

Types of Corpora and the Significance of Corpus-Based Approaches in Teaching English

Hamidova Nafisa¹ and Pulatova Durdona Ravshanovna¹✉

Institute of International School of Finance Technology and Science, Tashkent, Uzbekistan¹

ABSTRACT

Corpus-based methodologies have gained increasing attention in English as a Foreign Language (EFL) education because they provide access to authentic language data and enable learners to examine linguistic patterns in meaningful contexts. This study aims to examine the significance of corpus-based methodologies, identify major types of corpora, and explore their pedagogical applications in English language teaching. A Systematic Literature Review (SLR) approach was employed to identify, examine, and synthesize relevant literature on corpus-based language teaching, learner corpora, and Data-Driven Learning (DDL), with emphasis on recent studies published between 2015 and 2024. The findings indicate that corpus-based approaches can support vocabulary acquisition, grammar instruction, writing development, learner autonomy, and communicative competence. Corpora also provide opportunities for integrating authentic language resources into digital learning environments. However, effective implementation requires adequate corpus literacy, teacher training, appropriate corpus selection, and sufficient learner guidance. Overall, corpus-based methodologies represent a valuable complementary approach to EFL teaching by connecting classroom instruction with authentic language use and promoting more learner-centered and evidence-based learning.

Keywords: *Corpus-Based Approaches, Data-Driven Learning, English as a Foreign Language, Learner Autonomy, Learner Corpora.*

CORRESPONDING AUTHOR:

Pulatova Durdona Ravshanovna
Institute of International School of Finance Technology and Science,
Tashkent, Uzbekistan
contact: durdonapulatova88@gmail.com orcid.org

ARTICLE HISTORY

Received : March 12, 2025
Final Revised : October 04, 2025
Accepted : December 17, 2025
Published : December 30, 2025

1. | INTRODUCTION

The use of corpora in English language teaching (ELT) has become increasingly important because corpora provide extensive collections of authentic written and spoken language that allow learners and teachers to examine how English is actually used in different communicative contexts. Unlike conventional language teaching approaches that may rely heavily on prescriptive rules and decontextualized examples, corpus-based approaches enable learners to identify recurring linguistic patterns, word frequencies, collocations, grammatical structures, and contextual variations. In English as a Foreign Language (EFL) education, these characteristics make corpora valuable resources for connecting linguistic knowledge with authentic language use and communicative needs (Granger et al., 2015; Meunier, 2016). The development of learner corpora has further expanded the pedagogical potential of corpus linguistics. Learner corpora consist of language produced by learners and can provide evidence of recurring errors, linguistic preferences, and developmental patterns that may be difficult to identify through conventional classroom observation.

Such evidence can support the development of teaching materials, assessment practices, and pedagogical interventions that are more responsive to learners' actual language difficulties (Granger et al., 2015; Meunier, 2016). Consequently, corpus-based approaches can contribute not only to language description but also to more informed decision-making in EFL instruction. One prominent pedagogical application is Data-Driven Learning (DDL), in which learners investigate corpus data and formulate their own observations about language patterns. DDL encourages learners to become active participants in the learning process rather than passive recipients of grammatical or lexical rules (Boulton, 2017). Evidence from previous research indicates that corpus-based learning can support different aspects of language development. For example, Luo (2016) found positive effects of DDL activities on EFL learners' writing development, while Birhan et al. (2021) reported improvements in academic writing through corpus-based instructional mediation. Similarly, Barabadi and Khajavi (2017) demonstrated the usefulness of a data-driven approach for vocabulary learning. A meta-analysis by Boulton and Cobb (2017) further provided broader evidence of the effectiveness of corpus use in language learning.

Corpus-based approaches also have applications in grammar instruction and broader classroom practices. Kim (2019) examined exploratory corpus-based grammar instruction and highlighted the perceptions of both teachers and learners toward this approach. Corpora can additionally function as open educational resources, increasing learners' access to authentic language materials through digital learning environments (Vyatkina, 2020). These developments are particularly relevant as language education increasingly incorporates digital technologies and learner-centered approaches. Despite these benefits, successful corpus integration requires adequate corpus literacy and pedagogical preparation among teachers. Callies (2019) emphasizes the importance of integrating corpus literacy into language teacher education, while Ma et al. (2022)

demonstrate the potential of structured teacher training and online collaboration for developing corpus-based language pedagogy.

