



COMPARISON OF UZBEK AND RUSSIAN LANGUAGES

O‘ZBEK VA RUS TILI KORPUSLARI QIYOSI

СРАВНЕНИЕ УЗБЕКСКОГО И РУССКОГО ЯЗЫКОВЫХ КОРПУСОВ

Geldiyeva Khursandoy

Tashkent State University named after Alisher Navoi University of Uzbek Language and Literature Faculty of Translation Theory and Practice 305-group student
xursanoygeldiyeva@gmail.com

Abstract

Nowadays, many languages have their own international, national, or educational corporations. This, in turn, indicates a growing focus on language. It is known that the first national corpus included samples of fiction of artistic value, with linguodidactical significance (not translated). Later, language corpora created for various purposes, including the International Corpus of English and the American National Corpus, led to the emergence of some other types of corpora. In addition, more than a dozen studies on Uzbek corpus linguistics, which has not yet had a long history, have been conducted, and in a short period of time the Uzbek language corpus has been established at Tashkent State University named after Alisher Navoi. Uzbek language and literature. This article discusses the possibilities of the existing Russian National Corpus and the Uzbek Language Corpus.

Keywords: Uzbek language corpus, national corpus, academic corpus, language corpus, corpus linguistics.

Annotatsiya

Bugungi kunda ko‘plab tillar o‘zining xalqaro, milliy yoki ta’limiy korpuslariga ega. Bu esa o‘z o‘rnida tilga bo‘lgan e’tiborning oshganligidan dalolat beradi. Ma’lumki, ilk milliy korpusga badiiy qimmatga ega, lingvodidaktik ahamiyatga ega (tarjima qilinmagan) badiiy adabiyot namunalari kiritilgan. Keyinchalik turli maqsad doirasida yaratilgan til korpuslari, jumladan, ingliz tili xalqaro korpusi (international corpus of english) va Amerika milliy korpus (American national corpus)lari muayyan darajada boshqa tur korpuslarining yuzaga kelishiga turtki berdi. Shu qatorda hali uzoq tarixiga ega bo‘lmagan o‘zbek korpus lingvistikasida ham o‘ndan ortiq tadqiqot ishlari qilindi va qisqa vaqt ichida Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetida o‘zbek tilining ta’limiy korpusi yaratildi, ayni damda uning imkoniyatlari takomillashtirilmoqda, bazasi boyitilmoqda. Mazkur maqolada rus tili milliy korpusida mavjud ta’limiy korpus va o‘zbek tilining ta’limiy korpuslari imkoniyatlari yoritildi





Kalit so‘zlar: *o‘zbek tilining ta’limiy korpusi , milliy korpus, ta’limiy korpus, tili korpuslari, korpus lingvistikasi.*

Аннотация

В настоящее время многие языки имеют свои международные, национальные или образовательные корпорации. Это, в свою очередь, свидетельствует о растущем внимании к языку. Известно, что в первый национальный корпус вошли образцы художественной литературы, имеющие лингводидактическое значение (непереводимые). Позже языковые корпуса, созданные для различных целей, в том числе Международный корпус английского языка и Американский национальный корпус, дали начало нескольким другим типам корпусов. Позже языковые корпуса, созданные для различных целей, в том числе Международный корпус английского языка и Американский национальный корпус, дали начало нескольким другим типам корпусов. Кроме того, было проведено более десятка исследований по корпусной лингвистике узбекского языка, которая пока не имеет долгой истории, и за короткое время в Ташкентском государственном университете имени Алишера Навои был создан корпус узбекского языка. Узбекский язык и литература. В данной статье рассматриваются возможности существующих Национального корпуса русского языка и Корпуса узбекского языка.

Ключевые слова: *корпус узбекского языка, национальный корпус, академический корпус, языковой корпус, корпусная лингвистика.*

Introduction

The corpus is a language source with a large volume and a set of structured texts. In corpus linguistics, they are used to analyze statistical analysis, views, linguistic phenomena, or theoretical rules within a particular language or language department. The corpus can contain textual information in one language or in several languages. The corpus is usually understood as a text corpus, but the corpus is no longer just text. So we use the word corpus instead of the word corpus. To make language research more effective, corporations will be explained (more on this later). For example, one type of body comment is POS-tagging. This includes tags based on word categories and categories in that category. That is, the word books includes the following information: noun, plural, future tense. This is the information that touches the word. Another form of annotation is lemmatization, which means the basic form of a word.





For example, the word books, book's, to book has three different meanings, but they have the same basis - book. This is called lemmatization. We should not confuse basic and basic concepts here. Simply put, a lemma is the part of a word that is beyond the formative affixes [<https://shokiryuldash.blogspot.com/2020/06/>].

