Alisher Navoi Author's Corpus: Relevance, Necessity and Significance

Manzura Abjalova[1] [a], Nargiza Gulomova[2] [b] and Gulrux Xasanova[3]

[1]Tashkent State University of Uzbek Language and Literature, Tashkent, Uzbekistan
[2]Navoi Innovations University, Uzbekistan
[3]Samarkand State Institute of Foreign Languages, Samarkand, Uzbekistan

Abstract: The corpus of authorship is a system that consists of a set of texts belonging to a certain author. It is an electronic database based on the semantic classification and research of the writer's writing style, language, and text features in his works using information technologies. The author's corpus is different from other corpora. In its base, the works of a certain author are collected, texts are processed, grammatically and semantically tagged, necessary information and linguistic expressions are determined from the author's materials through the search system, statistical data is provided based on the texts, and metadata is provided. Thus, the author's corpus has a wide-ranging search system, it is an electronic database that covers all types and genres of works created by the author, has the possibility of searching based on special parameters, is not limited in size and it is possible to obtain sources easily and quickly related to the author and his work. The article provides information on the importance of Alisher Navoi author's corpus in the educational process, the need to create a corpus, the information available in the semantic base of the corpus, and the features of semantic tagging of 650 gazels in the collection "Badoye' ul-vasat" written by Navoi.

1 INTRODUCTION

Corpus (Latin means "body") is a complex of language units stored electronically, a source of solving various problems for linguists, and a system of linguo-didactic and educational value for users. Corpus linguistics is a branch of computer linguistics whose object is natural language texts and language corpora. The type of corpus consisting of a base of almost all types of texts of a particular language, processed with linguistic annotations for different purposes is the corpus of the national language, and special corpora intended for a specific purpose are its special / structural type. Also, accentological corpus, parallel corpus, author's corpus, newspaper corpus, educational corpus, artistic text corpus, etc.

Relevance Of Alisher Navoi Author's Corpus. Nowadays, the creation and development of author's corpora has become one of the most advanced areas of modern corpus linguistics. Author's corpora can be used to identify the authors of even anonymous works of art.

Decree No. PF-5850, issued by President Sh. Mirziyoev of the Republic of Uzbekistan on October 21, 2019, aims to significantly enhance the status and prominence of the Uzbek language as the state language. The decree includes objectives such as elevating the status and prestige of the Uzbek language, ensuring its proper integration into information and communication technologies, especially the Internet, within the global information network, and developing computer programs tailored for the Uzbek language. Furthermore, to actively integrate the state language into modern information technologies and communications, as outlined in the Concept of Uzbek language development and language policy improvement for 2020-2030, approved by the President's Decree on measures to develop the Uzbek language and improve language policy in the country dated October 20, 2020, there is a need to create extensive electronic resources containing comprehensive scientific, theoretical, and

[a] https://orcid.org/0000-0002-1927-2669
[b] https://orcid.org/0000-0002-7716-1799

practical information about the Uzbek language. It is also crucial to promote the Uzbek language on the Internet and ensure its significant presence in the global information network. As experts in this field, we bear a great responsibility to achieve these objectives.

To effectively implement these objectives and enhance the oral and written proficiency of students and learners, it is essential to explore the linguistic richness of the Uzbek language from the 15th century. This involves conducting scientific research and engaging educators in educational institutions. Additionally, promoting Alisher Navoi's creative legacy globally requires widespread preservation of our national heritage in modern information and communication systems. To achieve this, it is crucial to promote the works of our ancestors among young people and create philological corpora that present examples of classic literature in a fluent and understandable manner. This includes establishing an Alisher Navoi author's corpus and semantically tagging Navoi's creative heritage, which are important tasks in this regard. Therefore, in the first step, in most of the educational literature issued for secondary education, examples of works of Navoi "Badoye' ul-vasat" in the "Khazayin ul-maoni" collection are created so that students can read and understand the works of Alisher Navoi, understand Navoi and familiarize students with Navoi. His author's corpus was created based on the semantic tags of 650 ghazels in the "Badoye' ul-vasat" divan.

