

INCLUDING TOPONAMES OF SAMARKAND REGION INTO THE CORPS

*N. Norpo'lotova,**PhD student at SamDCHTI.*

Abstract: This thesis, based on scientific sources, discusses in detail the criteria and principles for including toponyms of the Samarkand region in the Uzbek language corpus, the existence of several specific principles for creating a language corpus of toponyms, the fact that if it is created, it will bring many benefits, and as a result, it will be possible to quickly and easily obtain information about place names through a simple program. It can be used by all specialists interested in this field.

Key words: computer technologies, computer linguistics, corpus linguistics, mathematical, linguistic corpus, database, toponyms, toponymy, place names, geography, ethnography, distinguishing feature.

Annotatsiya: Ushbu tezisd Samarqand viloyati toponimlarini o'zbek tili korpusiga kiritish mezonlari va tamoyillari, toponimlarning til korpusini yaratishning o'ziga xos bir qancha tamoyillari mavjudligi, agar u yaratilsa, ko'plab yengilliklarni keltirib chiqarishi va buning natijasida oddiy dastur orqali joy nomlari haqida tez va oson ma'lumot olish imkoni yaratilishi haqida ilmiy manbalarga tayangan holda atroflicha fikr yuritilgan. Undan bu sohaga qiziquvchi barcha mutaxassislar foydalanishi mumkin.

Tayanch so'zlar: kompyuter texnologiyalari, kompyuter lingvistikasi, korpus lingvistikasi, matematik, lingvistik korpus, ma'lumotlar bazasi, toponimlar, toponimiya, joy nomlari, geografiya, etnografiya, farqlovchi belgi.

Аннотация: В данном тезисе, опираясь на научные источники, подробно рассматриваются критерии и принципы включения топонимов Самаркандской области в корпус узбекского языка, наличие ряда специфических принципов создания языкового корпуса топонимов, если он будет создан, это принесет множество удобств и, как следствие, позволит быстро и легко получить информацию о названиях местностей через простую программу. Им могут воспользоваться все специалисты, интересующиеся этой областью.

Ключевые слова: компьютерные технологии, компьютерная лингвистика, корпусная лингвистика, математика, лингвистический корпус, база данных, топонимы, топонимия, топонимы, география, этнография, отличительный знак.

Introduction. In the 21st century, linguistics, like all sciences, is developing rapidly. In particular, the emergence of many new areas in it is a clear proof of this [1:7-10]. The most modern of these areas is computational linguistics, on the basis of which corpus linguistics was formed in linguistics and is currently considered the most developed new area. In this context, it is natural to ask why a corpus is needed and what is its significance in linguistics. First of all, computer technologies play an important role in solving various problems of linguistics and creating their basis [2:645]. In the 20th century, which is considered the century of information technology, the use of electronic corpora was

widespread in a number of areas, including computational linguistics. In the world community, the software aspects of creating mathematical and linguistic corpora have been reflected in the works of many scientists [3:90-98]. If all the information on linguistics is placed in the corpus, it will have many advantages and conveniences for users. If a large database of all areas is created in the Uzbek language corpus, it will be easier to study this area and make many innovations in it. As is known, corpus linguistics was formed after the development of computer linguistics. However, over the past decades, this area has been studied separately from computer linguistics. Our goal is to create a corpus of Uzbek place names of the Samarkand region. There are several specific principles for creating a linguistic corpus of toponyms, which, if created, will bring many conveniences and, as a result, will create the opportunity to quickly and easily obtain information about place names through a simple program.

Toponyms are a rich source of information for the humanities, social sciences and a number of other disciplines, and are especially valuable for chronological research in history, geography, ethnography, and linguistics. Indeed, no discipline can be imagined without place names. In particular, they are a rich source of processes studied in literature and linguistics. That is why we set ourselves the goal of including place names in the corpus. Toponymy is also in the relationship between language and geography. In addition to culture and linguistics, the science of place names also uses the methods of all other disciplines related to it [4:140-163]. Toponymy is a field located at the intersection of linguistics, geography, and history, and other disciplines and their subfields are also used in studying the naming process.

The criteria and principles for including place names in the corpus refer to the methodological standards and the procedure for their application, their description in the system of linguistics or other disciplines. Speaking about these criteria, they include the following norms among themselves. Before including a particular territory in the corpus, it is necessary to adhere to its own criteria. The first of these criteria is: Official level:

- this includes only toponyms that are in official documents or included in maps and legal sources;
- names used among the people are separately marked.

