



A CORPUS-BASED ANALYSIS OF LEXICAL BUNDLES IN ACADEMIC WRITING OF EFL LEARNERS

Yusupova Shahnoza Abduxafizovna

Associate Professor of Samarkand State Institute of Foreign Languages

Abstract: This study investigates the use of lexical bundles in academic writing produced by English as a Foreign Language (EFL) learners using a corpus-based approach. The main aim of the research is to identify frequency patterns, structural forms, and functional categories of lexical bundles in learner academic essays and compare them with native speaker academic writing. A learner corpus consisting of 120 essays was compiled from undergraduate students' written assignments and analyzed using corpus analysis software. The study applies both frequency-based and functional classification frameworks to examine formulaic language use in academic discourse. The findings reveal that EFL learners tend to overuse basic and general-purpose lexical bundles while underusing more advanced academic structures, which negatively affects coherence, cohesion, and academic style quality. The study highlights the importance of formulaic language instruction and suggests integrating corpus-informed pedagogical strategies into EFL writing instruction to improve academic writing competence and discourse awareness.

Key Words: Corpus linguistics, lexical bundles, academic writing, EFL learners, formulaic language.

This study adopts a corpus-based research design to examine lexical bundle usage in academic writing produced by EFL learners. The research corpus was compiled from 120 essays written by undergraduate students in an English language program. The corpus was digitized, cleaned, and processed using corpus analysis software, which enabled frequency extraction and concordance analysis of multi-word sequences. Lexical bundles were identified based on a minimum frequency threshold and further categorized according to structural patterns (e.g., noun phrase-based, verb phrase-based, prepositional phrase-based bundles) and functional roles (stance expressions, discourse organizers, referential expressions).

The analysis revealed significant discrepancies between EFL learner writing and native-speaker academic writing. Learners demonstrated a strong preference for high-frequency but functionally limited bundles such as “on the



other hand,” “in my opinion,” and “it is important that,” while more sophisticated academic bundles such as “it can be argued that” or “from the perspective of” were rarely used. This indicates restricted lexical repertoire and limited exposure to academic formulaic language. The comparative analysis with a reference native corpus confirmed that native academic writing exhibits greater lexical bundle diversity, higher structural complexity, and more precise functional usage.

From a discourse perspective, the overreliance on repetitive bundles resulted in reduced textual cohesion and weaker argumentative structure in learner essays. The findings also suggest that learners often rely on memorized chunks rather than flexible formulaic expressions, which limits their ability to adapt language to different academic contexts. This pattern reflects partial acquisition of academic discourse conventions and highlights the need for more targeted instructional intervention.

Overall, the results demonstrate that lexical bundle competence is a key factor in academic writing proficiency, and that corpus-based exposure can significantly enhance learners’ awareness of authentic academic language patterns.

The study concludes that lexical bundle knowledge plays a fundamental role in the development of academic writing competence among EFL learners. The findings clearly show that limited lexical bundle variety and overuse of basic formulaic expressions negatively affect coherence, fluency, and academic style. This suggests that learners require explicit instruction in academic phraseology and greater exposure to authentic academic texts.

Corpus-based pedagogy is recommended as an effective instructional approach to improve learners’ awareness of lexical bundles and their functional use in academic discourse. Integrating corpus tools into writing instruction can help learners identify real language patterns, develop more sophisticated academic expressions, and improve overall writing quality. The study also highlights the importance of bridging the gap between theoretical knowledge of academic writing and actual language use in learner production.

Future research should expand corpus size, include longitudinal analysis of lexical development, and explore the role of instructional interventions in improving formulaic language acquisition in EFL contexts.

REFERENCES

1. Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.



2. Coxhead, A. (2000). A new academic word list. *Applied Linguistics*, 21(2), 213–238.
3. Ellis, N. C. (2008). The dynamics of language use, language knowledge, and language acquisition. *The Modern Language Journal*, 92(3), 337–346.
4. Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*.
5. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
6. O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge University Press.
7. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
8. Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.