

**LEXICAL RELATIONSHIP BETWEEN KHOREZM REGIONAL DIALECTS AND THE UZBEK LITERARY LANGUAGE: A SOCIOLINGUISTIC AND ETYMOLOGICAL ANALYSIS****Norbayeva Shukurjon Khayitbayevna**

The department of uzbek language

Urgench state university named after Abu Raykhon Beruni

E-mail: [norbayeva1975@mail.ru](mailto:norbayeva1975@mail.ru)

ORCID: 0000-0002-0629-1107

**Abstract**

This article examines the lexical relationship between the regional dialects of the Khorezm oasis — specifically the Urgench (Urg.), Yangibazar (Y.bzr.), Gurlan (Grln.), Khanka (Xnq.), and Khiva (Xv.) subdialects and the contemporary Uzbek literary language. Drawing on an original corpus of approximately 60 dialect lexical items compiled from field research data, the study conducts a systematic etymological, semantic, and functional analysis of dialectal vocabulary that either lacks equivalents in, has been marginalized by, or exists in parallel with, the standard literary norm. The findings demonstrate that Khorezm dialectal lexis constitutes a rich, internally stratified layer of the Uzbek national language, encompassing archaic Turkic roots, Arabisms, Iranisms, indigenous cultural nominations, and onomatopoeia — many of which encode cultural realities absent from the literary standard. The study further argues that systematic integration of productive dialectal lexis can enrich the Uzbek literary language without destabilizing its normative coherence. These findings have practical implications for Uzbek lexicography, language planning, and dialect documentation efforts.

**Keywords**

Khorezm dialects, Uzbek literary language, dialectal lexicology, sociolinguistics, etymology, language norm, Turkic studies.

**1. Introduction**

Dialect lexicology occupies a foundational position within Turkic linguistics, serving as both a record of historical language development and a living repository of cultural knowledge that often escapes codification in standard literary norms. The Uzbek language, spoken by over 35 million people across Central Asia, encompasses a rich dialectal landscape that has been systematically investigated since the early Soviet period, yet the lexical wealth of peripheral dialect zones — particularly those of the Khorezm oasis in northwestern Uzbekistan — remains insufficiently represented in the national lexicographic tradition.

Khorezm, one of the oldest civilizational centers of Central Asia, has historically served as a crossroads of Turkic, Iranian, and Arabic linguistic influences. This multicultural convergence has produced a dialectal lexis of remarkable depth, preserving archaic Turkic vocabulary, culturally specific nominations for agricultural practices, material culture, human behavior, and natural phenomena — elements that have largely been supplanted or neutralized in the standardized Uzbek literary language that evolved primarily around the Tashkent and Fergana urban speech communities.

The relationship between regional dialects and the literary standard constitutes one of the core theoretical problems of Uzbek linguistics. This relationship is neither static nor purely hierarchical; it is characterized by bidirectional influence, selective adoption, semantic transformation, and, in many cases, functional complementarity. Dialectal lexemes may survive in literary usage in altered phonetic form, acquire expanded or narrowed meanings, fall into disfavor as archaisms, or be entirely excluded from normative dictionaries despite active use in their home communities.

The Khorezm subdialects — distributed across administrative districts including Urgench, Yangibazar, Gurlan, Khanka, Khiva, Bagat, Hazorasp, and Shovot — present a particularly compelling case for lexicological investigation. These subdialects share a core phonological and morphological system while exhibiting notable lexical divergence both from one another and from the literary standard. Previous scholarship has addressed phonological features of Khorezm speech (Borovkov, 1954; Shoabdurahmonov, 1962; Doniyorov, 1979), yet detailed lexical-semantic analyses drawing on systematic field data remain comparatively rare in contemporary research.

The present study addresses this gap by analyzing a corpus of dialectal lexical items collected from native speakers across the Khorezm region, examining their semantic structure, etymological provenance, geographic distribution, and functional relationship to the Uzbek literary standard. Three guiding research questions structure the investigation: (1) What semantic domains are most prominently represented in Khorezm dialectal lexis not reflected in the literary norm? (2) What etymological strata underlie this vocabulary? (3) Under what conditions might dialectal lexical items be productively integrated into — or recognized as enriching — the standard literary language?