Bennett and Uí Dhonnchadha (2023) similarly propose an in-service framework for helping EFL teachers become corpus literate. However, teachers may still perceive corpus-based instruction as technically challenging or difficult to implement, particularly when they lack sufficient training and experience (Poole, 2022). A systematic review by Lusta et al. (2023) also highlights both the pedagogical benefits and implementation challenges associated with corpus-based and DDL practices. Therefore, this study examines the significance of corpus-based methodologies in EFL education, identifies major types of corpora used in English teaching, and explores their pedagogical applications in supporting vocabulary, grammar, writing, communicative competence, and learner autonomy. It also considers the challenges of incorporating corpus-based approaches into classroom practice and the potential for integrating corpus resources into contemporary digital learning environments.

2. | RESEARCH METHOD

This study employs a Systematic Literature Review (SLR) approach to examine the significance, types, and pedagogical applications of corpus-based methodologies in English as a Foreign Language (EFL) teaching. The method was selected because the study aims to identify, synthesize, and interpret findings from previous research concerning the use of corpora in English language instruction, particularly in vocabulary, grammar, writing, communicative competence, learner autonomy, and teacher development. The literature review focused on academic publications discussing corpus linguistics, corpus-based language teaching, learner corpora, Data-Driven Learning (DDL), corpus literacy, and corpus-assisted language learning. The literature was identified through academic databases and search engines using combinations of relevant keywords, including corpus-based approaches in EFL, corpus linguistics and English teaching, learner corpora, Data-Driven Learning, corpus-based vocabulary learning, corpus-based grammar instruction, and corpus literacy.

The reviewed literature primarily emphasized publications published between 2015 and 2024 to provide a contemporary perspective, while foundational corpus-linguistics works were considered where necessary to establish the conceptual background. The selection process involved several stages. First, relevant publications were identified based on their titles, abstracts, and keywords. Second, studies were screened according to their relevance to corpus-based approaches and EFL instruction. Publications were included when they examined the pedagogical use of corpora, learner corpora, DDL, corpus literacy, or corpus-assisted language learning. Publications that did not directly address language teaching or learning were excluded. The selected studies were subsequently reviewed and categorized according to their main contributions.

The analysis used a thematic synthesis to identify recurring themes across the literature. The themes included the use of corpora for vocabulary acquisition, grammar instruction, academic writing, communicative and pragmatic competence, learner

autonomy, teacher training, digital learning, and the challenges of implementing corpus-based instruction. Findings from the selected studies were compared to identify common patterns, pedagogical benefits, limitations, and potential directions for integrating corpus resources into contemporary EFL classrooms. This methodological procedure provides a structured basis for synthesizing existing evidence and developing an overview of how corpus-based methodologies can support effective and authentic English language teaching.

3. | RESULTS AND DISCUSSION

The literature reviewed in this study indicates that corpus-based methodologies have become increasingly relevant to English as a Foreign Language (EFL) teaching because they provide learners and teachers with access to authentic language data. Rather than presenting language exclusively through predetermined rules, corpus-based instruction allows learners to observe how vocabulary, grammar, collocations, and discourse patterns occur in naturally produced language. This evidence supports the view that corpora can connect linguistic knowledge with actual language use and provide pedagogical information that is difficult to obtain from conventional teaching materials alone (Granger et al., 2015; Meunier, 2016). One of the major findings concerns the diversity of corpora available for EFL instruction. The literature distinguishes between general corpora, spoken corpora, learner corpora, and specialized corpora, each of which can serve different instructional purposes. General corpora provide broad evidence of language patterns, while spoken corpora are particularly useful for examining conversational language and oral communication.

Learner corpora provide evidence of the language produced by students and can therefore assist teachers in identifying recurrent errors and developmental patterns. Specialized corpora, meanwhile, can provide language relevant to particular disciplines and professional contexts. The pedagogical value of these different corpus types depends largely on the learning objectives and the ability of teachers and learners to interpret corpus evidence (Granger et al., 2015; Meunier, 2016). Vocabulary learning represents one of the most frequently identified applications of corpus-based approaches. Corpus data allow learners to investigate word frequency, collocations, lexical patterns, and contextual meanings rather than simply memorizing isolated vocabulary items. Barabadi and Khajavi (2017) found that a data-driven approach could contribute positively to English vocabulary learning, demonstrating the potential of corpus-informed activities to provide learners with richer lexical information. This finding is consistent with the broader argument that corpus-based learning encourages students to understand vocabulary through repeated and contextualized language use (Boulton, 2017). Consequently, corpus activities can help learners distinguish between words that appear semantically similar but differ in their collocational or contextual behavior.

