European Journal of Pedagogical Initiatives and Educational Practices
ISSN (E): 2938-3625
Volume 1, Issue 3, June, 2023

Principles of Creating the Perfect Corpus of "Baburname"

Dilafruz Mukhammadiyeva
Shukhrat Hayitov

Abstract
This article is focused on the principles of creating a perfect "Baburname" corpus. Gathering the achievements in Babur studies, defining the problems, showing its role in the development of world culture, the growing interest in Babur's personality, activities, and creativity on a global scale, and the emergence of new researches in Babur studies require the creation of this corpus. The facts that creating an electronic database on Babur's life and activities, processing texts on the basis of artificial intelligence; compiling a corpus of parallel texts related to the translations of "Baburname", conducting a search based on various symbols, explaining the social-political, cultural-educational features of the "Baburname" text; the importance of studying the translations, researches accomplished on Uzbek, Turkish, Azerbaijani, Kazakh, Kyrgyz, Russian, English, German, French languages were all the basis for creating "Baburname" corpus were analyzed.

Keywords: Zahiriddin Mukhammad Babur, "Baburname", corpus, electronic platform, indexing, tagging, manuscript, translation, site, artificial intelligence, text corpus, database.

Introduction

Today's fast-paced time requires people to be speedy too. On the basis of new innovative ideas, electronic forms of scientific and artistic works are being created and systematization of the conducted researches are being achieved. In the age of information technologies, the creation of all works in the form of an electronic platform or corpus creates much more convenience for further research.

In many countries of the world there are separate electronic platforms of poets and writers who are the pride of the nation. For example, in Russia there are sites dedicated to Dostoevsky and Tolstoy (http://tolstoy.ru, https://fedordostoevsky.ru), in Germany there is an electronic platform of Goethe, in Italy there are sites created even for the work of Marco Polo (digital Marco Polo). Having a site dedicated to "Baburname", the encyclopedic work of Zahiriddin Mukhammad Babur, who is the pride of our people, is the need of the present. After all, computer and corpus linguistics are the main tools that give immunity to languages in the context of globalization [3. b. 27].

Corpus (corpus) means "body" in Latin. "Corpus is a collection of texts in electronic format that means finding words, phrases, grammatical forms, word meanings through a specific search engine" [http://rusorpora.ru]. The term "corpus of texts" is used side by side with the concept of corpus. A corpus of texts is a whole that can consist of electronically stored phonemes, graphemes, morphemes, lexemes, sentences and texts. Corpora are in fact a collection formed as a database to solve linguistic problems and serve as material for conducting research in various directions [1.p. 61].

A corpus is a set of texts submitted to a search program in order to determine the characteristics of language units, a set of written or spoken texts in natural language stored in electronic form, placed in a computerized search system on the basis of software. Language corpora are an undeniable tool for language research and solution of practical tasks [4.p. 4].

Corpus linguistics is a relatively new direction both in Uzbek linguistics and in the modern information technology system. Linguistic corpora are the main source and powerful information resource for building large vocabularies. A language corpus provides rapid computer-aided construction and processing of vocabularies. With the development of corpus linguistics in Uzbek linguistics, a lot of work was done to create the corpus of the Uzbek language.

The incomparable scientific and creative heritage of Zahiriddin Mukhammad Babur, one of the unique figures who left an indelible mark in world history, has a special place not only in the formation of our national culture and literary and aesthetic thinking of our people, but also in the history of world literature, science and statecraft.

In the years of independence, extensive work was carried out on the comprehensive study of the life and activities of Zahiriddin Muhammad Babur, and the promotion of his works in our country and in foreign countries. Acknowledging all the work that is being done, it can be said that creating an electronic version of the poet's works, gathering the world's research work into one system and one place is one of the urgent issues of today. By creating a perfect corpus of “Baburname”, it is possible to collect all manuscript copies, translations of the work, as well as large and small research conducted on them, all photo (miniature), video and audio materials related to Babur and “Baburname”, and turn it into a corpus.

The Main Part

Creation of a perfect corpus of “Baburname” serves to systematize the creative heritage of the writer, define the progress of Babur studies, direct new researches, show the collective nature of Babur's work, shed light on his attitude to his mother tongue and language skills. Because turning the Uzbek language into a modeled, formalized language that can be understood by computers like world languages, digitizing and archiving the works of Uzbek national literature, as well as creating electronic books and spreading them to the general public is one of the main tasks facing our specialists.[1.p. 61].

Once a complete corpus of the “Baburname” is created, it will contain all the manuscripts of the “Baburname”, and it will open the ground for new textual studies, since until now no scholarly critical text has been compiled from all the manuscripts of the work. Another point is that the translations into other languages until today are based on one or two of the manuscript texts, so we cannot say that all translations are perfect.

Secondly, since all the international and republican research on “Baburname” is gathered together, the science of Babur studying will be enriched with new sources and will serve as a rich source for future scientific research. Researchers can conduct basic research based on data from a single perfect corpus. There are very few studies on the translation and language features of the work in Babur studies today.

