SCIENCE CONNECTS PEOPLE

INTERNATIONAL SCIENTIFIC JOURNAL OF

2026 / No. 1(2)

Contemporary Pedagogical and Philological Research

# THE EFFECT OF CORPUS-BASED VOCABULARY INSTRUCTION ON EFL STUDENTS' LEXICAL DIVERSITY

**Bozorova Mahliyo Ulugbekovna**

Senior student, SamSIFL, Samarkand

**Abstract.** Lexical diversity is an important dimension of second language writing proficiency because it reflects the range of vocabulary learners use to communicate ideas effectively and precisely. However, many English as a Foreign Language (EFL) students rely on a limited range of familiar vocabulary, which may result in repetitive expressions and restricted lexical development in academic writing. Corpus-based vocabulary instruction offers an alternative approach by allowing learners to explore authentic language use, identify frequently occurring lexical patterns, and develop a deeper understanding of vocabulary in context. The present study examines the potential effect of corpus-based vocabulary instruction on EFL students' lexical diversity. A qualitative literature-based research design is employed to analyze theoretical perspectives and previous empirical studies concerning corpus linguistics, data-driven learning, vocabulary acquisition, and second language writing development. Particular attention is given to the use of concordance lines, frequency lists, collocation analysis, and authentic language examples in vocabulary instruction. The analysis suggests that corpus-based learning can support lexical development by increasing learners' awareness of vocabulary variation, encouraging independent discovery, and providing exposure to authentic lexical patterns. It may also help students develop more precise vocabulary choices and reduce unnecessary repetition in academic texts. Nevertheless, lexical diversity should not be considered the only indicator of writing quality, since vocabulary accuracy, appropriateness, coherence, and grammatical control are equally important. The study concludes that corpus-based vocabulary instruction can be a valuable pedagogical approach for developing lexical

diversity among EFL learners when combined with explicit teaching, contextualized practice, and regular academic writing activities.

**Keywords:** corpus-based instruction, lexical diversity, EFL students, vocabulary acquisition, corpus linguistics, academic writing, data-driven learning, lexical competence, vocabulary development.

## INTRODUCTION

Vocabulary knowledge is one of the fundamental components of successful second language communication. Students learning English as a Foreign Language need a sufficiently wide vocabulary to express their thoughts, understand academic materials, participate in discussions, and produce effective written texts. However, vocabulary development involves more than memorizing individual words and their translations. Learners must also understand how words are used in different contexts, how they relate to other lexical items, and how they can select appropriate expressions for specific communicative purposes. One important aspect of vocabulary competence in academic writing is lexical diversity. Lexical diversity generally refers to the range of different lexical items used in a spoken or written text. Writers with greater lexical diversity may demonstrate the ability to express ideas through a wider range of vocabulary rather than repeatedly relying on the same words and expressions. For EFL students, developing lexical diversity can be challenging because their vocabulary knowledge is often influenced by limited exposure to authentic English, traditional vocabulary teaching methods, and insufficient opportunities for independent lexical exploration. As a result, students may repeatedly use common words such as "good," "important," "interesting," and "bad" without exploring more precise alternatives. For example, an academic text may become repetitive when a student repeatedly uses the word "important" instead of selecting contextually appropriate alternatives such as "significant," "essential," "crucial," or "valuable." Although these expressions are related in meaning, they cannot always be used interchangeably. Therefore, developing

