

## PROSPECTS OF UZBEKI DIALECTOLOGICAL DICTIONARY AND PRINCIPLES OF CREATING AN ELECTRONIC CORPUS

Norova Gulsanam Maksutovna

associate professor, doctor of philosophy (PhD) in philological sciences.

[gulsanam1981ndpi@umail.uz](mailto:gulsanam1981ndpi@umail.uz)

(+99890) 7304595

### Annotation

This article provides a comprehensive analysis of the prospects of Uzbek dialectological lexicography and the concept of corpus creation. The study systematically collects lexical materials across various regional dialects of the Uzbek language, standardizes them within a corpus system, and demonstrates their effective use for scientific research, dictionary compilation, language teaching, and pedagogical purposes. The article also discusses the methodological foundations for creating lexical corpora using modern computer technologies, metadata standards, search algorithms, and linguistic analysis tools. The research results confirm the importance of developing Uzbek dialectological lexicography, systematically studying regional dialects, unifying terminological units, and their scientific significance in linguistics, history, anthropology, and pedagogy. Furthermore, the corpus creation concept has been developed based on international standards and scientific-practical requirements, thereby laying the groundwork for future automated linguistic analysis and database development.

### Keywords

Uzbek language, dialectology, lexicography, corpus, regional dialect, lexical resources, linguistic analysis, terminology, modern technologies, lexical unit, electronic corpus, linguoculturology, lexical research, semantics, phraseology, database.

In recent years, the comprehensive study of the internal potential of the Uzbek language—particularly its lexical and dialectological richness—has become one of the priority directions in Uzbek linguistics. In the context of globalization, socio-cultural integration, and the rapid development of digital technologies, the preservation of regional features of national languages, their scientific documentation, and their integration into the digital environment have gained particular relevance.

From this perspective, the systematic collection, classification, and inclusion of dialectal lexical units into electronic corpora in order to introduce them into scientific circulation constitutes one of the pressing tasks of contemporary linguistics. A lexical resource in the form of an electronic corpus serves as a reliable scientific source not only for dialectology and lexicology, but also for sociolinguistics, ethnolinguistics, cognitive linguistics, and applied language technologies.

The scientific-theoretical study and systematization of the dialectological lexical wealth of the Uzbek language, as well as the development of an electronic corpus concept, form the basis for creating a comprehensive lexical platform that meets the requirements of modern linguistics. This also makes it possible to bring Uzbek dialectology and lexicography to a new scientific stage by consolidating fragmented dialectal vocabulary into a unified scientific system.

Dialectology is an important branch of linguistics that studies the regional, territorial, and social variants of a language—namely dialects and subdialects—based on their phonetic, lexical, grammatical, and semantic features. Through dialectological research, the historical development of the language, its ethnic layers, cultural-national worldview, and speech diversity are revealed. Therefore, dialectology is considered not only a linguistic discipline, but also one closely connected with ethnography, history, and cultural studies.



Lexicography, in turn, is a theoretical and practical discipline concerned with the systematic collection, description, and standardization of a language's vocabulary, encompassing both the art and methodology of dictionary compilation. Thus, while dialectology studies the regional forms of a language, lexicography is the science and art of dictionary-making [1, p. 34]. Within lexicography, particular importance is attached to identifying the semantic structure of words, their stylistic features, spheres of usage, and historical and regional characteristics. Dialectological lexicography, in particular, documents lexical units formed outside literary norms but actively used in living speech.

The scientific interrelation between dialectology and lexicography lies in the fact that it is impossible to create a complete and comprehensive lexicographic source without dialectological research. By recording dialectal units in dictionaries, the regional diversity of the language is preserved, and historical memory and cultural consciousness are transmitted through language across generations. From this perspective, the lexicographic documentation of dialectal vocabulary becomes an essential component of national language policy.

According to Valentina Maslova, who interprets language as "a semiotic system expressing the culture of a nation" [2, p. 57], language units—particularly dialectal lexicon—represent a sign system reflecting a nation's cultural code, worldview, and mentality. This approach necessitates the study of dialect units not merely as linguistic phenomena, but as linguocultural phenomena. For example, certain regional terms, names of household items, and dialect words related to customs and rituals reflect the historical lifestyle and cultural experience of the people.

