Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

INTEGRATING CORPUS LINGUISTICS IN EFL CLASSROOMS AND DATA-DRIVEN LEARNING FOR LEXICAL ACCURACY

Elmirzayeva Maftuna Dusmurod qizi
Karshi State University, Foreign language faculty
Teacher of Practical English department
Email: maftunaelmirzayeva7@gmail.com

Husanova Laylo Shuhratovna
Karshi State University Student of
Foreign Language faculty
Email: shuhratovnalaylo003@gmail.com

Abstract
This article explores the integration of corpus linguistics and Data-Driven Learning (DDL) into EFL classrooms to enhance learners' lexical accuracy. By analyzing real-life language use through learner-friendly corpora, students gain insights into word usage, collocations, and contextual appropriateness. The study investigates upper-intermediate learners' experiences with tools such as COCA, BNC, and SkELL, comparing outcomes with traditional vocabulary instruction. Results suggest that corpus-based activities improve lexical awareness and promote independent learning. The article also discusses challenges faced by learners and offers practical recommendations for implementing DDL effectively in classroom settings.

Keywords: Corpus linguistics, Data-Driven Learning (DDL), lexical accuracy, EFL learners, collocation, vocabulary instruction, concordance tools

Introduction

Lexical accuracy is a key component of communicative competence in second language acquisition. Traditional vocabulary instruction often falls short in

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

helping learners internalize natural patterns of usage, resulting in frequent lexical errors such as misused collocations and inappropriate word choices. Corpus linguistics offers a solution through authentic language data. This paper examines the use of corpora in EFL contexts, particularly through Data-Driven Learning (DDL), where learners engage with real language examples to notice patterns and improve accuracy. The goal is to explore whether corpus tools can foster better vocabulary use among EFL learners and complement traditional teaching methods.

Methodology

The study was conducted in an upper-intermediate EFL classroom with learners aged 18–25. Students were introduced to three corpus tools COCA, BNC, and SkELL and trained in their basic use. Over four weeks, learners completed vocabulary-focused tasks involving identifying collocations, comparing word usage in different registers, and correcting lexical errors. Feedback was collected through post-activity reflections and questionnaires to evaluate learners' perceptions and experiences. Additionally, student writing samples before and after corpus activities were analyzed to measure changes in lexical accuracy.

Findings

Learners showed increased awareness of lexical patterns, particularly in collocation use and word choice appropriateness. Common errors such as "do a decision" were corrected after observing authentic examples like "make a decision" in the corpus. Most students reported that using corpora made vocabulary learning more meaningful and contextual. While some initially found the tools difficult to use, the majority adapted quickly and appreciated the exposure to real-life usage. Improvements were noted in both written and spoken production, with students displaying greater confidence and independence in vocabulary use.

Overview

This study highlights the pedagogical value of corpus linguistics in EFL classrooms. By moving beyond rote memorization to contextual exploration,

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

learners engage more deeply with vocabulary and develop a better understanding of how language works in real use. Corpus tools also foster learner autonomy and critical thinking, essential skills for long-term language development. While the approach requires teacher support and training, it presents a powerful complement to traditional instruction.

Materials

COCA (Corpus of Contemporary American English)
BNC (British National Corpus)
SkELL (Sketch Engine for Language Learning)
Vocabulary-focused worksheets and guided concordance tasks
Pre- and post-task writing prompts

Results

Quantitative analysis of students' writing showed a notable decrease in lexical errors and an increase in appropriate collocations after corpus-based instruction. Qualitative feedback indicated enhanced learner motivation, greater interest in vocabulary learning, and a stronger sense of learner control. The integration of DDL activities also encouraged peer collaboration and discussion, further reinforcing understanding of language patterns. Quantitative analysis of students' writing showed a notable decrease in lexical errors and an increase in appropriate collocations after corpus-based instruction. Qualitative feedback indicated enhanced learner motivation, greater interest in vocabulary learning, and a stronger sense of learner control. The integration of DDL activities also encouraged peer collaboration and discussion, further reinforcing understanding of language patterns.

