



## ISSUES OF REPRESENTING LINGUISTIC TERMS IN THE UZBEK NATIONAL CORPUS

<https://doi.org/10.5281/zenodo.18880315>

**Tilovova Guzal Rustamovna**

*Senior Lecturer at the University of Economics and Pedagogy*

**Abstract:** *In recent decades, the creation of linguistic corpora and their application in research and education have become increasingly relevant. This article examines the issues related to the representation of linguistic terms in the Uzbek National Corpus (UNC), their systematization, frequency, contextual usage, and standardization. The study employs descriptive analysis, corpus linguistic methods, comparative analysis, and a historical-typological approach to investigate how linguistic terminology is reflected in the corpus. The findings reveal that although a considerable number of linguistic terms are present in the Uzbek National Corpus, certain key terms remain insufficiently represented. Variability in terminology (e.g., parallel usage of synonymous forms) indicates the need for further standardization. Frequency analysis shows that highly frequent terms such as “word,” “sentence,” “linguistics,” “phonetics,” and “morphology” play a central role in educational and scientific discourse. The study concludes that improving the terminological database of the Uzbek National Corpus will contribute not only to linguistic research but also to lexicography, translation studies, and language teaching methodology.*

**Keywords:** *Uzbek National Corpus, linguistic terminology, corpus linguistics, standardization, frequency analysis, contextual usage.*

### INTRODUCTION

In recent decades, the development of linguistic corpora and their use in scientific research and educational practice have become highly actual issues in linguistics. In global practice, large-scale national corpora of English, Russian, German, Chinese, Turkish, Arabic, and other languages have been developed. These corpora serve as important resources not only for linguistic research but also for lexicography,

translation studies, language teaching methodology, and terminology standardization. In Uzbek linguistics, this process began to develop gradually after independence. Since 2020, the first version of the Uzbek National Corpus (UNC) has been made publicly available. Currently, the development of its theoretical foundations, the expansion of its lexicographic potential, and particularly the comprehensive representation of linguistic terms within the corpus remain pressing scientific



tasks. The system of linguistic terminology reflects the intellectual, cultural, and scientific development of a national language. Significant contributions to the study of terminology in Uzbek linguistics have been made by scholars such as B. Mengliyev, S. Usmonov, G. Abdurahmonov, N. Mahmudov, Sh. Shukurov, Sh. Safarov, and R. Rasulov. In particular, B. Mengliyev emphasizes the importance of analyzing Uzbek linguistic terminology through modern corpus-based approaches. From this perspective, the present article aims to examine the theoretical and practical aspects of representing linguistic terms in the Uzbek National Corpus.

## Methodology

This study employs a multi-methodological approach to investigate the representation, usage, and variation of linguistic terminology in the Uzbek National Corpus (O'zbek milliy korpusi). The methodological framework integrates both traditional linguistic analysis and modern corpus-based techniques, ensuring a comprehensive examination of terminological phenomena across diverse text types. The following research methods were systematically applied:

Descriptive analysis was employed to systematically describe the linguistic terms attested in the corpus in terms of form, meaning, and usage. This method allowed for the identification of:

- canonical forms of linguistic terms,
- semantic scope and definitional nuances,

-functional deployment in different text types (educational, academic, and popular science). Through this approach, the study not only cataloged individual terms but also contextualized their usage patterns within the Uzbek linguistic and pedagogical tradition.

Corpus linguistic methods – to identify frequency indicators, contextual usage, and morphological-syntactic features of terms. These corpus techniques enable an empirical and quantitative understanding of how Uzbek linguistic terms are distributed, standardized, and integrated into educational and academic discourse. They also facilitate the detection of inconsistencies, variant forms, and emerging terminology trends.

Comparative method – to compare Uzbek linguistic terminology with terms found in other national corpora (e.g., the Russian National Corpus and the British National Corpus). By situating Uzbek terminology in a broader international context, the study highlights unique features of Uzbek linguistic practice as well as areas where cross-linguistic alignment or borrowing has occurred.

Historical-typological approach – to examine the stages of formation of Uzbek linguistic terminology and its current representation in the corpus. The methodological framework is based on corpus linguistics theories (T. McEnery, A. Wilson, S. Sharoff), terminology theory (V. Danilenko, D. S. Lotte, B. Mengliyev), and research on Uzbek terminological studies. This approach enabled the study to trace the



development of core and specialized linguistic concepts from early pedagogical and scholarly texts to modern corpus entries, highlighting patterns of continuity, innovation, and standardization.

