

THE NECESSITY OF CREATING THE AUTHORIAL CORPUS OF NUSRATULLO JUMAKHOJA

Shohista Akramova Islom kizi

Lecturer,
Asia International University, Uzbekistan

Abstract: This article explores the scientific, methodological, and cultural necessity of developing an authorial corpus of Professor Nusratullo Jumakhoja, a distinguished Uzbek philologist, literary critic, and Navoiy scholar. The study justifies the importance of digitizing his rich scholarly heritage consisting of over 300 articles published in various national and international journals. It emphasizes the relevance of corpus linguistics and digital humanities in modern philology and highlights how the creation of an authorial corpus enables objective, data-driven analysis of Jumakhoja's linguistic and stylistic individuality. The article also outlines the stages of corpus development and the potential benefits for Uzbek philological research.

Keywords: authorial corpus, Nusratullo Jumakhoja, corpus linguistics, idiolect, stylistics, digital humanities, Uzbek philology.

The 21st century has marked a profound transformation in philological research, characterized by the integration of digital technologies and linguistic databases. Corpus linguistics has emerged as a central methodology in modern language studies, enabling scholars to analyze large amounts of authentic textual data with high precision. Within this context, the creation of authorial corpora—collections of all known works by a single author—has gained increasing importance.

In Uzbek linguistics, corpus-based studies are still developing. While several national and specialized corpora have been established, the systematic creation of authorial corpora remains limited. Among Uzbek scholars, Professor Nusratullo Jumakhoja occupies a prominent position due to his exceptional contributions to literary criticism, Navoiy studies, and linguistics. He is the author of more than 300 articles published in academic journals, conference proceedings, and literary magazines. His works reflect a unique synthesis of scientific reasoning, literary aesthetics, and linguistic precision. The objective of this article is to substantiate the necessity of developing a digital authorial corpus of Nusratullo Jumakhoja's works, define its theoretical foundations, and demonstrate its value for linguistic, stylistic, and cultural research.

1. Theoretical and Methodological Foundations of Authorial Corpus

The concept of a corpus in linguistics refers to a large, structured collection of texts that represent a language or a specific linguistic variety. According to McEnery & Hardie (2012), corpus linguistics is not merely a method of collecting texts but a systematic approach to analyzing language empirically. Within this framework, authorial corpora play a significant role as they provide a controlled dataset representing an individual writer's linguistic behavior.

Authorial corpora allow researchers to identify idiolectal patterns, recurrent lexical choices, collocations, and syntactic constructions unique to the author. Sinclair (1991) emphasizes that corpus-based studies offer empirical evidence, reducing subjectivity in literary and linguistic interpretation. In the Uzbek linguistic tradition, the concept of authorial corpus has not yet been fully integrated into research methodology. Therefore, constructing Nusratullo Jumakhoja's

authorial corpus would mark a pioneering step toward advancing computational philology in Uzbekistan.

2. Nusratullo Jumakhoja's Scholarly Legacy and Linguistic Individuality

Nusratullo Jumakhoja (Doctor of Philology, Professor) is a recognized scholar, critic, and linguist whose contributions extend to diverse fields—literary criticism, textology, linguistic theory, and Navoiy studies. His articles exhibit a profound sense of linguistic awareness and stylistic artistry. Through his academic writings, he interprets classical literature, modern literary trends, and issues of language development in Uzbekistan.

His discourse is characterized by several notable linguistic features:

Lexical richness: frequent use of abstract nouns and terminological vocabulary.

Intertextuality: references to classical authors such as Alisher Navoiy and other Central Asian thinkers.

Analytical precision: use of logical connectors, parallel structures, and complex syntactic constructions.

Stylistic duality: balance between academic rigor and artistic expressiveness.

A corpus-based study of his works would empirically demonstrate these stylistic features, revealing the systematic patterns that constitute his individual linguistic portrait.

There are several reasons why the development of Jumakhoja's authorial corpus is both urgent and scientifically justified:

The digital corpus will serve as a long-term archive of his linguistic and intellectual legacy, ensuring that future generations can access and study his writings.

Corpus tools (e.g., AntConc, Sketch Engine, Voyant Tools) will enable researchers to study word frequency, collocations, and stylistic tendencies with quantitative precision.

The corpus will serve as a teaching and research resource for students and scholars in philology, linguistics, and literary criticism.

By digitizing Jumakhoja's works, Uzbek linguistics will join international trends in computational and corpus-based humanities.

As Biber et al. (1998) note, linguistic research in the digital age must rely on empirical data. The authorial corpus provides exactly this foundation, transforming textual analysis from subjective interpretation into objective measurement.

Creating an authorial corpus involves several key stages:

Text collection: Gathering all available works by the author, including journal articles, monographs, reviews, and conference papers.

Digitization: Scanning printed materials and converting them into machine-readable formats (e.g., UTF-8 encoded text).

Text normalization: Correcting orthographic inconsistencies and ensuring standardization across all texts.

Annotation: Adding metadata such as title, publication year, genre, and thematic category.

Linguistic tagging: Employing automatic tools to mark parts of speech, lemmas, and syntactic relations.

Database construction: Organizing texts in a searchable digital repository that allows frequency and concordance analysis.

Following these stages ensures the corpus's scientific reliability, transparency, and usability. According to Hunston (2002), the methodological rigor of corpus design determines the accuracy and validity of subsequent analyses.

The creation of Nusratullo Jumakhoja's authorial corpus will yield several important outcomes: Identification of his idiolectal features, including recurrent words, phrases, and grammatical structures.

Compilation of frequency dictionaries and concordances reflecting the most typical lexical units. Stylistic profiling — discovering how linguistic choices differ across thematic fields (e.g., literary criticism vs. Navoiy studies).

Comparative studies — enabling contrastive analysis between Jumakhoja's language and that of other Uzbek scholars.

Contribution to the National Uzbek Corpus — enriching national linguistic resources and promoting the integration of Uzbek data into international linguistic databases. Ultimately, this project will contribute to the development of digital philology and computational linguistics in Uzbekistan, providing a foundation for future corpus-based studies of other Uzbek authors.

The establishment of an authorial corpus of Nusratullo Jumakhoja's works is a scientifically grounded, culturally significant, and technologically feasible endeavor. It will preserve the intellectual heritage of a leading Uzbek scholar while offering powerful analytical tools for researchers and educators. The corpus will enable empirical linguistic and stylistic studies, promote interdisciplinary collaboration, and align Uzbek philology with global trends in digital humanities. The initiative reflects the growing recognition that the future of linguistic and literary research lies in data-driven, corpus-based methodology, which ensures both accuracy and innovation. Nusratullo Jumakhoja's authorial corpus will therefore not only immortalize his academic legacy but also serve as a model for similar projects in Central Asian linguistics.

References

1. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
2. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
3. Leech, G. (2016). *The State of the Art in Corpus Linguistics*. Routledge.
4. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
5. Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell Publishers.
6. Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins.
7. Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.
9. Akramova, SH. (202). *Authorial Corpora in Uzbek Philology: Methodological Perspectives*. Asia International University, Uzbekistan.
10. Sharipova, G. (2022). *Digital Humanities and Corpus-Based Research in Uzbek Philology*. Tashkent: Fan Publications.