TOPONYMS AS A COMPONENT OF THE NATIONAL LANGUAGE CORPUS

Kholmukhamedov B.F.

Associate Professor of SamSU

Abstract

This article provides information about toponyms as a component of the national language corpus. It demonstrates that toponyms are an integral part of linguistic culture and highlights their role in the national language corpus. The article suggests that including toponyms in the national language corpus will facilitate access to information about them and enhance users' ability to acquire more convenient and comprehensive facts.

Keywords

card index, toponym, corpus, language corpus, corpus linguistics, Internet system, specialized search engine.

In linguistics, toponyms, as an object of study in onomastics, particularly in toponymy, are examined as linguistic units that carry significant historical, national-cultural, and social information. The study of toponyms holds both theoretical and practical importance in terms of investigating the development and improvement of the language system, as well as the dialect, history, geography, ethnography, nature, and other aspects of the regions associated with them.

Toponyms are created by people in a specific language and provide information about the social life of that period. Toponyms are linguocultural units that preserve the people's past, customs, and various socio-political and ethnocultural processes they have experienced. In this sense, toponyms represent the historical and linguistic wealth of a language, an integral heritage of the people's spirituality. Indeed, preserving, studying, and passing on historical heritage to future generations is one of the most important priorities of our state policy [1, 26].

Today, objective conditions have been created for the fair and truthful portrayal of our history, values, rich national culture and spirituality, and scientific heritage. In particular, the attitude towards place names in the territory of the Republic of Uzbekistan has fundamentally changed; they are now considered an integral part of our national spirituality and culture. In this regard, a legal and regulatory framework for Uzbek toponymy has been established and systematized based on relevant documents [2.].

H. Hasanov, S. Qorayev, Z. Dosimov, T. Nafasov, S. Gubayeva, B. Orinboyev, Sh. Qodirova, L. Karimova, Y. Xojamberdiyev, T. Raxmatov, J. Latipov, N. Begaliyev, A. Turopov, N. Oxunov, T. Enazarov, N. Uluqov, S. Boriyev, O. Oripov, Sh. Temirov, and other scholars made significant contributions to the scientific study of Uzbek toponymy [7, 8-9; 4, 3; 5. 7.]. Through their research encompassing various aspects of toponymy, they established the scientific and theoretical foundations for studying Uzbek language toponymy, laid the groundwork for toponymy as an integral part of Uzbek onomastics, and made crucial contributions to its development. These studies examined the naming of geographical objects in the Uzbek language, including naming motives and principles, lexical-semantic, etymological, morphemic, and grammatical features, word formation characteristics, and distribution areas.

Today, with the advancement of computer technologies, it has become increasingly important to preserve toponyms as specific language units, expand their scope of study, and

investigate them as part of the national language corpus.

In world linguistics, attention to computational linguistics intensified from the second half of the 20th century. Corpus linguistics emerged as its main branch. The first corpora appeared in the form of card catalogs, and later as written and electronic dictionaries. To date, real language corpora have been created, encompassing millions of words and word forms [10. 44-45;].

In recent years, research in the field of corpus linguistics has been expanding. A number of scientific studies have been conducted in this area, both in world linguistics and Turkology. In Uzbek linguistics, B. Mengliev, S. Karimov, Sh. Shahobiddinova, Z. Kholmanova, L. Raupova, Sh. Hamroyeva, M. Abjalova, G. Toirova, G. Ikromova, J. Jumabayeva, G. Ergasheva, and A. Eshmominov have been contributing to the development of corpus linguistics through their research. Additionally, the conceptology of the Uzbek National Corpus is being developed by a team of scientists led by B. Mengliev [3. 7-8.].

At this point, let us answer the question: what is a corpus or a language corpus? V. Zakharov defines the term corpus as follows: A corpus is a collection of language units that make up a set of texts compiled for a specific purpose. V. Plungyan considers a corpus as "a collection of linguistic data based on recorded oral and written texts of a particular language" [8.].

