Importance of Interface in Creating Corpus

Toirova Guli Ibragimovna , Yuldasheva Mavjuda Rakhimovna, Elibaeva Lola Suleymanovna

Annotation: ***The article discusses the author's corps and its significance in modern glossary, the world of Pushkin's author's corps, the Czech writer's corps, Shakespeare's author's corps and their shortcomings. The interface of the author's corps is made up of different designs and structures, and the author is responsible for its completeness, the interface should be attractive and impressive. The creation of the interface is based on the design of the national or modern features, the interface should involve the life and works of the artist in photoes. The Corpus of Linguistics is a very rapidly developing branch of the world of computational linguistics, which has achieved great success in this regard.***

The Corpus of Linguistics is also taught as a science in world universities. The subject of this discipline is the theory and practice of building a corpus, such as body features and the basics of programming. The Corpus of Linguistics deals with general theory and practice of computational linguistics, the formation of the language body, and computer technologies. The article tells about modern information technologies that have created tremendous opportunities for language functionality. Computer translation, editing, analysis, electronic dictionary and thesaurus are proof of our opinion. Especially the creation of modern electronic dictionaries and the culture of their use is one of the effective ways of learning a language. In particular, the role of language buildings created and developing at a fast pace throughout the world when demonstrating the ability and ability to master the language is very large. The purpose of the article is to study the linguistic foundations of the Uzbek language corpus, to study the linguistic value of the linguistic corpus, the history of corpus linguistics, to study the author's linguistics of the corpuses, its features in the social, lexicological, educational and other fields.

The article gives an idea about the interface, the content of the corpus, its flawless functioning and at first glance the importance of the author's personality, creative heritage, classification.

Keywords**: *Interface, the author's corps, mathematical modeling, morphologic and semantic annotation, information, linguistic base, artificial intelligence, computer linguistics, corpus linguistics, language corpus, special software, e-library, lexical, morphological, grammatical, semantic symbols, problems with linguistic markup.***

I. INTRODUCTION

Corpus linguistics in the world linguistics became the subject of study in the 60s of the last century. Back in the 60s, R.G. Petrovsky said that "Reliable linguistic information can be obtained from a large array of texts," but case studies in the field of corpus were initiated by Bloomfield, Fritz and Bonders in the 40s. The creators of the Brown Corps (1961–1964), Nils Francis and Henry Kuzer, first developed the principles for the creation of a corps. The works of John Sinclair, author of the Bank of England (1980), deserve special attention. In Russian linguistics, V. Zakharov, A. K. Kutuzov, E. V. Nedoshivina, V. Rykov, V. Pungane studied the corpus, its varieties, features, social significance of the corpus and the principles of corpus construction. Special studies on copyright cases can be found in the works of O. Kukushkina, A. Polikarpov and E. Surovtseva. Although this issue has been studied in the linguistics of the World Corpus, there are no monographic studies of corpus linguistics in Uzbek linguistics. The studies of the aforementioned linguists have been carefully reviewed, studied and used during the dissertation research.

In State nationwide program is paying a great attention as main direction to the communicating of public education schools. According to this program educational establishments are provided with techniques of modern computers. Public education schools and educational establishments are connected with Internet and ZiyoNET completely. Modern people master so much information that it is not possible to use or treat information without Information and Communication Technology (ICT). Year by year in our life it has been developing computer and Information and communication technology. Nowadays the main goal of educational policy is directed at the learners, important and necessary for the future developing of modern education that satisfying demands of society and government. It is important to draw educators and leaders of schools and high educational establishments in developing professional skilful and from the first day in additional pedagogical education. There is the truth that impossible to refuse, if the representative of present time unable to use nowadays technology and unable to use them for their life, job and handicraft is considered drawback.

It is important to emphasize that, efficient using of possibilities of modern ICT by educators testifies that they are skilful specialists.