Material and Methods

Most of the world's languages have their own national corporations, which are distinguished by their level of excellence and ability to scientifically process text. Time has shown the importance and place of the national corps in the development of the national language and the state language. If we think about the first national corpus of the Russian language, the national corpus of the Russian language covers, first of all, the period from the middle of the XVIII century to the beginning of the XXI century. Whether the period is past or new, it is a socio-linguistic, artistic, colloquial, lively and partly dialectical text. The corps contains original (untranslated) fiction (prose, drama, and later poetry) that has artistic value and is of interest in language teaching. In addition to fiction, other examples of written literature include journalism, popular science literature, personal speeches (lectures), personal correspondence, diaries, documents, and more.

A relatively new direction as part of computer linguistics has been actively developing since the 1980s. A text conference is a collection of books, magazines, newspapers, and more designed for automated processing and automated processing. One of his first texts was written in English from 1962 to 1963 at the University of Bravan Corps under the direction of W. Francis. In Russia, since the early 2000s, V.V. Vinogradov's texts in Russian are estimated to be close to 100 million. In addition to building data barriers, he is also involved in the development of computer tools (computer programs) to extract various information from linguistic text buildings. From the user's point of view, the text contains the words representation (representation and efficiency requirements, completeness and efficiency).

Computer linguistics is actively developing in Russia and abroad. The flow of publications in this area is very large. Since 1984, Competitive Linguistics ("Computer Linguistics") has been published quarterly, with the exception of theatrical states. Excellent organizational scientific work, the Association of Accounting Linguistics (Association of Accounting) has the regional structures that exist in the world (especially the European branch). The international conference is held every two years (in 2008 the conference was held in Manchester).





The main directions of computer linguistics will be discussed at the International Conference "Dialogue" in Russia, as well as at the International Conference "Dialogue", organized by the Faculty of Philology of Moscow State University, Yandex and other organizations. The relevant problem is also widely demonstrated at international conferences on artificial intelligence at various levels.

Results

Methods of teaching Uzbek to foreigners are not being developed. There are no quality guides. They can't even be written. The quality / low quality, effectiveness / inefficiency of language applications is not immediately apparent. After all, native speakers do not learn their mother tongue only through education. She is a language teacher for her family, community and radio and television. But foreigners need perfect resources that will give clear results.

It is important to scientifically determine the stages of complexity of our language and on this basis to create a series of manuals at different lexical and grammatical levels. First of all, it is necessary to form a linguistic base. This requires systematic, in-depth scientific research. Language is the main means of preserving and enriching the spiritual heritage, so language is the existence of a nation.

Creation of an electronic corpus of the Uzbek language in electronic form, approved by the Decree of the President of the Republic of Uzbekistan dated October 20, 2020 "On measures to further develop the Uzbek language and improve language policy in the country." The "Concept of Improvement" includes all scientific, theoretical and practical information on the Uzbek language in the priority areas of modern information technologies and communications of the state language, ensuring the active integration of the Uzbek language in the world. promotes scale. We have a great responsibility to make sure that the Internet is an information network.

Language corps is a tool for learning a language and solving problems. This is different from a normal electronic library. The purpose of the e-library is to provide extensive coverage of works of art and journalism that reflect the socio-political, spiritual and economic life of the people. Electronic library texts are inconvenient for research because they are not linguistically processed. Because the e-library is not designed to create a database of research materials, but to collect national and spiritual heritage. The language corpus, unlike an electronic library, is a collection of useful and interesting texts for language learning and research.

"What is the National Corps?" The question is still of interest to many. The formation of the language corps in our country is a novelty.





The National Language Corps is an electronic search system for identifying the characteristics of language units, a collection of digitized written and spoken texts of a natural language. The National Corps is a programmed system with a wide range of possibilities for changing the language unit, obsolescence, the emergence of new ones, the expansion and contraction of meaning, the emergence of new expressions, the creation of traditional and modern dictionaries. The national word in the corpus means not only the language but also the structure and composition of the corpus. The National Corpus differs from other types of language corporations in that it includes internal corporations such as the academic corps, the author's corps, the poetic texts corps, the scientific and formal texts corps, the oral corps, the literary texts and newspapers, and the dialect corps. It is widely used by linguists, lexicographers, computer linguists, programmers, editors, translators, journalists, publishers, scientists, teachers, educators and all other professionals.

In general, a multi-purpose and multi-functional language corpus is the National Corpus of a particular language.

The National Corps must meet the following requirements:

- 1) The presence of large volumes of texts that apply to all styles of language;
- 2) Linguistic, morphological and syntactic definition of texts;
- 3) The availability of corpus metadata is a necessary requirement [Abjalova, 2022: 75].