lexical diversity requires not only learning additional vocabulary but also understanding the meanings, functions, and contextual restrictions of different lexical items. Traditional vocabulary instruction frequently emphasizes textbook definitions, translations, memorization, and isolated vocabulary exercises. These methods can help students acquire basic word meanings, but they may provide limited opportunities to observe how vocabulary functions in authentic communication. Corpus-based vocabulary instruction offers a different approach. It uses electronic collections of authentic spoken or written texts, known as corpora, to help learners investigate real language patterns. Through corpus analysis, students can examine how frequently words occur, which words commonly appear together, and how vocabulary choices differ across contexts and genres. Corpus linguistics has become increasingly important in language teaching because it provides opportunities to connect classroom instruction with authentic language use. Rather than depending exclusively on textbook explanations, students can explore actual examples of English and develop their own understanding of lexical patterns. Johns (1991) introduced the concept of data-driven learning, which encourages learners to act as language researchers by examining authentic linguistic evidence. This approach allows students to identify patterns and formulate conclusions based on real language examples. Similarly, Biber, Conrad, and Reppen (1998) demonstrate the importance of corpus-based approaches for examining linguistic variation and understanding how language differs according to context and communicative purpose. For EFL students, corpus-based vocabulary instruction may provide valuable opportunities to develop lexical diversity because it exposes learners to a wider range of vocabulary and demonstrates how words are used in authentic contexts. Students can investigate synonyms, collocations, word frequency, and contextual meanings through digital corpus tools. For example, they may examine how the words "significant," "considerable," and "substantial" are used in academic texts and identify differences in their typical contexts. This process may help learners

develop more precise vocabulary choices and avoid repetitive expressions in academic writing. The relationship between corpus-based instruction and vocabulary acquisition is also connected with the importance of noticing in language learning. When students examine concordance lines and observe recurring lexical patterns, they become more aware of features that might otherwise remain unnoticed during ordinary reading activities. Nation (2001) emphasizes that vocabulary knowledge involves several dimensions, including form, meaning, and use. Corpus-based learning can contribute to these dimensions by providing contextual information about vocabulary items and their relationships with other words. Furthermore, corpus tools may encourage learner autonomy. Students can independently search for vocabulary examples, compare expressions, and investigate language patterns when preparing essays and research assignments. This is particularly relevant in university EFL contexts, where students are expected to develop independent academic writing skills. In Uzbekistan, English is widely studied as a foreign language in schools and higher education institutions. University students are increasingly expected to produce academic essays, reports, presentations, and research papers in English. Nevertheless, limited lexical diversity remains a challenge for some learners, particularly when they have difficulty moving beyond basic vocabulary. Corpus-based vocabulary instruction may provide an effective way to address this challenge by introducing students to authentic lexical resources and encouraging independent vocabulary exploration. The purpose of the present study is to examine the potential effect of corpus-based vocabulary instruction on EFL students' lexical diversity. The study focuses on the relationship between corpus consultation, vocabulary development, lexical variation, and academic writing performance. It also explores the pedagogical strategies that can support the integration of corpus-based activities into EFL vocabulary instruction.

SCIENCE CONNECTS PEOPLE

## METHODS

The present study adopts a qualitative literature-based research design to investigate the relationship between corpus-based vocabulary instruction and lexical diversity among EFL students. This research approach is based on the analysis and synthesis of theoretical publications and previous empirical studies related to corpus linguistics, vocabulary acquisition, data-driven learning, and second language writing development. The theoretical framework draws on the contributions of researchers such as Johns (1991), Nation (2001), Biber, Conrad, and Reppen (1998), Sinclair (1991), and Schmitt (2000). These publications provide important perspectives on authentic language data, vocabulary learning, lexical knowledge, and the pedagogical applications of corpus analysis. The literature was examined in relation to the central research focus, namely the potential contribution of corpus-based vocabulary instruction to the development of lexical diversity in EFL writing.

Particular attention was given to the pedagogical use of frequency lists, concordance lines, collocation analysis, and contextualized vocabulary exploration. The analysis also considered different dimensions of lexical development, including vocabulary range, lexical variation, word selection, collocational knowledge, and contextual appropriateness. Corpus-based instructional activities were examined in terms of their potential application in university-level EFL classrooms. These activities include searching for academic vocabulary in electronic corpora, comparing synonyms, identifying collocations, analyzing authentic sentences, and applying newly discovered expressions in written assignments. The study also considered the role of learner autonomy and independent language discovery. Students who use corpus tools can investigate vocabulary beyond classroom materials and develop strategies for solving lexical problems independently. Previous research was analyzed thematically to identify common findings and theoretical arguments concerning the benefits and limitations of corpus-based vocabulary learning. The relationship between lexical

SCIENCE CONNECTS PEOPLE

diversity and writing quality was also considered. Particular attention was given to the distinction between using a wide range of words and using vocabulary accurately and appropriately. Because this study is based on previously published research rather than a newly conducted classroom experiment, it does not report original participant data, experimental scores, or statistical improvements. The results are therefore presented as a literature-based synthesis and pedagogical analysis.