## 2. Materials and Methods

The primary data source for this study is an original dialect lexicographic corpus comprising approximately 60 lexical entries gathered through systematic field collection across the Khorezm region. Geographic distribution markers attached to each entry — using abbreviations for Urgench (Urg.), Yangibazar (Y.bzr.), Gurlan (Grln.), Khanka (Xnq.), Khiva (Xv./Xiv.), Hazorasp (H.asp.), Bagat (Bg't.), Shovot (Shvt.), Karakalpakstan border areas (Q.kpr.), and Yangiariq (Y.ariq.) — allow for subdialectal mapping of individual lexemes.

Entries follow a standardized format: headword(s) in bold with phonemic variants separated by the double slash (//) convention indicating phonological alternation rather than semantic difference; geographic distribution in parentheses; a concise definitional gloss in Uzbek; and at least one authentic illustrative sentence in the target subdialect. Where items are marked *esk.* (*eskicha*, 'archaic') or *etn.* (*etnografik*, 'ethnographic'), the corpus recognizes their status as specialized lexis requiring particular analytical attention.

The analytical methodology integrates three complementary approaches. First, etymological analysis traces the historical origins of dialectal lexemes by reference to classical Turkic, Old Uzbek, Arabic, and Persian lexicographic sources, as well as comparative Turkic dictionaries. Second, semantic-field analysis groups lexemes according to the cultural-conceptual domains they populate, revealing the distributional patterns of dialectal vocabulary across semantic space. Third, comparative normative analysis systematically evaluates each dialectal item against the corpus of the Explanatory Dictionary of the Uzbek Language (*O'zbek tilining izohli lug'ati*) and related normative references, categorizing the relationship as: (a) equivalent present in the literary language with identical meaning; (b) equivalent present but with semantic shift; (c) partial equivalent (hyperonym or hyponym present); or (d) no equivalent — a lexical gap in the literary standard.

Data authenticity is ensured through the use of genuine illustrative sentences provided by native speakers, which are transcribed in a modified Uzbek Latin orthography sensitive to the phonemic distinctions of Khorezm speech, including vowel harmony variants and consonant alternations. Ethical considerations were observed in accordance with standard sociolinguistic fieldwork protocols; all informants participated voluntarily.

### 3. Results

***Semantic Domain Distribution.*** Analysis of the corpus reveals that Khorezm dialectal lexis is concentrated in seven primary semantic domains, which can be ranked by density of representation in the dataset:

The largest cluster encompasses material culture, craft, and household artifacts. This domain includes terms for traditional clothing accessories, vessels, implements, and textile materials. Examples include *baltirliq* (a strip of thin felt worn around the calf to support boot shafts), *bükläm* (a white cotton fabric used in place of velvet), *bürmä dörmiyän* (a velvet drawstring purse), *dägdä* (a decorated boot with ornamental shaft), *dägdi* (inner boot insulation made of thin felt), and *biltsur* (karakul fur with a bluish-grey underside and whitish top surface). These items carry irreplaceable cultural information about the pre-industrial material culture of the Khorezm oasis and have no direct equivalents in the Uzbek literary standard.

The second major domain covers agriculture, irrigation, and landscape. Khorezm's identity as an ancient oasis civilization dependent on intricate irrigation systems is powerfully encoded in its dialectal vocabulary. The term *batil ariq* denotes a canal that has not carried water for several years (distinguished from *batil* alone, which refers to any abandoned or disused object or land left unirrigated). *Büvät/bögät* names the system of twigs and other materials placed in a channel to impound water for redirection — a precise technical concept lacking a monolexic equivalent in the literary standard. *Daldarya/daldiryä* refers to a baffle or weir placed in a small irrigation channel or furrow to divert water flow. *Biygär* denotes the traditional communal labor practice (hashar) specifically organized for the digging of canals and aqueducts, encoding both the collective labor concept and its specific hydraulic application.

The third cluster relates to human physical appearance, character, and social behavior. Khorezm dialects possess a rich inventory of evaluative lexemes targeting social types. *Badabad* and *biysämmäk/buvsanmaq* both denote a boastful or self-aggrandizing person, but with different connotative emphases: the latter specifically implies arrogance based on reliance on a patron's or parent's power. *Bätçündük* characterizes a sharp-tongued person. *Bädqiliq* labels a person of bad character or malicious disposition. *Dapäng* describes a person of tall, large, imposing stature. *Gämşik* denotes someone with missing front teeth causing a speech impediment and protruding lower jaw — a hyperspecific social label with no literary equivalent.

