



## LEXICAL DENSITY AND NOMINALIZATION IN STEM ACADEMIC ENGLISH: A CORPUS-BASED PERSPECTIVE

**Rislikboyev Sardorbek**

Executive director of Presidential School in Andijan, Uzbekistan

### ABSTRACT

This thesis explores the use of lexical density and nominalization in STEM academic English. These two features contribute to the formal and compact style of scientific writing. The study uses a corpus-based approach to analyze how often nominalizations are used and how dense STEM texts are across disciplines. It also compares writing styles in fields such as physics, biology, and engineering. The research aims to understand how language shapes meaning in scientific texts and to support clearer academic writing in STEM fields.

### KEYWORDS

Lexical density, nominalization, STEM writing, scientific English, corpus-based analysis, academic language.

### INTRODUCTION

Academic writing in STEM (science, technology, engineering, and mathematics) fields is known for its abstract, compact, and formal style. One reason for this complexity is the frequent use of lexical density and nominalization. These features allow writers to compress complex ideas into fewer words and present scientific information in a more formal tone. However, these same features can also make scientific texts difficult for students, readers, and even non-native English-speaking authors to process or produce.

This short thesis presents a research plan that aims to explore how lexical density and nominalization function in STEM academic English, based on a corpus-based approach. The study compares texts from different STEM disciplines and analyzes how these two features contribute to the complexity of scientific language.

### LITERATURE REVIEW

Nominalization is the process of turning verbs and adjectives into nouns (e.g., develop → development, active → activity), which helps package information more densely. Lexical density refers to the percentage of content words (nouns, main verbs, adjectives, and adverbs) in a text. Together, these features are important in shaping the style of academic writing.

Istiqomah and Basthomi (2024) analyzed article abstracts from high-ranking applied linguistics journals. They found that nominalization played a major role in creating lexical density, with results showing lexical density ranging from 45% to 72%. This supports Halliday's view of nominalization as a "grammatical metaphor" that helps scientific language sound more formal and abstract.

Jalilifar and Heidari Kaidan (2018) compared 60 research articles from physics and applied linguistics. Their study showed that physics articles used fewer nominalizations than linguistics ones, but both relied on certain patterns of nominalization to create academic tone and

structure. Their findings confirm that different disciplines apply nominalization in different ways, reflecting their field-specific writing conventions.

Lu et al. (2018) examined over 150,000 STEM articles published in PLoS journals and measured linguistic complexity through syntactic and lexical indicators. Their results suggested only small differences in lexical density between native and non-native English writers, but confirmed the central role of lexical and syntactic complexity in professional STEM writing.

### RESEARCH QUESTIONS

This study will be guided by the following questions:

1. What is the average lexical density of STEM academic texts across selected disciplines?
2. How frequently are nominalizations used in scientific writing?
3. How do lexical density and nominalization vary across different STEM fields (e.g., physics, biology, engineering)?
4. What do these patterns tell us about the stylistic nature of scientific English?

### METHODOLOGY

This study will use a corpus-based approach to analyze lexical density and nominalization in STEM academic writing. A digital corpus of approximately 100,000 to 150,000 words will be created using published research articles from platforms such as arXiv, Springer, and PLOS. The selected texts will cover 3 to 4 STEM disciplines, including fields like physics, computer science, and biology. Analytical tools such as AntConc or Sketch Engine will be used to measure lexical density and word class frequency. Nominalization will be examined by identifying common suffixes (such as -tion, -ment, -ity, and -ance) and confirming them through manual analysis. The study will compare texts across disciplines to calculate average lexical density, identify the number and types of nominalizations per 1,000 words, and examine variations based on subject field or authorship background, where possible.

### EXPECTED FINDINGS AND SIGNIFICANCE

It is expected that:

- STEM articles will have high lexical density (above 50%)
- Nominalization will be common, especially in methods and discussion sections
- Fields like physics and engineering may use more compressed nominalized language than biology or environmental science

This research contributes to the understanding of how scientific meaning is shaped by language, especially in English academic writing. It may also help researchers, translators, and students better understand why some scientific texts feel dense or difficult.

### CONCLUSION

Lexical density and nominalization are key features of modern academic English, especially in STEM writing. By analyzing how they function across disciplines, this research hopes to clarify the linguistic strategies that make STEM writing both powerful and challenging. The results may inform future work on improving academic writing support, translation tools, or AI-generated scientific texts.

**REFERENCE**

1. Istiqomah, F., & Basthomi, Y. (2024). Exploring nominalization and lexical density deployed within research article abstracts: A grammatical metaphor analysis. *International Journal of Language Education*, 8(1), 14–28.
2. Jalilifar, A., & Heidari Kaidan, G. (2018). A cross-disciplinary study of nominalization in research articles: A comparative study of applied linguistics and physics. *International Journal of Applied Linguistics and English Literature*, 10(2), 83–118.
3. Lu, X., Yoon, H., & Kisselev, O. (2018). A corpus-based comparison of syntactic complexity in published and student writing. arXiv preprint, arXiv:1807.08374