

**THE EFFECTIVENESS OF USING CORPUS LINGUISTICS
TECHNOLOGIES IN TEACHING ENGLISH**

Sabirova Gulnoza Xidirbayevna

PhD, Acting Associate Professor, Ma'mun university

gulnoza_s@mamunedu.uz

Masharipova Adolat Odilbekovna

Assistant teacher, Ma'mun university

masharipova.a@mamunedu.uz

Abstract: *This article examines the effectiveness of corpus linguistics technologies in teaching English as a foreign language. Structured according to the IMRAD model, the study analyzes contemporary approaches to corpus-based language pedagogy, including Data-Driven Learning (DDL) and concordance-assisted instruction, and synthesizes empirical evidence on their pedagogical outcomes. Drawing on recent international research and meta-analytic data, the paper explores how corpus tools affect learners' vocabulary acquisition, collocational competence, grammatical accuracy, and writing proficiency. The findings indicate that corpus-informed instruction yields measurable, though moderate, positive effects on English language learning, particularly when corpus consultation is systematically embedded in the curriculum and accompanied by adequate teacher and learner training. The article concludes that corpus linguistics technologies constitute a valuable, evidence-based resource for improving the efficiency of English language teaching, provided that appropriate pedagogical mediation is ensured.*

Keywords: *corpus linguistics, English language teaching, data-driven learning, concordancer, corpus-based pedagogy, language technology, learner autonomy, EFL.*

Introduction

The rapid development of information technologies over the past three decades has substantially transformed the theory and practice of foreign language teaching. Among the numerous innovations that have entered the field of applied linguistics, corpus linguistics occupies a particularly significant place. Corpus linguistics is a branch of applied linguistics that relies on computer-assisted analysis of large, systematically compiled collections of authentic spoken and written texts, known as corpora, in order to describe how language is actually used¹. Since the so-called “corpus revolution” of the 1990s, researchers and practitioners have increasingly argued that corpus-based methods can transform materials design, syllabus construction, and classroom practice by grounding pedagogical decisions in authentic language data rather than in intuition alone.

The integration of corpus technologies into English language teaching (ELT) has been facilitated by the emergence of user-friendly corpus analysis tools, such as concordancers, which allow teachers and learners with limited technical background to search corpora for patterns of word use, collocation, and grammatical structure. One of the most widely used freeware tools of this kind is AntConc, developed specifically to make corpus analysis accessible within ordinary classroom settings². Alongside such software, pedagogical approaches such as Data-Driven Learning (DDL) have been developed, in which learners themselves act as “language researchers,” inductively discovering patterns of usage through direct engagement with corpus data.

Despite the growing body of research on corpus-based pedagogy, questions remain regarding the actual effectiveness of these technologies in real classroom conditions, the extent to which they improve specific language skills, and the practical difficulties associated with their implementation. The present article aims

¹Curry N., McEnery T. Corpus linguistics for language teaching and learning: A research agenda // Language Teaching. – 2025. – Vol. 58, № 2. – P. 232–250.

²Anthony L. AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom // Proceedings of the IEEE International Professional Communication Conference (IPCC). – 2005. – P. 729–737.

to examine the effectiveness of corpus linguistics technologies in teaching English by systematically reviewing recent empirical and theoretical research on the subject. The study addresses the following research questions: (1) which corpus-based technologies and methods are most frequently applied in ELT; (2) what empirical evidence exists regarding their effect on learners' language competence; and (3) what pedagogical conditions determine the successful implementation of corpus technologies in the English classroom.

Materials and methods

The present study employs an analytical-descriptive method based on a systematic review of scholarly literature published between 2005 and 2025. The material for analysis was drawn from peer-reviewed journal articles, conference proceedings, and empirical studies indexed in international academic databases, selected according to the following criteria: (a) direct relevance to the application of corpus linguistics or corpus-based technologies in language teaching; (b) availability of empirical data or a clearly described pedagogical methodology; and (c) publication in internationally recognized, peer-reviewed sources.

Seven principal sources formed the empirical and theoretical basis of the review, encompassing a meta-analysis of corpus-based instruction³, an empirical classroom experiment on data-driven writing instruction⁴, a critical review of DDL practice in academic writing classrooms⁵, a study of teacher trainees' attitudes toward corpus linguistics⁶, a foundational theoretical account of data-driven learning⁷, a description of corpus analysis software used in language classrooms⁸,

³Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // *Language Learning*. – 2017. – Vol. 67, № 2. – P. 348–393.