A fourth domain encompasses natural phenomena and environmental observation. *Düväläñ/dübäläy* refers to a severe sandstorm accompanied by rain and strong wind — a meteorological phenomenon specific to the arid landscape. *Gävüm* designates the dim, uncertain light of dusk before darkness falls, a temporally precise concept. *Doñmaq* describes the physical sensation of cold penetrating the body, causing pain and shivering — a more specific and experiential verb than the literary standard's general terms for feeling cold.

The remaining three domains — toponymic and directional terminology, zoology and agriculture, and onomatopoeic and expressively marked vocabulary — are less densely represented but no less linguistically significant. *Bändärgä/bändirgä* preserves an archaic term for crossroads or a stopping place (cf. classical Uzbek and Ottoman usage). *Gäççi* names the

domestic goat. *Gävdäş* denotes a large bowl or date variety. *Dükürdi/dübürdi* is an expressive onomatopoeia for noise and commotion. *Düsämmä* similarly denotes chaos and disorder.

**Etymological Stratification.** Etymological analysis of the corpus reveals four principal origin layers, consistent with the historically documented linguistic ecology of the Khorezm region:

The archaic Turkic substrate is the largest stratum. Lexemes such as *dal* (a smooth branch of a mulberry tree, also used as a silkworm feeding branch), *da:* (the ten-day pre-cocooning dormancy of the silkworm), *diyr* (millstone), *doñir/toñir* (a tin vessel for oil or fuel), and *dörämaq* (to be born, to emerge, to bring something into existence) represent Turkic lexical forms that either shifted in meaning in the literary language or were replaced by borrowings. The verb *dörämaq* is particularly notable: cognates exist across numerous Turkic languages (cf. Kazakh *myy*, Kyrgyz *мылы*) in the sense of birth and emergence, but the Khorezm form preserves a productive causative-semantic range that has been narrowed in Uzbek literary usage.

The Arabic borrowing stratum is evident in lexemes such as *batil* (from Arabic *bāṭil*, 'void, invalid, abandoned') and *biygär* (from Arabic/Persian *bīgār*, referring to unpaid or compulsory labor). These Arabisms have undergone semantic specialization in Khorezm dialects: *batil* has narrowed from the general 'invalid/void' sense to denote specifically 'disused, untended' objects and land. *Biygär* has similarly narrowed to the specific domain of canal labor, whereas in classical sources it covered compulsory labor more broadly.

The Perso-Tajik stratum contributes lexemes related to textile culture, household economy, and ornamentation. *Darayi* (a handwoven, hand-dyed cloth) derives from Persian *dārā* 'ī. *Dudar* (a feather decoration on a woman's skullcap, or a scarf) has Persian-Turkic hybrid morphology. *Däräçä* (a small door or wicket gate) reflects Persian *daricha*, a diminutive of *dar* ('door'). *Gäşir* (carrot) is a well-documented regional form of Persian *gazhar/gazar* widespread across Central Asian Turkic dialects.

A local innovation and expressive-onomatopoeic stratum includes items that appear to be regionally generated through sound-symbolic processes. *Dübdürüs/düpdürüs* (completely intact, in perfect working order) is a reduplicative intensifier with no clear external etymon. *Dükürdi* and *düsämmä* belong to an expressive-affective register characteristic of colloquial speech.

**Normative Comparison: Lexical Gaps and Functional Parallelisms.** Cross-referencing the corpus against normative Uzbek lexicographic sources reveals three patterns of normative relationship:

Category I — Full lexical gap: The majority of Khorezm dialect items in the corpus (approximately 72%) have no direct equivalent in the Uzbek literary standard. Terms such as *büvät/bögät*, *daldarğa*, *biltsur*, *dägdä*, *dägdä*, *bürmä dörmiyän*, *baltirliq*, *gävgüm*, *gämşik*, and *biygär* in its specialized sense are unrepresented as distinct lexemes in standard dictionaries. This represents a significant lacuna in the normative lexicographic record, particularly for terms encoding agricultural, textile, and social-behavioral knowledge.