Revised Manuscript Received on September 25, 2019

Toirova Guli Ibragimovna *Bukhara State University Department of Uzbek linguistics, associate professor, Doctor of philosophical sciences, Uzbekistan, Bukhara, e-mail: tugulijon@mail.ru*

Yuldasheva Mavjuda Rakhimovna, *teacher of the department of pedagogics, Bukhara State University, e-mail: mavjuda79@mail.ru*

Elibaeva Lola Suleymanovna, *Senior Lecturer, Department of Pedagogy, Bukhara State University, e-mail:elibayeva@mail.ru*

II. MATERIALS AND METHODS

The Computer translation, editing, analyzing, electron dictionaries, thesaurus are evidence of our opinion. Especially, creating e-dictionaries and forming the culture of using them is effective way of owning the possibilities of language. Particularly, the role of creating language corps according to representing and mastering the possibilities of language is great. As our president says : "It has shown necessity of supporting scientific and creative researches overall , marking as task creating necessary conditions for them, for this purpose working out and implementing definite measures by our government, according to the field of every subject doing profound researches, including linguistics. As the developing of information –communication system, it is appeared new branches in the subject as "Corps linguistics". B. Mengliyev and his apprentices have raised the problem of this subject firstly (Hamroyeva Sh. , 2018). Corps is an electron library, dictionary and linguistic grammar of internet system. It is a collection of texts as electron form of real language that is situated in the program of search. It is created Pushkin author corps and Chekhov author corps, Shakespeare author corps; national corps of Russian language, modern American English corps, Oxford English corps in the World. In Uzbekistan it has not created the corps of linguistic base yet. Nowadays though there is an electron library Ziyonet, but it doesn't work in the system of working on the text automatically and implement searching on the base of different signs from text. It doesn't outline to the program of learning language and creating dictionaries. It is impossible to listen audio listening of text. There is a system of working on the text automatically and implement searching on the base of different signs from the text in authoring and national corps program which we create. It is possible to find out the words, phrase, combinations that used less and to learn the use and the orthography of them and it gives the chance of directional teaching of education and the leaner is able to listen to the text. Linguistic analyzing does the main task for corps. Linguistic analyzing is linguistic and extra linguistic separating of special tags into texts and its component parts. Now there are following types of linguistic analyzing: morphological, semantic, syntactical, anaphoric, prosodic, discursive and etc. In some corps it is used the next component analyzing degrees.

Especially, some little corps is connected on the base of syntactic analyzing completely. Such kind of state is usually explained deeply or it has syntactical structure. For instance, syntactic analyzing likes big tree. Analyzing of text in hand takes long time. Now it presents programs in Russian and foreign sites which is possible to enter straightly and to analyze. They are divided into independent and websites. The last years producers' directing to web attachment deserves attention. There are some advantages of these systems: at the same time several users are able to analyze the same document, it doesn't demand to set additional programs but browser apart from it, entering right is limited, can be observed analyzing process. Word form, lemma and tag belong to morphological analyzing system. Word form is morphological unit of chosen text. The first step of analyzing word form is to lemming or to give the lexeme form of word form. The most difficult step of settings of inflectional languages is lemming or lexeme form of word to connect to word form as tag. Because in inflectional languages grammar meaning of word form is mixed to root of word. Differently from inflectional languages lemming is much easier in agglutinative language. [7,10]. The part of without grammar form of word form is equal to root or basis foundation lemma. In settings lemma is given inside of following sign :<*>. In all parts of speech lemming is as following or if it is based on "the part of root-basis foundation of word is equal to lemma" trend, in verb group verb –lemma is given as the form of II person imperative mood. In the articles of dictionary they are given as infinitive <to go>. But it is not suitable for corps, because in text of corps it is searched not form <to go >, but the form<go>. According to this verb-lemma is given as following form: **teach** <read>, **doesn't be** <be>, **show** <see> (Hamroyeva Sh., 2017). During the setting it demands to write from 5 till 10 morphological tags (explanations), sometimes more than them for each word form.

