JOURNAL OF ADVANCED LINGUISTIC STUDIES
VOL. 13, NO. 2, JUL-DEC 2026 – PART 2 (ISSN 2231-4075)

Mapping Social Vocabulary in the Uzbek-Tajik Dialects of Eastern Surkhandarya

SAYYORA SAIDOVA
Termez State University, Uzbekistan

INTRODUCTION

The Uzbek language is a Turkic language comprising three dialects, with a long historical tradition linked to the Karluk, Kipchak, and Oghuz dialects. Turaeva emphasizes that the classification of Uzbek dialects into three groups remains a subject of debate, noting that a number of dialects exhibit transitional features, making their assignment to a specific group somewhat arbitrary [1]. The eastern districts of the Surkhandarya region – Denov, Sariosiyo, Uzun, Sho'rchi, and Oltinsoy – serve as a clear example of this process of mixing: in this area, Karluk, Kipchak, and Oghuz dialects coexist, bordering one another and sometimes blending together. Turdialiyev analyzes the Karluk, Kipchak, and mixed dialects of the Denov sub-dialect zone in detail and points out the existence of differences in their phonetic-morphological structures. In his subsequent research, the author notes that the Denov Karluk dialect shares genetic and areal relationships with the dialects of neighboring regions [2]. In his specialized study dedicated to the Karluk dialects of Eastern Surkhandarya, D. Eshankulov scientifically recorded the lexical and morphological distinctive features of this zone [3]. While Babayazov and Ismailov applied a digital approach to dialectological research by creating a methodology for automated analysis of the Oghuz dialect [4], Kholova M.'s work on the dialectal corpus of Surkhandarya dialects is commendable, as the scientist advocates a morpholexic approach to digitizing dialects, which is one of the new approaches in practice that is entering Uzbek linguistics [5]. Mamatismoilova demonstrates historical ethnolinguistic

processes using the example of the Namangan Kipchak dialect, the transition of a dialect from one dialect zone to another, and these studies clearly demonstrate the relevance and necessity of systematic dialectological research for Eastern Surkhandarya [6].

THE MAIN FINDINGS AND RESULTS

The Tajik dialect spoken by the population in the territory of Uzbekistan. In Uzbekistan, especially in its southern regions, Tajik dialects are the native language of many local residents. In his study of lexical-semantic interference in the speech of the Tajik population in the Fergana region, Zokirov demonstrated the depth of the Uzbek-Tajik linguistic connection through empirical data. Mennecier et al. document the language varieties (Turkic or Indo-European) spoken by 88 informants from major ethnic groups in Kyrgyzstan, Tajikistan, and Uzbekistan (Karakalpaks, Kazakhs, Kyrgyz, Tajiks, Uzbeks, and Yagnobis) across 23 studies; they then measure the phonological diversity of 176 words from an expanded Swadesh list using Levenshtein distance (*ā*na-*e*na=1). It notes that, within the Turkic language family, the status of Kazakh and Karakalpak as distinct languages is not supported – suggesting instead the existence of multiple separate Karakalpak varieties – whereas Kyrgyz and Uzbek dialects exhibit a high degree of uniformity and, in areas of interaction, an above-average level of mutual assimilation, frequently characterized by bilingualism and the potential for shifting political borders [7].

Analyzing the relationship between Persian-Tajik and Turkic languages from a historical perspective, Perry notes that a historical-linguistic approach – focusing on the mutual influence resulting from their long-standing coexistence – serves as the most appropriate methodological foundation for understanding the interaction between the Uzbek and Tajik languages [8]. As a continuation of this view, we note that the social and linguistic factors influence the dynamics of the development of bilingualism, using the example of the Usmanov bilingualism phenomenon in the village of Varzik [9].

