuit of authoritative sources that index education research field, the current study utilizes two prominent and international electronic databases: Web of Science and Scopus, due to their influential positions in research community and a great number of publications they own in the linguistic domain.

In order to design a suitable framework on paper retrieval, this study adopted PRISMA flow chart diagram to conduct publications retrieving and screening (Moher et al., 2009). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement (PRISMA) is an appropriate method to retrieve relevant publications with accuracy and comprehensiveness in various academic fields (Qin, 2023). In this study, a three-step framework (see Figure 1) was designed to illustrate the paper retrieval and selection process. Firstly, aiming to yield the most appropriate results in each of the databases, the search terms were designed to furthest enlarge the emergence of potential publications on linguistics domain by using search parameters of “corpus linguistics” OR “corpora” OR “corpus” OR “corpus-based/assisted/informed” OR “corpus technology”. The retrieved articles were published from 2020 to 2024. The search was conducted on January 1, 2025. It was worth noting that the initial retrieved articles were ready to be exported as .ris files and then imported to EndNote to use its automatic tool to delete duplicates. Furthermore, a large quantity studies about using corpus-based analysis on language features was screened out in advance. Finally, the initial searching produced over 32324 potential papers.

Subsequently, the screening procedure was the second stage. Potential publications on the corpus linguistics domain were identified in terms of the keywords, titles and abstracts. However, some of the contents seemed to be unrelated to the relevant topics, like corpus-based analysis. Given the volume of data, it was imperative to implement procedures to exclude extraneous articles. Articles in relation to book reviews, editorials, review and conference proceedings were filtered out. Articles reported the English language in English with full text were included. After records screened, leaving a total of 2908 publications for further visual examination. In order to extract educational area of corpus linguistics parts from the whole corpus linguistics studies, the criteria for including or excluding these included papers should be explicitly and implicitly outlined. In this paper, as can be seen in Table 1, nine main inclusion and exclusion criteria, together with their subsets have also been outlined, like partially related. While this search takes place, unsuitable articles were removed from visual examination. This search strategy yielded 20 studies for inclusion in the systematic keyword map.

Table 1 Inclusion/Exclusion Criteria for Visual Examination

Inclusion/Exclusion	Criteria	Explanation
Inclusion	Databases (DB)	Web of science; Scopus
	Search terms (ST)	“Corpus linguistics” OR “corpora” OR “corpus” OR “corpus-based/assisted/informed/aided” OR “corpus technology”
	Time span (TS)	Publications between 2020 and 2024
	Search engine reason (SER)	Search terms can be found in title or abstract or keywords

	English language teaching (ELT)	A paper on the corpus linguistics domain must be related to ELT.
	Partially related (PR)	PR-1: A paper focuses on attitudes, perceptions, and so on, but with stated any aspect of pedagogical practices or challenges. PR-2: A paper focuses on topics other than language teaching, but with stated pedagogical implications.
Exclusion	Without methodology	A paper without methodology to be assessed
	Without full English text	A paper without full English text to be assessed
	Non-related (NR)	NR-1: A paper is not an academic article. For example, editorial materials, conference reviews, contents, or forewords NR-2: A paper is not related to empirical study. NR-3: A paper is not related to English language studies.

In general, some researchers may regard systematic reviews as having a limited focus in comparison to traditional reviews. Certainly, due to the strict adherence to particular criteria for inclusion and exclusion in this paper, it was inevitable that some important research that may have been published in less prominent academic journals and unindexed dissertations would be missed. Nevertheless, the above rigorous criteria has a positive result of diminishing prejudice, which in turn enhances the reliability and perhaps the significance of the study's findings and recommendations (Rose et al., 2018, p. 153).