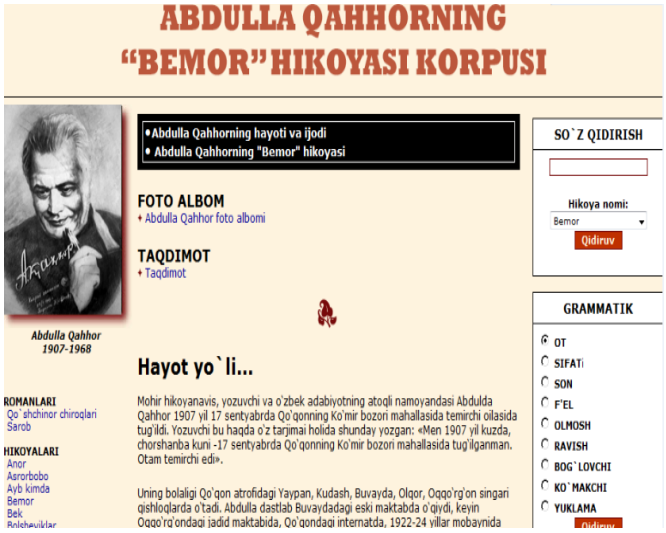
Mathematical modeling of natural languages is based on the creation of artificial intelligence around the world. The results of mathematical modeling serve as the basis for creating an artificial intelligence program. It is a pleasure to note that there are over 30 million speakers in the Uzbek language. However, it should be noted that technologies do not go up to the level of the Internet and, in particular, the conceptual, theoretical, practical, and organizational work in relevant research institutions and centers, and indicates that the list remains on the list of languages. Modern technology has created enormous opportunities for language functionality. Computer networks and information communication technologies (ICTs) have created opportunities for the education system, first of all, to obtain the required information quickly from anywhere in the world. Particularly, the fact that the Internet is accessible through global computer networks at the moment of access to the world's information resources. 2 Modern technology, which is the result of development, is designed to help people. Especially, development of information and communication technologies contributes to the development of each sphere. The development of information and communication systems has led to the emergence of new areas in science. One of such new areas is corpus linguistics. Creating a corpus to achieve the unique language of our mother tongue, to inherit the wealth of our future generations, and to achieve our place in the world's global network, defines the current problems of today.

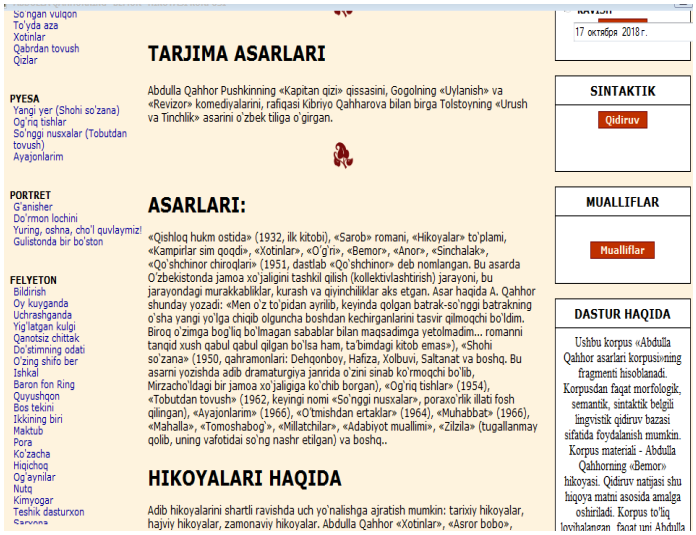
Corpe consists of a collection of electronic texts, working on a special search engine and millions of keyword contexts. The author's corpe is based on the texts containing fiction, journalism, and epic genres created by a particular artist. It is a series of lexicographical sources that include various dictionaries (basic, frequency, toponymy, grammar, phrase, etc.) The author's corps is of great importance in modern design.

This corps is a resource for creating large of dictionaries. Over time, the corpus has become a powerful informational resource, which is important for various linguistic trends. Due to computer-based corpses, dictionaries are created and processed quickly compared to the previous one. This way the dictionary reflects the language from the beginning to the end of the dictionary (until publication), and the dictionary does not have an "outdated" word.