Language	Country	Region	Place	Latitude	Longitude	Questionnaires
Karakalpak	Uzbekistan	Karakalpakstan	Kokdarya	43,09	58,78	4
			Shege	43,77	59,02	4
Karakalpak			Halqabad	42,94	59,78	3
Kazakh			Raushan	43,04	58,84	4
		Bukhara	Gazli	40,08	63,56	7
Kyrgyz	Kyrgyzstan	Issyk Kul	Tamga	42,16	77,57	2
			Barskoom	42,17	77,64	
		Naryn	Kulanak	41,36	75,50	5
			Akmuz	41,25	76,00	3
Uzbek	Uzbekistan	Andizhan	Orday	40,77	72,31	4
			Soy Mahalla	40,77	72,31	3
		Bukhara	Zarmanak	39,73	64,27	3
			Novmetan	39,73	64,27	
		Karakalpakstan	Hitoy	43,04	58,84	3
	Tajikistan	Penjikent	Urtoqqishloq	39,49	67,54	4
Tajik	Uzbekistan	Bukhara	Zarmanak	39,73	64,27	4
			Novmetan	39,73	64,27	
		Fergana	Kaptarhona	40,25	71,87	5
			Rishtan	40,36	71,28	3
		Samarqand	Kamangaron	39,50	67,27	5
			Agalik	39,54	66,89	
	Tajikistan	Ayni	Shink	39,28	67,81	3
			Urmetan	39,44	68,26	4
		Penjikent	Nushor	39,11	70,86	3
		Tajikabad	Navdî	39,11	70,45	3
			Nimich	39,12	70,67	4
Yagnobi	Tajikistan	Dushanbe	Safedorak	38,57	68,78	3
			Dugova	Next to above	Next to above	2
						= 88

Figure 1. *Measurement of Levenshtein distance between languages*

Contact between Uzbek and Tajik dialects: bilingualism and convergence. In the field of linguistic dialectology, the study of long-term contact between the Uzbek (Turkic) and Tajik (Iranian) languages has emerged as a distinct area of scholarly inquiry. In this regard, Johanson proposes a new conceptual model for the process of language acquisition in Iranian-Turkic language contact, working on the theory of "code-copying." Incorporation – a process in which a speaker introduces a lexeme or suffix from another language into their native language or dialect; for example, when Tajik speakers – during conversations involving numbers – inadvertently use numerals from the Uzbek language: *Tovishqon chan sum buday, ikkiyuzming buday*' (*Quyin qancha ekan, ikki yuzming ekan*). Acceptance, adaptation, the adoption of Tajik elements into one's native language, for example, Uzbek. For example: *Dilbayori das ba kor* (*Ko'ngling Ollohda, qo'ling ishda bo'lsin*). One of the studies serving as a basis for our work is that conducted by Kılıç, who analyzed the sociogenetic factors

involved in the formation of the Uzbek (Turkic) language – specifically, the interplay between social and bio-linguistic factors – and argued for the necessity of studying language development within a broad social context. The scholar noted that significant elements spanning the evolution of Iranian languages – from their ancient structure through Middle Persian to Modern Persian – are also reflected in Turkic languages.

During the Islamic era, the influence of Persian on the Turkic language – a result of linguistic convergence – is clearly evident in its syntax and grammatical structure; for instance, the absence of grammatical gender distinctions and the use of the singular form for nouns following a numeral. At the same time, a similar influence is observed among Tajik-speaking speakers of Uzbek dialects [10]. Beyond the lexical borrowings specific to the oasis, if we consider the dialect area as a whole, it is evident that the Uzbek and Tajik sound systems have converged across a large part of the region. Mennecier et al., evaluating the statistics of words transferred from one language family to another, draw attention to the high quantitative indicators of the intensity of Uzbek-Tajik contact [7]. Hamroyeva emphasizes that digitizing languages involved in bilingualism within bilingualism-based language corpora offers significant opportunities for studying social characteristics [10]. Although the theoretical foundations of contact linguistics have been developed in these studies, the almost complete absence of research on language contact in the field of social lexis in the geographical zones of the Surkhandarya region indicates that there is still a gap in this area that awaits its study.