B. Mengliyev asserts that a national corpus of the Uzbek language should be created, stating: "A corpus is a collection of texts subjected to a search program to determine the characteristics of language units, a compilation of written or oral texts stored in electronic form in natural language, integrated into a computerized search system based on software. Language corpora are an indispensable tool for solving research and practical tasks in linguistics" [9]. This implies that creating a database of toponyms, which are an integral part of the language corpus, provides certain conveniences in determining their linguistic features both diachronically and synchronically. For this purpose, a specific set of toponyms is collected, and their linguistic features are incorporated into the language corpus. Users can then access the necessary information through a specialized search system. Sh. Hamroyeva offers the following thoughts on the special search system created for language corpus users: "A corpus is a collection of electronic texts that operates on a specialized search engine, referencing contexts of millions of words. It is an electronic library, dictionary, and linguistic grammar within the Internet system. The corpus is an electronic form of natural (real) language that can be expanded, corrected, and edited" [6].

Corpus is the most modern, powerful programmed system for language research, language learning, and dictionary compilation. In this system, it is possible to compile a dictionary by collecting toponyms, indicating their scope of use, and, of course, analyzing their linguistic features. In the created corpus of toponyms, the following linguistic features of names can be examined:

The range of use of toponyms and their frequency;

Correct spelling of toponyms;

Grammatical structure and structural models of toponyms;

The usage status of toponym bases, topoformants, and toponymic identifiers involved in the formation of toponyms;

The semantic and etymological features of toponyms.

It is also possible to showcase the historical development, socio-cultural, geographical, ethnographic, and statistical features of toponyms existing in the language corpus, thereby providing users with the opportunity to access such information. Indeed, through place names, we learn about ancient trade routes, the distribution of tribes and languages, the locations of fortresses and settlements, the properties of water sources and mountain passes, mineral resources, the presence of wells in the steppes, the vegetation, fauna, and climatic characteristics of the area, crafts, historical events, legends, and other information. All of this is crucial for the history of our national economy and culture" [11.37.]. Furthermore, it should be emphasized that through toponyms, one can understand the history of a region, events that occurred in different periods associated with it, the ethnic characteristics of historical figures and the peoples who lived there, their natural and geographical conditions, location, map, population distribution, social status, national values, as well as the processes by which the name of the region has changed and been replaced by other names.

In conclusion, creating a database of toponyms in the national language corpus serves as a necessary software system for users to determine the extent of each name's presence in other regional names, its grammatical structure, semantic features, and etymological meaning. This facilitates the exchange of information in today's information age, allowing for quick and easy assimilation of data.

References

1. Mirziyoyev Sh.M. We will resolutely continue our path of national development and raise it to a new level. Volume I. - Tashkent: Uzbekistan, 2017, p. 26.

2. Law of the Republic of Uzbekistan "On the Names of Geographical Objects" / People's Word, 2011, October 13, No. 198. - P.1.

3. Abdurahmonova N. Computer models of the electronic corpus of the Uzbek language. - Toshkent, 2021, pp. 7-8.

4. Begmatov E., Uluqov N. Explanatory Dictionary of Uzbek Onomastic Terms. - Namangan, 2006, p. 3.

5. Begimov O. Historical-linguistic study of oronyms of Southern Uzbekistan. philol. sci. doc. diss. abstract. Samarkand. 2021, p. 7.

6. Hamroyeva Sh. Linguistic foundations of compiling the author's corpus of the Uzbek language: abstract of the dissertation of Doctor of Philosophy (PhD) in Philological Sciences. - Qarshi, 2018.

7. Karimov S., Buriev S. Uzbek toponymy at the stage of development. - Tashkent: Fan, 2006. Pp. 8-9.

8. Plungyan V. Why We Make Corpora. https://infopedia.su/1x77a7.html

9. https://kun.uz/uz/nevs/2020/10/18/bakhtiyor-mengliyev-uzbek-tilining-ham-milliy-korpusi-yaratilishi-lozim

10. Kholmukhammedov B. The Concept of Language Corpus / Language and Literature