Grammar instruction is another important area in which corpus-based approaches demonstrate pedagogical value. Corpus consultation enables learners to examine

authentic examples of grammatical structures and identify recurring patterns from actual language data. Kim (2019) demonstrates the relevance of exploratory corpus-based grammar instruction by examining the perceptions of teachers and learners in an EFL context. Such an approach can shift grammar instruction from the simple presentation of rules toward an exploratory process in which learners investigate linguistic evidence and develop their understanding of grammatical usage. This learner-centered orientation is consistent with the principles of Data-Driven Learning (DDL), which position learners as active investigators of language data (Boulton, 2017).

The findings also demonstrate a meaningful relationship between corpus-based instruction and writing development. Luo (2016) reported positive effects of DDL activities on EFL learners' writing development, indicating that corpus consultation can assist learners in making more informed linguistic choices during writing. Similarly, Birhan et al. (2021) found that corpus-based instructional mediation could improve EFL students' academic writing skills. These findings suggest that corpus resources can support writing not only by providing vocabulary but also by exposing learners to authentic patterns of grammar, phraseology, and academic language. The value of corpus-based writing instruction is therefore particularly relevant for learners who need to produce formal or academic English.

At a broader level, the effectiveness of corpus use is supported by Boulton and Cobb's (2017) meta-analysis, which synthesized evidence from previous research on corpus use in language learning. Their findings provide empirical support for the potential effectiveness of corpus-based learning while also indicating that outcomes may depend on instructional design and implementation. This is important because corpus technology itself does not automatically guarantee learning improvement. The pedagogical tasks, degree of learner guidance, corpus selection, and learning objectives all influence how effectively corpus data can be transformed into meaningful learning experiences.

Another important finding is the contribution of corpus-based approaches to learner autonomy. DDL requires learners to interact directly with language evidence, formulate observations, compare examples, and draw conclusions about linguistic usage. This process can encourage students to develop greater responsibility for their own learning rather than depending entirely on teacher explanations. Boulton (2017) emphasizes the pedagogical potential of DDL as an approach that encourages learners to explore authentic language data. Similarly, the use of learner corpora allows students to identify patterns in their own language production and become more aware of areas requiring improvement (Granger et al., 2015).

The reviewed literature also indicates that corpus resources are increasingly connected with digital learning environments. Vyatkina (2020) highlights the potential of corpora as open educational resources for language teaching, particularly because digital corpus resources can provide learners with access to authentic language materials beyond the physical classroom. This development is consistent with the

increasing availability of online corpus interfaces and digital tools that allow learners to search concordances and investigate linguistic patterns. Therefore, corpus-based learning can potentially be integrated into blended, online, and technology-assisted EFL instruction.

However, the literature reveals that corpus integration also presents significant challenges. One major issue is the development of corpus literacy, which refers broadly to the knowledge and skills required to understand, search, interpret, and pedagogically apply corpus data. Callies (2019) argues for integrating corpus literacy into language teacher education because teachers need more than technical access to corpus tools; they must also understand how corpus evidence can be translated into appropriate classroom activities. Ma et al. (2022) similarly demonstrate the importance of structured teacher training and online collaboration in developing corpus-based language pedagogy.

Teacher attitudes and confidence constitute another important limitation. Poole (2022) reports that teachers may perceive corpus use as difficult or technically challenging. This suggests that simply providing access to corpus technology is insufficient. Teachers require practical training, appropriate pedagogical materials, and institutional support to integrate corpus activities effectively. Bennett and Uí Dhonnchadha (2023) further emphasize the importance of developing corpus literacy among in-service EFL teachers through a structured teacher education framework. Thus, teacher professional development should be considered an essential component of corpus-based language teaching.

The systematic review conducted by Lusta et al. (2023) reinforces these findings by demonstrating the growing body of research on corpus and DDL applications in language classrooms while also identifying challenges associated with implementation. The evidence suggests that corpus-based approaches are most productive when learners receive appropriate guidance rather than being expected to independently interpret complex corpus data. This observation is particularly important for EFL learners who may have limited linguistic proficiency or limited experience with corpus interfaces.