## RESULTS

The analysis of theoretical literature and previous research identifies several important ways in which corpus-based vocabulary instruction may contribute to the development of lexical diversity among EFL students.

One of the most important potential benefits of corpus-based vocabulary instruction is the opportunity to encounter a wider range of vocabulary in authentic contexts. Traditional vocabulary learning often introduces a limited number of words through textbooks and teacher explanations. Although this approach provides a foundation for vocabulary acquisition, it may not expose students to the variety of expressions used in real academic communication. Corpus-based activities allow learners to explore large collections of authentic texts and identify vocabulary items that frequently occur in particular contexts. For example, students researching the vocabulary of academic writing may encounter expressions such as "demonstrate," "indicate," "illustrate," "suggest," and "highlight." These verbs can perform related functions, but their use depends on the meaning and structure of individual sentences. By examining authentic examples, students can observe how these verbs function in academic discourse and develop a broader range of vocabulary choices.

This process may help learners move beyond the repeated use of basic vocabulary and develop greater flexibility in written expression.

Another important finding concerns the relationship between corpus-based learning and contextual vocabulary awareness. Knowing the definition of a word does

SCIENCE CONNECTS PEOPLE

not necessarily mean that a student can use it appropriately in every situation. Vocabulary items often have different meanings and patterns depending on their contexts. Corpus tools allow students to examine multiple examples of a particular word and observe how its meaning and grammatical behavior change across different sentences. For instance, students can compare the academic uses of "issue," "concern," "problem," and "challenge." Although these words are sometimes used in similar contexts, they communicate different shades of meaning. By examining authentic sentences, learners can develop a more accurate understanding of these differences. This may contribute to greater lexical precision and help students select vocabulary according to the intended meaning rather than relying on direct translation.

Corpus-based instruction can also support lexical diversity through the development of collocational knowledge. Collocations are combinations of words that frequently occur together in conventional language use. Examples include "conduct research," "strong evidence," "significant impact," and "achieve success."

Students who understand collocations can produce more natural and varied academic expressions. Corpus analysis makes it possible to identify words that frequently occur together and examine their usage in authentic texts. For example, learners searching for the word "evidence" may discover combinations such as "provide evidence," "supporting evidence," "strong evidence," and "empirical evidence." These combinations can help students expand their lexical resources and produce more natural academic sentences.

Repetition is a common characteristic of writing produced by learners with limited vocabulary resources. Students may repeatedly use the same words because they are unfamiliar with alternative expressions or uncertain about their contextual appropriateness. Corpus-based vocabulary instruction may help address this problem by introducing learners to a wider range of lexical alternatives. For example, students can investigate how academic writers use expressions such as "important,"

SCIENCE CONNECTS PEOPLE

"significant," "essential," and "crucial." Through corpus consultation, they can observe differences in usage and identify expressions that are appropriate for specific contexts. However, lexical variation should not be understood as replacing every repeated word with a synonym. Repetition may be necessary when maintaining a clear topic or referring to a specific concept. The main objective should be to develop appropriate lexical flexibility rather than eliminate repetition completely.

The analysis also highlights the potential of corpus-based learning to promote learner autonomy. Traditional vocabulary instruction often depends heavily on teachers and textbook explanations. Corpus-based activities, in contrast, encourage students to investigate vocabulary independently. Learners can use online corpus resources and concordance tools to search for unfamiliar expressions, compare word combinations, and examine authentic examples. This process encourages active learning because students participate directly in discovering language patterns. Johns (1991) describes this approach as data-driven learning, in which learners investigate linguistic evidence and develop their own understanding of language. Such activities may be particularly valuable for university students who need to develop independent strategies for academic reading and writing.