⁴Luo Q. The effects of data-driven learning activities on EFL learners' writing development // *SpringerPlus*. – 2016. – Vol. 5. – Art. 1255.

⁵Chen M., Flowerdew J. A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom // *International Journal of Corpus Linguistics*. – 2018. – Vol. 23, № 3. – P. 335–369.

⁶Florou K. Harnessing the power of corpus linguistics in language education: A student-centered approach // *Journal of Contemporary Language Research*. – 2025. – Vol. 4, № 2. – P. 16–22.

⁷Boulton A. Data-driven learning: Taking the computer out of the equation // *Language Learning*. – 2010. – Vol. 60, № 3. – P. 534–572.

⁸Anthony L. AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom // *Proceedings of the IEEE International Professional Communication Conference (IPCC)*. – 2005. – P. 729–737.

and a recent research agenda outlining directions for corpus linguistics in language education⁹.

The analysis proceeded in three stages. First, the selected sources were classified according to the language skill or competence they addressed (vocabulary, grammar, writing, or general learner attitudes). Second, the reported outcomes and, where available, quantitative effect sizes were extracted and compared. Third, the pedagogical conditions and limitations reported across studies were synthesized to identify recurring patterns regarding the effective implementation of corpus technologies in English language teaching.

Results

The review indicates that corpus-based instruction produces a positive, statistically significant, though moderate effect on language learning outcomes. A large-scale meta-analysis of controlled studies comparing corpus-based instruction with traditional methods reported a moderate overall effect size favoring corpus use, with somewhat stronger effects observed for the learning of lexical and grammatical patterns than for general proficiency measures¹⁰.

With regard to writing instruction specifically, an experimental study comparing learners who used a corpus interface (BNCweb) with learners who relied on a general internet search engine found that the corpus-assisted group demonstrated significantly greater gains in writing fluency and lexical accuracy, although no significant difference emerged in syntactic complexity¹¹. This suggests that corpus consultation is particularly effective for addressing local, word-level and phrase-level errors rather than for developing more global aspects of writing competence.

At the same time, a critical review of data-driven learning practice in academic writing classrooms identified several recurring implementation

⁹Curry N., McEnery T. Corpus linguistics for language teaching and learning: A research agenda // Language Teaching. – 2025. – Vol. 58, № 2. – P. 232–250.

¹⁰Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // Language Learning. – 2017. – Vol. 67, № 2. – P. 348–393.

¹¹Luo Q. The effects of data-driven learning activities on EFL learners' writing development // SpringerPlus. – 2016. – Vol. 5. – Art. 1255.

challenges, including learners' limited corpus-searching skills, the time required for effective corpus consultation, and the uneven quality of pedagogical mediation provided by instructors¹². These challenges indicate that the benefits of corpus technologies are not automatic but depend heavily on how the tools are introduced and scaffolded.

Regarding teacher attitudes, a recent study involving teacher trainees found that initial skepticism toward corpus linguistics decreased substantially after trainees were given the opportunity to engage practically with corpus tools and data-driven activities, resulting in a measurable positive shift in perceptions of corpus-based teaching¹³. This finding underscores the importance of hands-on teacher training as a precondition for the successful classroom application of corpus technologies.

Skill / Focus	Technology / Method	Key Finding	Source
General proficiency / lexico-grammar	Corpus-based instruction (meta-analysis)	Moderate, statistically significant positive effect	Boulton & Cobb, 2017
Writing fluency and accuracy	BNCweb corpus interface (DDL)	Significant gains vs. search-engine group; no gain in complexity	Luo, 2016
Academic writing (DDL practice)	Corpus consultation, concordancing	Positive but constrained by skills/time/mediation gaps	Chen & Flowerdew, 2018
Teacher attitudes	Practical DDL training for trainees	Skepticism reduced; perceptions became more positive	Florou, 2025

Table 1. Summary of reviewed empirical findings on corpus-based English language teaching.

Discussion

The synthesized results confirm that corpus linguistics technologies can

¹²Chen M., Flowerdew J. A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom // *International Journal of Corpus Linguistics*. – 2018. – Vol. 23, № 3. – P. 335–369.

