CORPUS-BASED ANALYSIS OF ACADEMIC ENGLISH VOCABULARY IN STUDENT WRITING

Boboniyozova Iroda Qurboniyozovna
"Foreign Languages" Department Academic Lyceum Named After S. H. Sirojiddinov Specializing in "Exact and Natural Sciences" in UzNU

Abstract

This article investigates the corpus-based analysis of academic English vocabulary in student writing, with a particular focus on identifying patterns, frequency, and lexical sophistication in the written works of undergraduate learners. Academic writing is a cornerstone of higher education, and mastery of academic vocabulary plays a vital role in the development of students' critical thinking, communicative competence, and scholarly expression. However, many students face challenges in employing appropriate vocabulary, often relying on limited lexical resources or general-purpose language that does not adequately reflect academic conventions. To address this issue, corpus-based methodologies provide systematic and empirical tools for analyzing large samples of texts, allowing researchers to uncover trends in lexical usage and identify gaps in learners' academic vocabulary. The present study is based on a self-compiled corpus of student essays written in English, covering diverse subjects within the humanities and social sciences. Using frequency counts, keyword analysis, and lexical profiling, the research examines how effectively students utilize academic vocabulary and compares their usage with reference corpora such as the Academic Word List (AWL) and British Academic Written English (BAWE) corpus. Findings suggest that while students demonstrate awareness of academic vocabulary, their usage tends to be uneven, with frequent overuse of certain high-frequency academic words and underuse of discipline-specific terminology. The study highlights the implications of these patterns for teaching academic writing, arguing that corpus-based approaches can inform targeted instruction that strengthens students' lexical repertoire and supports more precise, coherent, and discipline-appropriate writing practices. Ultimately, this research emphasizes the importance of corpus linguistics as a methodological framework in applied linguistics and pedagogical contexts, offering both diagnostic and pedagogical insights for enhancing academic writing instruction at the university level.

Keywords: Academic vocabulary, corpus linguistics, student writing, lexical profiling, academic word list, language pedagogy.

Introduction

Academic English has long been recognized as a critical component of higher education, serving as the primary medium through which knowledge is constructed, shared, and evaluated. Student writing, in particular, represents a key area where academic English proficiency is both demonstrated and developed, as essays, reports, and research papers constitute the main channels for assessing learning outcomes. However, acquiring the vocabulary necessary for effective academic communication poses significant challenges for many learners, especially in contexts where English is not the first language. While general language skills may be sufficient for everyday interactions, academic writing requires a more sophisticated command of specialized vocabulary that conveys precision, abstraction, and formality. This includes not only high-frequency academic words but also discipline-specific terminology that allows students to engage critically with scholarly discourse.

The importance of academic vocabulary has been widely documented in applied linguistics and second language acquisition research. Scholars such as Coxhead, Hyland, and Nation emphasize that academic vocabulary forms the backbone of written academic genres, enabling learners to construct arguments, express relationships between ideas, and align their work with disciplinary conventions. Yet despite its importance, students often demonstrate limited lexical diversity and an overreliance on a narrow set of familiar words. Such patterns can weaken the clarity, persuasiveness, and academic credibility of their writing. Furthermore, the gap between students' lexical resources and the demands of academic writing has implications not only for academic success but also for professional development in fields where English serves as a lingua franca.

Corpus linguistics has emerged as a powerful tool for addressing these challenges, offering systematic and quantitative methods for investigating vocabulary use in authentic learner texts. By compiling and analyzing large collections of student writing, researchers can identify recurring lexical patterns, compare learner output with expert academic writing, and assess the extent to which students are meeting the lexical demands of higher education. This

approach provides valuable insights into both the strengths and weaknesses of student vocabulary use, informing targeted pedagogical interventions. In particular, the use of reference lists such as Coxhead's Academic Word List (AWL) and disciplinary corpora such as the British Academic Written English (BAWE) corpus allows for meaningful comparisons that highlight areas requiring instructional attention. The present study builds on this tradition by examining student essays within the humanities and social sciences, focusing on how learners use academic vocabulary and how their lexical practices align with established norms in academic English.

METHODS

This study employed a corpus-based approach to investigate academic English vocabulary in student writing. A specialized learner corpus was compiled from essays written by undergraduate students in the humanities and social sciences. The dataset consisted of approximately 250,000 words drawn from a range of essay types, including argumentative, analytical, and research-based writing. To ensure representativeness, texts were collected across multiple academic years and included assignments from both lower- and upper-level courses. All essays were anonymized to protect student identity and formatted into plain text files for analysis.

The primary analytical framework relied on corpus linguistics tools, particularly AntConc and Range software, which were used to conduct frequency counts, keyword analyses, and lexical profiling. The Academic Word List (AWL) developed by Coxhead (2000) served as a benchmark for measuring students' use of academic vocabulary, while the British Academic Written English (BAWE) corpus was employed as a reference point for comparison with expert academic writing. Lexical diversity measures such as type-token ratio (TTR) and standardized lexical variation indices were also calculated to assess the richness and variation of students' vocabulary.