Corpus Linguistics and DDL (Data-Driven Learning)

Corpus linguistics is the study of language based on real-life examples collected in large databases, called *corpora*. These corpora contain written or spoken texts that are systematically compiled and analyzed using computer software. Corpus linguistics allows researchers, teachers, and learners to observe how language is actually used, rather than relying solely on intuition or prescriptive rules. One

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

important application of corpus linguistics in language learning is Data-Driven Learning (DDL). In DDL, learners explore authentic language data themselves—usually through concordance tools—to notice patterns, collocations, grammar structures, and meanings. Rather than receiving pre-selected examples from a teacher, students "discover" rules and usage by observing many real instances of a word or structure in context. This method promotes learner autonomy, deeper processing, and awareness of usage nuances. Vocabulary remains one of the most difficult aspects of language learning for EFL (English as a Foreign Language) learners. While students may learn isolated word meanings, they often struggle with using vocabulary *accurately and appropriately* in context. Key challenges include:

1. Collocational difficulties: Learners often combine words unnaturally (e.g., saying strong rain instead of heavy rain).
2. Polysemy and context: Many English words have multiple meanings. Understanding which meaning fits a particular context can be confusing.
3. Register and style: Learners may use informal words in formal situations or vice versa, due to limited exposure to contextual usage.
4. Word formation and derivatives: EFL students might know a base word (e.g., decide) but struggle to use related forms (decision, decisive, indecisive).
5. False friends and L1 interference: Learners often transfer vocabulary use from their first language, which may lead to errors.

Corpus-based tools and DDL approaches can help address these challenges by exposing learners to authentic patterns of use, enhancing their lexical accuracy and depth of vocabulary knowledge. Lexical accuracy refers to the correct use of words in terms of form, meaning, and collocation. This means choosing the right word, in the correct grammatical form, and using it in combinations that sound natural to native speakers. High lexical accuracy is essential for effective and fluent communication in a second language. For example, a common lexical inaccuracy made by EFL learners is: "do a decision" instead of «make a decision". In this case, the verb "do" is incorrectly collocated with the noun "decision." While both "do" and "make" are general-purpose verbs, only "make" is conventionally used with "decision" in English. This is a collocational error, a type of lexical inaccuracy. Corpus Linguistics is the study of language based on

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

real-life examples of texts, both written and spoken, stored in structured collections called corpora (singular: corpus). These corpora are analyzed using software tools to observe how words and phrases are actually used in authentic contexts. Rather than relying on rules alone, corpus linguistics provides *evidence-based insights* into language patterns, frequency, usage, and collocation. By using corpus tools, learners can discover for themselves that native speakers consistently say "make a decision"—not ″do a decision″. This process is part of Data-Driven Learning (DDL) and helps improve lexical accuracy through real usage exposure.

The classroom consists of upper-intermediate EFL learners, typically aged 18–25, studying English for academic or professional purposes. They have a solid grasp of grammar and can understand general spoken and written English but often struggle with precise word choice, natural collocations, and register. Many of these learners rely on translation from their first language and tend to use familiar or overgeneralized vocabulary. The focus of the course is on improving lexical accuracy and fluency through authentic language exposure. In this setting, the teacher introduces corpus-based activities to raise learners' awareness of natural language use and to help them notice patterns in real contexts. These activities are usually integrated into vocabulary, writing, or speaking lessons.

COCA (Corpus of Contemporary American English): A large, balanced corpus of over one billion words, covering spoken, fiction, magazine, newspaper, and academic texts. It's useful for exploring word frequency, collocations, and context of use in American English.

BNC (British National Corpus): A 100-million-word corpus of British English, covering both written and spoken texts. It helps learners distinguish between British and American usage and observe formal vs. informal contexts.

SkELL (Sketch Engine for Language Learning): A simplified, learner-friendly interface that provides quick access to example sentences, word sketches (collocations), and similar words. Ideal for classroom use due to its clarity and ease of navigation. These tools allow students to explore how native speakers use language in context, encouraging independent learning and improving lexical awareness. For example, learners can search for "make a decision" and observe multiple authentic examples, reinforcing correct usage. Many EFL learners report

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

positive experiences when using corpora in the classroom. They find it engaging to explore real-life language examples and appreciate the opportunity to become more independent and reflective in their learning. Corpus tools help learners notice natural collocations and avoid common lexical errors. For example, one student noted, "Now I understand why it's 'make a decision' and not 'do a decision'—I saw it many times in SkELL." Learners also feel more confident choosing words appropriately for different contexts, often commenting that corpus use shows them differences between formal and informal usage more clearly than traditional materials. Additionally, they discover patterns and nuances that are rarely explained in textbooks, moving beyond memorization to genuine understanding. However, some learners initially struggle with unfamiliar corpus interfaces or feel overwhelmed by too much data, highlighting the importance of teacher guidance, especially at the beginning.

