

## **Creating a Database of Uzbek Dictionaries Related to the Text of "Boburnoma"**

*Dilafroz Muhammadiyeva*

**Abstract.** *Automated information technologies and computers are gaining importance in our society today. This is causing the emergence of new directions. Corpus linguistics is considered an important branch of computer linguistics, and all research related to the science is inextricably linked with software and lexicography. National lexical wealth and possibilities of natural language are shown in dictionaries. Therefore, dictionaries are characterized by being in a machine-readable format, having a search engine, indexing, categorizing, clustering, and the ease and breadth of the user's ability to use it. The software provides a linguistic resource base with such perfect dictionaries. This article discusses the importance of dictionaries in these languages in creating parallel corpora, in particular, the Uzbek, Turkish and Kazakh parallel corpora of "Boburnoma".*

**Key words:** *"Boburnoma", Turkish, Kazakh, parallel corpus, dictionary, electronic dictionary, lexicography.*

Recently, lexicography has been proven once again to be the fastest growing and fastest changing field of linguistics. The dictionary is a product of time, a reflection of a certain period in the history of the people. Its occurrence depends on certain historical and cultural conditions, the state of theoretical and practical linguistics, the development of various fields that need dictionaries, and many other aspects.

Dictionary also shows the status and development of languages. In a particular language, depending on various factors related to the period, the emergence of various forms of dictionaries, in general, the development of lexicography - with the development of the language and the nation that owns it. depends, allows you to talk about his bright future.

The emergence of a certain dictionary in a certain language, in a certain historical period, is the official status of the language in which the dictionary was written, the development of the areas that require a dictionary, more precisely, the society related to social, political, economic, cultural development. Therefore, unlike other departments of linguistics, lexicography "flourished" during the period of development of the country, "faded" when there was calmness in the life of the country, every stage of its development at one time it served to realize the interests of the ideology of the time, the national idea. The history of vocabulary is a witness to this. More precisely, the emergence and development of lexicography is a phenomenon related to the socio-political, economic, cultural development of the country, as well as the renaissance of the national language [Abdovakhabov, 2024:460].

Dictionaries are works that systematically collect information about words, phrases, their meanings, syntax, grammatical features and other linguistic aspects of a particular language. Dictionaries reflect the lexical system of each language, its own vocabulary, context of use of words, synonyms and antonyms. They are important tools in linguistic research, translation, lexicography, and language learning. A dictionary is a work that systematically stores and explains the words of a language, their meanings, usage and other linguistic features. Each dictionary contains the vocabulary of a particular

language and is annotated with the meanings, synonyms, antonyms, grammatical forms, usages and other relevant information of these words.

Dictionaries essentially have the following types: encyclopedic dictionaries (encyclopedic) and linguistic dictionaries (morphological). The most common dictionaries are linguistic dictionaries. Within a specific language, linguistic dictionaries are divided into spelling, orthoepic, descriptive, phraseological, etymological, etc. The object of morphological dictionaries is a word. They carry information on the correct pronunciation, spelling, semantics, etymology, and interpretation of grammatical features of words. Linguistic dictionaries contain almost all the words of a given language, that is, all independent and auxiliary words are recorded in the dictionary of linguistic dictionaries. The problem of automatic natural language processing has been relevant for more than half a century.

Today, the emergence of different cultures in the world, the development of languages and the expansion of technical possibilities for their mastery have fundamentally changed the worldview of people in society. There has been a special demand and need for the language from society, which is related to the issue of creating more electronic dictionaries. These demands and needs led to the rise of the lexicographic theory to a new level.

For more than two decades, the concept of electronic dictionary has been understood in the field of linguistics in a narrow sense, only within the framework of computer linguistics, as a result of the extremely narrow scientific views that the problems of the field are related only to the introduction of electronic technologies, computer lexicography has been insufficiently studied by linguists. was ignored. In fact, this term and related concepts serve the development of the field of linguistics and its full response to the needs of today's society. In this case, the field of lexicography of applied linguistics underwent great changes - improvement. The concept of an electronic dictionary has influenced linguistics in several ways.

