



CLASSIFICATION OF DIALECTAL CORPSES

Azizbek Vosiljonov, Maxsudjon Abdunabiyev

Fergana State University

It is important to emphasize the role of regional and social dialects in the formation of the corpus. Dialect words are lexical units used in the local speech of a certain region, which do not correspond to the norms of the literary language. When talking about dialects, first of all, it should be noted that these lexical units are more typical for the artistic style of the literary language. They primarily use this style of literary language. Dialects also play an incomparable role in enriching the language corpus.

Bringing the national language and its dialects to the information world, its electronicization is an important issue facing not only linguistics, but also linguistic researchers. In addition to corpora that represent the functioning of national languages in various fields of communication, corpora that simulate communication in separate linguistic communities that are distinct within a national language are also needed. The most important language formations of this type are dialects. Ideas about the need to create a machine fund of dialect texts were first put forward in the 1980s by A.S. Gerd and V.E. Goldin. Currently, in world linguistics there are a number of corpora representing individual elements of dialectal speech: foreign corpora of dialectal texts (for example, the Helsinki corpus of English dialects, Kirk's corpus of Northern Irish transcribed speech (NITCS), IViE (intonational variations in English) corpus, BBC Voices); dialect subcorpus of the National Corpus of the Russian Language (NCRL); Pustosha - Moscow region, Shatursky district, including texts - examples of speakers' speech in the dialect.

Dialect subcorpus of the national corpus of the Russian language. Work on the creation of the corpus of the Russian language began in the 2000s. The national corpus of the Russian language is organized in the form of an information system of Russian language texts collected in electronic form. This corpus is intended for everyone who is interested in the Russian language and is looking for answers to various questions. Researchers, teachers, students, linguists, students and others from the corpus. can use. The national corpus of the Russian language has been maintained since 2003 within the "Program of Philology and Informatics" of the Russian Academy of Sciences. The national corpus of the Russian language contains more than 140 million words, and it is planned to increase its volume to 200 million in the future.

Mahmudov states in his work: "One of the most important features of the national corpus of the Russian language is that it has a proportionate text corpus and is able to adequately express the language. This means that the corpus includes, whenever possible, all types of written and spoken texts available in the language. These texts are represented in the corpus according to their share in the language. The national corpus of the Russian language meets all the requirements of a modern corpus. It embodies various genres - official-business, journalistic, religious, artistic, everyday speech and others. The work can be divided into two parts: modern and diachronic. The texts belonging to the modern corpus represent the years 1951 – 2010. In the other part, texts covering the first half of the 18th, 19th and 20th centuries are collected. Most of the texts collected in the national corpus of the Russian language cover the last 200 years. Therefore, it is

not difficult to learn the small changes introduced here.

V.P. Zakharov divides the national corpus of the Russian language into the following subcorporations:

1. newspaper corpus - covers texts published in mass media in 2000;
2. syntactic corpus;
3. corpus of dialect texts - dialect materials of different regions of Russia were obtained while preserving grammatical features;
4. The corpus of poetic texts - covers poetry from the 18th century to the present day;
5. Corpus of the history of Russian accent - about the history of accents in words collected data and texts;
6. Oral language corpus - audio recordings, individual oral speeches, films is a corpus from which transcripts are collected. The corpus covers the years 1930 – 2000;
7. Multimedia corpus A corpus created on the basis of fragments of films shot in the 1930s - 2000s. In addition to language texts, non-verbal tools and gestural texts used in films are also housed in this building.

The history of active study of Russian folk dialects goes back more than two centuries. During this time, the regional classification of the Russian language was fully described, linguistically integrated regional communities (dialect groups, dialect zones) of different levels were identified. A detailed study of the phonetic, grammatical, lexical dialectal features that make up the system of inter-dialectal correspondences (dialectal differences) of different nature made it possible to develop the dialectal language theory as a special language formation. The second half of the 20th century "Dialectological Atlas of the Russian Language", "General Slavic Linguistic Atlas", "Orwas distinguished by large-scale dialectological projects such as "Atlas of Russian dialects along the Volga and Lower Volga", "Dialectological atlas of the Russian language". The publication of atlases and a number of dialect dictionaries, including the multi-volume "Dictionary of Russian Folk Dialects" (edited by F.P. Filin), created a good resource base for studying the various linguistic features of Russian folk dialects. The development of the cognitive-discursive scientific paradigm in linguistics led to the formation of a new direction in the study of Russian folk dialects.

The development of modern information technologies allows the creation of machine-processable corpus of dialect texts. Work on the creation of an electronic corpus of Russian dialect speech is carried out within the framework of the projects "National Corpus of the Russian Language" (Institute of the Russian Language of the Russian Academy of Sciences) and "Textual Representation of the Dialect". The corpus of dialect texts is part of the National Corpus of the Russian Language (NCRL) and is publicly available at <http://www.ruscorpora.ru/search-dialekt.html> since December 2006 available for searching. All dialect texts are provided in orthographic form and the newly introduced standard user not only with fragments of texts (as in other NCRL subcorpora) but also with whole texts. provided the opportunity to work.

References:

1. Vosiljonov, A. (2022). Basic Theoretical Principles Of Corpus Linguistics. *Academicia Globe*, 3(02), 173-175.
2. Vosiljonov, A. (2022). Lingvistik Tadqiqotlarda Korpus o 'Rganish Obyekti Sifatida. *Ijtimoiy Fanlarda Innovasiya Onlayn Ilmiy Jurnali*, 2(11), 176-182.
3. Vosiljonov, A. (2022). Pragmalingvistika Va Uning Tahliliy Shakllanish Tarixi. *Science And Innovation*, 1(B8), 99-105.
4. Vosiljonov, A., & Isaqova, X. (2023). Effectiveness Of Mother Tongue Education In The Primary Grades. *International Journal Of Advanced Research In Education, Technology And Management*, 2(2).
5. Vosiljonov, A. (2022). Pragmalinguistics And The History Of Its Analytical Development. *Science And Innovation*, 1(8), 99-105.

6. Khalimboyeva, F., & Vosiljonov, A. (2023). Maktabgacha Yoshdagi Bolalar Diqqatini Rivojlantirish Muammosini Nazariy O'rganilishi. *Journal Of Pedagogical And Psychological Studies*, 1(5), 94-98.
 7. Dilshodbek o'g'Li, R. S., & Boxodirjon o'g'Li, V. A. (2022). Xorij Psixologlarining Ishlarida Shaxsning Tadqiq Etilishi. *Innovative Developments And Research In Education*, 1(12), 39-47.
 8. Vosiljonov, A., & Abdullazizova, R. (2023). High-Spiritual Maturity, Ideality And Pedagogical Views Of The Class Leader. *Modern Science And Research*, 2(6), 1182-1186.
-