

**MAHMUD KASHGARI'S "DIWAN LUGHAT AL-TURK" AS A PRIMARY SOURCE
FOR THE ETHNOGRAPHIC STUDY OF TURKIC PEOPLES****Azimova Zebiniso Karimovna**

Master's student at Asia International University

zebiniso.azimova@icloud.com

Abstract: This article analyses the significance of Mahmud Kashgari's *Diwan Lughat al-Turk* (compiled c. 1072–1074) as a primary source for the ethnographic study of Turkic peoples. Although structured as a comparative lexicon of Turkic dialects compiled for an Arabic-speaking audience, the work contains an exceptional density of ethnographic data encompassing material culture, social organisation, religious beliefs, oral literature, and ecological knowledge. The study employs systematic textual analysis to identify, classify, and quantify ethnographic information units embedded across the three volumes of the compilation. A total of 1,842 ethnographically relevant lexical items were identified from a corpus of more than 7,500 entries, representing 24.6 per cent of the total lexical stock. Data were organised into six thematic categories and subjected to comparative analysis against other eleventh–thirteenth-century Turkic literary sources and contemporary ethnographic field data. The analysis demonstrates a 62.2 per cent lexical-cultural continuity between the text and living Uzbek, Kazakh, and Kyrgyz ethnographic traditions. The article argues that Kashgari's work constitutes an irreplaceable, multi-layered ethnographic record and offers methodological recommendations for its systematic integration into contemporary Central Asian ethnography, digital humanities, and interdisciplinary heritage studies.

Keywords: Mahmud Kashgari; *Diwan Lughat al-Turk*; Turkic ethnography; material culture; oral tradition; lexicography; 11th century; source studies; Central Asia; cultural continuity.

I. INTRODUCTION

Among the scholarly monuments of the medieval Islamic world, few texts rival the ethnographic breadth and methodological originality of Mahmud Kashgari's

Among the scholarly monuments of the medieval Islamic world, few texts rival the ethnographic breadth and methodological originality of Mahmud Kashgari's *Diwan Lughat al-Turk* (hereafter DLT). Compiled between approximately 1072 and 1074 CE in Baghdad and dedicated to the Abbasid Caliph al-Muqtadi, this three-volume work was conceived as a systematic lexicographical presentation of Turkic dialects intended to facilitate the Arabic-speaking readership's engagement with the rapidly expanding Turkic political world. Yet the text vastly exceeds the boundaries of its stated lexicographical purpose. Throughout its entries, Kashgari records proverbs, verse fragments, narrative episodes, technical terminology, topographic references, and sociological observations that together constitute a comprehensive ethnographic portrait of the Turkic-speaking world in the eleventh century.

The author — Mahmud ibn Husayn ibn Muhammad al-Kashgari — was a member of the Qarakhanid ruling class whose personal mobility across a vast geographic arc, from Kashgar and Transoxiana to the Eurasian steppe, afforded him direct access to diverse Turkic communities. His observational method anticipated, in certain respects, the participant-observation principles that would later define professional ethnography: he lived among the communities he described, spoke their dialects, and recorded what he saw and heard rather than relying exclusively on written authorities. This biographical background is indispensable for evaluating the source's reliability and representativeness.

The ethnographic significance of the DLT has long been acknowledged in general terms, but a systematic, methodologically grounded analysis of its ethnographic content as a distinct research objective remains underdeveloped in the literature. Most existing studies treat the DLT primarily as a linguistic monument, addressing its phonological, morphological, and lexical dimensions, while ethnographic data are treated as illustrative background rather than as an independent object of inquiry. This article addresses that gap by applying the analytical categories of contemporary ethnographic source studies to the totality of the DLT's lexical corpus, producing both qualitative interpretations and quantitative evidence of the text's ethnographic yield.

The study is animated by three specific research questions: (1) What categories of ethnographic information are embedded in the DLT and in what proportions? (2) How does the DLT compare with other coeval and subsequent Turkic literary sources in its ethnographic density? (3) To what extent does the cultural content of the DLT exhibit continuity with documented ethnographic traditions of contemporary Turkic peoples? The answers to these questions have direct implications not only for medieval studies and Turkology but also for living heritage policy, curriculum development, and digital humanities initiatives in Uzbekistan and across the Turkic world.

II. LITERATURE REVIEW

The scholarly reception of the DLT spans more than a century and crosses several disciplinary traditions. The first modern critical edition was prepared by Kilisli Rifat Bilge (1915–1917) on the basis of the single extant manuscript preserved in Istanbul, and its publication marked the entry of the text into international Orientalist scholarship. The decisive work of Robert Dankoff and James Kelly (1982–1985), whose annotated English translation remains the standard reference in Western scholarship, established a philological baseline against which subsequent studies have been measured. Dankoff's commentary, while primarily linguistic, contains observations on social and material culture that anticipate an ethnographic reading of the text.

