

USING CORPUS-BASED LINGUISTICS TO IMPROVE VOCABULARY LEARNING

Erejepova Nilufar – a student of the University of innovation technologies

Abstract. This paper explores the effectiveness of using corpus-based methods to improve vocabulary learning in language education. Drawing on principles from Corpus Linguistics, it examines how authentic language data can enhance learners' understanding of word usage, collocations, and contextual meaning. In contrast to traditional vocabulary teaching methods based on memorization, corpus-based approaches provide meaningful exposure to frequently used lexical patterns. The paper also discusses the benefits and challenges of integrating corpus tools into the classroom, emphasizing the need for teacher guidance and appropriate training. Overall, the findings suggest that corpus-based linguistics offers a practical and innovative strategy for developing vocabulary proficiency in modern language learning environments.

Keywords: Corpus Linguistics, vocabulary learning, lexical development, language teaching, authentic.

Vocabulary acquisition is a fundamental component of language learning, influencing learners' ability to communicate effectively in both spoken and written contexts. Traditional methods of vocabulary instruction often rely on memorization and isolated word lists, which may not fully reflect how language is used in real-life situations. The introduction of corpus-based linguistics has resulted in an important shift in teaching the English language. This move indicates a recognition of the transformative significance that actual language data has in influencing language teaching. With the spread of digital corpora and advances in computational linguistics, educators now have unparalleled access to enormous collections of real-world language use. In contrast, corpus-based approaches offer learners exposure to authentic language data [1,44].

Drawing on principles from Corpus Linguistics, this paper explores how corpus-based methods can enhance vocabulary learning by providing contextualized, frequency-based, and usage-driven insights into language.

Corpus-based linguistics is an approach that uses large, structured collections of real-world texts, known as corpora, to analyze language patterns. These corpora can include spoken conversations, written texts, academic articles, and digital communication. Tools such as concordancers allow learners and researchers to observe how words are used in authentic contexts, including their collocations, grammatical patterns, and semantic nuances.

Unlike traditional dictionaries, corpora provide dynamic examples of language in use, enabling learners to see how vocabulary functions across different contexts and registers.

One of the key advantages of corpus-based learning is the emphasis on frequency and real usage. Learners can identify high-frequency words and prioritize them in their studies. Additionally, corpora help learners understand collocations (e.g., “make a decision” vs. “do a decision”), phraseology (common expressions and idioms), register differences (formal vs. informal usage), contextual meaning (how meaning changes depending on usage).

There are some useful Corpora in the field of teaching English for learners, and Sketch Engine for Language Learning (SKELL) is one of them [1,26].

SKELL is a corpus, where can be found clear information about words, collocations, examples, and thesaurus in English for English language learners. As SKELL is a simple tool to use and understand both for students and teachers. Firstly, words can be easily checked in a particular word or phrase, that is used by real English speakers. There are given examples with the words/phrases, that can be easy to learn how to use the words in the context. Secondly, SKELL is free to use, no registration or payment is needed. And also, SKELL gives good examples of the word or phrase for learners rather than Google search. And then, nowadays SKELL is widely used in other languages, like Russian, Italian, etc. That means it

is more beneficial for my students to use it in their Russian classes to find information about a word/phrase.

Moreover, working on SKELL is suitable for learners' age and level, of course. Because, choosing age suitable technology for English classes is very essential part of the lesson, and teachers are always responsible in implementing technology into their classes. As it is stated in Kiddle (2013) technology that we are implementing in our classes should be equal to the cultural competencies, social skills, as well as learners' needs, and all of these requirements should be involved in using digital technology in language classes.

Apart from this, learning new words, working with them always cause issues for students. In most cases students do not want/eager to learn new words, because almost in all lessons they have to work with new vocabulary. Here SKELL is one of the effective and engaging ways of not only in teaching new vocabulary to adolescent learners, but also it helps to increase learners' writing skills, creativity and outlook. Because, while creating their own short stories, learners improve their writing skills, and they have to think creatively. This method also promotes learning from each other, that is one of the effective methods in teaching English, and demonstrates learner-autonomy [1,37].

For these purposes it is highly recommended to work with SKELL for teachers, who guides their students' different language exams/tests, because knowing and using synonyms are very important in those exams.

Due to Corpora, instead of being passive recipients of knowledge, students become researchers who analyze authentic data. For example, learners can use concordance lines to investigate how a word like "run" is used in different contexts (e.g., "run a business," "run fast," "run out of time"). This process promotes deeper cognitive engagement and helps learners develop analytical skills alongside vocabulary knowledge.

Also, corpora provide real examples of language as it is naturally used, helping learners avoid unnatural or outdated expressions. And, learning vocabulary in context enhances memory retention, as words are associated with meaningful

usage rather than isolated definitions. Corpus tools encourage independent learning, allowing students to explore language patterns on their own. Learners can observe differences between British and American English, formal and informal registers, and spoken versus written language. By analyzing language data, learners develop analytical and problem-solving skills, as well [1,10].

In recent years, the field of corpus linguistics has witnessed a growing intersection with computational methods and natural language processing. These advancements have opened up new avenues for analyzing and understanding language on an even larger scale (Kennedy1998). For example, researchers can now employ machine learning algorithms to automatically annotate and extract linguistic features from massive corpora. This allows for more efficient and comprehensive analysis of language patterns and structures.

Corpus-based linguistics offers a powerful and effective approach to vocabulary learning by emphasizing authentic language use, contextual understanding, and learner engagement. By integrating corpus tools into language education, teachers can move beyond traditional memorization techniques and foster deeper, more meaningful vocabulary acquisition. Although challenges exist, careful implementation and teacher support can maximize the benefits of this approach. As technology continues to evolve, corpus-based methods are likely to play an increasingly important role in language teaching and learning.

References

1. Kiddle, T. (2013) *Developing Digital Language Learning Materials*. In B. Tomlinson (Ed.). *Developing materials for language teaching* 206 p. Bloomsbury Publishing.
2. Hanson-Smith, E. (2018). *CALL (computer-assisted language learning) materials development*. In J. I. Lontas 9Ed.), *TESOL encyclopedia of English language teaching* 1-7 p.
3. [Sketchengine.eu/skell/](https://sketchengine.eu/skell/)
4. O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press.

5. Stubbs, M. (2001). Words and Phrases: Corpus Studies of Lexical Semantics. Blackwell Publishers.
6. Kennedy, G. (1998). An Introduction to Corpus Linguistics. Longman.
7. Tagliamonte, S. (2006). Analysing Sociolinguistic Variation. Cambridge University Press.