

# BUKHARA DIALECT AS A NEW LINGUISTIC DATA SOURCE FOR THE UZBEK LANGUAGE CORPUS

**Soliyeva Farida Bobir qizi**

National University of Uzbekistan named after Mirzo Ulugbek

Faculty of Journalism and Uzbek Philology

First-year master's student in the Computer Linguistics program.

**Abstract:** This article analyzes the linguistic features of incorporating Bukhara dialect into the Uzbek language corpus. The study examines the phonetic, morphological, syntactic, lexical, and sociolinguistic characteristics of Bukhara dialect, highlighting the differences and similarities between it and the Uzbek language and other dialects. The regional and social significance of the dialect, its contribution to the language system, as well as its scientific importance in the fields of linguistics and dialectology, are clarified. The article emphasizes that the inclusion of Bukhara dialect in the Uzbek language corpus plays a significant role in enhancing the diversity and richness of the Uzbek language, creating a new resource for scientific research.

**Keywords:** Bukhara dialect, Uzbek language, dialectology, phonetic features, morphological features, syntactic features, lexical features, sociolinguistics, linguistics, Uzbek language corpus.

## INTRODUCTION

Uzbek language is an attractive and rich language belonging to the Turkic language family. Its regional variations and dialectal characteristics play a significant role in its development and formation process. Every region, including cities and villages, has its own unique forms of speech, which reflect changing and diverse linguistic systems. Bukhara dialect, as one of the important dialects in Uzbekistan, stands out with its distinct characteristics. This dialect, possessing lexical, phonetic, morphological, and syntactic features consistent with the general system of the Uzbek language, offers the possibility of contributing new linguistic data to the Uzbek language corpus.

The distinctiveness of Bukhara dialect is closely linked to its history and culture. These dialectal features hold significant importance not only for linguistics but also for cultural studies, ethnography, and history. For linguists, Bukhara dialect provides an opportunity not only to study unique grammatical and lexical constructions but also to analyze them in relation to the overall structure of the Uzbek language and to identify new linguistic categories. Additionally, it is important to examine how the linguistic structure of Bukhara dialect dynamically changes in relation to time and space and how it is used by specific regional social groups.

This article analyzes the unique characteristics of Bukhara dialect, focusing on its phonetic, morphological, and syntactic aspects. It also discusses the place of Bukhara dialect in the Uzbek language corpus and explores the potential for opening new research directions in linguistics. The article emphasizes that through the scientific study of dialectal differences, the Uzbek language corpus can be further enriched and expanded, benefiting not only linguists but also a broader audience studying the language. Bukhara dialect reflects not only the linguistic features of the region but also its cultural wealth, making it a valuable source for not just linguistics but also the study of cultural heritage.

### LITERATURE REVIEW

The study of Bukhara dialect holds significant importance in Uzbek linguistics, as it is not only unique in its phonetic and lexical aspects but also in its morphological and syntactic features. Numerous studies have been conducted on this dialect in Uzbekistan. Bukhara dialect is studied not only as a distinct dialect but also as a valuable source of new linguistic data that contributes to the development of the Uzbek language corpus.

There are phonetic changes present in various dialects of the Uzbek language, and Bukhara dialect exhibits distinctive phonetic differences. S.Ashirboyev, in his "Uzbek Language Dialectology" handbook, analyzes the phonetic system of Bukhara dialect, highlighting how the vowel "o" changes to "o'" in this dialect [1]. For example, the word "yo'l" is pronounced as "yo'l'" in Bukhara. These phonetic changes reflect not only the linguistic features of the dialect but also the cultural and social structures of the region.