Compared to traditional vocabulary teaching methods, corpus-based learning through Data-Driven Learning (DDL) offers a more dynamic and contextual approach. While traditional methods typically involve memorizing word lists, learning definitions, and using artificial example sentences, corpus-based approaches encourage learners to explore authentic language use. Traditional instruction is often teacher-centered, with limited focus on real usage patterns. In contrast, corpus-based methods are learner-centered, requiring active participation and discovery. This promotes deeper processing, greater retention, and a better understanding of how vocabulary functions in actual communication. Though traditional methods provide useful structure and controlled input, learners often prefer a blended approach—using both teacher guidance and corpus tools—to develop lexical accuracy and fluency more effectively.

Conclusion

Integrating corpus linguistics through DDL provides EFL learners with authentic, discovery-based vocabulary learning experiences. It supports the development of lexical accuracy by exposing students to real-world usage, helping them overcome typical challenges such as collocational errors and inappropriate register. Although some initial training is needed, corpus tools can be powerful

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

instruments for enhancing language competence when used alongside traditional methods.

Recommendations

Begin with user-friendly corpus tools like SkELL for lower cognitive load. Provide explicit training and modeling before independent exploration. Use guided tasks that focus on specific lexical issues such as collocations or word choice. Blend corpus-based activities with traditional vocabulary exercises for reinforcement. Encourage learner reflection to deepen understanding and track progress. Integrate DDL tasks regularly into writing and speaking lessons for practical application.

References

1. Xue, Liya. "Using data-driven learning activities to improve lexical awareness in intermediate EFL learners." Cogent Education 8.1 (2021): 1996867.
2. Xue, L., 2021. Using data-driven learning activities to improve lexical awareness in intermediate EFL learners. Cogent Education, 8(1), p.1996867.
3. Lusta, Amel, Özcan Demirel, and Behbood Mohammadzadeh. "Language corpus and data driven learning (DDL) in language classrooms: A systematic review." Heliyon 9, no. 12 (2023).
4. Namozova, Saodat Bakhtiyorovna. "THE IMPORTANCE OF LEARNING A FOREIGN LANGUAGE." Mental Enlightenment Scientific-Methodological Journal 6.03 (2025): 297-304.
5. Elmirzayeva, Maftuna D. "Formation and Development of Social Protection Terms in English and Uzbek Languages." Mental Enlightenment Scientific-Methodological Journal 5.03 (2024): 131-138.
6. ELMIRZAYEVA, Maftuna. "NGLIZ VA O ‘ZBEK TILLARIDA IJTIMOIY HIMOYAGA OID TERMINLARNING LEKSIKOGRAFIK TAHLILI." News of the NUUz 1.1.6. 1 (2024): 406-408.
7. Elmirzayeva, M. (2024). Study of the Evolution of Social Protection Terms and Comparative Study in English and Uzbek Languages. American Journal of Philological Sciences, 4(05), 61-67.

Modern American Journal of Linguistics, Education, and Pedagogy
ISSN (E): 3067-7874
Volume 01, Issue 06, September, 2025
Website: usajournals.org

8. Elmirzayeva, Maftuna, and Nigina Mustafoyeva. "Grammar in Spoken and Written Communication: Practical Aspects–the Use of Communication and Common Errors in Grammar Usage." MAKTABGACHA VA MAKTAB TA'LIMI JURNALI 3.3 (2025).
9. Guan, Xiaowei. "A study on the application of data-driven learning in vocabulary teaching and learning in China's EFL class." Journal of Language Teaching and Research 4, no. 1 (2013): 105-112.
10. Cangır H. Using corpus tools for academic writing in EFL settings: A data-driven learning approach. InNew directions in technology for writing instruction 2023 Jan 2 (pp. 171-191).