Corpus-based vocabulary instruction may also contribute to academic writing development by exposing students to vocabulary commonly used in scholarly communication. Students can investigate academic expressions related to research, analysis, discussion, and evaluation. For example, they may encounter phrases such as "the findings indicate," "the results demonstrate," "the study suggests," and "further investigation is required. "These expressions can provide useful linguistic resources for writing research reports and academic essays. By examining authentic examples, learners can understand how academic vocabulary functions in different sections of a text. This may help them produce writing that is more varied, precise, and appropriate to academic conventions.

SCIENCE CONNECTS PEOPLE

Another important aspect of corpus-based instruction is the use of frequency information. Word frequency lists can help students identify vocabulary items that are commonly used in particular genres and contexts. For example, an academic corpus may contain frequently occurring words and expressions associated with research, analysis, evidence, and argumentation. This information can help teachers select vocabulary that is relevant to students' academic needs. Authenticity is equally important. Corpus examples demonstrate how language is used by actual writers and speakers rather than presenting only artificial textbook sentences. Such exposure may help students develop a more realistic understanding of English vocabulary and its contextual variation.

## DISCUSSION

The findings of the literature-based analysis suggest that corpus-based vocabulary instruction can be a valuable approach to developing lexical diversity among EFL students. One of the central arguments is that vocabulary learning becomes more meaningful when students examine words in authentic contexts rather than memorizing isolated definitions. This perspective is consistent with Sinclair's (1991) work on corpus linguistics, which emphasizes the importance of examining language through naturally occurring textual data. Biber, Conrad, and Reppen (1998) also demonstrate how corpus-based analysis can reveal patterns of language use and variation across different contexts. These theoretical perspectives support the use of corpus resources in EFL classrooms because they provide students with opportunities to observe authentic vocabulary patterns. The findings also suggest that corpus-based instruction may contribute to lexical diversity by increasing learners' awareness of alternative expressions. When students examine academic texts, they encounter different ways of expressing similar ideas. This may help them move beyond a limited vocabulary repertoire and develop greater flexibility in writing. However, lexical diversity should not be treated as the only measure of vocabulary development. A

SCIENCE CONNECTS PEOPLE

student may use many different words but still produce inaccurate or inappropriate sentences. Therefore, lexical diversity must be considered together with vocabulary accuracy, contextual appropriateness, grammatical control, and overall coherence. This distinction is important because students sometimes assume that sophisticated vocabulary automatically improves writing quality. In reality, effective academic writing requires precise word selection and an understanding of how vocabulary functions within particular contexts. The role of data-driven learning is another important issue. Johns (1991) emphasizes the value of allowing learners to investigate authentic linguistic data and develop their own conclusions. This approach may encourage students to become more independent language learners. Instead of immediately asking teachers for definitions, they can examine examples and identify patterns themselves. Nevertheless, corpus-based learning can be challenging for students with limited language proficiency or insufficient experience using digital resources. Teachers may therefore need to provide guidance when introducing corpus tools. Students should first learn how to search for words, interpret concordance lines, recognize common patterns, and distinguish relevant examples from less useful ones. Another important consideration is the relationship between corpus-based vocabulary instruction and academic writing genres. Different academic disciplines use vocabulary in different ways. Scientific writing, for example, may emphasize terms related to experiments, measurement, and results, whereas social science writing may focus more heavily on concepts associated with society, behavior, and institutions. Therefore, corpus-based activities should reflect students' academic interests and educational needs. The findings also highlight the importance of combining corpus-based learning with explicit instruction. Although students can discover many lexical patterns independently, teacher explanations remain valuable for clarifying subtle differences in meaning, grammatical restrictions, and contextual appropriateness. An effective instructional approach should therefore combine independent corpus exploration with

SCIENCE CONNECTS PEOPLE

guided practice and regular writing assignments. The development of lexical diversity also requires repeated exposure and active vocabulary use. Students should not only identify new expressions but also apply them in their own sentences, paragraphs, and academic essays. Overall, the literature suggests that corpus-based vocabulary instruction can support lexical development by connecting vocabulary learning with authentic language use, contextual analysis, and independent discovery. However, further empirical research involving Uzbek EFL students is needed to establish the measurable effects of corpus-based instruction on lexical diversity and overall writing performance.