S.Choriyeva, in her research titled "Differences between Modern Uzbek Literary Language and the Bukhara Region Dialect," analyzes the impact of certain words from the Romitan district dialect of Bukhara on the Uzbek literary language. Through the dialect, phonetic, lexical, morphological, and syntactic differences are identified. Dialectal features play an important role in enriching the literary language and in the formation of new lexical units. This study contributes to understanding the dialectal richness of the Uzbek language and its cultural and historical transformations. [2] The word "gavra" lexically, differs from the literary word "beshik." "Gavra" refers to a special cradle made of reed or mulberry wood, used for rocking a baby to sleep, while "beshik" in the literary language refers to a traditional vessel or piece of furniture used to lull a baby to sleep. An example from Bukhara dialect illustrates the use of "gavra": "Ona gavrani asta tebratib, uning ichidagi bolasini tinchlantirdi." This example reflects the distinctive lexical characteristics of the dialect and the differences from the literary language.

Sh.Fozilova, in her research "Inconsistencies in the Present-Day Uzbek Literary Language and the Kogon District Dialect of Bukhara," dedicated to the analysis of dialectal features, has revealed the phonetic, morphological, syntactic, and lexical characteristics of the Kogon district dialect of Bukhara. [3] The unique appeal and richness of each language, as well as the dialect words, shape its linguistic system. This article presents the words of the Kogon district dialect in the new Uzbek transcription, shedding light on the distinct features of the dialect.

When analyzing the words of Bukhara dialect, the phonetic features hold particular importance. Changes in pronunciation, the sounds and their transformations, and alterations in the form of words are analyzed. Let us consider the following examples:

1. supurgi → jo'rop: We can observe changes in pronunciation in these words. In the word "supurgi," the "s" and "p" sounds are pronounced differently in Bukhara dialect as "jo'rop," where "j" and "r" sounds are added. These pronunciation differences indicate phonetic features specific to the region.
2. ona → nana: In Bukhara dialect, the word "ona" is pronounced as "nana." Here, the "o" sound is replaced by "a." This change is a distinctive phonetic feature that sets the dialect apart from others.
3. ho'p → hay: In these words, the "p" sound in "ho'p" changes to "y." Such pronunciation changes are common in Bukhara dialect and contribute to the uniqueness of the words.
4. paypoq → jo'lop: The word "paypoq" is pronounced as "jo'lop" in Bukhara dialect. Here, the "p" sound changes to "l," altering the pronunciation of the word. This change plays an important role in identifying phonetic features.



5. cho‘mich → choygardon: Pronunciation changes in these words also demonstrate the phonetic distinctiveness of Bukhara dialect. The "ch" sound in "cho‘mich" is pronounced as "ch" in "choygardon," reflecting the regional phonetic variations.

The changes in sounds and pronunciation in Bukhara dialect are among the primary factors studied in phonetic features. Along with its distinct phonetic features, the dialect also reflects lexical and morphological characteristics. Studying these processes is significant in linguistics and helps in the deeper analysis of the dialect-specific linguistic features of the Bukhara region. The changes in sounds and pronunciation in Bukhara dialect are among the primary factors studied in phonetic features. Along with its distinct phonetic features, the dialect also reflects lexical and morphological characteristics. Studying these processes is significant in linguistics and helps in the deeper analysis of the dialect-specific linguistic features of the Bukhara region. The inclusion of Bukhara dialect into the Uzbek language corpus is bringing its linguistic analysis to new theoretical and practical directions. Local researchers emphasize the importance of adding dialectal data to the Uzbek language corpus and discuss the role of Bukhara dialect in this process. The phonetic, morphological, syntactic, and lexical features of Bukhara dialect are viewed as a new source of data that aids in the expansion of the Uzbek language corpus. Through studying this dialect, not only in linguistics but also in understanding the social and cultural context of regional changes, significant insights are gained.

The process of adding Bukhara dialect as a new linguistic data source to the Uzbek language corpus opens up opportunities for broader analysis of the dialectal features of the Uzbek language. Research by Uzbek linguists clearly demonstrates the phonetic, lexical, morphological, and syntactic characteristics of this dialect. The scientific significance of Bukhara dialect in linguistics not only reflects the changes in the Uzbek language system but also the cultural and historical transformations. This, in turn, contributes to enriching the Uzbek language corpus and opens up new perspectives in the field of linguistics.