2 LITERATURE REVIEW

Principal Vocabularies Used

Alisher Navoi author's corpus contains thousands of lexemes related to the author's ghazels, semantic tagging of the texts collected in the corpus, providing the lexical meanings of the words by explaining them, allowing the reader to quickly and easily understand the contextual meaning of the word, analysing the lexical compatibility of various word combinations, their combinatorial ability, allows you to determine whether a certain syntactic construction is acceptable or not.

"The language of Navoi's" works were defined by P.Shamsiyev and S.Ibrohimov in explaining the current lexical meaning of the words in the ghazels of the "Badoye ul-vasat" divan, the 4-volume explanatory dictionary of the language of Alisher Navoi's works under the editorship of E.Fazilov, Yu. Berdak's Navoi language dictionary and the dictionary of literary works of Mumtaz and the "Dictionary of works of Alisher Navoi" by H. Muhammad were used.

The Need to Create this Corpus

In world and Uzbek philology, a lot of research has been done on Alisher Navoi's work, his personality, translation of his works. Till now, there are sites that collect Alisher Navoi's biography and the texts of his works, various mobile applications and works of Navoi's works in pdf format, as well as several explanatory logs created on the language of Alisher Navoi's works. But this alone is not enough for the modern student to study Alisher Navoi's work, understand his vocabulary, and understand Navoi.

Information on the life and work of Alisher Navoi in world literature and in the field of information technologies is covered on the web pages. However, neither in world linguistics nor in literary studies, developments have been made that could create Alisher Navoi author's corpus. This corpus is expected to be one of the innovative and practical developments.

Today, there are websites, various mobile applications, and Navoi's works in PDF format that contain his biography and texts. However, this information is insufficient for modern readers to fully grasp Navoi's work, appreciate its quality, and understand the man himself. To enhance the study of Navoi's corpus, a semantically tagged database is necessary. This database would provide explanations for words that may be challenging for contemporary readers, helping them understand the meaning of such words in the context of his poetry.

Semantic Base of Alisher Navoi Author's Corpus

650 ghazels in "Badoye' ul-vasat" book, proverbs, expressions, archaism, historical words, poetic arts such as telmeh, tashbih, tanosub, irsoli masal, etc. Alisher Navoi author's corpus in a new format was created for special secondary and higher education students, and a search interface in the corpus was also developed for the convenience of users. As a result of this, users will be able to develop the skills of working independently with Navoi's work, generally, the sources of the 15th century, and the ability to understand the vocabulary of that period.

In order to turn Navoi's work into a readable process, to understand Navoi's philosophy, and to learn and understand the grammar of the 15th century language, it was determined that the text of Navoi's works must be grammatically equalized. As the ghazels, which are an invaluable treasure of Alisher Navoi's work, are semantically equalized in the

corpus of the authorship, the content and importance of the educational, socio-political, philosophical-ethical issues expressed by the creator, along with the gloss of the artistic symbols presented in the genre, are reflected in the variety of poetic arts, artistic image, logical reasoning. the importance is trivial. 21, 226 explanatory words, more than 5,000 archaisms, 170 historicisms, 43 proverbs, 30 expressions, 236 words related to the art of talmekh, 164 words with the opposite meaning (tazod) for Alisher Navoi MK in the "Badoye ul-vasat" divan. The tagging of lexical units related to 806 ratios and 124 tashbih poetic art became the basis for the creation of Alisher Navoi AC.
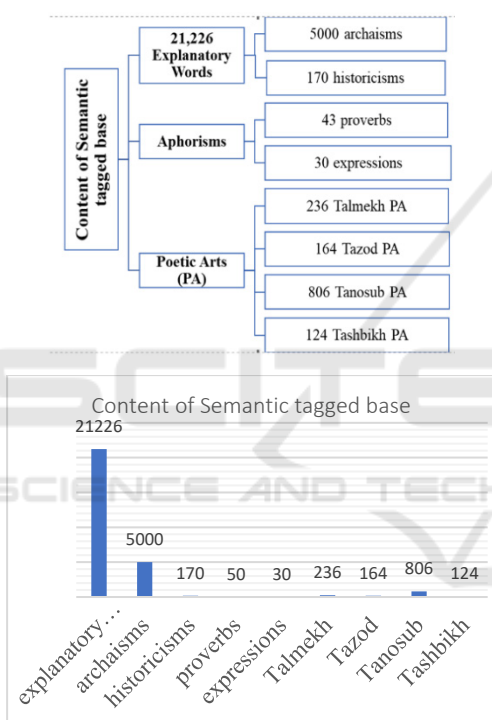

Figure 1: Content of Semantic tagged base of Alisher Navoi AC.

