Spectrum Journal of Innovation, Reforms and Development
Volume 19, September, 2023 | ISSN (E): 2751-1731
Website: www.sjird.journalspark.org

METHODOLOGY OF COMPILING THE NATIONAL CORPUS OF THE UZBEK LANGUAGE AND ITS PROBLEMS

Toirova Guli Ibragimovna-
BukhSU Usbekische Linguistik und Journalismus
Professor der Abteilung, Doktor der Philologischen Wissenschaften

Aziza Navro'zova Aslonovna
Staatliche Universität Buchara
Doktorand der 1. Stufe

Abstract
The article discusses the types of interfaces and the importance of the search window of the corpus in the creation of the national corpus of the Uzbek language. The interface of the national corpus consists of various designs and structures, the author is responsible for their completeness, the interface should be attractive and effective. The creation of the interface is based on national or modern features, and the interface should focus on the national color. Linguistic corpora are a very fast-growing branch of the world of computational linguistics that has achieved great success. An interface is a communication system between a technology and a user. Interface types such as visual, gestural and linguistic were analyzed.

Keywords: interface, author's corpus, mathematical modeling, morphological and semantic annotation, information, linguistic basis, artificial intelligence, computational linguistics, corpus linguistics, language corpus, specialized software, electronic library, lexicon, morphological, grammatical, semantic signs and problems with linguistic signs.

Introduction

Corpus linguistics is the most developed field of world science, including linguistics, and the corpus is a necessary working tool for linguists, a source of information on oral and written monuments and national cultural heritage. It is a collection of searchable texts, a well-structured corpus that serves as a stable linguistic basis to ensure the effectiveness of linguistic research[1,2,4,5]. Artificial intelligence products include electronic dictionaries, translation portals, terminological databases, virtual (electronic) libraries, electronic text corpora, electronic administration, electronic publications, electronic textbooks and manuals

Linguistic electronic resources, which are products of artificial intelligence, are considered raw materials for the creation of a specific language corpus We provide a visual interface (interface) for creating the national corpus of the Uzbek language This type of interface (interface) does not cause difficulties in creating the first case [20]. This is because it is a standard computer interface that transmits information using visual images displayed on a

monitor. In the future, it will be possible to offer a language version that takes blind people into account .

One of the latest trends in this area is the touch interface. The principle of operation is based on the physical interaction between the user and the machine, which takes place via certain objects[34]. We can say that this is an attempt to provide the user with the information that he previously received graphically through the monitor.

Figure 1

The interface of the corpus has a different design and structure and its perfection is the responsibility of the author who creates the corpus. Because the surface is the first impression of the case, the attractive overview. When designing an interface (interface), it is necessary to take into account decorations that reflect the national color and signs that reflect classicism or modernity.

This is called the search window of the body, i.e. the interface. "Search from the corpus" – at this point it is possible to search for a word or phrase. The result will look like this. The following features of the word can be found using the "Lexico-grammatical search" button of the corpus: word meaning, grammatical status, synonymy, homonymy, paronymy, antonymy, variant of the word, period of use of the word, method of use of the word, etc[25, 28].

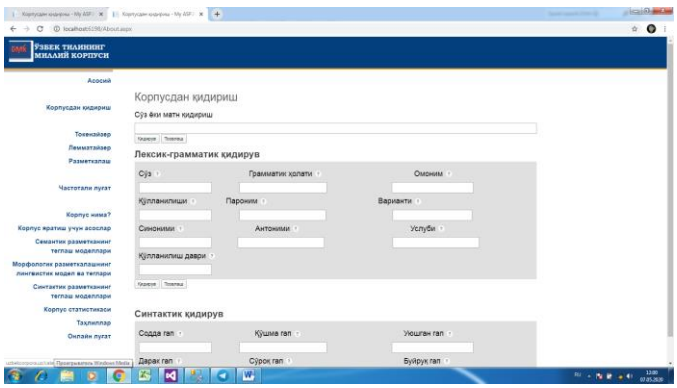

Figure 2

For example, if the word "ingichka" ("thin") is written, we get the following information:
• the meaning of the word: 1. The cross-section is smaller than the norm; 2 Chiyildoq, sharp (over clay);
• Grammar status: [adjective];
• Synonym: soft;
• Homonym: no;
• Paronym: no;
• Antonym: thick;
• Option: no;
•Period:
• Method of application: neutral;

Using the "Syntactic Search" button, we distinguish between simple and compound, organized sentences according to the purpose of the sentence: declarative, interrogative, command type and according to the structure of the sentence.

