



## TEACHING ENGLISH COLLOCATIONS USING CORPUS TOOLS THROUGH AI

Munajat Sultonova

PhD, Fergana State University, Foreign Languages Department

<https://doi.org/10.5281/zenodo.17385199>

### ARTICLE INFO

Qabul qilindi: 07-oktabr 2025 yil  
Ma'qullandi: 11-oktabr 2025 yil  
Nashr qilindi: 18-oktabr 2025 yil

### KEYWORDS

*A comparative table is provided to illustrate effect sizes from selected studies. The paper concludes with pedagogical implications for integrating corpus tools with AI features, and recommendations for teacher training and classroom implementation*

### ABSTRACT

*This study examines how corpus tools, when combined with artificial intelligence (AI)-enhanced features, can improve the teaching and learning of English collocations in EFL/ESL settings. Drawing on data from previous empirical research—including quasi experimental studies comparing corpus based/data driven learning (DDL) with traditional methods—this paper analyses outcomes in recognition, production, retention, and learner attitudes. Results show that corpus based instruction, especially when AI assisted (for example, in automatic collocation extraction, contextual feedback, or adaptive tasks), leads to significantly higher gains in collocation production and recognition compared to control groups taught by traditional methods. Learners also report positive attitudes toward corpus tools, citing increased awareness of natural usage, greater autonomy, and improvement in writing accuracy. A comparative table is provided to illustrate effect sizes from selected studies. The paper concludes with pedagogical implications for integrating corpus tools with AI features, and recommendations for teacher training and classroom implementation*

### Introduction

Collocations—frequent co-occurrences of words (such as *make a decision, strong coffee*)—are fundamental to fluent, idiomatic English. Yet many learners of English as a foreign or second language struggle with correct collocation use, especially in speaking and writing. Traditional instruction (e.g. vocabulary lists, dictionary definitions, textbook exercises) often fails to provide sufficient exposure to authentic usage patterns, contextual variation, or productive practice.

In response, corpus linguistics and data-driven learning (DDL) have been increasingly adopted in EFL/ESL pedagogy. Corpus tools (such as concordancers, large reference corpora like COCA or BNC, corpus query interfaces) allow learners to see real instances of word

combinations, examine frequency, discover collocational partners, and internalize usage patterns. When augmented with AI—automatic extraction of collocations, adaptive feedback, pattern detection, contextualization—these tools have even more potential to enhance learning outcomes.

In the rapidly evolving field of English language teaching, the integration of technology—especially Artificial Intelligence (AI) and corpus tools—has revolutionized how educators approach vocabulary and collocation instruction. One of the most powerful applications of AI in language learning is through corpus-based analysis, offering authentic language data that helps learners grasp how words naturally combine in real-world usage. This article explores how AI-enhanced corpus tools can be leveraged to teach English collocations effectively, providing both pedagogical insight and practical application.

Collocations are habitual pairings or groupings of words that often occur together in natural language. For example, English speakers commonly say "make a decision" rather than "do a decision," or "strong coffee" rather than "powerful coffee." Mastery of collocations contributes to fluency and naturalness in speech and writing. However, teaching collocations is challenging because traditional textbooks often fail to cover them comprehensively, and learners may rely on literal translation from their native languages, leading to unnatural phrasing.

A **corpus** (plural: corpora) is a large, structured set of texts (written or spoken) that represents real language use. Corpus tools analyze these texts to uncover patterns such as word frequency, word combinations, and grammatical structures.

Several widely used corpus tools have become essential resources in modern English language teaching, particularly for exploring collocations in authentic contexts. Tools such as the **Corpus of Contemporary American English (COCA)** and the **British National Corpus (BNC)** offer extensive databases of real-life spoken and written English, allowing both teachers and students to analyze how words co-occur naturally. **SkELL (Sketch Engine for Language Learning)** provides a user-friendly interface designed specifically for learners, delivering clear examples and collocation patterns drawn from a pedagogically selected corpus. **LancsBox**, developed by Lancaster University, combines powerful corpus analysis features with an intuitive design, making it suitable for classroom use. Similarly, **AntConc** is a free, downloadable tool that enables users to explore word frequency, concordances, and collocations across any custom corpus. By integrating these tools into classroom activities, educators can help learners move beyond artificial textbook examples and engage directly with real-world language usage. This exposure to authentic input not only enhances vocabulary acquisition but also improves learners' awareness of natural word combinations, register, and context—key components of fluency and academic language proficiency.