The "Badoye' ul-vasat" divan contains 5001 ghazals, totalling 10002 lines and 66539 words. Each ghazal consists of 6 to 13 stanzas: 1 stanza with 6 stanzas, 437 with 7 stanzas, 8 with 8 stanzas, 187 with 9 stanzas, 1 with 10 stanzas, 14 with 11 stanzas, 1 with 12 stanzas, and 1 with 13 stanzas. This information provides a comprehensive overview of the structure and contents of the divan, which is valuable for studying Alisher Navoi's poetry. [12]. Such statistical analyses of Alisher Navoi MK are considered important, they give the user a clear account and show the scope of work on the database.

The wise words and phrases created by Navoi have become folk proverbs, while his figurative artistic expressions have enriched the phraseology of the Uzbek language. Through such combinations and expressions, he summarized the typical characteristics of events in social life, characteristic of folk wisdom, and showed the possibilities of short, concise, and meaningful expression. More than 50 proverbs and stanzas, each featuring a fixed number of combinations of phrases, are being scientifically analysed and incorporated into column 5 of the ghazels found in the "Badoye ul-vasat" divan.

Creating a semantic database of explanatory words found in Alisher Navoi's ghazels and integrating them into the author's corpus plays a pivotal role in the meticulous examination of rare instances within our classical literature, as well as in the cultivation of our national spirituality. Throughout the process of semantic equalization of the ghazels, it becomes apparent that certain words convey more than ten distinct meanings. This polysemy not only underscores Navoi's mastery of music but also highlights his exceptional skill in crafting rich, multifaceted verses. In order to understand deeply, catch and understand the essence of Navoi's ghazels, scientific and artistic analysis of every stanza and every word used in the ghazel is important that helps in quick and easy understanding, ensures its fluency.

We are currently working on the practical project "Creation of Alisher Navoi Author's Corpus" numbered AL-662205561 to fill the base of this corpus with semantic tagging of other works of Alisher Navoi. In the project, 1950 ghazals from Navoi's collection "Khazayin ul-Maoni" will be semantically tagged.

The "Khazayin ul-maoni" is indeed a remarkable collection, encompassing four divans that offer unique insights into various stages of life: "Garayib us-sigar" (Extraordinaries of youth), "Navodir us-shabab" (Rarities of youth), "Badoye ul-vasat" (Arts of middle age), and "Favoyid ul-kibar" (Benefits of old age). These divans not only provide a glimpse into the poet's profound understanding of human experience but also showcase his mastery of language and poetic form.

Today there are websites, various mobile applications, and including works of Navoi's works in pdf format, which contain Alisher Navoi's biography and the text of his works. But this little bit is not enough for the modern reader to study Alisher Navoi's work, understand his quality, and understand Navoi. In order to complete this chapter in Navoi's corpus, a semantic tagged base is necessary to know

the explanation of the words that the reader has difficulty understanding today and to understand the meaning of such a word in the line.

3 RESULTS

The corpus consists of 8 columns:

Column 1: Alisher Navoi's biography section is a database about the thinker's life and creative activity.

Column 2: Simple and special search by corpus. When you type the de-sired word in the search page, all ghazel verses that use this word will be dis-played in the search window. By referring to any of the ghazels, the text of the ghazels and metadata related to the ghazel can be read.

Column 3: 8 divan texts belonging to Alisher Navoi (including the first di-van compiled by his fans).

Column 4: Alisher Navoi's works (odes; written on scientific, artistic, religious, historical, religious topics) are collected and can be used as a database.

Column 5: 650 ghazels and poetic works in Badoye ul-vasatdevan are presented to the user with semantic tags.

