JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS

VOLUME 05, ISSUE 01
MOPNTHLY JOURNALS

ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

THE IMPACT OF BUILDING THE NUSRATULLO JUMAKHOJA AUTHOR CORPUS ON THE DEVELOPMENT OF THE UZBEK NATIONAL CORPUS

Shohista Akramova Islom kizi

Lecturer,

Asia International University, Uzbekistan

Abstract: This study investigates the linguistic, technological, and methodological impact of developing the Nusratullo Jumakhoja Author Corpus on the structure and expansion of the Uzbek National Corpus (UNC). The paper outlines the principles of corpus compilation, identifies author-specific lexical, semantic, and stylistic features, and evaluates their contribution to the enrichment of the national linguistic database. The findings demonstrate that author corpora play a significant role in enhancing semantic diversity, stylistic variability, and terminological precision in modern Uzbek corpus linguistics.

Keywords: Author Corpus; Uzbek National Corpus; Idiolect; Corpus Linguistics; Stylistics; Lexical-Semantic Analysis; Digital Philology; Nusratullo Jumakhoja.

The rise of corpus linguistics has transformed modern linguistic research, creating new opportunities for empirical analysis, statistical modelling, semantic exploration, and digital preservation of linguistic heritage. In Uzbek linguistics, the establishment of the Uzbek National Corpus represents a major shift toward technologically enhanced research practices. However, the effectiveness of a national corpus depends not only on the quantity of texts collected but also on the inclusion of genre-specific, terminologically rich, and stylistically diverse materials. In this regard, author-specific corpora serve as an essential component in expanding the representativeness and analytical capacity of the Uzbek National Corpus.

The Nusratullo Jumakhoja Author Corpus is built on the scientific and literary works of Professor Nusratullo Jumakhoja—an influential scholar, literary critic, linguist, and leading specialist in Navoi studies. His writings are characterized by terminological precision, poetic interpretation, and deep conceptual analysis, making his works particularly valuable for corpus-based research. The development of this author corpus involves systematic collection, digitization, annotation, and analysis of the author's complete texts. The process includes OCR-based text recognition, orthographic normalization, metadata labelling, morphological tagging, syntactic parsing, and semantic annotation in accordance with modern corpus construction standards. Special attention is given to extracting frequency dictionaries, concordance lines, collocations, stylistic markers, and idiolectal features.

The scholarly literature emphasizes the importance of author corpora in identifying linguistic individuality. International researchers such as Sinclair, Biber, McEnery and Hardie have demonstrated that author corpora allow for precise identification of stylistic tendencies, syntactic patterns, and conceptual preferences. In Uzbek linguistics, although national corpus development has accelerated in recent years, the role of author corpora—particularly those of literary critics and philologists—remains insufficiently explored. Jumakhoja's corpus provides a strong foundation for filling this gap, offering rich material for linguistic, stylistic, semantic, and philological research.

The linguistic significance of the Jumakhoja Author Corpus is reflected in its ability to reveal the author's idiolect. His texts contain recurrent syntactic structures, extended attributive

constructions, metaphorically saturated expressions, evaluative comments, and logically organized academic discourse. The corpus highlights author-specific collocations and phraseological patterns, including frequent combinations related to literary criticism, poetics, and Navoi studies. These features provide valuable stylistic data that enrich the Uzbek National Corpus with new layers of textual diversity.

The lexical-semantic contribution of the corpus is equally substantial. Jumakhoja's works include high-frequency terms such as poetic consciousness, aesthetic interpretation, conceptual depth, spiritual illumination, textual memory, and philosophical-aesthetic insight. These terms expand the semantic domains of the national corpus, especially in fields such as literary theory, philology, and cultural studies. Moreover, the author frequently creates original lexical units—occasionalisms—that carry significant semantic and emotional intensity. Units such as aesthetic radiance, creative origin, poetic memory phenomenon, and intellectual illumination enhance the creative and expressive potential of the national corpus.

From a methodological standpoint, the author corpus contributes to the advancement of Uzbek digital philology. It provides material for stylometric analysis, semantic network modelling, idiolect profiling, and corpus-based textual criticism. By integrating the author's texts into the Uzbek National Corpus, researchers gain access to a more nuanced linguistic ecosystem, enabling comparative studies between general language patterns and author-specific usages. This supports the development of advanced linguistic technologies such as automatic term extraction, semantic clustering, and genre-based language modelling.

In conclusion, the creation of the Nusratullo Jumakhoja Author Corpus significantly enriches the Uzbek National Corpus by expanding its lexical-semantic resources, strengthening terminological precision, enhancing stylistic diversity, and contributing to modern corpus methodologies. The project aligns Uzbek linguistics with international standards of corpus compilation and fosters the growth of digital humanities in the region. The integration of this author corpus into the national linguistic infrastructure promises new perspectives for scholarly research, philological analysis, and linguistic innovation in Uzbekistan.

References

1. Akramova, S. I. (2025). Nusratullo Jumaxo'ja mualliflik korpusi lingvistik ta'minoti. In Kompyuter lingvistikasi: Muammolar, yechim, istiqbollar (Vol. 1, No. 01, pp. 229–233). Toshkent Davlat O'zbek Tili va Adabiyoti Universiteti. (compling.navoiy-uni.uz)
2. Elov, B. B., Hamroyeva, S. M., Abdullayeva, O. X., & Uzoqova, M. T. (2022). Evaluation of Part of Speech tagging in Uzbek language: problems and proposals. Language and Culture: Applied Philology, 2(5), 51–68.
3. Khamrayeva, S. M. (2018). Morphological markup and linguistic model. American Journal of Research, 187–198.
4. Khamrayeva, S., & Alimbekova, K. (2025). Tagging the corpus of Fitrat by lexical layer [Conference paper].
5. Sharipov, M., Mattiev, J., Sobirov, J., & Baltayev, R. (2022). Creating a morphological and syntactic tagged corpus for the Uzbek language. arXiv. https://arxiv.org/abs/2210.15234
6. Madatov, K., Bekchanov, S., & Vičič, J. (2023). Uzbek text summarization based on TF-IDF. arXiv. https://arxiv.org/abs/2303.00461
7. Mansurov, B., & Mansurov, A. (2020). Development of word embeddings for Uzbek language. arXiv. https://arxiv.org/abs/2009.14384