¹³Florou K. Harnessing the power of corpus linguistics in language education: A student-centered approach // *Journal of Contemporary Language Research*. – 2025. – Vol. 4, № 2. – P. 16–22.

enhance the effectiveness of English language teaching, particularly in the areas of vocabulary, collocation, and lexico-grammatical accuracy, where authentic frequency and usage data provide learners with evidence that intuition-based teaching cannot easily replicate. However, the moderate size of the reported effects¹⁴ and the implementation difficulties documented in classroom-based studies¹⁵ indicate that corpus tools are not a self-sufficient teaching method but rather a resource whose value depends on careful pedagogical mediation.

This conclusion is consistent with the broader research agenda proposed for the field, which calls for closer collaboration between corpus researchers, curriculum developers, and classroom teachers in order to translate corpus findings into materials and activities appropriate for specific educational contexts¹⁶. Rather than expecting learners to interact directly with raw, unmediated corpus data, effective practice typically involves teachers pre-selecting concordance lines, designing guided discovery tasks, and gradually building learners' capacity for independent corpus consultation, an approach consistent with the original conception of data-driven learning as an inductive, learner-centered method¹⁷.

For the context of English language teaching in Uzbekistan, these findings suggest that corpus technologies could be productively integrated into university-level EFL courses, particularly in the teaching of academic writing, vocabulary, and English for specific purposes, where access to freeware concordancers and openly available corpora removes most financial barriers to implementation. At the same time, the review suggests that such integration should be accompanied by targeted teacher training, since instructors' own familiarity with corpus tools appears to be a decisive factor in determining whether corpus-based activities are used effectively

¹⁴Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // *Language Learning*. – 2017. – Vol. 67, № 2. – P. 348–393.

¹⁵Chen M., Flowerdew J. A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom // *International Journal of Corpus Linguistics*. – 2018. – Vol. 23, № 3. – P. 335–369.

¹⁶Curry N., McEnery T. Corpus linguistics for language teaching and learning: A research agenda // *Language Teaching*. – 2025. – Vol. 58, № 2. – P. 232–250.

¹⁷Boulton A. Data-driven learning: Taking the computer out of the equation // *Language Learning*. – 2010. – Vol. 60, № 3. – P. 534–572.

in the classroom.

The present review has certain limitations. It is based on a relatively small, purposively selected set of sources rather than an exhaustive systematic search, and the studies examined vary considerably in research design, sample size, and target learner population. Consequently, the conclusions drawn here should be regarded as indicative rather than definitive, and further empirical research, particularly in the Central Asian EFL context, is needed to confirm the transferability of these findings.

Conclusion

The analysis presented in this article demonstrates that corpus linguistics technologies, including concordancing software and data-driven learning activities, exert a measurable and generally positive influence on the effectiveness of English language teaching, particularly with respect to vocabulary, collocation, and writing accuracy. At the same time, the benefits of these technologies are not automatic: they depend on systematic curricular integration, adequate teacher training, and thoughtful pedagogical mediation between raw corpus data and classroom activity. It is therefore recommended that English language programs, including those in higher education institutions of Uzbekistan, consider the gradual incorporation of freely available corpus tools into teacher training curricula and course materials, while further empirical research is conducted to evaluate their long-term impact on learner outcomes in local educational settings.

References

1. Anthony L. AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom // Proceedings of the IEEE International Professional Communication Conference (IPCC). – 2005. – P. 729–737.
2. Boulton A. Data-driven learning: Taking the computer out of the equation // Language Learning. – 2010. – Vol. 60, № 3. – P. 534–572.
3. Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // Language Learning. – 2017. – Vol. 67, № 2. – P. 348–393.
4. Chen M., Flowerdew J. A critical review of research and practice in data-driven

learning (DDL) in the academic writing classroom // International Journal of Corpus Linguistics. – 2018. – Vol. 23, № 3. – P. 335–369.

5. Curry N., McEnery T. Corpus linguistics for language teaching and learning: A research agenda // Language Teaching. – 2025. – Vol. 58, № 2. – P. 232–250.

6. Florou K. Harnessing the power of corpus linguistics in language education: A student-centered approach // Journal of Contemporary Language Research. – 2025. – Vol. 4, № 2. – P. 16–22.

7. Luo Q. The effects of data-driven learning activities on EFL learners' writing development // SpringerPlus. – 2016. – Vol. 5. – Art. 1255.