Column 6: About Alisher Navoi's corps.

Column 7: research results for this corpus.

Column 8: information about the authors of the corpus.

The created corpus can serve as a source of satisfaction of spiritual needs for specialists, educational material at training stages, and a source of information related to fields.

Alisher Navoi's corpus of authorship has educational, historical, linguistic, social, educational, and spiritual significance, and the creation of this corpus creates the following opportunities:

- Studying Navoi's personality.
- Study of literary style.
- Linguo-poetic analysis of the poet's work.
- Researching the possibility and skill of the creator.
- creation of author's dictionaries.
- compilation of author's phrases.
- finding the author of anonymous works through parameters that show the personality and style of the creator in AC.
- author's paraphrase, summary of wisdom; it is possible to determine the scope of use of figurative expressions from the context of the creator.

Representation of Alisher Navoi Authors Corpus

Representativeness means that all information about the text is given. Represented texts have their source, style, period of writing, author, age of audience, type of text clearly indication [11]. 650 ghazels available in the divan were separately represented in the author's corpus created on the basis of the Badoye ul-vasat divan. A total of 20 types of metadata were formed according to the content of each ghazel (oshiqona, orifona, rindona), the age of the audience (15+, 18+), the number of word forms used in the ghazel.

Metadata for Alisher Navoi's "Badoye ul-vasat" divan ghazels include:

1. Title of the work ("Badoye ul-vasat").
2. Author (Alisher Navoi)
3. Gender of the author (male).
4. The author's year of birth (February 9, 1441).
5. The year of the author's death (January 3, 1501).
6. The time when the devan was created (1492-1498 years)
7. Year of publication (2011).
8. Publication parameter (number).
9. Publishing house ("Tamaddun" LLC publishing house).
10. Field of application (literature).
11. Literary type (lyric).
12. Genre (ghazel).
13. Time and place of the event (Herat).
14. Text style (artistic).
15. Text type (orifona, romance, rindona).
16. Audience age (15+, 18+ years).
17. Potential of the audience (for the public).
18. Type of internal corpus (author's).
19. Amount of word form (66 539).
20. Tagger (G'ulomova N.)

In the future, it is intended that Alisher Navoi's corpus of works will be covered as corpus units of all his writings in prose and verse, scientific, historical, philosophical, and religious works of the thinker.

Software of the Corpus

A corpus in its modern sense is a reliable database on a computer, in the pro-cess of its creation, special programs are used. HTML, CSS, JS, programming languages, Bootstrap5, JQUERY design frameworks were used to create the de-sign of Alisher Navoi's ghazels. Figure 2 shows the Business Process Model and Notation (BPMN) of Alisher Navoi author's corpus. Using the Python programming language and the Django framework, the general and special search part of Alisher Navoi's corpus ensures that the
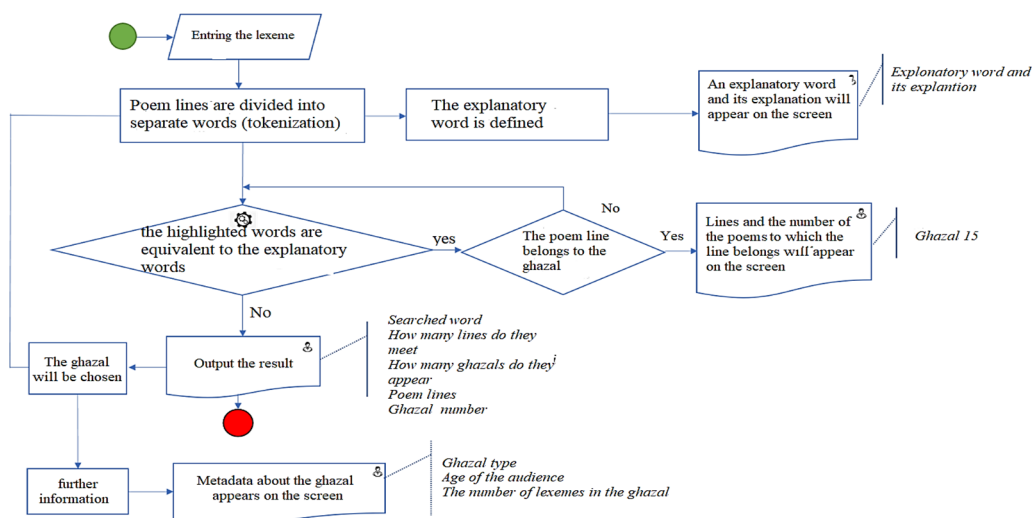
Figure 2: Business Process Model and Notation (BPMN) of Alisher Navoi author's corpus.