Thirdly, historical figures, toponyms, scientific terms related to various professions, proverbs and phrases in the text of “Baburname” are explained in the form of tags in Uzbek, Russian

and English languages. It is this aspect that makes the work interesting for a wide readership. Each unclear aspect is explained in a separate window in the hyperlink.
Research is conducted in two directions:
The first direction is part of scientific research which comprises facsimile of manuscript copies of “Baburname”, translations in different languages, compilation and classification of scientific works related to “Baburname”; collecting dictionaries related to the language of the sources of the period when “Baburname” was created, dictionaries of ancient Turkic language, old Uzbek language, explaining the meaning of lexical units in “Baburname”, distinguishing denotative and contextual meanings, own and acquired layer lexemes tagging; drawing conclusions based on factual materials, forming comments and parts of articles
The second direction is to create the form of "Digital Baburname creating a complete corpus" in cooperation with experts, work on its structure, software of the electronic platform, and place the materials collected in the first direction. In this case, the body consists of the following parts:
1. General information about the corpus.
2. All manuscripts of “Baburname” or their facsimiles.
3. Scientific-critical publications of “Baburname” in Uzbek (1960), (2002). Eiji Mano's Critical Text (1995). Scientific and critical editions of “Baburname” are issued in the Arabic alphabet, transcription, Latin alphabet and the current Uzbek tabdili form.
4. Translations of “Baburname”. All multiple translations in all languages are listed.
5. Researches on “Baburname” are published in the form of pamphlets, articles, and conference materials.
A) republic-wide sources, Uzbek language studies, monographs, artistic works
B) foreign sources, studies, monographs, artistic works.
6. All dictionaries created on “Baburname” are introduced.
7. Miniatures, movies, films, documentaries, videos, audio texts about Babur's personality and work.
The platform acquires practical importance in interpreting the views of Uzbek thinkers not for the sake of history and spirituality, but as a solution to problems aimed at solving a number of today's issues. Babur made an important theoretical contribution to the development of linguistics and literature. In addition, he expressed important ideas about mathematics, geography, chemistry, medicine, and seismology. The corpus provided for in the project includes information of an advertising nature, which will attract specialists of other fields.
This body will be in the form of a site, each column page will be created separately. The corpus has a special search system: search windows for written material, multimedia products are formed.
The published dictionaries of Babur's works and the annotated dictionary of “Baburname” will have separate windows. If the desired unit is entered in the query window, the user will be able to get the necessary information.
Information on socio-economic fields in “Baburname” interface will also have a separate search system. The smallest unit of search is a word. The interface of multimedia tools is created separately for each database. It houses a library of video and audio products. Online viewing of products is provided. Multimedia products work in a title-based search type.

The corpus database allows you to store each information in several languages (Uzbek, Turkish, English, Russian). It is possible to add new languages to the electronic platform, that is, it is possible to improve the system by adding a version of the information in one language. The database consists of two parts: source and research parts. The source part is the text part of the project, which includes manuscript copies of "Baburname", scientific-critical text and translation. The text is organized according to the search engine.

A part of the database is a scientific part, which includes scientific studies devoted to Babur's work and creativity.

Until now, sources, researches, works of art related to "Baburname" have not been systematized and formed in the form of an electronic database. The electronic platform contains a very large database of materials related to "Baburname" and will serve as a resource for future scientific research of this type.

The first steps were taken by the scientific team on this research, a dictionary of Babur's works was compiled, monographs and scientific articles were published.

Summary

In the future, it is planned to spread Babur's works, research on Babur studies, culture and values of the Uzbek people on the world scale as classical sources through the materials from the excellent corpus.

The project "Digital Baburname creating a complete corpus" opens a wide way to gather readers from all over the world who are interested in Babur's work.

Through this, it opens a wide path to inter-literary relations and international friendships. Considering the absence of a web page dedicated to "Baburname", this national corpus solves the problems related to the shortage of scientific and educational literature in Uzbek linguistics and literary studies, because the corpus contains Turkish, Kazakh, Azerbaijani, Russian, English Scientific works on "Baburname" in several languages such as French, German are also posted. The possibility of organizing webinars on this platform makes it possible to commercialize this platform later. Also, even when the project is over, it would be continuously widely available to realize the literature on the basis of financial contracts in the framework of the project.

References:

1. Abdurakhmonova M. 2021. Architecture of the author corpus of Uzbek national literature. International scientific-practical conference on "Theoretical and practical issues of creating Uzbek national and educational corpora". – Tashkent.
2. Daniyarov B. 2019. The issue of providing lexical synonyms in the national corpus of the Uzbek language. Foreign philology #4. – Tashkent.
3. Menliev B. 2021. "Creation of the national corpus of the Uzbek language": achievements, problems and tasks. International scientific-practical conference on "Theoretical and practical issues of creating Uzbek national and educational corpora". – Tashkent.
4. Mengliev B. 2021. National corpus of the Uzbek language. New Uzbekistan newspaper. – Tashkent, April 7. No. 69.
5. Rahimov A. 2011. Fundamentals of Computational Linguistics. – Tashkent.

ISSN (E): 2938-3625
Volume 1, Issue 3, June, 2023

6. Kholmanova Z. 2021. The importance of corpora as linguistic research material. “International scientific-practical conference on theoretical and practical issues of creating Uzbek national and educational corpora”. – Tashkent.