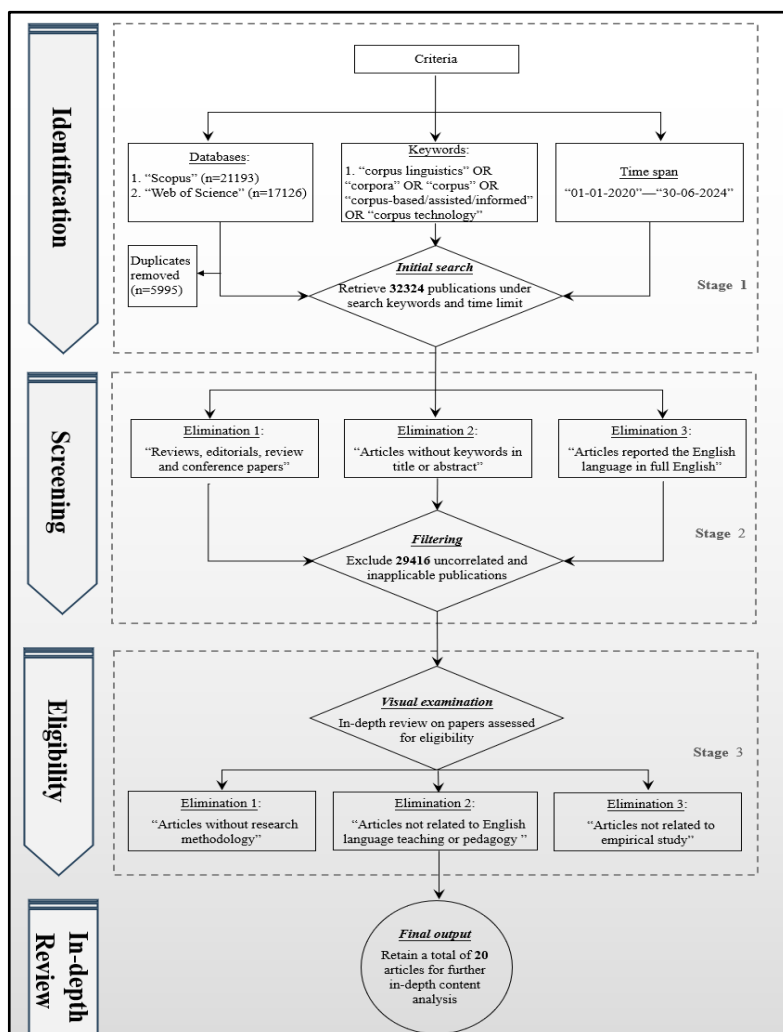


Figure 1 PRISMA Flow Diagram of Literature Search and Selection

Data Analysis

The analytical framework of the present review was adapted from the study of Littell et al. (2008), where five steps were outlined to analyse data. The review included: 1) coding relevant information from the studies into an analytical table; 2) screening data related to corpus technology, tools, activities, challenges, hurdles, and barriers influencing using corpus and software; 3) comparing codes and coded data within categories to identify relationships, similarities, and differences; 4) describing the results; and 5) interpreting the findings. Table 2 illustrated the structure of the codebook used for preparing data analysis of this study.

Table 2 Coding Framework for the Six Reviewing Parameters

Variable	Coding scheme	Value
Publication information	Open	Release date and time, publication outlet, etc..
Research methods	Open	Non-empirical, qualitative, quantitative, and mixed methods
Data collection methods	Open	Questionnaire, interview, survey, etc..
Challenges	Pre-constructed	Time-consuming, unfriendly software, etc..
Pedagogical practices	Pre-constructed	Student-centered, group activities, inductive learning, etc..
Effectiveness	Pre-constructed	Evaluation, feedback, and peer observation, etc..

FINDINGS AND DISCUSSION

Overall Summary of the Included Studies

This review analyzed 20 included studies on the integration of corpus technology into ELT classes. Research methods served as a structured approach that guides researchers in conducting their studies in a rigorous and methodical manner, which facilitated researchers to ensure the reliability and validity of the research results (Creswell, 2012).

In this review, there was a slight difference among three major research methods (see Table 3). These included studies adopted mixed-methods approaches (35%) and quantitative methodologies (35%) equally, followed by qualitative methodologies (30%). Notably, the majority of studies (50%) employed experimental research designs, including pretest-posttest designs and quasi-experimental designs, to explore the differences after intervention. Overall, this experimental approach is commonly used to evaluate new educational methods, technologies, and interventions. Furthermore, data collections methods or techniques were also examined in the study.