The second is historical validity:

- names found in historical sources can also be included.

The third is stability:

- when talking about the toponymy of a certain territory, the selected place should be included in the constant speech of the population, in the media or in relevant literature.

The fourth is territorial accuracy:

- the geographical location should be clear.

The fifth is compliance with linguistic norms:

–it should comply with grammatical norms.

The fifth is absence of repetition:

–if the language corpus contains information on the selected territory, a distinguishing mark should be included.

When a program of new toponyms is created in the Uzbek language corpus based on these principles, the user can easily and quickly obtain information about the territory he belongs to through this program. Therefore, it is necessary to operate on the basis of these inclusion conditions. Along with the implementation of the criteria for including toponyms in the corpus, there are also specific principles. These main approaches are as follows:

The first is scientificity:

–when using toponyms, each toponym must be scientifically based on sources or be included based on certain official documents.

Systematicity:

–toponyms must be divided into territories in a certain order.

Normativeness:

–it is required to include them based on spelling rules.

Based on the historical-etymological approach:

– the origin of toponyms is studied separately.

Multilingualism:

–if a place name is used in many languages, each of them is indicated separately.

Technical suitability:

–each toponym used must be in electronic form, in accordance with technical standards.

Based on the above principles of entry, the user can quickly and easily form the toponym he needs, and this has a number of advantages. Because creating it electronically is an effective method not only for the user, but also for historians and travelers from other regions. Therefore, when entering toponyms into the corpus, it is necessary to fully adhere to all criteria and principles.

There are also stages in compiling a corpus of toponyms, which are as follows:

1) collecting information;

2) compiling a list of toponyms;

3) annotation.

There are also special symbols used in compiling a corpus of toponyms. For example:

<name>Samarkand</name>

<historical name>

There are also specific possibilities for using the corpus:

- a) linguistic research;
- b) sources based on history and geography;
- c) studying how it is used in dialects and phonetic variants;
- d) identifying and introducing changes in geographical names over time.

Based on these principles and criteria, we will present a sample corpus of Samarkand toponymy in tabular form. That is, we will compile a corpus of the Urgut place name related to this region as an example [5:72-75]:

T/r	Toponim nomi	Ma'nosi va etimologiyasi	Tuman/ shahar	Kategoriya asi	Tarixiy nomlari	Til variantlari	Manba
1.	Urgut	"balanddagi shahar"	Urgut tumani va shahri ham mavjud	Tuman/ shahar	Urgut	Urgut, Urkach, Urkut, Urkent, Orkent	Tarixiy manbalar

The above table can also be included in the corpus in Excel or Word, XML formats. This, in turn, brings a number of other conveniences to users.

In the current era of rapid change, every developing field cannot fail to use computers in its scientific activities [6: 120-123]. It can be said that corpus linguistics is a science with its own standards, because this field creates much more convenience for scientists with its research object. Therefore, the inclusion of toponyms in the corpus is extremely relevant. In the development of its criteria and principles, it is necessary to work based on specific evidence.

REFERENCES:

1. Rahimov A. Fundamentals of computer linguistics / A. Rahimov; editor-in-chief A. Nurmonov; Ministry of Higher and Secondary Specialized Education of the Republic of Uzbekistan, Andijan State University. - T.: Akademnashr, 2011. - 160 p.
2. Abdurahmonova N. Urazaliyeva M. Theoretical and practical issues of creating a corpus of oral texts in the electronic corpus of the Uzbek language//Academic Research in Educational Sciences Volume 3 | Issue 3 | 2022. Bet-650
3. Patrick Jentsch, Stephan Porada. Digitization, Text Analysis and Corpus Linguistics Book Title: Digital Methods in the Humanities 2021 page 128
4. Aidana Ainabek¹ and Bekzhan Abdualiuly L.N. Gumilyov Eurasian National University, Republic of Kazakhstan Journal of Ethnic and Cultural Studies 2024, Vol.11, No. 2, 229 Effects of Language Learning Strategies on Teaching Toponyms and Folk Geography Terms in Kazakh and Nogai Languages
5. Boriev S. Lexical-semantic analysis of microtoponyms of Urgut district. elephant. science. name... diss. -Tashkent, 2009. -B. 72-75. 140)
6. Jurayeva M.B. MAIN FEATURES OF CORPUS LINGUISTICS// "Economy and Society" №5(96)-2 2022 www.iupr.ru