Corps must have a unique interface. What is Interface? The term "interface" is derived from

English and is used to mean "appearance". This word is often used in computer technology. It is the only communication system that provides a variety of communication between a person and a car. Interface is a communication system between different blocks and complex blocks, as well as technology and user. It is logical (information representation system) and physical (data transfer). Interiors of the author's corps are of different design and structure, and their perfection is entrusted to the creator. Because the interface is the first impression that the body is about. In the creation of the interface, it is necessary to take into consideration the national decorative designs and the characteristic that reflects the classic or modernity. The creator's life and activities, the works created by him in a separate window, and the photogallery must be on the interface. The above-mentioned materials allow to use the body as an electronic library. In the "Search" system, the creator's life, activities, works created by him in a separate window, and windows such as photos use the body's capabilities. These windows are software that searches for corpus units. The author's body corpe should be friendly, optimally versatile, and perfect. The global network are characterized by the presence of large volumes of materials (the plurality of materials) and the depth of it. For example, Pushkin and Chekhov's corpuses are the perfect ones with morphological and semantic annotations. But it does not responce modern programming requirements. Because they were 8 or 10 years old. Further, the excellent design of the Shakespeare's corps differs by the design of the whole design, but it is not morphological or semantic. Each of the existing corpses has a distinct minority. For example, the Tajik language can only function as an electronic library, photo and video gallery(Wang W., Liu Y., Harper M. P. , 2002). This body is completely unmatched, and the intermediate system is imperfect and incomplete. But it can only point to a single piece of work. Pushkin and Chekhov's corpse, created by Russian linguists, do not have syntactic significance. The author's corpse is Abdulla Qahhor's e-database on the Patient Story, which is based on artificial intelligence: the processing of the Uzbek language by means of textual editing, automatic editing and translation, searching for different characters. Created body - can determine the period and frequency of the word, analyze the lively construction of the language, the ability to combine words, automatically process text (eg, translate) and various search engines automatically. It saves time. Note: Users are currently using a simple card reader rather than a linux database. It has the following interfaces:





Thus, the interface plays a key role in the creation of the corpse. It is necessary to create an easy-to-use working environment that meets the requirements of modern software design.

III. CONCLUSION

In conclusion, linguistic modeling of tags is expedient, because in linguistic modeling morphological tag turns into conventional abbreviation form. It is produced forms of special linguistic model for settings of each part of speech. Setting is linguistic modeling. The scientific novelty of the research is as follows:

1) for the first time in Uzbek linguistics, the term corpus, corpus features, theoretical foundations, linguistic, practical and educational meaning of the language corpus is revealed;
2) the formation and development of corpus linguistics, features of the first and subsequent generations, the history of language corpses in Russian and English, the state of corpus linguistics, features of modern Russian, English, Turkish and Tajik languages, their general and various aspects, classification of corpus types;
3) general principles of building a

body: the technological process of designing and the stages of building a body, the importance of forming a body, the fact that the tag is a linguistic tool characterized by the characteristic features of the body manager and its types;

4) the goals and objectives of the author's corpus, the characteristics and structure of the author's corpus, the structure and composition of the author's corps, similar and distinctive aspects of A. S. Pushkin, F. D. Dostoevsky, A. Griboedov, and U. Shakespeare;
5) developed the principles of creating the author's case.