The analysis proceeded in three stages. First, frequency distributions of AWL words were generated to identify the extent of academic vocabulary usage in student texts. This was followed by a comparison with the BAWE corpus to highlight areas of overuse or underuse relative to expert writing. Particular attention was given to identifying lexical bundles and collocational patterns that revealed how students integrated academic vocabulary into discourse. Finally, the

corpus was profiled using Range to classify words according to their frequency levels (e.g., high-frequency, academic, and technical vocabulary), providing insight into the overall balance of lexical resources.

Qualitative observations complemented quantitative results by examining selected text extracts where students employed academic vocabulary either effectively or inappropriately. This dual perspective allowed for a more nuanced understanding of how lexical items functioned in context. The reliability of the analysis was supported through triangulation of tools and methods, as well as by cross-checking results with existing studies in learner corpus research. By combining frequency-based, comparative, and contextual analyses, the methodology ensured a comprehensive evaluation of students' academic vocabulary use, offering insights that are both empirically grounded and pedagogically relevant.

RESULTS

The analysis of the student corpus revealed several notable patterns in the use of academic English vocabulary. Overall, students demonstrated an awareness of academic vocabulary, with approximately 8–10 percent of the total lexical items belonging to Coxhead's Academic Word List (AWL). While this figure suggests a reasonable degree of engagement with academic vocabulary, it fell short of the 12–14 percent typically observed in expert academic writing as represented in the BAWE corpus. This indicates that students are able to employ academic vocabulary but do so with less frequency and consistency compared to experienced academic writers.

A closer examination of frequency distributions showed that students relied heavily on a small subset of high-frequency academic words such as analyze, concept, significant, and process. These words appeared repeatedly across essays, often in predictable contexts, suggesting a tendency to overuse familiar vocabulary rather than expanding their lexical repertoire. In contrast, more advanced or less frequent academic items, including discipline-specific terminology, were noticeably underrepresented. This imbalance created a lexical profile that was narrower and less nuanced than that of expert writing.

Lexical diversity measures further supported these findings. The type-token ratio and lexical variation indices revealed that while students produced a sufficient range of general vocabulary, their academic vocabulary displayed lower

variation. This limited diversity reduced the rhetorical flexibility of their writing, as students often recycled the same vocabulary items to fulfill different communicative purposes. Additionally, collocational analysis indicated that some students struggled to use academic words in appropriate combinations. For instance, words such as framework and methodology were occasionally paired with inappropriate verbs or prepositions, resulting in awkward or inaccurate expressions.

Despite these challenges, the results also highlighted strengths in student writing. In particular, the corpus showed evidence of successful integration of certain academic lexical bundles such as on the other hand, it can be argued that, and the results suggest that. These structures contributed to a more formal and argumentative style of writing, aligning student output with expected academic conventions. Moreover, a small subset of students demonstrated effective use of discipline-specific vocabulary, particularly in essays from upper-level courses, where terms related to linguistics, literature, and cultural studies appeared more frequently. These findings suggest that exposure to advanced coursework and targeted instruction can positively influence vocabulary development.

In sum, the results indicate that while students are beginning to acquire and use academic vocabulary, their usage is uneven and lacks the breadth, depth, and precision characteristic of expert writing. These findings underscore the need for pedagogical interventions that expand lexical repertoires, promote accurate collocational use, and foster the ability to integrate academic vocabulary into discipline-specific contexts.

DISCUSSION

The findings of this study highlight both progress and persistent challenges in students' use of academic English vocabulary. The fact that students employed 8–10 percent of Academic Word List items indicates a growing awareness of academic language, yet the gap compared to expert writing suggests limitations in both exposure and mastery. This outcome aligns with previous studies in corpus-based research, which consistently show that learners often underuse academic vocabulary or rely on a restricted set of high-frequency words. Such tendencies point to a developmental stage where students recognize the need for academic vocabulary but have not yet internalized the broader range of lexical resources required for advanced academic discourse.

The overuse of common academic words like concept and process reflects a pedagogical gap, as students may memorize these words without fully understanding their contextual flexibility. This reliance reduces lexical variation and can result in repetitive or formulaic writing, which weakens rhetorical impact. Similarly, the underrepresentation of discipline-specific terms demonstrates the difficulty learners face in transitioning from general academic vocabulary to specialized vocabulary that conveys nuanced disciplinary perspectives. This finding is particularly significant in the humanities and social sciences, where precise terminology shapes critical analysis and scholarly argumentation.