Figure 3

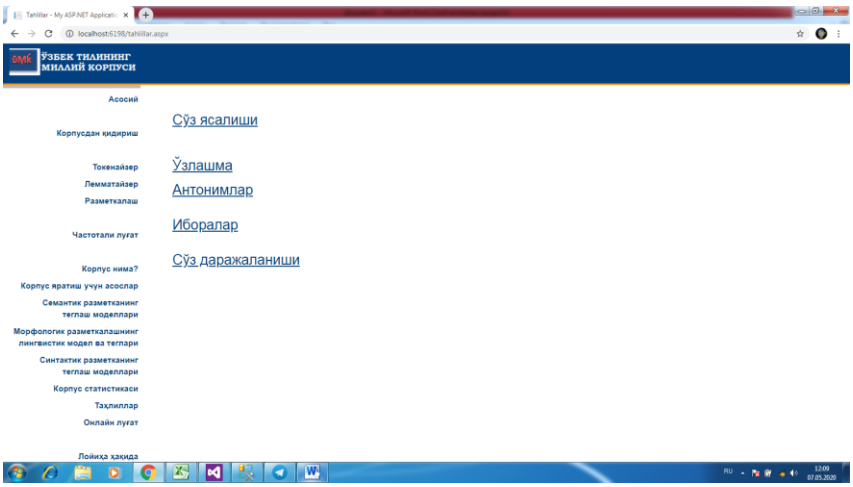

Figure 4

In the "Word formation" button, about 200 word formations are analyzed.

илож	но	лан	лик					
илтифот	сиз	лик	бе	лик				
инсоф	ли	сиз	лик	бе	лик	но	лик	она
инсон	лик	парвар	лик	сифат				
интизом	ли	лик	сиз	лик	бе	лик		
ипак	ли	лик	фуруш	чи	лик			
иқлим	лаш	шунос	лик					
иқтидор	ли	сиз						
иқтисод	ий	от	чи					
исбот	ла	сиз	бе-					
ил	гак	гич						
илгари	ги	ча	ла					
жабр	ла	ли						
жадал	ла	лик						
жадид	изм	чилик						
жафо	каш	лик	кор	ли	сиз			
жаҳолат	ли	параст						
жаҳон	бахш	гир	лик	намо	шумул			
жилов	бардор	дор	ла	хона				

Figure 5

When extracting the database, Excel spreadsheets are used, in which the first column contains the file name (exact path), and the other columns contain metatext attributes and technological information**.** This action allows you to effectively use the built-in tools of the Excel program and provides greater convenience in the search engine for example, search, filtering, analysis and data processing (action list, autocomplete, statistics) **.** In this case, the tables must be saved in text format, and this format must be understood by Excel**.** As a result of this process, the file saved in tabular form can be taken over not only by Excel, but also by other spreadsheet programs, providing an opportunity to increase the efficiency of the work situation**.**

In conclusion, it is worth noting that when compiling the online version of the national corpus of the Uzbek language, the words contained in the "Annotated Dictionary of the Uzbek Language" are analyzed. The interface should meet the requirements of modern software design, be understandable for the user and be comfortable to use**.** An interface is a communication system between a technology and a user. The user interface is visual, gestural and audible. A national corpus should consist of a system and an application programming interface.

References

1. Charlez, Meyer, (2004). English corpus linguistics: An introduction. Cambridge University Press, UK, 168 p.
2. Eshmuminov, A., (2019). Synonymous database of the Uzbek language national corpus. Dissertation of PhD in Philology, Tashkent.
3. Fries, Ch.C., (1969). The structure of English. An introduction to the construction of English sentences, London.
4. Hamroeva, Sh., (2018). Linguistic bases of creation of the author's corpus of the Uzbek language: Author’s Abstract of the Dissertation of PhD in Philology, Tashkent.
5. Zakharov, V.P., (2011). Corpus linguistics: a textbook for students of humanitarian universities, Irkutsk, 161 p.