### RESEARCH METHODOLOGY

This research is aimed at the linguistic analysis of the incorporation of Bukhara dialect into the Uzbek language corpus, utilizing several research methods. The objective of the study is to identify the phonetic, morphological, syntactic, and lexical characteristics of Bukhara dialect, while also demonstrating its scientific and practical significance for inclusion in the Uzbek language corpus. The descriptive method is employed to analyze in detail the unique phonetic, lexical, morphological, and syntactic features of Bukhara dialect. Through this method, changes in pronunciation, word forms, and sentence structures are described. The verbal and written materials collected from the Bukhara region are used to illustrate the speech system and grammatical structures of this dialect.

Additionally, the features of Bukhara dialect are compared with other dialects of Uzbek and the standard Uzbek language. This comparative method identifies dialectal differences and common features, examining their impact on the language system. The sociolinguistic method is applied to analyze the social context of Bukhara dialect in society. Through this method, the differences between social groups, youth, women, and men in the use of the dialect are explored. It highlights how social identification, cultural changes, and the social functions of language are conveyed through the dialect.

Corpus research is used to collect data on the lexical and syntactic characteristics of Bukhara dialect, such as the use of sentences and word combinations, and the phonetic structure of the language. New materials related to Bukhara dialect are incorporated into the existing database



of the Uzbek language corpus. Through this method, the place of Bukhara dialect in the Uzbek language corpus and its linguistic richness are determined.

Qualitative and quantitative analysis methods are applied together in the research. Qualitative analysis aims to provide in-depth insights into the structural features of Bukhara dialect and its role in linguistics, while quantitative analysis allows for the measurement of the distribution and frequency of dialectal features. Quantitative data is gathered through surveys, interviews, and dialectal materials, and analyzed using statistical methods. Furthermore, the use of surveys and interviews provides more precise information on the practical application of Bukhara dialect and its understanding by various social groups in the region. Participants in these surveys and interviews share their perspectives on the unique features of Bukhara dialect, its social aspects, and its place in linguistics.

This methodology offers a comprehensive approach to the process of incorporating Bukhara dialect into the Uzbek language corpus and its linguistic analysis. Each methodological approach serves to explore dialectal differences, changes in the language system, and the influence of social and cultural factors. Bukhara dialect contributes significantly not only to the expansion of the Uzbek language system but also to strengthening the connection between different social groups and regions. The analysis and findings of the study on the inclusion of Bukhara dialect in the Uzbek language corpus have led to several significant conclusions. The research examined the phonetic, morphological, syntactic, lexical, and sociolinguistic characteristics of Bukhara dialect, and through comparisons with other dialects of Uzbek and the standard form, the following key results were drawn:

The phonetic system of Bukhara dialect differs from other Uzbek dialects. For instance, certain words are pronounced differently, such as the word "*qora*" (black), which is pronounced as "*kora*" in Bukhara. Other phonetic changes include the swapping of vowels such as "o" and "u," indicating the dialect's distinctive phonetic characteristics. These differences contribute a new phonetic layer to the Uzbek language corpus.

While the morphological structure of Bukhara dialect largely aligns with standard Uzbek, certain morphological forms are distinct. For example, the dialect uses forms like "*kelo'pti*" or "*kelibdi*" instead of "*kelayapti*" (come). Some morphological forms in the dialect are simplified, which establishes a unique morphological system within the dialect. The syntax of Bukhara dialect follows the general rules of Uzbek but displays some unique structural traits. Sentences in the dialect tend to be simpler and more concise. For example, instead of saying "*Men uyingizga boraman*" (I will go to your house), Bukhara speakers might say "*Men uyga kelaman*" (I will come home). These syntactic simplifications are characteristic of the dialect. The lexical composition of Bukhara dialect significantly differs from that of the standard Uzbek language. Some words used in Bukhara are either not used in other regions or carry different meanings. For example, "*do'st*" (friend) in Bukhara can mean "*helper*" or "*partner*," whereas in other areas, it only means "*friend*". Additionally, certain archaic words, such as "*tepasi*" have been preserved in Bukhara, reflecting elements of Old Uzbek.

