

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

Manzura Abjalova* and Nargiza Gulomova†

Tashkent State University of Uzbek Language and Literature named after Alisher Navoi, Tashkent, Uzbekistan

Keywords: Alisher Navoi, Devon (Collection), Author Corpus, Gazel, Interface, Representation, Metadata.

Abstract: The author's corpus of authorship is a system comprised of a set of texts belonging to a specific author. It is an electronic database grounded on the semantic classification and research of the author's writing style, language and text features using information technologies. The article offers insights on the importance of Alisher Navoi's author corpus in the educational process, the necessity for its creation, the information contained in the corpus' semantic base, and the distinct features of semantic tagging of 650 gazels in the "Badoye' ul-wasat" collection written by Navoi.

1 INTRODUCTION

An author's corpus is a system that consists of a set of texts belonging to a specific author. It is basically 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 differs from other corpora in that its base collects the works of a certain author, processes the texts, tags them grammatically and semantically, determines the necessary information and linguistic expressions from the materials related to the author through the search system, provides statistical information based on the texts, and clear metadata information 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. It allows for searches based on special parameters, isn't limited by size, and enables easy and quick access to resources related to the author and their work.

Today, the creation and development of author corpora has become one of the most advanced directions of modern corpus linguistics. Corpora of authorship can be used to identify the authors of even anonymous works of art.

In the Decree № PF-5850, issued by the President of the Republic of Uzbekistan, Sh.Mirziyoyev, on October 21, 2019, titled "On measures to fundamentally increase the prestige and position of

the Uzbek language as a state language", defined to the role of Uzbek language in the social life of our people and tasks such as fundamentally increase its reputation at the international level, to increase the position and prestige of the state language at the international level, to ensure that the Uzbek language occupies a significant place in the information and communication technologies, particularly on the Internet and within the global information network. As a part of this, the creation of Uzbek language computer programmes has been defined as a priority. "Concept of Uzbek language development and improvement of language policy in 2020-2030" approved by the President of the Republic of Uzbekistan with the Decree of October 20, 2020 "On measures to further develop the Uzbek language and improve the language policy in our country", was a big responsibility for specialists, with the assignment of tasks such as creation of a large electronic resource that includes all scientific, theoretical and practical information on the Uzbek language in the direction of ensuring the active integration of the state language into modern information technologies and communications, as defined in the concept, and the Uzbek language Internet world information popularization in the network, the tasks of ensuring that it occupies a worthy place in it.

To ensure the implementation of these tasks and to develop the oral and written speech of students, researchers conducting scientific research, pedagogues working in educational institutions,

* Corresponding author

revealing the rich linguistic layer of the 15th century Uzbek language can open up linguistic possibilities, promote the creative heritage of Alisher Navoi on a global scale, and to achieve broader use of our national heritage in the modern information and communication system. Efforts to promote the works of our ancestors among young people, and to present examples of classic literature in a readable and understandable way is an important task to create philological corpora, including Alisher Navoi's authorship corpus. To accomplish this, semantic tagging of Navoi's creative heritage is imperative. Therefore, in the first step, since the most of the educational literature issued for secondary education was created based on the "Khazayin ul-Maani" collection of "Badoye' ul-wasat" devon of the of Alisher Navoi, to enable students to read, understand, and familiarise themselves with the works of Alisher Navoi, Alisher Navoi's author corpus based on the semantic tags of 650 gazels (Abjalova M., Gulomova N., Rashidov H. (2022)) in the "Badoye' ul-wasat" collection (Alisher Navoi,3) was created.

In the world and Uzbek philology, a lot of research has been done on the works of Alisher Navoi, his personality, translation of his works. So far, there exist websites featuring the biography of Alisher Navoi and the text of his works (14,15), various mobile applications, works of Navoi's works in pdf format (16,17,18,19), and several glossaries centered around the language of Alisher Navoi's works. (Dictionary of Muhammad H. (2017),8,9,13). However, this is not enough for modern students to study Alisher Navoi's work and understand his vocabulary. Information on the life and work of Alisher Navoi in world literature and in the field of information technologies (20,21,22,23,24) is covered on the site pages. However, until now, neither in world linguistics nor in literary studies, developments have been made to create Alisher Navoi's author corpus. In this respect, this corpus is expected to be one of the innovative and practical developments. In order to fill this gap in Navoi's corpus, a semantically tagged base is necessary for the reader to know the explanation of the words that are difficult to understand today and to understand the meaning of such a word in the verse.