6. Karimov, R., (2021). Linguistics and programming issues of creating a parallel corpus of Uzbek and English, Author's Abstract of dissertation of PhD, Bukhoro, 151 p.
7. Leech, G., (1991). The State of Art in Corpus Linguistics, English Corpus Linguistics, London.
8. Melchuk, I.A., (1985). Word order in the automatic synthesis of the Russian word (preliminary messages), Scientific and technical information, 12:12-36.
9. Mengliev, B., (2018). Is the Uzbek language corpus being created? Ma'rifat newspaper. April 3, 2018, retrieved from: http://marifat.uz/marifat/v_pomosh_uchitelu-marifat/savol/1142.htm.
10. Mengliev, B., Bobojonov, S., Hamroeva Sh., (2018). Uzbek National Corpus. April 26, 2018, retrieved from: http://marifat.uz/marifat/ruknlar/fan/1241.htm.
11. Pulatov, A. Q., (2011). Computer Linguistics. Tashkent, Akademnashr, 520 p.
12. Toirova, G., (2019). The Role of Setting in Linguistic Modeling. International Multilingual Journal of Science and Technology, 4(9):722-723, available at: http://imjst.org/index.php/vol-4-issue-9-september-2019/.
13. Toirova, G., (2020). About the technological process of creating a national corpus. Foreign languages in Uzbekistan, 2(31):57-64, available at: https://journal.fledu.uz/uz/2-31-2020.
14. Toirova, G., (2020). The importance of the interface in the creation of the corpus. Internauka, 7, DOI: https://doi.org/10.25313/2520-2057-2020-7-5944.
15. Тоирова Г. Корпус лингвистикасининг атамалар луғати. Изоҳли луғат. Германия: «GlebeEdit» номли халқаро нашриёт. 2020, –68 б.
16. Toirova G.Uzbek tili milliy corpusini yaratishning nazariy va amaliy masalari. Halkaro monograph. Germany: "GlebeEdit" by Nomli Halkaro, 2020. – 169 p.
17. Toirova G.,Yuldasheva M., Elibaeva l. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. –P.352-355. (scopus)
18. Toirova G., Jurayeva O., Abulova Z., Norova M., Norova F. Application of Innovative Technologies in Teaching Process. // International Journal of Psychosocial Rehabilitation, Vol. 24, Special Issue 1, 2020. ISSN: 1475-7192.–P.386-390. (scopus)
19. Toirova G., Hamroyeva N. The importance of linguistic models in the development of language bases // Sciences of Europe. vol 2, No 59 (2020) ISSN 3162-2364. –P. 57-64
20. Toirova G. The importance of the interface in the creation of the case. International Scientific Journal «Internauka», International Scientific Journal «Internauka». – 2020. – №7.Online journal. https://doi.org/10.25313/2520-2057-2020-7-5944(Impact Factor –8,758)
21. Mengliev B., Toirova G., Hamroeva Sh. Uzbek tilining milliy va mualliflik korpi. Guvohnoma No. DGU 05735 Uzbekiston Respublikasining Intellectual Mulk Agency. – Toshkent, 2018
22. Toirova G. Importance of interface in creating corpus. // Хоразм Маъмун академияси ахборотномаси. –Урганч, 2020, – №2/2. –Б. 49-52.

23. Тоирова Г. Миллий корпусни яратишда интерфейснинг аҳамияти. //Қарақалпақ давлат университети хабаршысы. – Нукус, 2019, ISSN 2010-9075–№ 4(45). – С.195-198.
24. Toirova G. Milliy corpus yaratishning tekhnologik zharayoni hususida //Uzbekistonda khorizhiy tillar. Electron ilmium-techniques journal. – Toshkent. 2020, –No 2 (31), – B.57– 64.
25. Тоирова Г. Ўзбек тили миллий корпусни яратишда интерфейснинг аҳамияти. // Сўз санъати халқаро журнали, – Тошкент, 2020, № 3, – Б.100-105.
26. Toirova G. The importance of linguistic module forms in the national corpus/ Current problems of modern science, education and training (Current Problems of Modern Science, Education and Training in the Region) (Electronic Scientific Journal), – Urgnch. 2020, –No. 5 , –B.155–166.
27. Toirova G. The importance of linguistic models in the development of language bases. // Бухоро Давлат университети илмий ахбороти. – Бухоро, 2020. –№6. – Б.98-106.
28. Тоирова Г. Лингвистик базани тузишда модуллаштиришнинг аҳамияти // Наманган давлат университети илмий ахборотномаси. –Наманган, 2021. – №3. - Б.377-386.