The history and development trends of linguistic geography and dialect mapping. Research linking dialects with geography began with early studies in Germany (Georg Wenker & Ferdinand Wrede) and later in France (Jileron). Geolinguistics and dialectometry – by the 20th century, dialectometric analysis based on Goebbel's *Atlas linguistique de la France* data proved that isoglosses can be used to more accurately delineate dialect boundaries using quantitative methods, and this method served as the foundation for subsequent mapping studies [11]. That is, as a model experiment in digitizing paper atlases, in the GeoDialect

project of the University of Grenoble, the *Language Atlas of France* (ALF) population centers are geocoded and lexical data are brought into a single model, isoglosses are automatically converted to the Thyssen population area method. That is, as a model experiment in digitizing paper atlases, in the GeoDialect project of the University of Grenoble, the *Language Atlas of France* (ALF) population centers are geocoded and lexical data are brought into a single model, isoglosses are automatically converted to the Thyssen population area method [12]. By 2017, the scholar Montgomery substantiated the combined application of theoretical dialectology and GIS technologies, adopting a comprehensive approach to mapping methodology in dialectology [13]. While dialectological perspectives grounded in a theoretical approach within Uzbek linguistics were advanced by Sh. Shoabdurahmonov, V. Reshetov, Kh. Doniyorov, and F. Abdullayev, the initial practical applications of this approach are evident in various studies by M. Kholova [14-15]. Meanwhile, Mengliev and Abdurakhmonova propose a digital approach and modern NLP methods for analyzing dialectal variations in Uzbek texts. Although these works could serve as a global methodological basis for dialect mapping, no such comprehensive linguistic atlas has yet been compiled for Central Asia, and specifically for the Surkhandarya region.

The literature we have analyzed highlights significant research within Uzbek dialectology regarding phenomena of bilingualism, contact linguistics, and social vocabulary. However, most of these studies rely on general dialect classifications – such as phonetic, morphological, or grammatical approaches – and do not encompass analyses related to contact linguistics or the study of the social lexicon of a specific region. That is, from a real perspective, we have witnessed a lack of focused research dedicated to studying GIS mapping.

The need for a unique linguistic laboratory arises due to the constant interaction of Uzbek and Tajik dialects in the eastern Surkhandarya districts – Denov, Sariosiyo, Uzun – and their coexistence, side by side.

We emphasize that, in this region, there has been no systematic study to date regarding the mapping of social

vocabulary across dialectal groups (Karluk, Kipchak, Oghuz), the tendency for archaic dialectal forms to persist over time, the mutual transfer of lexemes and structures, or processes of convergence. We have tried to evaluate the reflection of social relations in dialects, the influence of bilingualism, and the processes of adoption and transition based on sociolinguistic analysis using the example of just one proverb.

A digital analysis of the proverb *Бо моҳ шини, моҳ шави, бо дег шини, сиёҳ шави* in the Eastern Surkhandarya region.

DATA SOURCE

We conducted our analysis on the areal.dialect.uz platform [16] over the course of one week in 2025, specifically from March 19 to March 26. The survey was titled “Occurrence of Eastern Surkhandarya Proverbs" and consisted of two main sections: the first comprising 10 sub-questions and the second containing 6. A total of 571 people participated in the survey; 555 responses were validated, while errors were observed in 16 cases. Data was collected from 11 districts of the Surkhandarya region. In the questionnaire, informants were asked whether proverbs and some socially significant lexemes are found, not found, or found in altered form in the regions where they live. Two types of platform statistics were used in the analysis. First, data was collected from 11 districts of the Surkhandarya region. No surveys were conducted in the Shoʻrchi, Jarqoʻrgʻon, Angor, and Bandixon districts, as bilingualism was not observed there. The second aspect involved analyses based on six social indicators – dialect, sub-dialect, ethnicity, education, gender, and age – conducted within the Sariosiyo, Uzun, and Denov districts.

It should be noted that the number of responses and the number of respondents in the two sources differ, as the platform focuses on the number of responses. Therefore, the results were interpreted in terms of percentage shares rather than absolute numbers. Differences between dialectal units were determined using Pearson's chi-square test, while the strength of proximity and association was assessed via Cramér's coefficient. In the initial approach, residents of the area were given a questionnaire regarding the use of a Tajik proverb: *Oy bilan yursang oy*

bo'lasan, qozon bilan o'tirsang qarayasan. Naturally, drawing upon the Uzbek people's way of life and universal perspectives, there exists an educational variant of this concept – either in the form of the saying *yaxshi bilan yursang, yetarsan murodga, yomon bilan yursang, qolarsan, uyatga* or its equivalent, *Qozonga yaqin yursang qorasi yuqar*. In Russian, the form *С кем поведёшься, от того и наберёшься* can be observed.