Table 3 Research Methods of Included Publications

	Research Methods	Frequency
1	Quantitative	35%
2	Qualitative	30%
3	Mixed methods	35%

Additionally, as shown in Table 4, surveys and assessments were the most popular technique (respectively 50%), followed by interviews (45%). Under assessment methods, various techniques like pretest and post-test, delayed-test, and revision checklists were employed. Analyzing lesson designs and fieldnotes in classroom observation were administered in 40% of the studies. The rest of employed methods were distributed in using journals (15%) and open-ended questions (10%). All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

Table 4 Distribution of the Reviewed Studies by Data Collection Method

Methods	Instruments or techniques	Proportions (%)
survey	questionnaires, surveys	50%
assessment	tests; revision checklists	50%
reports	journals	15%
interviews	discuss with stakeholders	45%
observation	fieldnotes; lesson design	40%

Challenges of integrating corpus technology into ELT classes

As technology advances and corpora grow, corpus technology has increasingly revolutionized English language teaching and learning. In this regard, possessing high level techniques of synergizing multiple knowledge seems to present various challenges to both English language teachers and learners. Indeed, using corpora and corpus tools in everyday classrooms may trigger several difficulties and barriers. Of the 20 articles included, these challenges have been identified, where 90% of studies reported that both teachers and students may face unavoidable difficulties in using corpora and corpus tool in everyday classrooms. Specifically, the majority of studies (50%) indicated that technical issues and paucity of training are the top two challenges to both teachers and students. Many teachers may not have the necessary training or experience to navigate and interpret corpus data, which can hinder their ability to effectively incorporate corpus technology into their teaching practice. Similarly, students may face difficulties in maintaining interest and understanding corpus queries due to lack of computing and corpora skills. Certainly, there is no denying that technical expertise is required to be a pre-requisite to effectively use corpus tools and resources. Nevertheless, fully mastering a corpus database can be a time-consuming and effort-consuming process, thus 35% of studies mentioned that time and effort required to find and select appropriate corpus resources for language teaching and learning. Another 30% of studies reported that teachers may feel struggled to find relevant and up-to-date corpus data due to limited resources accessibility and inappropriate materials selection, while 20% of studies showed that teachers had difficulty in incorporating corpus technology into current curriculum that aligns with their teaching objectives and the needs of their students. In fact, some corpus tools and resources may require a subscription or licensing fee, which can be a barrier for teachers and institutions with limited financial resources. Furthermore, 15% of studies mentioned insufficient IT infrastructures and logistic issues in some areas when introducing corpus technology into classrooms, while only a small percentage of studies (5%) reported that internet connection issue may interrupt the progress of integrating corpus technology into teaching classroom.

To illustrate, Liu and Gablasova (2023) clearly identify five challenges that teacher encounter when integrating corpus technology into ELT classes, low technical issues both for teachers and students, time-consuming to interpret data, shortage of available computers, unstable internet connection, lack of pre-training. Apart from that, a high level group of teachers is scarce due to the complexity of synergizing multiple layers of knowledge, including pedagogical, technological and content knowledge (Schmidt, 2023). This complexity can hinder their ability to implement corpus-based methods into classroom teaching successfully (Crosthwaite, 2020). Moreover, effort-consuming is another crucial problem, because teachers may struggle to implement DDL effectively in a large size class. Teachers not only have to select appropriate materials the respective learning context but also possess comprehensive knowledge to integrate them meaningful in the classroom (Almegren, 2022). The careful material selection allows students to build their familiarity with corpus knowledge progressively (Crosthwaite, 2020). Hence, teachers need to balance the use of corpus tools with traditional teaching practices under the current educational system (Tsai, 2021). In addition, teachers may find it difficult to align these activities with the learning objectives and outcomes already established in their courses, leading to potential conflicts with traditional teaching methods (Almegren, 2022). Thus, improving teachers' confidence through long term

continuous training can translate into more effective teaching practices in applying direct corpus methods in their lesson.