The sociolinguistic analysis reveals the differences in the use of the dialect across social groups, including youth, women, and men. The dialect is particularly prevalent among the youth and rural populations, reflecting social stratification. The use of the dialect signifies social identity and regional distinction, with differences observed between urban and rural areas based on social structure and lifestyle. Bukhara dialect was compared with other Uzbek dialects, such as those from Samarkand, Tashkent, and Fergana. The phonetic changes, morphological forms, and lexical features of Bukhara dialect differ from those of other dialects. However, there are



also shared linguistic features across all dialects, which are integrated into the broader Uzbek language system.

The inclusion of Bukhara dialect in the Uzbek language corpus was facilitated through corpus linguistics methods, with new material collected. Statistical analysis of this data highlights the importance of Bukhara dialect in expanding the Uzbek language corpus. When compared with other dialectal features, the dialect's inclusion contributes to the development of new analytical approaches and methods in linguistics.

The results of the research underscore the cultural and social significance of Bukhara dialect in addition to its scientific value. The linguistic features of the dialect and its inclusion in the Uzbek language corpus provide new opportunities for further exploration and deepen the understanding of the evolution of the language. The distinctive lexical, morphological, syntactic, and phonetic features of Bukhara dialect enhance the diversity and richness of the Uzbek language. The inclusion of Bukhara dialect in the Uzbek language corpus is not only scientifically significant but also holds cultural and social value, offering new possibilities for understanding the development of the Uzbek language and its linguistic diversity.

### CONCLUSION

The inclusion of Bukhara dialect into the Uzbek language corpus creates new opportunities in linguistics and significantly contributes to the development of the Uzbek language. The research findings have revealed the unique phonetic, morphological, syntactic, lexical, and sociolinguistic characteristics of Bukhara dialect. Through the dialect, one can observe the regional differences, dialectal features, and changes in the language. The phonetic system, morphological and syntactic forms, as well as lexical characteristics of Bukhara dialect add new diversity and depth to the Uzbek language corpus.

All the identified features, as well as the distinctions and unique characteristics of Bukhara dialect compared to other dialects, allow for a deeper study of the development of the Uzbek language society. Understanding the social role of Bukhara dialect in the language, its unique usage among youth and rural populations, and its acceptance as a form of social identification helps to better understand the language changes.

The preservation of archaisms and dialectal features in Bukhara dialect, particularly the differences in lexical, morphological, and syntactic aspects, strengthens the uniqueness and richness of the Uzbek language. Additionally, the inclusion of Bukhara dialect into the Uzbek language corpus provides a new resource for linguistic research, contributing to a deeper understanding of the diversity of the Uzbek language. In general, the inclusion of Bukhara dialect into the Uzbek language corpus opens the way for new research and methodological approaches in linguistics, dialectology, and linguistic studies. This will enable a broader and more comprehensive study of the Uzbek language and its dialects.

### REFERENCES:

1. Ashirboyev Samixon "O'zbek dialektologiyasi" O'quv qo'llanma. [www.ziyouz.com](http://www.ziyouz.com) kutubxonasi Toshkent-2013. 86b.
2. Sabinabonu, C. (2023). Hozirgi O'zbek Adabiy Tili Va Buxoro Viloyati Shevasi O'rtasidagi Farqlar. *Journal of Innovation, Creativity and Art*, 2(1), 105–107.
3. Sh.Fozilova Hozirgi o'zbek adabiy tili va buxoroning kogon tumani shevasidagi nomutanosibliklar. *Barqarorlik va yetakchi tadqiqotlar onlayn ilmiy-amaliy jurnali* Vol. 5 No. 1/ 2025. 419-423b.