



Morphological Differentiation Of Forming Words In The National Corpus Of The Uzbek Language

Abror Kurbonmuratov

Lecturer, Termiz State University, Uzbekistan

Journal Website:
<http://usajournalshub.com/index.php/tajssei>

Copyright: Original content from this work may be used under the terms of the creative commons attributes 4.0 licence.

ABSTRACT

The article discusses the problems of correct identification of problems in Uzbek computer linguistics, in particular, the understanding of the need to create corpus linguistics, as well as recommendations for solving today's problems in the process of teaching a second (foreign) language through the National Language Corpus.

KEYWORDS

Creation of the National Corpus of the Uzbek language, problems, solutions.

INTRODUCTION

To correctly identify the problems of computational linguistics at a time when Uzbek computational linguistics entered a new path of development, study and assimilation of the advanced traditions of world computational linguistics, in particular, understanding the need to create corpus linguistics is a significant step towards great success.

One of the problems that the process of teaching a second (foreign) language is facing today is the lack of pedagogical text materials that meet the demand. Therefore, language teachers have to spend a lot of time on daily tasks such as preparing new materials that help to motivate students. In this case, the problem

can be solved using the methods of corpuslinguistics.

The corpus of language and the science associated with it, that is, corpus linguistics, is a new direction that has rapidly entered the field of linguistics and language teaching methodology in the late XX and early XXI centuries. The creation of the language corpus and the development of corpus linguistics is one of the most pressing issues in the field of linguistics, language teaching methods. Although it is not yet recognized by all experts that corpus linguistics is a separate modern branch of linguistics, the role and importance of this field is growing day by day.

Corpus linguistics is a branch of linguistics that deals with the development of text corpus. The term was introduced to science in the 1960s in connection with the beginning of the process of building corpus. Since the 1980s, however, improvements in computational techniques have led to the further development of corpus linguistics. The term corpus linguistics was first used in 1984. Although not a very long history, during this period corpus linguistics has become a leading branch of modern linguistics.

Corpus linguistics emerged in the 1960s mainly from material in English, but corpuses of other languages soon began to appear. In 1963, the first text corpus was created by scientists such as U. Francis and G. Kuchera at Brown University in the United States, [1] which consisted of 500 texts. Each of these texts contained 2,000 words, and were the English texts written in the 15 most popular genres of prose genre in the United States. The case also includes a frequency indicator and some statistics.

A set of concepts related to corpus linguistics has been formed since the 1990s, and various sources have begun to provide similar definitions of this new direction. For example, J. Sinclair describes the corpus as a collection of naturally occurring texts selected to show the diversity of language, to describe different types of language materials [2]. In Russia, corpus linguistics emerged relatively late, and now this direction is developing rapidly. The creation of the National Corpus of the Russian Language in 2004 was an undoubtedly important event in Russian philology and Russian language education. A corpus is an electronic complex of texts in one or more languages, compiled by a number of Russian scholars on the basis of certain parameters. For example, according to N. Kozlova, the corpus is a collection of written and oral texts, which differs from the set of ordinary texts by being numbered, that is, it is analyzed electronically, consists of a set of texts marked and stored on a computer, as well as data and linguistic annotations that regulate this information [3]. In our opinion, the most complete explanation of the concept of “corpus” was given by V.P. Zakharov. According to the scientist, the corpus is a large volume of structured (having a structure) in electronic form, marked / special electronic symbols, a philologically comprehensive and a set of linguistic data designed to solve specific linguistic problems [4]. In general, the following is recognized in every definition given to the corpus:

- 1) many texts must be submitted electronically (online or on disk);
- 2) In order to carry out linguistic analysis, special marks should be made on the language material;
- 3) it should be possible to distribute the language material collected as a result of

the analysis based on different principles (for example, genre, date of text creation, topic, etc.).

One of the main principles of corpus formation is the collection of unedited texts, i.e. language should be included in the corpus in the same way as it is expressed in speech (written or spoken). However, there may be many language “options”, even appearances that are not part of the literary language, rather than patterns that exist in the language to correctly construct sentences from the corpus.

As you know, a simple bank of texts is not enough to solve linguistic problems. Another important requirement for creating a corpus is to put special characters on the text and its components, that is, to perform text marking - to put tags. These symbols are introduced to distinguish specific information objects and represent external, extralinguistic, structural, and linguistic specificity. They represent the lexical, grammatical, and other descriptions of the text[5]. The meta-metric consists of information about the author and the text in general. Let us consider the purely linguistic types of layout in the example of existing corpus. For example, morphological markup is the most common type of designation, which takes into account features that belong not only to word groups but also to grammatical categories. Morphological marking is carried out using special automatic morphological analysis programs. In syntactic notation, the syntactic structure of each sentence is explained using symbols, but such designations require a lot of time and effort. In addition, the elements in the corpus can be provided with semantic, prosodic, graphic and other types of symbols.