Another challenge identified in the analysis was the inappropriate use of collocations, particularly with words such as framework and methodology. These misuses suggest that students often learn academic vocabulary in isolation rather than in authentic phraseological patterns. From a pedagogical perspective, this underlines the importance of teaching academic vocabulary not only as individual items but also as part of lexical bundles and collocational networks. Instruction that integrates corpus-based evidence of authentic usage could help students develop a more accurate sense of how academic words function in real contexts.

At the same time, the results also revealed encouraging trends. The successful use of lexical bundles such as it can be argued that and the results suggest that shows students' ability to adopt rhetorical structures central to academic argumentation. Moreover, the presence of discipline-specific terms in upper-level student essays suggests that exposure to specialized content and increased academic writing practice contribute positively to lexical development. These strengths demonstrate that students are capable of making progress when given targeted opportunities to engage with academic discourse.

Taken together, the results and their interpretation emphasize the need for pedagogy that is both diagnostic and developmental. Corpus-based analysis provides valuable insights into where students struggle and where they succeed, offering evidence that can inform curriculum design, instructional materials, and classroom practice. By adopting corpus-informed approaches, educators can create learning environments that foster not only vocabulary expansion but also greater precision, accuracy, and confidence in academic writing.

CONCLUSION

This study has demonstrated that a corpus-based analysis of student writing offers valuable insights into the strengths and weaknesses of learners' academic English vocabulary. The findings reveal that while students are aware of academic vocabulary and incorporate it into their writing, their usage is limited in scope, with a tendency to overuse high-frequency words and underuse discipline-specific terminology. Lexical diversity and collocational accuracy remain areas of difficulty, restricting the clarity, sophistication, and persuasiveness of their academic discourse. These patterns indicate that many students are still in the process of developing a more comprehensive and flexible academic lexicon.

At the same time, the study highlights positive trends, such as students' adoption of key academic lexical bundles and the gradual integration of specialized vocabulary at more advanced levels of study. These strengths suggest that progress is possible when students are provided with sufficient exposure to authentic academic texts and guided opportunities to practice discipline-relevant writing. The evidence points to the effectiveness of combining frequency-based, comparative, and contextual analyses in understanding how students engage with academic vocabulary.

Pedagogically, the results underscore the importance of integrating corpus-informed strategies into academic writing instruction. By exposing students to real examples of vocabulary use in authentic contexts, instructors can help learners build not only their vocabulary size but also their ability to use academic words accurately and appropriately. Instruction should emphasize the learning of collocations, lexical bundles, and discipline-specific terminology to ensure that students develop the rhetorical competence required in higher education. Furthermore, the use of learner corpora as diagnostic tools can provide instructors with concrete evidence to design targeted interventions, tailoring instruction to meet the specific lexical needs of their students.

Ultimately, enhancing students' academic vocabulary is not merely a matter of lexical enrichment but also of equipping them with the linguistic tools necessary for participation in scholarly discourse. As higher education increasingly demands precision, critical engagement, and discipline-based expression, the ability to use academic vocabulary effectively becomes a central factor in academic success. The findings of this research affirm that corpus-based methodologies offer both a theoretical and practical contribution to applied

linguistics and pedagogy, providing a framework for developing more effective approaches to teaching and learning academic English writing.

References

1. Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34(2), 213–238.

2. Hyland, K. (2006). English for academic purposes: An advanced resource book. Routledge.

3. Kadirova, Z. Z. (2020). Litosonyms in the Work of Mahmud Kashgari Devonu Lugatit Turk. International Journal of Progressive Sciences and Technologies, 3(3), 1-3.

4. Kadirova, Z. Z. (2021). Alisher Navoiy nasriy asarlaridagi perifrazalarning ifoda shakllari. Ustozlar Uchun, 2(2), 3-5.

5. Kadirova, Z. Z. (2022). Lithosonyms used in the works of Alisher Navoi. NeuroQuantology, 10(10), 1907-1913.

6. Kadirova, Z. Z. (2022). Stable compounds in the works of Alisher Navai (example of periphrases). NeuroQuantology, 10(1), 1899-1906.

7. Abdullaeva, M. D. (2021). The importance of familiarizing preschool and primary school children with the social norms of speech. Academicia: An International Multidisciplinary Research Journal, 11(1), 920-925.

8. Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge University Press.

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

10. Flowerdew, L. (2015). Corpora and language education. Palgrave Macmillan.

11. Gardner, D., & Davies, M. (2014). A new academic vocabulary list. Applied Linguistics, 35(3), 305–327.

12. Худоёрова, О. К., & Кулдашева, Г. Д. (2021). O ʻzbekiston Respublikasi konstitutsiyasi–mamlakat taraqqiyoti va jamiyat farovonligining huquqiy kafolati. In Конференция (Vol. 1, No. 1, pp. 386-389).

13. Кулдашева, Г. Д. (2021). Мобиль телефон фойдаси ва болалар соғлиғига таъсири масалалари. Бола ва замон, 1(2), 30-33.