In the time we live in, the pace of life has accelerated, and the concept that is relevant today will be obsolete tomorrow. The printed dictionaries that are being created are not able to cover the influx of words. Commentaries on things, concepts, and events in explanatory dictionaries are out of print and out of reach of readers. More than a century ago, the famous writer, academician, Nobel laureate Anatoly Franz said: "A dictionary is a whole universe arranged in alphabetical order! If you think about it, a dictionary is a book of books. It includes all the other books, you just need to be able to find what you need in it..." he wrote. This definition has not lost its validity since the emergence of the concept of "dictionary". A dictionary is a real treasure for someone who can use it regularly. Being able to quickly and conveniently search for the necessary information is an important indicator of a modern person. A person is not capable of knowing everything, but he has the ability to find out what he needs to know in any situation. The dictionary is created specifically for this purpose [ Bakhridinova, 2021 ].

Linguistics has long been developed in connection with the science of mathematics: their essence goes back to one root - logic. The development of applied linguistics has reinforced this idea and made it more effective. While emerging technologies have facilitated the process by enabling the translation of texts directly into electronic format, it is the user's responsibility to understand the unique nature of each language and the way texts are structured. increased.

that there are differences between traditional dictionaries and electronic dictionaries. These differences indicate the superiority of one or another form of vocabulary. That is, traditional dictionaries are preferable in one aspect, and electronic dictionaries are more effective in another aspect. For example, the materiality of using traditional dictionaries, the conditions and actions necessary to achieve it ( the need to go to a store or library to get a dictionary, find it, the capacity of the library base, or the material cost spent on its purchase). The fact that it is absolutely not required in electronic dictionaries has come in very handy for users during the pandemic. The convenience of using electronic dictionaries in learning the secrets of the benchmark language and learning foreign languages has been proven once again.

The fundamental changes taking place in lexicography require the integration of linguistics with other disciplines, the development of cooperation. The relationship between linguistics and technical sciences is required to find a solution to the situation and the existing problems. Based on the connection of linguistics with technical sciences, new terms engineering linguistics, computational linguistics, electronic linguistics appeared. Under this concept, a number of important issues surrounding language still remain relevant. Systematic language learning through computers, machine translation and lexicography issues are among the important issues in this field [ Abduvakhobov, 2024: 460 ].

A corpus dictionary is a dictionary in motion and keeps up with the times. Such dictionaries are constantly changing, improving and filling. It is reasonable to call them simply "living dictionary". Vocabularies, which are an important tool for developing national thinking and educating a person with high intellectual potential, cannot be created, they cannot take place in educational corpus bases, students' desks educational dictionaries and lexical minimums created specifically for it do not exist, for this educational lexicography is not systematically developed as a theoretical and practical field, and most importantly, with the preparation and publication of dictionaries unless a related scientific and financing system is created, our noble goal of developing the national language will not be fully realized, mother tongue education will remain alienated from its true essence.

A corpus serves as a resource for building large vocabularies. Over time, the corpus has become a powerful informational resource with relevance for various linguistic disciplines. Corpus-based computer dictionaries are created and processed faster than before. In this way, the work reflects the language from the beginning to the end (until publication), the dictionary article does not have time to "obsolete" [ Po -latova, 2024:429 ].

Creating a complete dictionary of that language plays an important role in creating a database of any language. A dictionary with great efficiency according to the requirements of the present time is a computer dictionary. A computer dictionary, unlike a traditional dictionary, consists of the following parts:

1. Widely used words and phraseological expressions in the Uzbek language with lexical meaning included in traditional dictionaries. The dictionary includes all basic words (herbaceous, computer, good, etc.), artificial words (rich, beautiful, worker, etc.), compound words (colic, stomach, sunflower, etc.), complex abbreviations (airport, microscope, etc.) is entered.
2. Words specific to all dialects.
3. Terms used in certain fields.
4. Vulgar words and phrases.
5. Personal names.
6. Geographical names.
7. Abbreviations [Polatov A., 2011: 213].

All the words in our language make up the vocabulary of the Uzbek language. This shows that lexemes form a special field in the linguistic support of programs that perform text processing, automatic text analysis and editing, and machine translation. Therefore, when creating a linguistic prediction of automatic translation programs, mobile applications and platforms, morphologically analyzed lexemes are included in the database as lexical units and their translation in other languages is given [ Fattahova , 2024: 201 ].

Today, the frequent repetition of terms such as computer technology, automatic translation, machine translation, google translator, amazon translator, microsoft translator, DeepL, parallel corpus in daily use, not only translation studies, but also machine translation (computer and linguistics since the 1950s which is the integration of the fields) indicates that many works have been carried out in the fields. As far as many people know, this is the first experience in this regard It was conducted in 1954 at Georgetown University in the USA, and in 1960 it was formed as an independent science [ Rahimov , 2011:160 ].