As we observe the proverb, we can see that the proverb, which reflects social attitudes, is formed in almost all countries based on their lexical attitudes. For instance, while in Tajik the concept of social closeness – *hamnishinlik* (sitting together) – and the sphere of interaction are expressed through an opposition (contrasting the moon, representing the positive, with the cooking pot, representing the negative, based on their colors), and in Uzbek proverbs this process is linked to personal character traits, in Russian, this phenomenon is evaluated against the backdrop of human interaction. For this reason, this unit can serve as a convenient indicator for mapping within the region as an element of social vocabulary. In paremiology, proverbs are considered a highly stable yet adaptable layer of language [17; 4] and the fact that the proverb we analyzed appears – either in its exact form or in variations – among the population indicates its widespread use in oral speech.

The use of proverbs in a region, either fully adapted or partially modified, indicates their vitality, and the distribution of such units in bilingual regions depends on the relationship between the language and its constituent dialects and its various factors [3].

Table 1. *Occurrence of the proverb by district (recorded responses)*

District	Not found	Unchanged	Partially changed	Total	Not found, %	Unchanged, %	Partially changed, %
Sariosiyo	108	657	111	876	12,3	75,0	12,7
Uzun	30	503	19	552	5,4	91,1	3,4
Denov	71	391	26	488	14,5	80,1	5,3
Muzrabot	5	84	3	92	5,4	91,3	3,3
Qiziriq	11	54	3	68	16,2	79,4	4,4
Oltinsoy	8	43	1	52	15,4	82,7	1,9
Sherobod	0	44	0	44	0	100	0

Qumqo'rg'on	19	12	1	32	59,4	37,5	3,1
Boysun	0	8	0	8	–	–	–
Termiz city	0	4	0	4	–	–	–
Termiz district	0	3	1	4	–	–	–
Total	252	1803	165	2220	11,4	81,2	7,4

Note: Percentages were not calculated for districts with fewer than 10 responses

Given the large volume of collected data, we compiled the screenshots into a single file and processed them using the Claude AI; the resulting analysis – categorized by district – yielded statistics of $\chi^2 = 160.97$, df = 12, $p < 0.001$, and Cramer's V = 0.19, with strong opposition observed in the Uzun, Sariosiyo, Denov, and Qumqo‘rg‘on districts. In other words, while the figure ranged between 75% and 91% in the first three districts, it stood at 59% in the second group – specifically in Qumqo‘rg‘on. We observe that the result aligns with the region's ethnolinguistic structure, as the area covered by this study comprises the northern foothill zone – characterized by a dense Tajik population – while Qumqo‘rg‘on, where the Kipchak dialect predominates, forms its periphery. It can be said that there is a basis for drawing a paremiological isogloss line between Uzun, Denov, and Sariosiyo, and the abundance of the proverb's "ayni" case in the western districts: Sherabad, Muzrobot, and Boysun indicates the need for additional research (Figure 1).

In our study, we focused on the distribution of social proverbs across various segments in Eastern Surkhandarya and observed the following results. Result 1: An analysis of the proverb *Oy bilan yursang oy bo‘lasan, qozon bilan o’tirsang qorayasan* across three variants – occurring, not occurring, and occurring in a partially modified form – the values are: $\chi^2 = 27.55$; df = 4; $p < 0.001$; V = 0.1. According to estimates, the canonical form (main form) of the proverb in Uzun district was 76.5 percent of the area where it was stably preserved. In the Sariosiyo district, the share of partially changed variants was 20.4 percent, which is the largest share, and we emphasize that this proverb is not disappearing, but rather that new variants are emerging in terms of folklore variation, thereby serving to enrich the language in some sense (Table 3).