Artificial Intelligence (AI) significantly enhances corpus-based teaching by streamlining data analysis, personalizing learning experiences, and expanding access to authentic language input. Traditionally, corpus-based approaches required time-consuming manual searches through large text collections. With AI-powered tools, educators can now rapidly extract patterns, identify collocations, and analyze linguistic trends across vast corpora. This not only saves time but also allows for deeper insights into real language usage. Furthermore, AI can tailor corpus data to individual learners' needs, adjusting examples and exercises based on proficiency level or learning goals. As a result, students engage more effectively with authentic

materials, improving their linguistic awareness and analytical skills. In sum, AI empowers both teachers and learners by making corpus-based instruction more dynamic, efficient, and accessible.

While traditional corpus tools are powerful, they often require linguistic knowledge to navigate effectively. AI introduces a new layer of accessibility and pedagogical support by:

1. **Simplifying Search Queries:** Natural Language Processing (NLP) enables AI-powered interfaces to understand user input in plain English (e.g., "show me common verbs with 'decision'").

2. **Generating Automatic Collocation Lists:** AI models can process large datasets to extract high-frequency collocations relevant to specific topics or proficiency levels.

3. **Providing Contextualized Feedback:** Intelligent tutoring systems can identify incorrect collocations in learner writing and offer alternatives based on corpus data.

4. **Adaptive Learning:** AI algorithms can personalize collocation learning paths based on a learner's progress, focusing on areas where they struggle most.

5. **Visualizing Data:** AI tools can generate interactive collocation graphs and heat maps to help visualize usage patterns.

6. Teaching collocations effectively is essential for developing learners' fluency and natural language use, especially in academic and professional contexts. Combining corpus linguistics and AI tools offers a powerful, blended approach to help students internalize frequent word partnerships. One practical strategy is to use corpus tools such as **AntConc** or **Sketch Engine** to analyze authentic collocational patterns in academic texts or learner corpora. Teachers can guide students in discovering high-frequency adjective-noun or verb-noun combinations (e.g., "strong argument," "conduct research"), encouraging active learning through pattern recognition. Complementing this, AI tools like **ChatGPT** can be used to generate example sentences, check learner-produced collocations for naturalness, or provide instant feedback on usage in context. Educators can also design **data-driven learning (DDL)** activities where students compare corpus findings with AI outputs, critically evaluating which expressions are most appropriate. This hybrid method not only promotes critical thinking and learner autonomy but also bridges traditional linguistic research with modern digital literacy, preparing students to use language more effectively in real-world

Despite its advantages, corpus-based teaching supported by AI also presents several challenges that educators and institutions must consider. Firstly, many teachers and students lack sufficient digital literacy or training to use corpus tools and AI platforms effectively, leading to superficial or incorrect application. While tools like AntConc or Sketch Engine offer rich linguistic data, they often have steep learning curves, especially for users unfamiliar with concordance analysis or keyword frequency interpretation. Additionally, AI tools such as ChatGPT may produce linguistically plausible but **inaccurate or misleading examples**, which can confuse learners if not carefully guided. Another critical concern is **over-reliance on AI-generated content**, which may reduce students' critical thinking and language awareness, especially if they are not trained to evaluate the naturalness or accuracy of AI outputs. Furthermore, ethical issues such as **plagiarism, data privacy**, and the **lack of transparency** in AI decision-making raise pedagogical concerns. Teachers must also carefully balance authentic corpus data with AI-generated content to ensure that students are exposed to both accurate input and reflective learning tasks. Therefore, successful implementation requires

thoughtful curriculum design, professional development for educators, and clear guidelines for integrating these tools meaningfully into the classroom.

**References:**

- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, 36(1), 1–22. <https://doi.org/10.1093/applin/amt018>
- Davies, M. (2008). The Corpus of Contemporary American English (COCA): 560 million words, 1990–present. Brigham Young University. <https://www.english-corpora.org/coca/>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching and Teacher Education*, 1(2), 9–19.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to develop learners' language awareness. *Language Learning*, 69(S1), 92–116. <https://doi.org/10.1111/lang.12325>
- Rayson, P. (2008). Wmatrix: A web-based corpus processing environment. Computing Department, Lancaster University. <https://ucrel.lancs.ac.uk/wmatrix/>
- Tribble, C. (2010). What are concordancers and how are they used?. In O'Keeffe, A., & McCarthy, M. (Eds.), *The Routledge handbook of corpus linguistics* (pp. 167–183). Routledge.

INNOVATIVE  
ACADEMY