650 gazels (gazel is a lyrical genre common in Eastern literature) from the "Badoye' ul-wasat" collection, along with proverbs, expressions, archaism, historical words, telmeh (in poetry, it is considered the art of talmeh to refer to a historical event or historical figure, to literary heroes), tashbih (tashbih is one of the poetic arts, comparing two or more things, phenomena or features expressed in

words from the point of view of some similarity, commonality that exists between them), tanasib (tanosib is the art of using words that denote interrelated and close to each other concepts in poetry), irsoli masal (irsoli masal is a method of using a certain set of Proverbs, sayings, proverbs in a poem) etc. was created Alisher Navoi's author corpus in a new format for higher education students, and a search interface was also developed in the corpus for the convenience of users. As a result of this, users will develop the skills of independent work with Navoi's work, in general, the sources of the 15th century, thereby enhancing their understanding of the language of that erae.

To make Navoi's work more readable, 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 gazels, which are an invaluable treasure of Alisher Navoi's work, are semantically equalized in the corpus of authorship, the content and importance of the educational-educational, socio-political, philosophical-ethical issues expressed by the creator, along with the gloss of the artistic symbols presented in the genre, are used in poetic art. The significance of the variety of horses, their artistic representation, and logical reasoning is not substantial. The basis for the creation of Alisher Navoi's author Corpus is found in the "Badoye' ul-wasat" collection, which includes 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 opposite meanings (tazad), 806 ratios, and 124 lexical units related to the poetic art of tashbikh for Alisher Navoi. The gazels in "Badoye' ul-wasat" collection consist of 5001 byte – 10002 lines, a total of 66539 words, each gazel is from 6 byte to 13 byte: 6 byte – 1, 7 byte – 437, 8 byte – 8, 9 byte – 187, 10 byte – 1, 11 byte – 14, 12 byte – 1, 13 byte – 1 (A.Navoi. (2011)). Such statistical analysis of the Alisher Navoi Author corpus is considered important and gives the user a clear account and shows 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 of the people. Through such combinations and expressions, we summarized the typical features of the events in social life, characteristic of folk wisdom, and showed the possibilities of short, succinct and meaningful expression. The lines containing compound phrases were scientifically analyzed and included in the 5th column.

Creating a semantic base of explanatory words found in Alisher Navoi's gazels and including them in the author's corpus plays an important role in the careful study of rare examples of our classical literature and in the development of our national spirituality. In the process of semantic equalization of gazels, we witness that some words express more than ten meanings. Of course, such polysemanticity indicates that Navoi is the sultan of the word property. In order to fully feel, realize and understand the meaning and essence of Navoi's gazels, it is important to scientifically and artistically analyze every line and every word used in the gazel, in today's rapidly developing era of "Artificial Intelligence", Alisher Navoi's author's corpus, with the help of computer programmes, facilitates the quick and easy understanding of the author's language and ensures its readability.

Remind: "Khazayin ul-ma'ani" collection is a perfect collection of books which consist of "Garayib us-sigar" ("Strange things of youth"), "Navodir ush-shabab" ("Poems of the rarity of youth"), "Badoye' ul-wasat" ("Middle-age Badias") and "Favoyidul-kibar" ("Benefits of old age") books. It is also referred to as "Chor Devan" (Four Devan).

The corpus consists of 8 columns (Alisher Navoi):

1st Column: "Biography of Alisher Navoi" section is a database about the poet's life and creative activity.

2nd Column: Simple and special search by corpus. When you type the desired word form into the search page, all gazel verses that use this word will be displayed in the search window. If you refer to any of the gazels, you can read the text of the gazels and metadata related to the gazels.

3rd Column: 8 devan texts belonging to Alisher Navoi (including the "First Devan" compiled by his fans).

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

5th Column: 650 gazels and poetic arts in the "Badoye' ul-wasat" collection are presented to the user with semantic tagging.

6th Column: About Alisher Navoi's corpus.

7th Column: Research results of this corpus.

8th Column: Information about the authors of the corpus.

The created corpus can serve as a source of spiritual fulfillment for specialists, educational material for different stages, and a source of information for related fields.

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

- To study Navoi's personality;
- Study of literary style;
- Linguopoetic analysis of the poet's work;
- Researching the ability and skill of the creator to use words;
- Creation of authorship dictionaries;
- Compilation of author's phrases;
- Identifying the author of anonymous works through the parameters that reveals the personality and style of the creator in the Authority Corpus;
- Summarizing the author's paraphrase, parema, wisdom; the extent of the use of figurative expressions can be determined from the context of the creator.