As a result, both learners and teachers must demonstrate sufficient control over corpus technology, otherwise this can be particularly daunting for novices who may not be familiar with the tools and platforms necessary for effective corpus consultation (Crosthwaite, 2020). To guarantee that the integration of corpus technology into English language teaching and learning becomes a more widespread and successful language pedagogy, it is essential to address these obstacles. Once we have a thorough understanding of how to combine corpus technology and pedagogy, it will greatly contribute to improving language teaching and learning via effective technical and pedagogical instruction.

Pedagogical practices of integrating corpus technology into ELT classes

Based on the above identified challenges of integrating corpus technology into ELT classes, developing appropriate language pedagogy was necessary to fully optimize the advantages and applications of corpus technology. Between the year 2020 and 2024, while most of the studies (45%) emphasized Data-Driven Learning (DDL), the latest research (25%) has shifted towards Corpus-Based Language Pedagogy (CBLP). The concept of CBLP is firstly introduced by Ma et al. (2022), which is somewhat different with DDL (Ma et al., 2024). Specifically, CBLP has paid much attention on teachers' pedagogical application of corpora apart from learners' self-exploration. As a broader concept, CBLP may greatly facilitate teachers to design more corpus-based materials in the context of everyday classrooms. Also, selecting available and accessible corpora and corpus software provide teachers with more appropriate scaffoldings to guide students effectively. Teachers can implement corpus-based activities through direct and indirect methods (Almegren, 2022). In direct use, students accessed the corpus themselves to explore language data, while in indirect use, teachers presented examples from the corpus to illustrate collocations and their usage in context. This dual approach catered to different learning styles and promoted active engagement with the material. Notably, out of a total of 25% studies focused on collaborative approach as a means to enhance learners' learning ability, while two studies mentioned that students may feel reluctant to join group work. For instance, Tsai (2021) previously supports that encouraging collaborative learning among students can significantly benefit the corpus-based instruction process. In contrast, Schmidt (2023) emphasizes that adopting collaborative approach and iterative lesson design approach could lead to continuous improvement in teaching methods and better alignment with student needs. It was interesting to note that despite the potential for collaborative learning to reduce the workload of corpus searches, students have previously expressed dissatisfaction with the dullness of these tasks (Crosthwaite, 2020; Tsai, 2021). Hence, a well-organized and invigorative may ensure students keep motivated without feeling overwhelmed.

In addition, Liu and Gablasova (2023) also highlight that DDL promoted an inductive, learner-centered approach in the classroom where students were encouraged to search language data independently. Teachers act as facilitators, guiding students in their discovery of collocational patterns. Most importantly, teachers need to prepare authentic materials and select appropriate corpus software well so that students feel more interactive and engaging based on their overall language competence for real-world language use. Assessment of learning outcome could be designed through quizzes, writing assignment, or reflective journal to provide deeper insights into the effectiveness of DDL use in the teaching. It was noteworthy that providing learners with access to vast amounts of authentic language data enabled them to develop their language skills, enhance their awareness of language usage, and become more critical thinkers.

Consistent with these findings, the pedagogical practices of conducting corpus-based activities included direct and indirect use of corpora, integration of authentic materials, collaborative learning, use of technology and analytical tools, and regular assessment and feedback. These practices not only enhanced learners' collocation competence but also contributed to a more engaging and effective language learning environment. Regular assessment through pre- and post-tests can help measure the effectiveness of corpus-based activities. Providing feedback based on these assessments allowed teachers to identify areas for improvement and adjust their teaching strategies accordingly. This practice ensured that students are making progress in their collocation learning. While DDL can be conducted independently, the presence of

scaffolding and support from instructors is vital for maximizing its effectiveness. Teachers played a crucial role in guiding students through the DDL process, helping them navigate the complexities of corpus data and enhancing their query skills. Consequently, more empirical evidences on testing different corpus technology to develop appropriate pedagogical practices may require to refer in the future education.