Category II — Semantic shift (approximately 18%): Some dialectal items have cognates in the literary language but with meaningfully different semantic scope. *Batil* in literary Uzbek refers broadly to 'invalid, void, null' (e.g., in legal or religious contexts), whereas the Khorezm dialectal usage narrows it specifically to abandoned land and disused material objects, generating the compound *batil ariq* (an unused irrigation canal), a culturally precise syntagma unavailable in the literary register.

Category III — Phonological variant (approximately 10%): Items such as *gāšir* (literary: *sabzi*) and *gāççi* (literary: *echki*) represent phonologically distinct regional forms of semantically equivalent literary items, functioning as full synonyms within the Khorezm speech community but absent from normative dictionaries.

#### 4. Discussion

The findings of this study contribute to three interlocking debates in Uzbek and broader Turkic linguistics: (a) the theoretical framing of dialect–standard relationships; (b) the practical question of dialectal vocabulary integration into the literary norm; and (c) the cultural-linguistic significance of preservation versus standardization.

Regarding theoretical framing, the Khorezm dialectal lexis examined here resists characterization as either a mere 'substandard' deviation from the literary norm or as an entirely separate lexical system. Rather, it constitutes what may be productively described as a complementary lexical stratum: a set of vocabulary items that fulfill referential functions in their speech community that the literary standard has not yet addressed. This perspective aligns with sociolinguistic theories of diglossia (Ferguson, 1959) and heteroglossia (Bakhtin, 1981) as applied to regional dialect situations, while also connecting to more recent frameworks in variational linguistics that reject strict hierarchical models of standard–dialect relations in favor of multidimensional continua (Trudgill, 2004; Milroy & Milroy, 2012).

The lexical gap data are particularly significant from a language planning perspective. If the purpose of a national literary standard is to provide a functional tool for all domains of public and private communication, then the absence of vocabulary for concepts as culturally and practically significant as irrigation management, traditional material culture, and environmental observation represents a real limitation. The historical trajectory of Uzbek literary language standardization — which was shaped by ideological pressures during the Soviet period and by post-independence normative reform — resulted in a standard that privileged urban, written, and formal registers at the cost of peripheral dialectal richness (Schlyter, 1998; Landau & Kellner-Heinkele, 2001). The Khorezm data suggest that this cost has been substantial in lexical terms.

The etymological findings reinforce the theoretical significance of Khorezm dialects as an archive of Turkic language history. The preservation of archaic Turkic vocabulary (*doñmaq*, *dörämaq*, *dal*, *da.*, *diyr*) alongside specialist Arabic borrowings (*batil*, *biygär*) and Perso-Tajik stratum items (*darayi*, *däräçä*, *gāšir*) mirrors the documented linguistic history of Khorezm as a multilingual contact zone. Crucially, the dialect has preserved forms that the literary standard lost during standardization — a process that frequently, though not always deliberately, favors lexical items associated with prestige urban centers over those of peripheral regions.

The practical implications of the normative comparison are significant for Uzbek lexicography. Several items in the corpus — particularly those encoding agricultural practices (*büvät*, *daldarğa*, *batil ariq*), material culture (*dägdä*, *darayi*, *biltsur*), and precise social categorizations (*biysämmäk*, *bätçündük*) — could productively be incorporated into a supplementary dialectal lexicographic dictionary, or indeed flagged as candidates for inclusion in expanded editions of the Explanatory Dictionary with appropriate regional usage labels. The principle of *lexical enrichment* from dialect sources has precedent in the development of other major literary languages: Turkish lexicographic reform in the twentieth century, for instance, drew extensively on Anatolian dialectal vocabulary in the construction of a post-Ottoman standard (Lewis, 1999).

The subdialectal distribution patterns are equally informative. Several items marked as distributed across all or most of the Khorezm subdialects (Urg., Xnq., Y.bzr., Grln., Xv., H.asp.,

Bg't.) may be considered stable pan-Khorezm regional lexemes, whereas items with narrower distributions (e.g., *gävgüm* attested only in Gurlan; *gäläk* only in Khanka) represent more localized micro-dialectal vocabulary. This internal stratification within the Khorezm dialect zone itself suggests that lexicographic documentation must operate at multiple geographic scales simultaneously — a methodological point with implications for ongoing dialect atlas work in Uzbekistan.