**Theoretical Foundations** The methodological framework is grounded in several theoretical perspectives:

1. **Corpus linguistics:** theories and methods by T. McEnery, A. Wilson, and S. Sharoff were used to operationalize frequency counts, contextual analysis, and collocational studies.

2. **Terminology theory:** approaches by V. Danilenko, D. S. Lotte, and B. Mengliyev provided principles for identifying term boundaries, semantic precision, and standardization requirements.

3. **Uzbek terminological studies:** research by leading Uzbek linguists was consulted to align corpus findings with national terminological norms and pedagogical practices.

**Integration and Validation.** To ensure methodological rigor, the study triangulated data from multiple sources:

- cross-referencing corpus evidence with dictionary entries, textbooks, and scientific articles,

- verifying term usage against internationally recognized standards,

- conducting internal consistency checks for variant forms and semantic overlap. This integrated approach not only enhances the reliability of the findings but also establishes a model for future

corpus-based terminological research in Uzbek linguistics.

In sum, the combination of descriptive, corpus-based, comparative, and historical-typological methods, grounded in both international and national theoretical frameworks, provides a robust foundation for investigating the representation, variation, and functional deployment of linguistic terminology in the Uzbek National Corpus. This methodological rigor ensures that the study captures both quantitative patterns and qualitative nuances, offering valuable insights for lexicography, pedagogy, and terminological standardization.

## Results

The analysis of linguistic terminology in the Uzbek National Corpus demonstrates that the corpus represents a comprehensive empirical resource for examining the distribution, variation, and functional load of linguistic terms across different genres and discourse types. The corpus includes data from school and university textbooks, peer-reviewed scientific articles, monographs, explanatory and terminological dictionaries, as well as popular science publications. This diversity of sources ensures representativeness and allows for comparison between educational and academic usage. The corpus contains a substantial number of core linguistic terms related to phonetics, morphology, syntax, lexicology, and general linguistics. Fundamental units such as word, sentence, phoneme, morpheme, and



syntax are consistently attested across all subcorpora. However, the analysis also reveals partial or inconsistent representation of several specialized terms. For instance, terms corresponding to morphonology, syntagm, and derivation appear less frequently and often in restricted contexts. In some cases, these terms are replaced by descriptive phrases rather than standardized equivalents. This indicates that advanced or narrowly specialized terminology has not yet achieved stable integration into broader educational discourse. The uneven coverage suggests the need for systematic terminological enrichment and harmonization within corpus resources. The findings indicate:

## Discussion

The representation of linguistic terms in the Uzbek National Corpus reveals several theoretical and practical issues. First, compiling terms in a corpus and observing their usage patterns makes it possible to identify trends in the development of linguistic terminology. As emphasized by B. Mengliyev, terms should be studied not merely as lexical units but as elements of a systematic conceptual category.

Second, the variability of terms within the corpus indicates that the process of standardization in Uzbek linguistics has not yet been fully completed. The experience of terminological commissions in Russian linguistics, which have established unified standards, may serve as a useful model. Third, corpus-based materials

provide opportunities to improve educational dictionaries, textbooks, and manuals. Statistical and contextual analysis of corpus data can help determine how terms should be presented and explained in teaching materials. Finally, research conducted at the Alisher Navo'i Tashkent State University of Uzbek Language and Literature highlights the growing importance of corpus linguistics and terminology studies. Expanding the terminological base of the Uzbek National Corpus and developing new electronic explanatory dictionaries remain priority directions.

## Conclusion

The study demonstrates that the representation of linguistic terms in the Uzbek National Corpus is not yet fully comprehensive. Therefore, the following tasks are recommended:

- expanding the terminological database of the corpus by incorporating new scientific sources;
- standardizing variant forms of terms and establishing unified norms;
- developing corpus-based educational explanatory dictionaries and manuals;
- conducting broader comparative studies with other national corpora. Thus, a more comprehensive representation of linguistic terminology in the Uzbek National Corpus will contribute significantly to linguistic research, lexicography, translation studies, and language education.



## REFERENCES:

1. Biber, D., Conrad, S., & Reppen, R. (2009). *Corpus linguistics: Investigating language structure and use* (2nd ed.). Cambridge: Cambridge University Press.
2. Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.
3. Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2014). *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press.
4. Rasulov, R. Tilshunoslik terminlarining izohli lug'atlarda berilishi masalalari. *Filologiya masalalari*, 2, 2019. 89–96.
5. To'xliyeva, G. O'zbek milliy korpusida terminlarning qo'llanishi. *Til va madaniyat*, 3, 2021. 65–72.
6. Tilovova G. The study of linguistic terms in world linguistics. *Science and pedagogy in the modern world: problems and solutions"* (volume-3. Issue4). 2025. 68-71-p