The results indicate that corpus-based methodologies have considerable pedagogical potential across multiple dimensions of EFL education. The strongest recurring applications identified in the literature involve vocabulary learning, grammar instruction, writing development, learner autonomy, digital learning, and teacher professional development (Luo, 2016; Barabadi & Khajavi, 2017; Kim, 2019; Birhan et al., 2021; Vyatkina, 2020). Nevertheless, the benefits of corpus use depend on appropriate corpus selection, purposeful instructional design, teacher corpus literacy, and adequate learner guidance (Boulton & Cobb, 2017; Callies, 2019; Poole, 2022; Ma et al., 2022).

Taken together, the reviewed evidence supports the integration of corpus-based methodologies as a complementary approach to conventional EFL teaching rather than as a complete replacement for teachers or established instructional practices. Corpora

provide authentic evidence and opportunities for discovery, but teachers remain important in selecting suitable materials, explaining complex patterns, providing feedback, and connecting corpus observations with communicative objectives. The future development of corpus-based EFL teaching therefore depends on combining technological accessibility with sound pedagogical design, teacher training, and learner-centered guidance (Granger et al., 2015; Meunier, 2016; Bennett & Uí Dhonnchadha, 2023; Lusta et al., 2023).

4. | CONCLUSION

This study concludes that corpus-based methodologies have significant potential to enhance English as a Foreign Language (EFL) teaching by providing learners and teachers with access to authentic and contextualized language data. The reviewed literature shows that different types of corpora, including general, spoken, learner, and specialized corpora, can support different instructional objectives and provide valuable evidence of actual language use. The findings indicate that corpus-based approaches can contribute to several important areas of English language learning, particularly vocabulary acquisition, grammar instruction, writing development, communicative competence, and learner autonomy. Data-Driven Learning (DDL) further encourages learners to actively explore linguistic patterns and develop greater awareness of how English is used in authentic contexts. Corpus resources also offer opportunities for integration into digital and technology-assisted learning environments, increasing access to authentic language materials beyond the traditional classroom.

However, the implementation of corpus-based methodologies is not without challenges. Limited teacher knowledge, insufficient corpus literacy, technical difficulties, and the need for appropriate learner guidance can restrict their effective application. Therefore, corpus-based instruction should be implemented through appropriate pedagogical planning and supported by teacher training and professional development. Corpora should be regarded as complementary resources rather than replacements for teachers, whose roles remain essential in selecting appropriate materials, interpreting complex language patterns, providing feedback, and connecting corpus evidence with learners' communicative needs. Corpus-based methodologies provide a valuable framework for making EFL instruction more authentic, evidence-based, learner-centered, and responsive to real-world language use. Future development should focus on improving the accessibility and usability of corpus tools, strengthening corpus literacy among teachers and learners, and integrating corpus resources into digital learning platforms. Such developments can further strengthen the contribution of corpus-based approaches to effective and sustainable English language education.

Acknowledgment

We gratefully acknowledge the contributions of individuals who supported the completion of this article.

Funding Information

This research did not receive any funding.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Ethical Approval and Originality Statement

Ethical approval was obtained for this study. The manuscript represents original work and has not been previously published, nor is it under consideration by another journal.

Data Disclosure Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- Barabadi, E., & Khajavi, Y. (2017). The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. *Cogent Education*, 4(1), 1283876.
- Bennett, C., & Dhonnchadha, E. U. (2023). Becoming corpus literate: An in-service EFL teacher education framework for integrating corpora into EFL teaching. *Applied Corpus Linguistics*, 3(1), 100048.
- Birhan, A. T., Teka, M., & Asrade, N. (2021). Effects of using corpus-based instructional mediation on EFL students' academic writing skills improvement. *Theory and Practice of Second Language Acquisition*, 7(2), 133-153.
- Boulton, A. (2017). Data-driven learning and language pedagogy. *Language, education and technology: Encyclopedia of language and education*, 3, 181-192.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language learning*, 67(2), 348-393.
- Callies, M. (2019). Integrating corpus literacy into language teacher education. *Learner corpora and language teaching*, 245-263.
- Granger, S., Gilquin, G., & Meunier, F. (2015). The Cambridge handbook of learner corpus research. Cambridge: Cambridge University Press.
- Kim, H. (2019). The perception of teachers and learners towards an exploratory corpus-based grammar instruction in a Korean EFL primary school context. *Primary English Education*, 25(1), 123-152.
- Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus*, 5(1), 1255.
- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9(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.
- Meunier, F. (2016). Learner corpora and pedagogical applications. *The Routledge handbook of language learning and technology*, 376-387. Routledge.
- Poole, R. (2022). "Corpus can be tricky": revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, 35(7), 1620-1641.
- Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals*, 53(2), 359-370.