

THE IMPORTANCE OF EMPIRICAL APPROACH IN COLLECTING LINGUISTIC DATA AND CREATING THE STRUCTURE OF BUKHARA DIALECT

Soliyeva Farida Bobir qizi

National University of Uzbekistan named after Mirzo Ulugbek

Faculty of Journalism and Uzbek Philology

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

Abstract: The article analyzes the theoretical and practical importance of studying the phonetic, morphological, and lexical features of the Bukhara dialect through an empirical approach. The research results demonstrate the dialect's relationship with the Uzbek literary language, regional and intergenerational differences, as well as its connection with historical and cultural context. Data collected based on empirical methodology serves as an important source in creating the linguistic structure of the dialect and in deep understanding of regional dialects.

Keywords: Bukhara dialect, phonetics, morphology, lexicon, empirical approach, speech sample, linguistic structure, regional dialect

Introduction: The Bukhara dialect occupies a special place in the Uzbek language dialect system. Its phonetic, lexical, and grammatical features are closely connected with the historical development of the local linguistic environment, ethnolinguistic factors, and regional socio-sociological processes. In the process of scientific analysis of dialect materials, the empirical approach acquires special importance, since data based on concrete facts serves as the primary source for subsequent linguistic research. This approach enables not only the collection of linguistic units specific to the dialect, but also their systematic processing, creation of databases, and scientific classification of regional linguistic features.

Literature Review: S.Choriyeva analyzed linguistic differences between the current Uzbek literary language and the Bukhara regional dialect in phonetic, lexical, and grammatical aspects. By selecting respondents based on various age, gender, occupation, and regional criteria, she enables the identification of intergenerational differences and regional pronunciation features in the dialect. N. Rajabov emphasizes the importance of identifying regional dialects and their phonetic, lexical, and morphological features. Q.Sodiqov demonstrates that historical sources, including Devonu lug'ati turk, serve as the main scientific source in studying the lexical and morphological features of the Bukhara dialect. H.Doniyorov helps achieve deeper understanding of the linguistic and historical context of the dialect through the integration of genealogical and historical information. These studies create theoretical and methodological foundations for identifying differences between dialect and literary language.

In Choriyeva's research, linguistic differences between the current Uzbek literary language and the Bukhara regional dialect have been thoroughly scientifically analyzed. According to the author, studying the phonetic, lexical, and grammatical features of the Bukhara dialect based



on an empirical approach enables accurate description of the current state of the regional speech system.[1] During the research, collection of live speech samples, speakers' pronunciation features, lexical richness, and grammatical structures were analyzed. Additionally, some archaic words found in the Bukhara dialect, specific morphological forms, and pronunciation differences were compared with literary language standards, and the reasons for their origin were scientifically explained. The author identified intergenerational differences in the dialect by selecting respondents based on various age, gender, occupation, and regional criteria, and also demonstrated regional pronunciation differences within Bukhara province. This approach enables systematic collection of linguistic data, their scientific classification, and deep understanding of the relationship between Bukhara dialect and Uzbek literary language. The scientific importance of Choriyeva's research lies in the fact that the author not only describes the phonetic, lexical, and grammatical differences between dialect and literary language, but also explains the factors of their formation. It is shown that some word forms found in the Bukhara dialect are inherited from ancient Uzbek language, while others were formed under the influence of historical-sociolinguistic factors. This approach enables encompassing not only linguistic, but also cultural and historical context in the process of studying the Bukhara dialect.

In his research, Rajabov emphasizes the importance of identifying regional dialects and their phonetic, lexical, and morphological features in studying issues of Uzbek dialectology.[2] According to the author, in the process of studying the Bukhara dialect, through collecting and analyzing speech materials, it becomes possible to identify differences and similarities between contemporary Uzbek literary language and dialect. Rajabov shows that studying dialects based on age, gender, occupation, and regional criteria plays an important role in identifying real features of speech. Additionally, analyzing written and oral sources together enables understanding the historical roots of the dialect and systematic description of regional layers of language.

In Sodiqov's work, historical sources, including Devonu lug'ati turk, are presented as the main scientific source in studying the lexical and morphological features of the Bukhara dialect. According to the author, analyzing ancient dictionaries enables identifying word combinations found in the dialect and their changes in modern pronunciation. When this approach is combined with empirical methodology, it serves as a theoretical foundation for systematic study of the Bukhara dialect and creation of databases.[3] Additionally, Sodiqov's work shows the influence of lexical units from ancient Turkic language on the Bukhara dialect, and it is emphasized that by comparing them with written and oral sources, the historical roots and modern linguistic structure of the dialect can be revealed more precisely.

Doniyorov studies genealogical and historical sources as a means of deeper understanding of the linguistic and historical context of the Bukhara dialect. He shows that the phonetic, lexical, and morphological features of the dialect demonstrate not only its linguistic uniqueness, but also illuminate the mutual influence of regional dialects and the historical development process of the Uzbek language.[4] In Doniyorov's research, the ancient linguistic layers of the dialect and changes in modern pronunciation are analyzed, and the relationship between the language's diachronic development and regional speech system is determined. By integrating genealogical information with phonetic, morphological, and lexical analysis, he enables systematic and scientifically-based description of the linguistic structure of the dialect. When this approach is combined with empirical research results, the position of the Bukhara dialect in the context of



theoretical linguistics, dialectology, and sociolinguistics becomes clearer, and strengthens the methodological foundation for studying regional linguistic layers.