semantic tag is displayed on the screen when you click on the explanatory words in the ghazels. Since the interface is considered the first impression of the case, ensuring its perfection in a unique design is a very important process. National and, at the same time, modern features were taken into account when creating the interface of Alisher Navoi's author's corpus.

4 DISCUSSION

Historical Significance of Alisher Navoi Author's Corpus

In Navoi's ghazels, there are many stanzas with the participation of historical, literary, mythical figures, geographical and ethnic place names, which create the art of talmekh from an artistic point of view, the most important thing is that in the process of determining the meaning of the words in the reading stanza, historical, artistic figures and geographical, ethnic place names will also have information about. For example, in the ghazels of "Badoye ul-vasat" divan, he referred to the word Kavsar 23 times, and this word expresses the meanings of material and spiritual fullness:
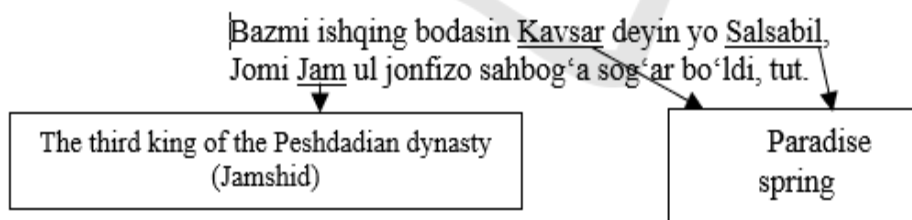

Figure 3: Understanding the meaning of words in the ghazel.

The Linguistic Significance of Alisher Navoi Author's Corpus

The author's corpus has the opportunity to show completely the author's language, in detail, and objectively, so such corpuses are distinguished from other information banks by their advantage. The corpus can serve as a basis, source, tool for various types of research. Another advantage of such corpora is that with their help it is possible to know the languages of not only a single word or sentence, but

also the entire work. The fact that the information in the author's corpora is edited on the basis of scientific sources guarantees the accuracy and reliability of the information provided in it and allows for comprehensive and objective organization of the entire spectrum of linguistic phenomena.

According to Navoi's author's corpus, the thinker receives quick, clear and complete information about all the linguistic features of word units, their changes from the point of view of the

modernity of the language and the disuse of words in today's social life, the cases of activation and passivation, as well as linguistic phenomena. Enables automatic processing of texts. The fact that the information in the author's corpora is edited on the basis of scientific sources guarantees the accuracy and reliability of the information provided in it and allows for comprehensive and objective organization of the entire spectrum of linguistic phenomena. From the corpus containing a large amount of spiritual treasure, in comparative-historical and cross-typological studies of linguistics, conducting lexicographic research, creating 15th-century frequency dictionaries with the help of Alisher Navoi's works, compiling historicism and archaisms, conducting research on the etymology of words, words that are difficult for today's fluent readers to understand understanding the explanation using the array of contextual examples provided in the corpus provides many linguistic possibilities, such as finding its semantic explanation. Regardless of the audience's age, this will serve to increase the number of those who are interested in our classic literature, who understand the language of Navoi's era, and who can enjoy Navoi's spiritual treasure by reading his works. The linguistic basis of the corpus of authorship is to proceed from the word lexeme to its content. In this case, the translation is carried out based on the morphological, syntactic, and semantic analysis of the language, dictionaries, grammatical rules, corpus of texts.