DODobrovolsky, Yu. Scientists such as Tao, AAKokoeva and V. Zakharov conducted research on the structure, composition and capabilities of the parallel corpus and created many works. According to them, in the field of machine translation, the creation of parallel corpora is important, because the original text and the translated text are included in bitex form. This makes it possible to determine the content of the text depending on the context in the interpretation of the text and to find the semantic compatibility for translation into another language, as well as to include a lot of necessary information about the semantic relationship of parallel texts.

Also, in our republic, serious studies aimed at the development of this field have been carried out, and special attention has been paid to the creation of linguistic support for the program that translates foreign texts into Uzbek and vice versa. Remembering that programmers and mathematicians have also paid attention to this area and conducted certain studies also means that this area is relevant [ Hakimov , 2009:297 ].

Corpus-based machine translation is recognized as one of the most widespread areas. In this case, each bitex (English - Uzbek text) included in the parallel corpus is automatically analyzed and certain conclusions are drawn for translation [Abdurakhmonova, N. 2021,2022 ] . The meanings of words and phrases in parallel texts translated with the help of skilled translators are determined depending on the context [ Sirojiddinova M. 2024:77 ].

A number of scientific studies on the language of Babur's works have also been carried out by world scientists. The first attempts in this regard began with the dictionaries compiled during the translation of the work. V. Erskin, J. Leyden, A. Kayser, Pave de Courteil, LUKing, Denison Ross, later A. Alparslon, M. Jonp o ' lat, Yu. Bilol, T. Seyhan, N. Jeylan and other linguists and source studies scientists Babur in his works a glossary of words used, with explanations and indicators. For example, LUKing, who was engaged in the translation of "Boburnoma", explained more than four thousand words. In the translations of U. Texton, R. Arat, and A. Beveridge, explanations and comments of many words are given. Also, thousands of words have been explained in Hindi and other language editions of the work. Especially Turkish scientists have done a lot of research on the language of Babur's works, including the lexical units in his book [ Haitov Sh, 2024 ].

was attached at the end of the first edition of "Boburnoma" made by P.Shamsiev and S.Mirzaev in 1948-1949 . After that, in 1972, H. Nazarova published "A brief dictionary for the works of Zahiriddin Muhammad Babur". In the 1960 edition of "Boburnoma" more than 600 words and other lexical units are explained. An explanatory dictionary of "Boburnoma" was compiled and published by F. Is'hakov in 2008. 2498 words are explained in it. This dictionary is based on the text of S. Hasanov's 2002 edition of the work. A. Ibrohimov commented on about four thousand lexical units used by the poet in the books "Indian Words in Boburnoma" (2002), "Urdu-Uzbek Common Words Dictionary" (2022).

It can be said that the "Annotated Dictionary of Babur's Works" created in 2024 brought the work in this direction to a new level. It contains an explanation of thousands of words used in the works "Boburnoma", "Babur Divan", "Mubayyin" . Such fundamental studies on the interpretation of words in Babur's works have been used as a source for the interpretation of Babur's vocabulary and the meaning of words used in all the poet's works. passes.

The role of dictionaries in creating a parallel corpus of Boburnoma texts and translations is very important, as dictionaries and parallel corpora are complementary and mutually supportive tools. Such a parallel corpus provides high efficiency in language learning, translation, linguistic research and other linguistic fields.

A parallel corpus is a collection of data composed of compatible versions of texts written in two or more languages. Such corpora help translators and linguists to determine the contextual translation of words, phrases and texts.

Relationships between dictionaries and parallel corpora:

- Dictionaries accurately describe the meanings and translations of Uzbek, Turkish and Kazakh words and phrases.

- A parallel corpus, on the other hand, contains texts translated to each other, which helps to verify the accuracy of the translation.
- The following steps are taken to create dictionaries:
- Data collection: Data of words is collected to create a dictionary. Linguists, translators and specialists participate in this process.
- Structure and formatting: Words are arranged alphabetically and their meanings are explained.
- Synonyms, antonyms and grammatical details: Add synonyms and antonyms, grammatical descriptions of words to the dictionary.
- Introduction and annotations: Before the main part of the dictionary, an introduction is prepared on words and annotations.