In Navoi's gazels, there are numerous expressions involving historical, literary, mythical figures, geographical and ethnic place names, which contribute to the art of talmeh from an artistic point of view and also provide information about historical, artistic figures and geographical, ethnic locations. For example, in the gazels of the "Badoye' ul-wasat" collection, he refers to the word "kavsar" 23 times, and this word expresses the meanings of "material and spiritual satiety":

As a result of the semantic tagging of the names of famous places, historical figures, holy saints and historical locations included in Alisher Navoi's authorship corpus, knowledge about the names of works, famous people, prophets mentioned in theological books, textile images, holy places, is formed and developed.

The corpus of authorship has the ability to display the author's language in full, in detail, and objectively, so such corpora are distinguished from other information banks by their advantage. The corpus can serve as a basis, source, and tool for various types of research. Another advantage of such corpora is that with their help, it is possible to learn the languages of not only a word or sentence, but also the entire work. The information in the author's corpora, being edited based on scientific sources, guarantees its accuracy and reliability, enabling a comprehensive and objective study of the entire spectrum of linguistic phenomena.

From the corpus of authorship, all linguistic features of the poet's word units, changes in them – from the point of view of that era, the update in the language and the disuse of words in today's social life, the cases of activation and passivation, and linguistic phenomena, are quick, clear and accurate, receives information, easily creates large vocabularies of various types, creates the possibility of automatic processing of texts. The fact that the information in the authorship corpus 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 study of the entire spectrum of linguistic phenomena. The corpus, a large spiritual treasure, provides numerous linguistic possibilities, including 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 word etymology, provides many linguistic possibilities, such as understanding the interpretation of words that are difficult for today's readers to understand, using the array of contextual examples provided in the corpus, and finding its semantic interpretation. This will serve to increase the number of those who are interested in our classic literature, understand the language of Navoi period, and enjoy the spiritual treasure of Navoi by reading his works, regardless of the age of the audience. The linguistic basis of the author's corpus is "going 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.

The authoritative corpus plays a vital role in the educational process, elucidating the educational, social, historical, linguistic and didactic significance of the author's works, for classifying the dialectal features of the words used in the author's works, and for 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 frequently used in the text (nouns, adjectives, key words, verbs, grammatical forms, sentence construction, in short, tools that indicate the writer's style) are determined using linguostatistical analysis. The comparative analysis of evidence from different texts allows us to determine the content of that text, the period in which the text was composed, the argumentative character of the evidence, and even the authorship. The study of Alisher Navoi's life and work at all levels of secondary schools increases the educational value of Alisher Navoi's Authority Corpus with the use of the poet's corpus in a broad sense. This corpus serves as a convenient source of information in the preparation of educational material during the educational process, the teacher will have the opportunity to quickly prepare fresh, meaningful, reliable educational material for the training session.

Authoritative corpora enable students to develop their research capacity and conduct minor investigations based on reliable evidence. It should be noted that the social significance of the author's corpora is considered comprehensive, and it can be used in linguistic, ethno-psycho-linguistic research, mother tongue, literature, foreign language education, automatic text processing, translation programs. Author corpora are also the most convenient means of monitoring the changes in the vocabulary of the language (neologism, historicism, archaism phenomena). In the process of analyzing the possibility of lexical-semantic combination of the words used by the author, the possibility of comparing the dictionary and grammar of the ancient and new generation expands. The diversity of the word author by means of the corpus of authorship provides practical help in differentiating the possibilities of belonging to several semantic categories at the same time. In addition to the poet's linguistic views, it will be possible to learn about the poet's attitude to social spheres and views on political issues through the author's corpora.

2 BEHAVIOURAL SIGNIFICANCE OF ALISHER NAVOI'S AUTHORITATIVE CORPUS

The preservation of our spiritual heritage, development of skills to honour the spirit of ancestors, and cultivation of a reading culture, all facilitate the acquisition of profound insights and broadened thought processes.

3 BASIC VOCABULARIES USED

Alisher Navoi's authoritative corpus, encompassing thousands of lexemes related to the poet's gazels, facilitates understanding through the semantic tagging of included texts. This provides lexical meanings of words with explanations, enabling the reader to quickly and easily comprehend the contextual meaning of various words. The lexical compatibility of compounds and the analysis of their combinatorial ability, allows to determine whether a certain syntactic construction is acceptable or not.