Effectiveness of integrating technology into ELT classes

To assess the effectiveness of incorporating corpus technology into English language teaching, most scholars adopt various research methods to seek the answers. Of the 20 articles reviewed, 5 studies directly reported the effectiveness of incorporating corpus technology into the current language pedagogy, while rest of the studies more or less mentioned the enhancement of language proficiency that corpus technology produce. For instance, Liu and Gablasova (2023) investigate the effectiveness of DDL on L2 learners' collocation performance by using a quasi-experimental research design. It is worth noting that the intervention was conducted in the classroom across the period of eleven weeks duration as suggested by (Boulton & Cobb, 2017). The length of the intervention could be associated with greater gains. The results showed that the DDL approach has significantly enhanced learners' collocation knowledge compared to those using traditional teaching methods. Of special note that this intervention could help learners retain knowledge according to a delayed test and post-treatment questionnaire after three months. Meanwhile, the effectiveness of DDL was evident in its ability to booster learner confidence and form a good habit of autonomous learning. Similarly, Almegren (2022) also highlights that the differences in outcomes between the experimental and control groups are statistically significant. This underscored the effectiveness of corpus-based activities in improving language learning outcomes, as the experimental group showed a marked improvement after the intervention. Compare to the traditional teaching methods, corpus-based activities indeed showed significant improvements in collocation competence.

In contrast, the majority of studies (45%) highlighted the process of how corpus technology integrated into language pedagogy, which could also help facilitate teachers effectively to use corpora and corpus tools in everyday classrooms. In order to present a comprehensive procedure of conducting CBLP, Schmidt (2023) conducts a case study on six university teachers who developed DDL activity plans in their classes during the fall semester. The findings also proved that corpus-based activities promoted language awareness among students. By engaging with authentic corpus data, learners can better understand language patterns, usage, and nuances, which can significantly improve their writing proficiency and critical thinking skills. Equally, Tsai (2021) explores effect of the corpus-based intervention on the business writing performance by using qualitative and quantitative approaches. The results from questionnaire and interview also reveal that the experimental group guided by corpus-based instruction shows a notable increase in their post-test scores compared to the control group.

Overall, while corpus technology held promise for enhancing English language teaching and learning, there were several challenges that need to be addressed in order to effectively integrate corpus technology into English teaching classes. Teachers and educators must receive adequate training and support to navigate and interpret corpus data, and efforts should be made to improve the accessibility and affordability of corpus tools and resources for language learners. Additionally, more research is needed to establish best practices for incorporating corpus technology into English teaching classes and to evaluate its impact on language learning outcomes.

In conclusion, studies examining the effectiveness of incorporating corpus technology into English language teaching can provide valuable insights into the benefits and limitations of this approach. Teachers can use a variety of assessment tools, such as pre-and post-tests, learner surveys, and classroom observations, to measure the impact of corpus technology on students' language proficiency and learning strategies. Despite the encouraging evidence generated from empirical studies conducted by researchers, the studies on synergizing multiple knowledge that include technological, content and corpus literacy for teachers remains a lot to be explored in the future research direction.

CONCLUSIONS

This review analyzed the studies on the integration of corpus technology into language classrooms, published from 2020 to 2024 and then discussed common challenges, current pedagogical practices and its effectiveness. The search strategy yielded 20 studies for inclusion in the systematic map. The results showed that the integration of corpus technology into language classroom revolutionized English language teaching (ELT) but still faced barriers that limits its application in a broader context. Specifically, according to the included studies, despite corpus-based activities offered significant benefits for learning vocabulary, collocation, writing, and so on, both teachers and students still faced various challenges, including unfamiliarity with corpus tools, time constraints, curriculum integration, resource availability, student resistance, and assessment difficulties. Addressing these challenges is also crucial for maximizing the effectiveness of corpus-based learning in EFL contexts.

Also, the findings of reviewed studies showed that while DDL activities as common methods have been frequently designed and adopted in the classroom, CBLP instruction increasingly arises as a new practice among teachers. Though studies reported the effective procedure of implementing CBLP in the classroom, teachers feel struggled to apply it flexible due to complex issues, such as technical problem, corpora selection, paucity of training, and so on. Hence, providing sufficient training for both teachers and students is crucial. Teachers should be well-versed in using corpus tools, and students need guidance on how to effectively consult and interpret corpora to identify linguistic phenomenon. Thus, by implementing corpus technology into the language classroom following best practices, teachers can create engaging and effective learning experiences that empower learners to succeed in their language learning journey. Furthermore, corpus technology can also be used to enhance language teaching by providing teachers with valuable insights into learners' language needs and challenges. Through analyzing the language data collected from learners, teachers can identify common errors and misconceptions, as well as track students' progress over time. This information can help teachers tailor their instruction to meet the specific needs of individual learners, thereby improving the effectiveness of language teaching. Ergo, sufficient training programs help in maximizing the benefits of corpus technology in the classrooms.