In modern linguistics, the integration of dialectology and lexicography has reached a new stage through the creation of electronic dictionaries and linguistic corpora. Electronic dialectological lexicography makes it possible to mark dialect units regionally, semantically, and functionally, and to analyze them interactively. This provides an important scientific foundation not only for theoretical research, but also for applied language technologies, translation systems, and educational processes.

The scientific systematization of lexical units existing in Uzbek dialects constitutes one of the important stages of dialectological research. Dialectal vocabulary represents the main layer reflecting the territorial diversity of the language. It is closely connected with the literary language, yet possesses distinctive phonetic, semantic, and phraseological features. Therefore, a comprehensive approach is required in classifying dialectal lexical units.

The use of the same lexical units in different phonetic forms across dialects is a common phenomenon. In such cases, the meaning of the word is preserved, but its sound composition changes according to regional pronunciation characteristics. For example:

- ko'cha → qochcha
- bola → bala
- ketdi → kitti

Such phonetic variants have developed in connection with historical phonetic processes, assimilation and dissimilation phenomena, and local pronunciation traditions. In dialectological lexicography, it is important to record these variants separately, as they reflect the real state of living speech. Phonetic variants may also serve as a source of enrichment for literary language norms.

In dialectal vocabulary, the semantic expansion of a word within a regional context or the emergence of figurative semantic layers leads to semantic variants. For example, in the literary language, the word yomg'ir ("rain") denotes only an atmospheric phenomenon. In some dialects, however, it may also refer to wind, cold weather, or unfavorable natural conditions.

This demonstrates the context-dependence of dialectal semantics. Semantic variants are closely linked with popular worldview, perception of natural phenomena, and cultural experience,



and therefore require analysis from a linguocultural perspective. Such units expand the figurative potential of language and reflect regional worldviews.

Dialectal phraseological units represent one of the most stable and culturally rich layers of language. Although often close to literary phraseological expressions, they display distinctive forms and semantic shades in regional variants. For example:

- quloq solmoq (“to listen”) – in some regions appears as quloq tutmoq;
- qo‘l tekkizmoq – used in variants such as qo‘l urmoq, qo‘l tiqmoq.

These phraseological units reflect everyday life, social relations, and national mentality. Dialectal phraseologisms often possess a high degree of figurativeness and emotional-expressive meaning [3, p. 102].

The systematization of dialectal lexical units from phonetic, semantic, and phraseological perspectives enables:

- scientific documentation of the territorial richness of the Uzbek language;
- creation of electronic dialectological corpora and dictionaries;
- preparation of a solid source base for historical-linguistic research;
- effective use of regional language materials in education.

Thus, systematizing dialectal vocabulary should be regarded not only as a lexicographic task, but also as a strategic scientific direction aimed at preserving the cultural memory of the language.

Today, dialect lexicography is no longer limited to traditional card indexes, expedition records, or printed dialect dictionaries. Instead, it is developing in integration with digital technologies, corpus linguistics, information and communication tools, multimedia platforms, and comparative-historical and areal linguistic approaches. This enables the study of dialectal units based on authentic speech samples, real communicative contexts, their historical roots, territorial distribution, and interaction with other languages.

In conclusion, the future of Uzbek dialectological lexicography lies in the harmony of scientific innovation, cultural responsibility, and technological opportunity. New-generation dialect dictionaries created on this basis will serve as a solid scientific foundation for preserving national linguistic wealth, transmitting it to future generations, and ensuring its integration into global linguistic scholarship.