To elucidate the current lexical meaning of the words in the gazels of "Badoye' ul-wasat" collection were utilised "Language of Navoi's Works" by P. Shamsiev

and S.Ibrokhimov (Shamsiev P., Ibrokhimov S. (1972).), 4-volume "Annotation of the Language of Alisher Navoi's Works" edited by E.Fazilov (Kayumov P., Shukurov Sh., Hayitmetov H., Bektemirov H., (1983).), Yu.Berdak's "Dictionary of Navoi language" (Berdak Y. (2018).) and "Dictionary of classic literary works" and "Dictionary of works of Alisher Navoi" (Dictionary of Muhammad H. (2017).) by H. Muhammad.

Representativeness means that all information about the text is given (Zakharov V., Bogdanova S. (2020).). Represented texts have their source, style, period of writing, author, age of the audience, text type clearly indicated (Kochetova L.A. (2016)). The 650 gazels in the divan were individually represented in the corpus, created on the basis of the divan "Badoye' ul-wasat". A total of 20 types of metadata were formed according to the content of each gazel (romantic, orifona, rindona), audience age (15+ - 18+), and the number of words used in the gazel.

Metadata of Alisher Navoi's "Badoye' ul-wasat" collection include:

- 1) title of the work ("Badoye' ul-wasat");
- 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) publisher ("Tamaddun" Ltd publisher);
- 10) field of application (literature);
- 11) literary type (lyric);
- 12) genre (gazel);
- 13) time and place of the event (Herat);
- 14) text style (artistic)
- 15) text type (romantic, orifona, rindona);
- 16) audience age (15+, 18+ years);
- 17) potential of the audience (for the general public);
- 18) type of internal body (authorship);
- 19) number of word forms (66,539);
- 20) tagger (G'ulomova N.)

The future aim is to include all of Alisher Navoi's works in the corpus, encompassing his prose and verse, as well as his scientific, historical, philosophical, and religious writings.

A corpus, in its modern context, refers to a reliable computer database that utilises specialised programs during its creation. HTML, CSS, JS, programming languages, Bootstrap5, JQUERY design frameworks were used to create the design of Alisher Navoi's authority corpus. Using the Python programming language and the Django framework, the general and

special search part of Alisher Navoi's corpus ensures that the semantic tag is displayed on the screen when you click on the explanatory words in the gazels. Since the interface is the first impression of the case, it is a very important process to ensure its perfection, creating a unique design. National and, at the same time, modern features were taken into account when creating the interface of Alisher Navoi's authority corpus.

The corpus will be accessible not only literary 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. has been developed and implemented with the aim of promoting Uzbek computer and corpus linguistics, enhancing educational processes and research, catering to the educational needs of specialists of all ages, and expanding vocabulary and literacy.

REFERENCES

- Author corpus of Alisher Navoi – <http://navoiykorpusi.uz/>
- Abjalova M., Gulomova N., Rashidov H. (2022). Semantic base of gazels in Navoi's "Badoye' ul-wasat" collection for AlisherNavoi's authority corpus. Certificate of authorship, No.00583. - Tashkent,
- Abjalova M., Gulomova N., Sadullayeva Sh. (2022). Authority corpus of Alisher Navoi. Certificate of authorship, No.18544. – Tashkent,
- Dictionary of Muhammad H. (2017). Alisher Navoi's works. – Tashkent: Akademnashr, - 407 p.
- Ochilov E. (2011) Wisdom of Navoi. - Tashkent: Uzbekistan,. – 23 p.
- Rahimov A. (2011). Fundamentals of computer linguistics. – Tashkent: Academy, – 156 p.
- Raupova L., Elov B., Abjalova M., Alayev R. The educational corpus of the Uzbek language and its possibilities. // Language and culture in Uzbekistan, – Tashkent:, 60-75 p.
- Explanatory dictionary of the language of Alisher Navoi's works. / under the editorship of E.I.Fazilov. Editorial board: Kononov A., Kayumov P., Shukurov Sh., Hayitmetov H., Bektemirov H., (1983). Karimov Q. 4 volumes. – Uzbekistan: Science,
- Berdak Y. (2018). Dictionary of Navoi language. Tashkent: Sharq, - 496 p.
- Zakharov V., Bogdanova S. (2020). Corpus Linguistics. – National University of St. Petersburg.– S. 23.
- Kochetova L.A. (2016). Statistical methods in corpus research. - Volgograd.– 49 p.
- A.Navoi. (2011). Badoe' ul- wasat. – Tashkent: Tamaddun,– 721 p.
- Shamsiev P., Ibrokhimov S. (1972). Dictionary of Navoi's works. – Tashkent: Gafur Gulam, - 784 p.