In addition, the integration of corpus technology into language pedagogy can also help promote learner autonomy and independence. By providing students with access to corpora and other language resources, educators can empower learners to take control of their own language learning process. This can encourage students to explore the language on their own, engage in independent research, and develop their language skills outside of the classroom.

Looking towards the future, the integration of corpus technology into language pedagogy holds great potential for transforming the way languages are taught and learned in educational settings. As technology continues to advance, educators can expect to see even more effective tools and resources for using corpora in language teaching. By harnessing the power of corpus technology, educators can create more engaging, effective, and personalized learning experiences for students, ultimately leading to improved language proficiency and communication skills.

ACKNOWLEDGMENT

The authors wish to express their utmost appreciation and gratitude to University Technology Malaysia (UTM) for supporting this study

REFERENCES

1. Aronoff, S., (1989). *Geographic Information Systems: A Management Perspective*. Ottawa: WDL Publications.
2. Almegren, A. (2022). Effect of corpus-based activities on learning verb–noun collocations in Saudi EFL classes. *Jordan Journal of Modern Languages and Literatures*, 14(2), 371-383.
3. Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics*, 1(1), 1-37.

4. Bettany-Saltikov, J. (2010). Learning how to undertake a systematic review: part 2. *Nursing Standard* (through 2013), 24(51), 47.
5. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language learning*, 67(2), 348-393.
6. Creswell, J. W. (2012). *Educational research*. Pearson.
7. Crosthwaite, P. (2020). Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics*, 43(2), 169-195.
8. Flowerdew, L. (2012). Exploiting a corpus of business letters from a phraseological, functional perspective. *ReCALL*, 24(2), 152-168.
9. Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of chiropractic medicine*, 5(3), 101-117.
10. Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.
11. Liu, T., & Gablasova, D. (2023). Data-driven learning of collocations by Chinese learners of English: a longitudinal perspective. *Computer Assisted Language Learning*, 1-26.
12. 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
13. Ma, Q., Yuan, R., Cheung, L. M. E., & Yang, J. (2024). Teacher paths for developing corpus-based language pedagogy: a case study. *Computer Assisted Language Learning*, 37(3), 461-492.
14. McCarthy, M., & Carter, R. (2001). Size isn't everything: Spoken English, corpus, and the classroom. *TESOL quarterly*, 35(2), 337-340.
15. McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
16. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*, t. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.
17. Qin, M. X. (2023). Across 40 Years: A Bibliometric Analysis of the Chinese English Scholarship (1980–2020). *SAGE Open*, 13(4), 21582440231210410.
18. Römer, U. (2011). Corpus research applications in second language teaching. *Annual review of applied linguistics*, 31, 205-225.
19. Rose, H., Briggs, J. G., Boggs, J. A., Sergio, L., & Ivanova-Slavianskaia, N. (2018). A systematic review of language learner strategy research in the face of self-regulation. *System*, 72, 151-163.
20. Roslim, N., Abdullah, M. H. T., Aziz, A., Nimehchisalem, V., & Almuddin, A. (2020). Corpus research applications in second language teaching: A look into corpus-informed materials development. *International Journal of Modern Languages and Applied Linguistics (IJMAL)*, 4(2), 15-37.
21. Schmidt, N. (2023). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, 35(1), 40-57.
22. Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
23. Tsai, Y. R. (2021). Exploring the effects of corpus-based business English writing instruction on EFL learners' writing proficiency and perception. *Journal of Computing in Higher Education*, 33(2), 475-498.
24. Yoon, H. (2005). *An investigation of students' experiences with corpus technology in second language academic writing* [The Ohio State University].