REFERENCES

1. Sh.M. Mirziyoyev (2017) Report of the President of the Republic of Uzbekistan at the 72nd session of the United Nations General Assembly September 19, 2017. http://www.uza.uz/ru/politics/prezident-uzbekistana-shavkat-mirziyeev-vystupil-na-72-y-ses-20-09-2017
2. Ahmedova M.B. (2018) Genetic and structural specifications of the "spirituality" nominative units in the Uzbek language // International Scientific Journal " Theoretical and Applied Science.- USA, Philadelphia, 2018.- Volume 66.-P. 331-333(Impact factor- 3.04)
3. 3. Vanyushkin A.S., Grashchenko L.A. (2017) Evaluation of key word extraction algorithms: tools and resources // New information technologies in automated systems. - 2017. - № 20. - pp. 95-102.
4. 4. Nikolaev I.S., Mitrenina O.V., Lando T.M. (2006) Applied and Computational Linguistics - M.URSS, 2016. - 320 p.
5. 5. Nedoshivina E.V. (2006) Corps Texts Programs: A Review of Core Corps Managers. Teaching manual. - St. Petersburg. - 2006. 26 p.
6. 6. Rakhilina E.V., Marushkina A.S. (2015) Corpus studies of the peculiarities of speech of non-standard speakers ("Russian hermetic") // Acta Linguistica Petropolitana. Proceedings of the Institute of Linguistic Studies. 2015. T. XI. № 1. S. 621-639.
7. Leech G.(1991) The State of Art in Corpus Linguistics // English Corpus Linguistics / Aimer K., Altenberg K.(eds.) – London, 1991. – P. 8-29.
8. Кутузов А.Б. (1968) Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf.
9. Bloomfield L. Language. - M .: Progress, 1968. - 608 p.
10. Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – L.,1969.-C.98
11. Bongers H. (1947) The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947.-C.74
12. Francis N., Kucera G. (1967) Computational analysis of modern American English. - M., 1967
13. Melchuk, IA (1985) Word order in the automatic synthesis of a Russian word (preliminary reports) // Scientific – Technical Information. 1985, №12. -C.12-36
14. Hamroyeva Sh. (2018) Linguistic basis for the creation of the Uzbek language. 2018.-52 б.
15. Hamroyeva Sh. (2017) Use in education from the corpus "Language and literary education" Journal. September 2017, № 9. Б.49-50.
16. Hamroyeva Sh. (2018) Corpus creation principles. Journal "Scientific Bulletin of Science". 2018. № 3.
17. Tairova G. (2015) Some of the differences between paradigmatic and discursive systems. // IMPACT: International Jurnal of Research in Humanities, Arts and Leteratura. **(**impact: ijrhal**)** Vol. 3, Issue 12, Dec 2015, 1-4. (№ 12 Index Copernicus Impact Factor - 1,7843)
18. Tairova G. (2016) Phatics - actual problems of linguistics uzbek research // Iranian Journal of Social Sciences and Humanities Research. UCT. J. Soc. Scien. Human. Resear. (UJSSHR). – Takestan, Iran, 2016, Volume 4, Issue 2. – P.16-19. (№5 Global Impact Factor, Impact Factor – 0,765).
19. Tairova G. (2017)Systematic and informative in uzbek discourse// UCT Journal of Social Sciences and Humanities Research.(UJSSHR). – Takestan, Iran, 2017, Volume 5 Issue 2 June. – P.1-6. (№5 Global Impact Factor, Impact Factor –2,758).
20. Tairova G. (2013) Izosign – graphic expression of the discourse as pragmatical situational system 10th International Conference on Crossroad of Civilzations: Aspeets of Lenguage, Culture and Society. – Japan, 2013. – P. 525-529.
21. Casares J.(1969) Dissionario ideologico de la lengua Espanola.– Barselona,1969. –887 c.
22. March F.A. (1958) March's Thesaurus Dictionary. – N.Y., 1958.– 1312 p.
23. Roget P.M. (1952) Thesaurus of English words and phrases. – Lnd., 1952. – 1258 p.
24. http://www.dialog-21.ru/media/2138/zakharov.pdf Zakharov V.P. Corps of the Russian language.
25. Mamontova V. V. (2008) Corpus of parallel texts and database for the study of translation correspondences: problems and procedures for the formation Address of the article: www.gramota.net/materials/1/2008/8-2/50.html
26. Antonova A., Alexey M. ()Building a Web-based parallel corpus and filtering out machinetranslated text. https://www.aclweb.org/anthology/W11-1218
27. Resnik, Philip and Noah A. Smith. (2003) The web as a parallel corpus. Computational Linguistics, 29:349– 380.
28. Wang W., Liu Y., Harper M. P. (2002) "Rescoring effectiveness of language models using different levels of knowledge and their integration", in Proc. ICASSP, Orlando, FL, May 2002.