The importance of dictionaries in creating a parallel corpus is as follows:

1. are an important tool in determining the correct translations of Uzbek words used in the text of "Boburnoma" in Turkish and Kazakh . Each word and phrase can have multiple equivalents, which ensures a high-quality parallel corpus.
2. Synonyms and antonyms: Dictionaries provide information about synonyms and antonyms, which help to define exactly how the words in the parallel corpus are used in different contexts. In "Boburnoma" some words may be used in Uzbek language with different meanings , which leads to confusion in translation. A dictionary helps to solve these issues when creating a parallel corpus.
3. Grammatical Descriptions: Dictionaries may provide grammatical descriptions of words (eg noun , verb, adjective, etc.), which is important for accurate and correct analysis of texts in a parallel corpus. The grammatical features of each word in the text must be preserved in its translation.
4. Contextual Information: In the text of "Boburnoma", words and phrases have a specific cultural and historical context. The role of dictionaries in a parallel corpus is that they help define these contexts. For example, it is important to understand how the Uzbek word "sultan" changes in its Turkish and Kazakh equivalents, and how the word was used in a historical context.

Dictionaries and parallel corpus work as complementary tools. Dictionaries help language learners or translators look up and understand words and their translations in a text, while parallel corpora confirm these translations in context. The following techniques facilitate the use of dictionaries in a parallel corpus:

- Uzbek word-to-word and Turkish/Kazakh translations into a parallel corpus: Turkish and Kazakh translations of each Uzbek word and its contextual usage are stored in a parallel corpus.
- Grammatical compatibility: The grammatical features of each language are matched to each other in a parallel corpus , which ensures the correctness and accuracy of the translation.
- Contextual translation: The Turkish or Kazakh translation of an Uzbek word in the corpus takes into account not only the meaning, but also the contextual differences.

Of using dictionaries in the process of creating a parallel corpus is as follows:

1. Accuracy of word translation: Dictionaries provide the correct translation of words used in a parallel corpus. It takes into account not only the exact meaning of the word, but also the context in which it is used.
2. Methodological principles: Dictionaries help in determining the grammatical features of words and their different levels of meaning. This creates a clear picture of word usage, interactions and their contextual interpretations in a parallel corpus.
3. Multilingual texts: Dictionaries help identify and apply equivalents in different languages in a parallel corpus. It is a powerful tool in multilingual translation and linguistic research.



4. Considering cultural differences: The text of "Boburnoma" contains words related not only to language, but also to culture. By including the dictionaries in a parallel corpus, it is possible to accurately translate the cultural context and at the same time mark the cultural differences.

In conclusion, it can be said that dictionaries related to the text of "Boburnoma" play an important role in creating a parallel corpus, because they help in the process of translating the text clearly and correctly. Dictionaries, Turkish and Kazakh translations of Uzbek words in a parallel corpus, synonyms, antonyms, grammatical features and contextual information are clearly defined. It is of great help not only to translators but also to linguists and language learners. The integration of parallel corpora and dictionaries are very important tools in ensuring the richness of the language and the accuracy of the translation.

## Literature

1. Abdurahmonova N. Linguistic foundations of machine translation. Manual. Tashkent: Akademnashr, 2012.
2. Abdurakhmonova, N. Formal-Functional Models of The Uzbek Electron Corpus. English. Journal of the Association-Institute for English Language and American Studies, 2021. 10(8), 59-66.
3. Abduvakhabov G'. The need to create electronic dictionaries in the Uzbek language Modern technologies of computer linguistics - s tcl.2024. Tashkent, 2024. -B.460.
4. Bakhriddinova B., Bahadirova G. National language renaissance and dictionary development in Uzbekistan // Theoretical and practical issues of creating Uzbek national and educational corpora. - Tashkent, 2021.
5. Fattakhova D., Jumanazarova D. "Linguistic maintenance of the morphological dictionary of the Uzbek language // Modern technologies of computer linguistics - s tcl.2024". Tashkent, 2024. - B.201.
6. Hayitov Sh, 2024
7. Rahimov A. Fundamentals of computer linguistics. - Tashkent: Akademnashr, 2011. -160.
8. Polatov A., 2011: 213
9. Polatova G. Creating a linguistic base of adjectives for the thesaurus of the Uzbek language // Modern technologies of computer linguistics - s tcl.2024". Tashkent, 2024. -B.429.
10. Sirojiddinova M. The linguistic importance of creating a bilingual phraseological dictionary for corpus-based machine translation (as an example of a work of art) // Modern technologies of computer linguistics - s tcl.2024". Tashkent, 2024. -B.77.
11. Khakimov M.Kh. Rasshiryaemyy input language matematicheskogo modelirovaniya estestvennogo zhyzika dlya mnogoyazychnoy situ a tsii mashinnogo perevoda // O'zMU reports, 2009. - No. 1. - P.297.