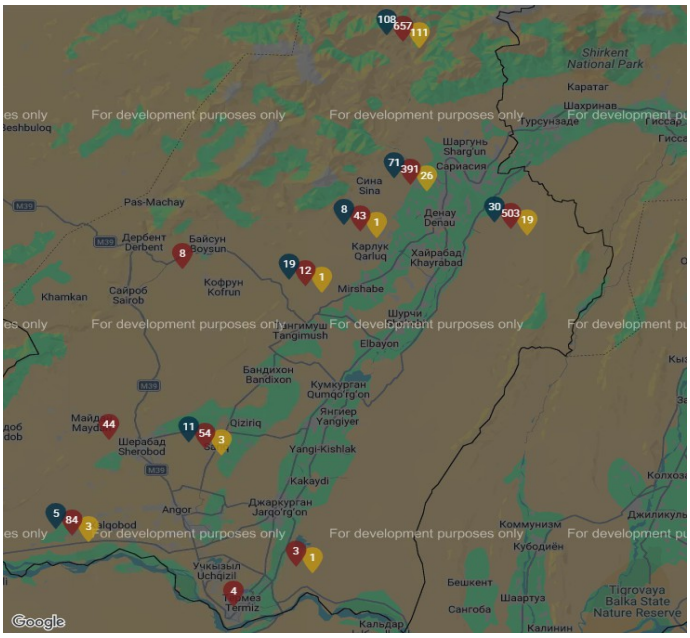


Figure 1. *Occurrence of the proverb by district (map from the areal.dialect.uz platform; blue – not found, red – found in this specific form, yellow – found in a partially modified form)*

Table 3. *Distribution of variants across key districts (responses by indicator)*

District	Responses	Not found, %	In this specific case, %	Partially changed, %
Sariosiyo	393	23,9	55,7	20,4
Uzun	179	13,4	76,5	10,1
Denov	178	23,0	66,3	10,7
Total	750	21,2	63,1	15,7

From a sociolinguistic perspective: Tajiks in the Surkhandarya region who are in regular contact with the Uzbek population, as well as Ethnic settlement patterns and the influence of dialects. Ethnic settlement patterns and various forms of interaction inevitably influence the distribution of dialects and the processes of convergence or integration. According to the survey regarding

the usage of the aforementioned proverb, while 20.4% of the population in Tajik-speaking areas indicated that it is not used, 45% of the population in Uzbek-speaking areas stated that this proverb is not found among them. Such a distinct difference is evident in the example of the city of Denov, which has a mixed population comprising two ethnic groups: $X^2 = 7.35$; $p = 0.025$.

At the same time, the figure for the population belonging to the Kipchak dialect group was 2.5 times higher (47.7% vs. 19.1%), and due to the social situation, this indicator stood at 44.6% in the relevant group as well. It is worth noting that this proverb is currently more characteristic of the spoken language among the Tajik-speaking population, whereas we have observed that equivalents of this proverb are widely prevalent within Uzbek dialectal environments.

CONCLUSION

The proverb *Bo moh shinī, moh shavī, bo deg shinī, siyoh shavī* is actively used in Eastern Surkhandarya; specifically, 81% of the local population reports using it in this exact form, and the area where this paremiological unit is most concentrated comprises the Uzun, Sariosiyo, and Denov districts.

The area of distribution for this saying corresponds to an ethno-linguistic boundary – with Kipchak-dialect speakers in the Qumqo‘rg‘on district forming its periphery – and the primary factor governing its use is ethnic affiliation; while its usage within the dialect is significant, factors such as gender and age have little influence on its prevalence. The observation of variation in the Sariosiyo local dialect – where the proverb appears in an altered form in actual usage – contrasted with the Uzun district, where the canonical form remains stable; consequently, this proverb, which embodies the concept of "companionship" and carries a social dimension, served as an effective indicator for mapping the social lexicon of Uzbek- and Tajik-speaking communities.

