

## ENHANCING STUDENTS' LEXICAL RANGE THROUGH CORPUS LINGUISTICS: A CORPUS-BASED APPROACH TO VOCABULARY DEVELOPMENT

**Yangiyeva Mohinur Asror qizi**

*“Tashkent Institute of Irrigation and  
Agricultural Mechanization Engineers”*

*National Research University*

**Abstract:** Lexical range, defined as the breadth and diversity of vocabulary that learners can understand and use appropriately, is a critical component of second language proficiency. A limited lexical range often restricts learners' ability to communicate effectively, comprehend authentic texts, and express complex ideas. Traditional vocabulary instruction frequently relies on memorization of isolated words, which may not adequately support long-term vocabulary growth. Corpus Linguistics offers an innovative approach by providing learners with access to authentic language data that illustrate how words are used in real contexts. This study explores the role of Corpus Linguistics in enhancing students' lexical range through the use of the Corpus of Contemporary American English (COCA). Employing a qualitative corpus-based methodology, the study analyzes lexical patterns, collocations, synonyms, academic vocabulary, and phraseological units. The findings indicate that corpus-based learning promotes vocabulary expansion, contextual understanding, collocational competence, and learner autonomy. The study concludes that integrating corpus-based activities into English language teaching can significantly improve students' lexical range and overall communicative competence.

**Keywords:** Lexical range, Corpus Linguistics, vocabulary acquisition, COCA, EFL, Data-Driven Learning, lexical competence.

Vocabulary knowledge is one of the strongest predictors of language proficiency. Learners with a broad lexical range can communicate more effectively, comprehend complex texts, and produce more sophisticated spoken and written discourse. Conversely, a restricted vocabulary often results in repetitive language use and communication difficulties.

The concept of lexical range refers to the variety of words and expressions available to a language user. It encompasses knowledge of synonyms, collocations, phraseological patterns, academic vocabulary, and contextually appropriate word choices. Developing lexical range is particularly important for students in higher education, where academic and professional communication requires precise and varied language.

Despite its importance, vocabulary instruction in many TEFL classrooms remains focused on memorizing individual words and definitions. Such approaches may help learners acquire basic vocabulary but often fail to develop deeper lexical knowledge.

Corpus Linguistics has emerged as a powerful tool for vocabulary teaching and learning. By analyzing large collections of authentic texts, learners can observe how words occur in different contexts, identify common collocations, discover synonyms, and understand register-specific usage. One of the most influential corpora available for educational purposes is the Corpus of Contemporary American English (COCA), developed by Mark Davies. COCA contains more than one billion words from spoken, fiction, magazine, newspaper, academic, blog, and web texts, providing extensive opportunities for vocabulary exploration. For example, the English word “run” may refer to physical movement (“run in the park”), management (“run a company”), or operation (“run a computer program”). Similarly, the word “cloud” may denote a weather phenomenon or a virtual data-storage environment in information technology. Such semantic flexibility often causes confusion among TEFL learners who are accustomed to learning vocabulary through isolated word lists and dictionary definitions.

Traditional vocabulary instruction frequently focuses on a single meaning of a lexical item, leaving students unprepared for the multiple meanings encountered in authentic discourse. Consequently, learners may misunderstand texts, misuse vocabulary, and experience difficulties in communication.

Corpus Linguistics provides an innovative approach to vocabulary learning by offering large collections of authentic language data known as corpora. Through corpus analysis, learners can observe how words are used in real contexts, discover collocations, identify semantic patterns, and infer meanings independently.

One of the most valuable corpora available for English language teaching is the Corpus of Contemporary American English (COCA), which contains more than one billion words from spoken language, fiction, magazines, newspapers, academic texts, blogs, and television transcripts. The present study explores how COCA can be used to enhance students’ awareness of polysemous words and facilitate deeper vocabulary learning.

The purpose of this study is to investigate how Corpus Linguistics can contribute to the enhancement of students’ lexical range through authentic language exposure and corpus-based learning activities.

Vocabulary acquisition has been a central concern in second language research for decades. Nation (2013) argues that successful language learning requires both vocabulary breadth and vocabulary depth. While breadth refers to the number of words known, depth involves understanding multiple aspects of word knowledge, including collocations, meanings, and usage patterns.

Corpus Linguistics has significantly influenced vocabulary teaching by providing empirical evidence of language use. Sinclair (1991) emphasizes that words derive meaning from their contexts and frequent co-occurrences rather than existing as isolated units. This perspective led to the development of corpus-based approaches to language teaching.

Johns (1991) introduced Data-Driven Learning (DDL), which encourages learners to discover language patterns independently through corpus investigation. According to Biber, Conrad, and Reppen (1998), corpus-based learning enables students to observe authentic language patterns and develop greater awareness of lexical relationships.

Research suggests that corpus consultation can improve learners' understanding of collocations, phraseology, and academic vocabulary (Kennedy, 1998). Furthermore, studies have demonstrated that exposure to authentic corpus examples facilitates vocabulary retention and promotes learner autonomy.

However, despite extensive research on corpus-based vocabulary instruction, there remains a need for studies specifically examining the role of Corpus Linguistics in expanding students' lexical range.

The findings demonstrate that Corpus Linguistics contributes to lexical range development in several ways. First, corpus examples expose learners to authentic language, helping them discover vocabulary beyond textbook word lists. Second, corpus analysis promotes contextual learning, enabling students to understand how words function in real communication. Third, learners acquire collocational competence by observing frequent word partnerships. Fourth, corpus consultation helps students expand their academic vocabulary and become aware of register differences. Finally, Data-Driven Learning encourages learner autonomy by allowing students to investigate language independently.

These findings support previous research by Sinclair (1991), Johns (1991), and Nation (2013), which emphasizes the importance of contextualized vocabulary learning and authentic language exposure.

This study examined the role of Corpus Linguistics in enhancing students' lexical range through authentic corpus analysis. The findings indicate that corpus-based instruction facilitates vocabulary expansion by exposing learners to synonyms, collocations, academic vocabulary, phraseological units, and register-specific language patterns.

Unlike traditional vocabulary teaching methods that emphasize memorization, corpus-based learning enables students to observe how words function in authentic contexts and develop a deeper understanding of lexical relationships. As a result, learners become capable of using a wider range of vocabulary accurately and appropriately.

The study therefore recommends the integration of Corpus Linguistics and Data-Driven Learning activities into TEFL and ESP classrooms, particularly in higher education contexts. Future research may investigate the impact of corpus-based instruction on students' lexical range through experimental studies involving pre-test and post-test assessments.

### References:

1. Biber D., Conrad S., Reppen R. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.
2. Davies M. *The Corpus of Contemporary American English (COCA)*. Brigham Young University, 2008–. Available at: <https://www.english-corpora.org/coca/> (accessed: 11.06.2026).
3. Johns T. Should you be persuaded: Two examples of data-driven learning // *English Language Research Journal*. 1991. No. 4. P. 1–16.
4. Kennedy G. *An Introduction to Corpus Linguistics*. London: Longman, 1998.
5. Nation I. S. P. *Learning Vocabulary in Another Language*. 2nd ed. Cambridge: Cambridge University Press, 2013.
6. Schmitt N. *Researching Vocabulary: A Vocabulary Research Manual*. London: Palgrave Macmillan, 2010.
7. Sinclair J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.
8. Stubbs M. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell, 2001.
9. Hunston S. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press, 2002.
10. McCarthy M., O'Dell F. *English Vocabulary in Use: Advanced*. 3rd ed. Cambridge: Cambridge University Press, 2017.