

## ON THE GENERAL PRINCIPLES OF LANGUAGE BODY CREATION

*Bo'ronova Malika Baxtiyorovna*

*English teacher of Bukhara Institute of Engineering  
and Technology Academic Lyceum*

**Abstract:** The article discusses the basic principles of corpus linguistics, a new field in applied Uzbek linguistics, as well as the process of designing and constructing corpus. Examples of achievements of world linguistics in the creation of corpus resources in the Uzbek language are given. Practical linguistic experience proves how to set up the first stage of corpus linguistics.

**Keywords:** corpus, corpus linguistics, genre-thematic structure, VP Zakharov, SY Bogdanova, LN Zazorina, frequency dictionary, computerization.

The corpus project is the stage of its creation and the ways of further improvement coverage is a key step in creating a perfect corpus. Corpus concept, a new form of traditional index cards for linguists; they are by the 20th century was computerized, it became possible to use it publicly.

Of course, the Internet plays a significant role in the transformation of card files into a corpus earned As a result; it is a big one that allows you to conduct various linguistic studies a universal version of voluminous texts appeared. A dictionary on this matter, the scope of language material that serves as the basis for grammars and the issue of balance is on the agenda, in particular, in the process of creating national corps stood crosswise. The issue of representativeness of the corpus is the adequacy and variety of texts solved with.

For example, whether a small advertising text in newspapers is treated as a separate text or not should they be combined into one text? Whether a newspaper article is considered a text or not should one issue of a newspaper be evaluated as a whole text? Each poem is one

Is it necessary to include the text or the poetry collection as a whole? In response to each other written, published letters discussing essentially the same subject are one text or should it be treated as a separate corpus unit? To these questions

The answer depends on the type of the composing body, the task it will perform later will give.

A corps unit is determined depending on whether it is a national corps or a special corps.

V.P. Zakharov and S.Y. Bogdanova is an important aspect of the design process of the case they also mention the issue of chronology. The beginning of the text in the

corpus what part is taken from the form and what is excluded is another important point is a matter. The pictures in the text belong to the language material excluding tables from the text included in the corpus due to its absence processing according to the case is also important.

They represent the content of the text important, but if it is left in the body, it creates difficulty in sorting.

Quotations, quotations, proper units (terms), measurement units are also separate requires attention. The listed issues are known at the design stage if it is solved based on the principle, some of them are the process of creating a corpus, the solution in using the corpus will be done. At the same time, return the case with the user before starting communication should also be taken into account. Specialists in the technological development of the corpus the following stages are distinguished in the process

1. Ensuring the entry of the text into the corpus in accordance with the specified source.

2. Text processing in the form of automatic reading. Included in the case

Text in electronic form can be obtained in different ways: typed by hand, scanned, copyrighted copy, gift, exchange, Internet, publishers original models given to the body builder by

3. Analysis, preliminary text processing. From various sources at this stage accepted texts undergo philological examination and editing.

4. Conversion, graphematic analysis. Some texts are recoding process the first machine that is realized repeatedly goes through the processing, non-material parts (Fig. table) is deleted or changed. Hyphenation in text, borders (MSDOS in texts) are canceled, hyphens, other symbols are made identical.

Graphematic analysis divides the text included in the corpus into parts (words, connectives), consists of performing actions such as deleting a non-textual element.

The next step in the design of the case is the selection of the source.

The importance of a corpus is that it gathers together a wide range of texts in the same language when the order is given, so work on the basis of several criteria when creating it will be seen. What is the basic unit of the body when sorting the body material, what is its size (preferably how many words it should have), what is the written text based on the source, how many there are, what language the text is included in

A solution will be found to the issues related to the field. The first of this question answer by R. G. Piatrovsky and his students between 1965-80 was given. They are a frequency dictionary, a text for linguostatistical research have drawn up the principles of selection.

Of the corps its main units are word form, root (base, lemma) and sentence. The size of the body to be built determined based on the purpose of the corpus. If it is a

letter, letter combination, sound, it doesn't have to be that big if it aims to study diphthongs.

Study of the text's lexical units, morphological phenomenon, syntactic, stylistic characteristics if it is built for the purpose, a large volume is required.

This is the body builder in the process, of course, to a linguist, linguostatistics specialist or questionnaire method applies. The author relied on his own experience in the process of creating the corpus the total size of the corpus, the publication time of the text, the number of texts, the size of the elementary selection, the main factor is the type of genre to be chosen.

After studying the results, a list of 19,000 books was compiled. Each on this basis 1045 texts consisting of 500 word forms were selected. The corpus as a conclusion selection of material (text) during the design process, sorting, its technical aspects it should be noted that adaptation to the body is the most basic stage.

#### References:

1. Shaykhislamov, N. (2020). Cognitive Linguistics, the Symbolic and Interactive Functions of Language. *Образование и наука в XXI веке*, 1(6), 390-393.
2. Shaykhislamov, N. (2020). Linguistic and Cultural Aspects of Bread Baking Terms in English and Uzbek Languages.
3. Shaykhislamov, N. Z., & Makhmudov, K. S. (2020). Linguistics and Its Modern Types. *Academic Research in Educational Sciences*.
4. Шофқоров, А. М. (2018). Fonetik takrorlar va ularning uslubiy xususiyatlari. *Ta'lim tizimida ijtimoiy gumanitar fanlar*, (4), 79-86.
5. Шофқоров, А. М. (2019). Kelishik turlarini o'rganishda interfaol ta'lim metodlaridan foydalanish.