The curriculum is equally important for learning a native language and a foreign language. The corpus is very useful in teaching the language, showing the richness of the vocabulary, and giving a series of examples to explain the use of the word through grammatical means. The importance of the model for language learning is constantly updated, and the ability and opportunity to reflect it is available only in the corpus. Here the teacher can find new, convincing, endless and colorful examples, assignments, exercises are not difficult to determine, in a few minutes can prepare assignments consisting of new examples on the topic. The educational corps is a language corps based on school textbooks and dictionaries, and is a special (structural) type of the Uzbek national corps. The Uzbek language educational corps AM-FZ-201908172 was established within the framework of the practical project "Creation of the Uzbek language educational corps", the first presentation of which took place on April 23, 2021, and its base is currently being enriched. and improved [Abjalova, 2022: 75].

It is recommended to rely on a corpus where the database is regularly updated so that language learning is not interrupted. The National Language Corps is one of the main tools for achieving the globalization of our native language, raising and enhancing our





national spirituality. We wish success to the team of the Uzbek National Corps [Manzura Abjalova, "The Uzbek National Corps is an important cultural event"].

Discussion

There are similarities and differences between the Russian National Corps and the Uzbek Language Corps. First of all, if we look at the similarities, they both have a "search" button, and the first section we see is both the home page and the "home" section. There are also sections for user manuals, corpus information, scientific publications, and morphoanalysts.

As for the differences between the two corporations, the Russian National Corps is a bilingual corps with English and Russian speakers. In addition to Uzbek, Uzbek can be used in Russian and English in the National Corps. The Uzbek language corps also has a language resources and thesaurus department, but the Russian national corps does not have such departments. The Russian corpus provides statistics, graphs, and word frequencies, while the Uzbek corpus does not. The Uzbek-language curriculum includes an author's corpus and a parallel corpus, while the Russian national corps has another corpus - the drug corpus.

In our curriculum, we can search for words by lexeme, lemma, compound, and consonant. We use this method to find the word we want. As an example, we went in to search for the word lemma and searched for the word book. When you click the search button, all the words related to the word in the book will appear. It also indicates the period and style in which the word book was used.

These features are reflected in the corners of the case. As an example, we see the result by typing the word "book" in the search bar in the "Dictionaries" section of the work:

1) "Word structure" - automatic morphological analysis determines that "book" is the basis;

2) "Word syllables" are separated in book form;

3) The word book is explained;

4) It turned out that the opposite meaning is the word notebook. There is even a hyperlink to see the linguistic details of the word notebook;

5) The meaning of the book lemma is given: manuscript;

Talking about the main difference between the two corpses, we went to the dictionaries section and found the word in the Uzbek corpus by clicking on one word. Areas of application of the word sought, as well as antonyms, paronyms, homonyms, synonyms of the word, appeared. This is an advantage of the Uzbek language education corps. The main advantage of the Russian national language corpus is the morphological analysis of the word used, as well as its morphological analysis.





Conclusion

In conclusion, all the organs created in the world make a great contribution to the development of science. It is also very useful for language learners and scholars. In fact, we write the word we want in the language we want, that is, in the language in which the language corpus exists, and find its meanings, grammatical analysis, and how it is used. It also indicates the period and style in which the word was used. With the development of time and the focus on language, language corporations have a special place in language learning, that is, in order to increase the prestige of the language, to increase the number of language learners. That is why we must not stop developing language corporations.

References

1. <http://marifat.uz/marifat/ruknlar/fan/O-zbek-tili-milliy-korpusi-muhim-madaniy-voqelik.htm>
2. Abjalova M. O'zbek tili korpusining dasturiy ta'minoti imkoniyatlari. // Kompyuter lingvistikasi: muammolar, yechim va istiqbollari. Xalqaro ilmiy-amaliy konferensiya to'plami. Elektron nashr / ebook. – Toshkent: ToshDO'TAU, 22.04.2022. – B. 73-81.
3. Chambers A. Towards the corpus revolution? Bridging the research–practice gap // Language Teaching. 2019. V. 52. P. 460–475 doi:10.1017/S0261444819000089 14.
4. Farr F., O'Keeffe A. Using corpus approaches in English language teacher education // Routledge Handbook of English Language Teacher Education (S. Walsh and S. Mann – Eds). London: Routledge. 2019. pp.268-282. 13.
5. Leńko-Szymańska A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education// ReCALL. 2014. V.26, No 2. P. 260–278. doi: 10.1017/S095834401400010X 9.
6. Leńko-Szymańska A. Training teachers in data-driven learning: Tackling the challenge// Language Learning & Technology. 2017. V. 21, No 3. P. 217–41. DOI: 10125/44628 10.
7. O'Keeffe A., McCarthy M. J., Carter R. From corpus to classroom: Language use and language teaching. Cambridge: Cambridge University Press. 2007. 315p. doi.org/10.1017/CBO9780511497650
8. Дикарева С.С. Корпусные технологии в режиме диалога «преподаватель — студентисследователь» // Труды международной конференции «Корпусная лингвистика – 2011», 27 – 29 июня 2011 г. СПб.: С.-Петербургский гос. университет. Филологический факультет. 2011. С. 157–162. 11.