#### REFERENCES:

1. Maslova V.A. *Lingvistika va madaniyat: nazariya va amaliyot*, Moskva, 2018, 34-b.
2. Sobirov A. *O‘zbek tili dialektologiyasi asoslari*, Toshkent, 2020, 57-b.
3. Rahmonov M. *O‘zbek frazeologik lug‘ati*, Toshkent, 2019, 102-b.
4. Karimov N. *Til va madaniyat: lingvokultrologik tadqiqotlar*, Toshkent, 2021, 88-b.
5. Севортян Э.В. *Этимологический словарь тюркских языков*. — Москва: Наука, 1974. — 215-217-бетлар.
6. Радлов В.В. *Опыт словаря тюркских наречий*. — Санкт-Петербург, 1893. — 498-500-бетлар.
7. Abdullayev F. *O‘zbek shevalari leksikasi*. — Toshkent: Fan, 1985. — 96-98-бетлар.
8. Makhsutovna, N. G. (2020). ALL-TURKIC LEXICAL LAYER IN THE UZBEK DIALECTS OF ZARAFSHAN OASIS. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, 9(9), 10–15. Retrieved from <https://www.anglisticum.org.mk/index.php/IJLLIS/article/view/2115>
9. Makhsutovna, N. G. (2020). All-Turkic lexical layer in the Uzbek dialects of Zarafshan oasis. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, 9(9), 10-15.



10. Norova, G. (2024). All-Turkish Lexical Layer in Turkic Languages and Uzbek Language Dialects. *Elmi Tədqiqat Beynəlxalq Elmi Jurnal*, 27-33.
11. Norova, G. M. (2026). THE LITERARY HERITAGE OF ALISHER NAVOI AND THE ROLE OF UZBEK DIALECTOLOGY AS NATIONAL CULTURAL HERITAGE: A THEORETICAL, TEXTUAL, AND TRANSLATION ANALYSIS. *Bulletin news in New Science Society*, 3(1), 5-15.
12. Norova, G. (2025). Specific Characteristics of Tense Nouns and Adverbs in Uzbek and Azerbaijan Dialects. *Turkology*, (1).
13. Norova Gulsanam, M. (2025). Description of Dialectal Units in Jadid Dictionaries. *31.03. 2025*, 40.
14. Maxsutovna, N. G. (2025). Description of Dialectal Units in Jadid Dictionaries. *Spanish Journal of Innovation and Integrity*, 40, 359-363.
15. Norova, G. (2023). MAQSUD SHAYXZODA ADABIY MEROSINING TIL XUSUSIYATLARI. *Uzbekistan language and culture*, 5(1), 138-146.
16. Maxsutovna, N. G. (2025). MILLIY KONTENT YARATISHDA TILIMIZDAGI MADANIY KONSEPTLARNING O ‘RNI. *Navigating Green Economy through Innovations, Policies, and Strategies on Sustainable Practices*, 1, 447-450.
17. Maxsutovna, N. G. (2025). Description of Dialectal Units in Jadid Dictionaries. *Spanish Journal of Innovation and Integrity*, 40, 359-363.
18. Norova, G. M. (2025). OZBEK VA INGLIZ TILLARIDA SODDA GAPDA SUBYEKT PREDIKAT ORTASIDA MAZMUNYIY SINTAGMATIK MUNOSABATLAR TURLARI, ULARNING UMUMIY VA XUSUSIY JIHATLARI. *CANADA-SCIENTIFIC REVIEW OF THE PROBLEMS AND PROSPECTS OF MODERN SCIENCE AND EDUCATION*, 1(10), 8-13.
19. Maxsutovna, N. G. (2025, October). O ‘ZBEK TILI TARIXI VA O ‘ZBEK SHEVALARINI TADQIQ ETISH MUAMMOLARI. In *Conferences* (Vol. 1, No. 6, pp. 186-192).
20. Anvarovna, A. S. (2023). Constructions (models) of social intelligence in future English language teachers. *Horizon: Journal of Humanity and Artificial Intelligence*, 2(4), 169-172.
21. Atakhujaeva, S. (2023). Constructions (Models) Of Social Intelligence In Future English Language Teachers. *Horizon: Journal of Humanity and Artificial Intelligence*.
22. Атаходжаева, Ш. А. (2025, October). Роль Информационных Технологий В Обществе. In *International Conference on Global Trends and Innovations in Multidisciplinary Research* (Vol. 1, No. 4, pp. 17-21).
23. Usarov, J. E., Khimmataliev, D. O., Kuvatova, N. B., Atakhujaeva, S. A., Chudakova, V. P., & Akbarova, N. N. (2022). Pedagogical competence.