The presence of onomatopoeic and expressive vocabulary (*dükürdi*, *düsämmä*, *dübdürüs*) points to an affective-expressive register of Khorezm speech that is closely associated with oral culture and is unlikely to be preserved by literary standardization alone. The documentation of such items is therefore particularly time-sensitive, given the accelerating effects of urbanization, media standardization, and formal education on dialectal speech patterns across Uzbekistan.

## 5. Conclusion

This study has examined the lexical relationship between Khorezm regional dialects and the Uzbek literary language through systematic analysis of an original field corpus. The principal conclusions are as follows:

First, Khorezm dialectal lexis constitutes a semantically rich, etymologically stratified, and culturally irreplaceable component of the broader Uzbek language system. The corpus demonstrates concentration in semantic domains — material culture, irrigation and agriculture, social characterization, environmental observation — that are either underrepresented or entirely unaddressed in the normative literary standard.

Second, the etymological composition of Khorezm dialectal vocabulary — spanning archaic Turkic, Arabic, Perso-Tajik, and locally innovative strata — reflects the historically documented linguistic ecology of the Khorezm oasis and provides independent evidence for contact-linguistic processes across Turkic language history.

Third, normative comparison reveals that approximately 72% of the analyzed items represent genuine lexical gaps in the Uzbek literary standard. This finding has direct implications for Uzbek lexicography and language planning, suggesting the need for expanded dialect-inclusive lexicographic resources and for a more accommodating normative stance toward the productive incorporation of dialectal vocabulary.

Fourth, subdialectal distribution analysis reveals internal stratification within the Khorezm dialect zone itself, necessitating geographically granular documentation approaches in future research.

These findings argue for a reassessment of the relationship between Uzbek dialectal diversity and the national literary standard — one that moves beyond a hierarchical model of correction and suppression toward a complementary model that recognizes dialects as resources for lexical enrichment, cultural documentation, and linguistic identity. Future research directions include: the compilation of a comprehensive annotated Khorezm dialectal dictionary; corpus-based investigation of dialect lexical items in oral and written usage; and comparative analysis with neighboring Karakalpak and Turkmen dialectal systems that share geographical and historical contact with Khorezm Uzbek.

## References

1. Абдуллаев Ф. (1961) Ўзбек тилининг Хоразм шевалари. I қисм. Луғат. II Хоразм шеваларининг таснифи. – Тошкент: Ўзбекистон Фанлар академияси нашриёти.

2. Bakhtin, M. M. (1981). *The dialogic imagination: Four essays*. (C. Emerson & M. Holquist, Trans.). University of Texas Press.
3. Боровков, А. К. (1954). *Узбекские диалекты: Фонетика и морфология*. Институт языкознания АН СССР.
4. Дониёров Х. (1979) *Қипчоқ диалектларининг лексикаси*. – Тошкент.
5. Ferguson, C. A. (1959). Diglossia. *Word*, 15(2), 325–340.
6. Landau, J. M., & Kellner-Heinkele, B. (2001). *Politics of language in the ex-Soviet Muslim states: Azerbaijan, Uzbekistan, Kazakhstan, Kyrgyzstan, Turkmenistan, Tajikistan*. Hurst & Company.
7. Lewis, G. (1999). *The Turkish language reform: A catastrophic success*. Oxford University Press.
8. Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard English* (4th ed.). Routledge.
9. Ражабов, Н. (1996). *Ўзбек шевашунослиги*. – Тошкент.
10. *O'zbek tilining izohli lug'ati*. (2006–2008). (Vols. 1–5). O'zbekiston Milliy Ensiklopediyasi.
11. Schlyter, B. N. (1998). Language policy reforms and social change in Uzbekistan. In B. N. Schlyter (Ed.), *Language change in Central Asia* (pp. 132–156). Mouton de Gruyter.
12. Шоабдурахмонов, Ш. (1962). *Ўзбек адабий тили ва халқ шевалари*. – Тошкент.
13. Trudgill, P. (2004). *Dialects* (2nd ed.). Routledge.