Foreign experience shows that with the advent of electronic language corporations, the opportunities to demonstrate the diversity of language and to study linguistic material have grown significantly.

If the search criteria are correctly specified by the researcher, language teacher or language learner when working with the corpus, the various characters listed above will directly collect factual material, simplifies and speeds up the process of performing a number of tasks, such as analyzing it, preparing material for the lesson, checking the practical application of the learned language units.

Currently, the issue of variability of language units in language education is very relevant. If linguistic variability is not taken into account in language education, regional, social, professional, physiological, sexual, individual differentiation of language is not taken into account, the results of research or language teaching will not be objective [6]. In this regard, the existence of a perfect corpus of language can be a ready resource to keep in mind the variability in linguistics.

The practice of using corpus data as an empirical component of lecture courses and student projects is becoming commonplace in the world's leading universities. At the same time, it became clear that the corpus-based approach is the best way to study aspects such as historical, geographical and social variability of language, as well as changes in the language system. At the same time, there is a real opportunity to master the corpus methods of linguistic analysis, to understand their basic principles [7].

Currently, corpus linguistics is becoming more and more popular in the collection and analysis

of practical (factual) material. This is a natural state of affairs as a result of the rapid development of information technology in linguistics.

Here are some ways to use cases.

A student can independently study the authentic material included in the corpus, and make assumptions and conclusions about the essence of a particular language unit, its meaning in the construction of speech, its capabilities. It is well known that the knowledge acquired using this inductive method is stored in the student's memory for a long time.

For example, a second language learner must explain the difference between two similar words using existing corpus materials to instantly find 20-30 sentences containing the above words and use them in the context analysis process. Understands the semantic difference between similar words. Today, computerized education is moving away from the method of presenting teaching material and towards methods that require creativity.

For example, Carol Chapel points out that in the process of grammar analysis, we involve the student in the analysis of the computer corpus, which in turn helps students acquire independent analysis skills, as well as their own understanding of how to explain grammatical rules in the future [8].

In conditions when the issue of improving the culture of self-study of students is relevant, it is very useful to perform the above tasks. Consequently, the student will have the opportunity to find answers to specific questions, to see the real natural state of the language being studied, to prove its historical, geographical, social variability.

In language education, corpus methods are also considered to be highly effective innovative methods that complement traditional teaching technologies. This method has such aspects as interdisciplinary communication, empirical adequacy, authenticity, flexibility, flexibility to specific tasks and objectives, the ability of the student to acquire independent knowledge, the use of the method of "invention" in education.

It seems that the development of corpus linguistics as a network of computational linguistics is one of the most urgent directions. The presence of language corporations provides a number of conveniences and benefits in the use of language, especially for linguists and language learners. The challenge for programmers is to create programs that allow AI to "understand, read, edit, and respond to the Uzbek language". Without such programs, Uzbek corpus linguistics can not develop. If an Uzbek language bridge is created, Google will create a search engine in Uzbek. Online translation (internet translation) or automatic translation programs will be developed. This will expand the possibility of creating various electronic dictionaries of the Uzbek language.

REFERENCES

1. Brown's Corpus, freely accessible from the University of Leeds website: <http://corpus.leeds.ac.uk/protected/> Sinclair J. Corpus, Concordance, Collocation. Oxford, 1991. - P. 171
2. Kozlova N. V. Linguistic corpuses: definition of basic concepts and typology. NSU Bulletin. Series:

-
- Linguistics and Intercultural Communication. 2013. T. 11. Issue 1. - p. 79.
3. Zakharov V.P. Corpus linguistics. – Sank Petersburg: Saint-Petersburg State University, 2005 .—P.3.
 4. Zakharov V.P. Corpus linguistics. – Sank Petersburgs : Saint-Petersburg State University, 2005 .—P. 6.
 5. Plungyan V. A. “Integrum” and the National Corpus of the Russian Language in Linguistic Research // Integrum: Exact Methods and Humanities. - Moscow: Summer Garden, 2006 .—P.P.76-77.
 6. Information and Communications Technology for Language Teachers. Introduction to the applied value of the enclosure, <http://www.ict4lt.org/en/enmod2-4.htm>
 7. Carol A. Chapelle // Essential teacher. 2003. Vol. 9. P.P. 5-11. - P. 6.