Educational Significance of Alisher Navoi Author's Corpus

The corpus of the author is considered important in the educational process of elucidating the educational, social, historical, linguistic, and didactic significance of the author's works, in classifying the dialectal features of the words used in the author's works, and in studying the semantic features of the National Language. The language style of the 15th century can be taught to students in comparison with today's language. As a result of the statistical analysis of literary texts, language units (nouns, adjectives, keywords, verbs, grammatical forms, sentence structure, in a word, tools that show the writer's style) that are frequently used in the text are determined by means of linguistic statistical analysis. Comparative analysis of evidence from different texts allows us to determine the content of the text, the period in which the text was composed, the argumentative nature of the evidence, and even the authorship. The organization of Alisher Navoi's life and work at all levels of secondary schools increases the educational

significance of Alisher Navoi with the broad use of the thinker's career. This corpus serves as a convenient source of information for the preparation of educational material during the educational process, the teacher will have the opportunity to prepare quickly fresh, meaningful, reliable educational material for the educational session.

Social Significance of Alisher Navoi Author's Corpus

Author's corpora allow students to develop research capacity and perform small research based on reliable evidence. It should be noted that the social significance of the author's corpora is considered extensive, it can be used in linguistic, ethno-psycho-linguistic research, mother tongue, literature, foreign language teaching, automatic text processing, translation programs.

Corpus of authorship is also the most convenient means of monitoring the changes in the vocabulary of the language (neologism, historicism, archaism phenomena). In the process of analysing the possibility of lexical-semantic com-bination of the words used by the author, the possibility of comparing the dictionary and grammar of the ancient and new generations will expand. The rich-ness of the author's word by means of the author's corpus provides practical help in differentiating the possibilities of obtaining a child belonging to several se-mantic categories at once. Through the author's corpora, it is possible to learn about the poet's linguistic views as well as his attitude to social spheres and political issues.

As a result of preserving our spiritual heritage, forming the skills of glorifying the spirit of ancestors, increasing the culture of reading, it is possible to acquire deep thoughts and expand thinking.

Corpus Users

Not only experts and linguists, but also representatives of all fields, as well as applicants and researchers, and foreigners interested in the personality and work of Alisher Navoi will have the opportunity to use the corpus. A modern author's corpus was created and put into practice aimed at promoting and developing Uzbek computer and corpus linguistics, improving the educational process and re-search, meeting the educational needs of specialists of all ages, increasing vocabulary and literacy.

5 CONCLUSION

This article discusses the features of the author's corpus that differ from other corpora, the technical capabilities of Alisher Navoi's author's corpus, the design with National Colorite in the interface, the composition of its semantically tagged base, the importance of the interactive system in the study of Navoi's works, and examples explain the educational, historical, linguistic and social significance of Alisher Navoi's AC. Explanatory dictionaries are not considered a sufficient source for reading and understanding 15th-century texts. A modern interactive technological tool for semantic tagging – language corpora - is now important for understanding the interpretation of words, especially their contextual interpretation.

REFERENCES

Abjalova M., Adalı E. and Iskandarov O., "Educational Corpus of the Uz-bek Language and its Opportunities," 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023, pp. 590-594, doi: 10.1109/UBMK59864.2023.10286682.

Abjalova M. (2021). The importance of language corpus in the construc-tion of lexicographic sources. Current Research Journal of Philological Sciences (2767-3758), 2(12), 161–166. https://doi.org/10.37547/philological-crjps-02-12-31. https://masterjournals.com/index.php/crjps

Abjalova M. Corpus Linguistics. [Text]: methodological manual / M.A. Abjalova. - Tashkent: Nodirabegim, 2022. - 110 p.

Abjalova M., Gulomova N. ALISHER NAVOI AND THE THIRD RENAISSANCE PERIOD // Procedia of Theoretical and Applied Sciences. Vol. 4 (2023). 28.02.2023. – pp. 111-115. ALISHER NAVOI AND THE THIRD RENAISSANCE PERIOD | Procedia of Theoretical and Applied Sciences

Abjalova M., Gulomova N. Author's Corpus of Alisher Navoi and its Se-mantic Database. // IEEE – UBMK – 2022: 7th International Conference on Com-puter Science and Engineering. 24-26 September 2022. – Diyarbakir, Turkey. – pp. 182-187. Impakt Factor 5.5. DOI: 10.1109/UBMK55850.2022.9919546