## PEDAGOGICAL IMPLICATIONS

The findings have several implications for English language teaching in Uzbek higher education institutions. Vocabulary instruction should include opportunities for students to explore authentic language through electronic corpora. Teachers can introduce accessible corpus tools and demonstrate how to search for vocabulary items and examine their contextual use. Students should be encouraged to compare synonyms and identify differences in meaning and frequency. This can help them avoid using alternative expressions without understanding their contextual restrictions. Collocation analysis should also be incorporated into vocabulary activities. Students can investigate common combinations and use them in academic sentences and paragraphs. Furthermore, teachers can introduce corpus-based writing assignments in which learners select a topic, identify relevant academic vocabulary, examine authentic examples, and apply newly discovered expressions in their own texts. Digital resources such as the British National Corpus, the Corpus of Contemporary American English, and the academic sections of other accessible corpora can provide useful materials for classroom activities, depending on access and availability. Finally, vocabulary assessment should consider not only the number of different words used but also their accuracy, appropriateness, and contribution to the overall quality of writing.

SCIENCE CONNECTS PEOPLE

## LIMITATIONS OF THE STUDY

The present study has several limitations. First, it is based on a literature review and theoretical analysis rather than a newly conducted experimental investigation involving EFL students. Second, lexical diversity is a complex concept that can be measured using different indicators. Simple measures such as type-token ratio may be influenced by text length and do not always provide a complete picture of vocabulary development. Third, the effectiveness of corpus-based instruction may depend on learners' proficiency levels, digital literacy, motivation, and familiarity with corpus tools. Another limitation concerns the differences between academic disciplines. The vocabulary patterns found in one academic corpus may not fully represent the language needs of students studying other subjects. Future research should employ empirical designs involving pre-tests and post-tests, student writing samples, corpus-based vocabulary tasks, and interviews. Such studies could examine whether corpus-based instruction produces measurable changes in lexical diversity, lexical accuracy, and vocabulary use over a specific instructional period. Further investigations involving Uzbek university students would also help identify the most appropriate corpus resources and teaching strategies for the local EFL context.

## CONCLUSION

The present study examined the potential effect of corpus-based vocabulary instruction on EFL students' lexical diversity through an analysis of theoretical perspectives and previous research. The analysis suggests that corpus-based vocabulary instruction can contribute to vocabulary development by exposing learners to authentic language, increasing awareness of contextual word use, and encouraging independent lexical exploration. The use of concordance lines, frequency lists, collocation analysis, and authentic academic texts may help students develop a wider vocabulary range and reduce unnecessary repetition in academic writing. Furthermore, corpus-based learning can promote learner autonomy by providing students with opportunities to investigate

SCIENCE CONNECTS PEOPLE

vocabulary independently and develop their own understanding of language patterns. However, lexical diversity should not be considered an isolated measure of writing quality. Effective academic writing requires a combination of lexical variety, accuracy, contextual appropriateness, grammatical competence, and coherence. In conclusion, corpus-based vocabulary instruction represents a valuable pedagogical approach for developing lexical diversity among EFL students. Its integration into university English language courses may help learners become more confident, independent, and effective academic writers. Further empirical research is recommended to establish its measurable impact on Uzbek EFL students' vocabulary development and writing performance.

## References

1. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

2. Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*(3), 301–315.

3. Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

4. Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *ELR Journal, 4*, 27–45.

5. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

6. Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

7. O‘Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

8. Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

9. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

10. Tribble, C., & Jones, G. (1997). *Concordances in the classroom: A resource book for teachers*. Athelstan.

11. Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.