REFERENCES

1. Turaeva, R. 2015. Linguistic ambiguities of Uzbek and classification of Uzbek dialects. *Anthropos*, 110/2, 463-476. doi: 10.5771/0257-9774-2015-2-463.
2. Turdialiyev, A. 2023. Some features of the Karluk, Kipchak and mixed dialects of the Denov sub-dialectal zone. Zenodo. doi: 10.5281/zenodo.8417213.
3. Eshankulov, D. Kh. 2025. Description of research on the Qarluq dialects of Surkhandarya. *American Journal of Philological Sciences*, 5/6, 259-264. doi: 10.37547/ajps/volume05issue06-68.
4. Babayazov S., Ismoilov Sh., Madaminov O., Babayazov U., Ruzimova N. & Xajiyeva D. 2025. Oghuz dialect analysis of the Uzbek language: Methodological approach and experimental study. *IEEE XVII International Scientific and Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE)* IEEE 2025. doi: 10.1109/APEIE66761.2025.11289360.
5. Kholova, M. & Bozorova, E. 2025. Areal approach in Uzbek dialectology: A new stage. *Research on Uzbek Folklore and Dialects: Practice, Methodology, and New Approaches: Proceedings of the International Scientific-theoretical Conference*. Tashkent: TSUULL. Available online: <https://conferences.tsuull.uz/index.php/conferences/article/view/320>.
6. Mamatismoilova, X. 2023. Historical ethnolinguistic processes and the national language of the Uzbek people (based on the Namangan Kipchak dialects. *Scientific Journal of the Fergana State University*, 28/3, 73. doi: 10.56292/SJFSU/ vol28_iss3/a73.
7. Mennecier, P., Nerbonne, J., Heyer, E. & Manni, F. 2016. A Central Asian language survey: Collecting data, measuring relatedness and detecting loans. *Language Dynamics and Change*, 6/1, 57-98. doi: 10.1163/22105832-00601015
8. Perry, J. R. 2001. The historical role of Turkish in relation to Persian of Iran. *Iran & the Caucasus*, 5, 193-200. doi: 10.1163/157338401X00224
9. Usmonov, M. & Hoshimova, M. 2025. Bilingualism in dialects (on the example of the village of Varzik). *Research on Uzbek Folklore and Dialects: Practice, Methodology, and New Approaches: Proceedings of the International Scientific-theoretical Conference* (pp. 792-795). Tashkent: TSUULL. doi: 10.52773/tsuull.conf.2025.24.05/SFPT7638.

10. Khamroeva, Sh. M., Muxamedova, S. & Ruziyeva, M. I. 2025. Analysis of social specificity features in bilingual speech corpora. *10th International Conference on Computer Science and Engineering (UBMK) – IEEE, 2025*. (pp. 1602-1606). doi: 10.1109/UBMK67458.2025.11206826.
11. Goebl, H. 2003. Dialectometric views on the data of the Linguistic Atlas of France (ALF): Relations between genetic affinity and language contact. *Géolinguistique*, 8, 71-100.
12. Gally, S., Chauvin, C., Davoine, P.-A., Demolin, D., Contini, M. 2013. GeoDialect: Exploration of geomatic tools for processing and analysing geolinguistic data. *Géolinguistique*, 14, 187-208. doi: 10.4000/geolinguistique.1002.
13. Montgomery, C. 2017. Maps and mapping in (perceptual) dialect geography. In C. Montgomery & E. Moore (Eds.), *Language and a Sense of Place: Studies in Language and Region* (pp. 147-170). Cambridge: Cambridge University Press. doi: 10.1017/9781316162477.009.
14. Kholova, M. A. 2020. Transcription in the corpus of the Uzbek national dialects (on the example of the "j" dialects of Boysun district). *Journal of Critical Reviews*, 7/5, 844-847.
15. Kholova, M. A. 2022. Linguistic Foundations of Compiling the Corpus of Uzbek National Dialects (on the example of the "j" dialects of Boysun district). PhD (Philol.) dissertation.
16. Survey on the occurrence of proverbs in Eastern Surkhandarya March 19-26, 2025. Available online: <https://areal.dialect.uz>.
17. Alimov, T. E. & Kupriyanova, M. E. 2024. Socio-cultural specificity of the etiquette forms of greetings and addresses in the Uzbek and Russian languages. *Philology. Theory & Practice*, 17/11, 3902-3911. doi: 10.30853/phil20240550.

SAYYORA SAIDOVA
TERMEZ STATE UNIVERSITY,
TERMEZ, UZBEKISTAN.