Research Methodology: The main tools of empirical approach in studying the Bukhara dialect are surveys, interviews, and audio recordings. Respondents are selected based on age, gender, occupation, and regional criteria, which enables identification of intergenerational and regional differences. Oral sources express the natural form of speech, while written sources reveal historical development and lexical richness. Collected data is processed with the help of analytical programs and a systematic linguistic base is formed. With the help of modern technologies, phonetic, morphological, and lexical specificities are analyzed in depth. The effectiveness of the empirical approach depends primarily on data collection methodology. In this process, tools such as surveys, interviews, and audio recordings are used as one of the most reliable sources. Through surveys, respondents' linguistic features, vocabulary richness, pronunciation specificities, and the level of use of grammatical structures are determined. The interview method helps observe the respondent's natural speech state during conversation. Audio recordings are considered an important tool for subsequent deep phonetic, prosodic, and lexical analysis of empirical materials. Data collected based on these processes are subsequently processed with the help of analytical programs and organized into a precise system. Respondent selection criteria directly affect the reliability of the research. Factors such as age, gender, occupation, regional affiliation play an important role in this process. Participation of representatives of different age groups enables identification of intergenerational differences in the dialect. Gender differences can be observed in the selection of lexical units and some features in pronunciation. The occupational environment significantly affects vocabulary richness, since representatives of certain professions use specific terms and word combinations in daily communication. The regional criterion covers micro-areas within Bukhara province, since the dialect of one area may differ significantly from another in phonetic or lexical aspects. Thus, by selecting respondents considering various factors, the scientific value of research results is increased by ensuring diversity of empirical materials. Forming a linguistic base is necessary for effective processing of collected data. Here, the combined use of written and oral sources yields high results. Oral sources mainly consist of live speech samples, and they express the natural form of the dialect. Such materials are usually formed based on audio recordings from interviews, conversations, and folklore samples. Written sources consist of historical documents, literary works, press materials, and other linguistic resources, and they help identify the historical development of dialect lexicon and grammar. The harmony of these two types of sources enables more perfect creation of linguistic structure.

Another advantage of the empirical approach in creating the linguistic structure of the Bukhara dialect is related to the possibility of relying on precise statistical data. Collecting data in digital form and creating a special corpus based on them increases the accuracy of research results. With the help of this corpus, phonetic changes, morphological specificities, and lexical richness are analyzed systematically. Modern linguistic technologies, including automatic speech recognition programs, morphological tagging tools for texts, and statistical models, enable deeper study of dialect features. This not only expands existing linguistic knowledge but also creates the possibility of drawing new scientific conclusions.

Analysis and Results: Phonetic specificities, morphological suffixes, and regional lexical units are clearly manifested in the Bukhara dialect. For example, in sentence form: In the sentence "*Men uygaman*" the suffix of the past tense verb is used in a softened form compared

to the literary language. In *"Bachalar bog'ga keldilar"* the plural suffix expresses regional morphological features. As examples of regional lexical units: *"Nana choyni qo'yib bering"*, *"U bacha apasi bilan ertalabdan bozorda yuribdi"* and *"Kichik bachalar hovlida o'ynashyapti"*[5] sentences preserve phonetic softness and lexical specificities. In *"Uyga ketdimku"* the suffix *"-ku"* is used as a regional morphological unit. Through this, phonetic, morphological, and lexical layers of the Bukhara dialect are analyzed systematically and serve as empirical data in creating the linguistic structure of the dialect.

Discussion: The empirical approach enables determining the position of the Bukhara dialect in the linguistic system and its influence on the Uzbek literary language. The research illuminates phonetic and morphological specificities, regional lexical units, and pronunciation differences. At the same time, the obtained data is useful in practical aspects of linguistics, for example, in developing textbooks, language standardization, and creating linguistic corpora. The research also helps in deep understanding of regional linguistic layers, historical roots, and modern development processes.

The relevance of the empirical approach is also explained by determining the position of the Bukhara dialect in the linguistic system and its influence on the Uzbek literary language. Through collecting, classifying, and analyzing dialect materials, regional lexical features, specific aspects in pronunciation, and grammatical structures become clear. Such research creates an important theoretical foundation for linguistics, and is also useful in practical fields—for example, in language standardization, developing textbooks, and expanding the Uzbek language corpus. From this point of view, deep study of the Bukhara dialect has special importance in understanding national language development, its historical roots, and modern development processes. The empirical approach enables not only data collection, but also their scientific systematization, creating linguistic models, and understanding the dynamic features of the dialect. The process of studying regional layers of language is multi-staged and is conducted in interconnection with sciences such as linguistics, sociology, history, and cultural studies. As a result, scientific knowledge about the Bukhara dialect becomes richer, the development prospects of Uzbek linguistics expand, and the possibility of deeper understanding of the national characteristics of the language emerges.

Conclusion: Collecting linguistic data of the Bukhara dialect through empirical approach and systematic processing has important significance in identifying the phonetic, morphological, and lexical features of the dialect. Based on collected data, it becomes possible to create linguistic structure, analyze regional and intergenerational differences, and deeply understand the relationship between dialect and literary language. Research results create an important scientific base for scientifically studying the Bukhara dialect and determining the development prospects of the national language in the field of linguistics.

References:

1. Choriyeva Sabinabonu. "Hozirgi O'zbek Adabiy Tili Va Buxoro Viloyati Shevasi O'rtasidagi Farqlar" *Journal of Innovation, Creativity and Art* Vol. 2, No. 1, 2023. 105-107p.
2. Ражабов Н. «Ўзбек шевашунослиги». «Ўқитувчи», Т.: 1996. – 304 б.
3. Содиқов Қ. Девону луғати турк. Фафур Ғулом номидаги нашриёт-матба ижодий уйи. Т.:2017.– 488 б.
4. Дониёров Х. Ўзбек халқининг шажара ва шевалари. – Тошкент:Фан, 1968. – 28-б.



5. <https://uztrj.uz/wp-content/uploads/2025/02/O%E2%80%98ZBEK-SHEVALARI-TADQIQOTLARI.pdf>