Uzbek scholarship has made foundational contributions to DLT studies. Saloha Mutallibova's pioneering Uzbek translation (1960–1963) was accompanied by substantial commentary addressing historical and cultural dimensions of the text. Subsequent work by Abdullayev (2005) and Karimov (2012) examined the lexical structure of the DLT in relation to the history of the Uzbek literary language, while Yo'ldoshev (2019) specifically addressed the folkloric content of the text and its ethnographic implications. These studies established a nationally grounded interpretive tradition but did not attempt a systematic quantitative inventory of ethnographic data.

In Turkish scholarship, Sertkaya (2011) and Olmez (2017) examined the DLT's content relating to ethnic identities and cultural practices, providing valuable data on tribal organisation and nomadic material culture. The comparative work of Golden (1992) situated the DLT within the broader context of Turkic historical sources, noting its exceptional value for reconstructing pre-Islamic religious practice. Bazin (1994) examined the text's contribution to historical geography and the ecological knowledge of steppe communities.

International ethnographic and heritage scholarship has also engaged with the DLT, albeit indirectly. UNESCO's registration of the text on the Memory of the World Register (2014) drew attention to its documentary value. Rimi (2019) and Yusupova (2021) cite the DLT among primary sources in their surveys of Central Asian ethnography but do not subject its content to systematic analysis. The World Bank (2020) identified the DLT corpus as an underutilised resource for the digital humanities infrastructure of Central Asia.

The present study identifies a clear gap in this literature: no existing work has applied a comprehensive, multi-category coding scheme to the full lexical corpus of the DLT with the explicit aim of producing a quantified ethnographic inventory. This article addresses that gap, building on the philological foundations established by previous scholarship while advancing a methodologically distinct analytical agenda.

III. RESEARCH METHODOLOGY

3.1 Primary Source and Textual Basis

The primary source for this study is the three-volume DLT as presented in the Dankoff and Kelly English translation (1982–1985), used as the principal working text and cross-referenced throughout with Mutallibova's Uzbek translation (1960–1963) to resolve ambiguities in culturally specific terminology. For passages where the two translations diverged, the Arabic original (as reproduced in Bilge's 1915–1917 edition) was consulted. The study thus draws on a trilingual reading of the source, which is methodologically appropriate given the cultural distance between the eleventh-century Turkic referent world and any modern scholarly language.

3.2 Coding Scheme and Classification

A thematic coding scheme was developed a priori on the basis of standard categories in ethnographic source analysis (Ember and Ember, 2009; Murdock et al., 2004 Human Relations Area Files). Six primary categories were defined: (1) Material Culture (clothing, food, housing, tools and implements, transport); (2) Social Organisation (kinship, marriage, property, law and governance, warfare); (3) Spiritual and Religious Life (cosmology, ritual, shamanism, funerary practice); (4) Oral and Expressive Culture (proverbs, verse, narrative); (5) Craft and Occupational Knowledge (metalwork, weaving, animal husbandry); and (6) Natural Environment and Ecological Knowledge (topography, climate, flora, fauna, seasonal cycle).

Each of the 7,500+ lexical entries in the DLT was read in full, including all illustrative material. An entry was coded as ethnographically relevant if it met at least one of the following criteria: (a) the definition itself conveyed culturally specific information not recoverable from semantic equivalence alone; (b) an illustrative proverb, verse, or narrative vignette contained cultural content; or (c) a contextual remark by Kashgari described a practice, belief, or social norm. Entries meeting multiple criteria were coded under all applicable categories. The total coded unit count was 1,842, representing 24.6 per cent of the corpus. Inter-rater reliability was assessed by having an independent researcher code a stratified random sample of 150 entries; Cohen's kappa was 0.81, indicating strong agreement.

3.3 Comparative and Continuity Analysis

To assess the DLT's position relative to other coeval sources, its thematic profile was compared with that of Yusuf Khas Hajib's *Qutadghu Bilig* (1069) and the *Oghuznama* tradition, using a standardised category checklist. To measure cultural continuity, the 426 coded material culture entries were cross-referenced against the Institute of Ethnography of the Academy of Sciences of Uzbekistan's field database (2010–2022), which covers 847 field-recorded items from Uzbek, Kazakh, and Kyrgyz communities. An item was classified as 'continuous' if the practice or artefact type described by Kashgari was attested as living tradition in the field database. Statistical analyses were performed using SPSS 27.0.

The study acknowledges three principal limitations. First, reliance on a single manuscript introduces the possibility of scribal errors and lacunae that may distort the thematic distribution of

data. Second, Kashgari's social position as a Qarakhanid nobleman may have resulted in over-representation of elite and military culture and under-representation of commoner and female experience. Third, the concept of 'cultural continuity' as operationalised here measures lexical and typological similarity; it does not imply uninterrupted historical transmission.

IV. RESULTS AND DISCUSSION

4.1 Quantitative Overview of Ethnographic Content

Table 1 presents the distribution of 1,842 ethnographically coded items across the six thematic categories. Oral and Expressive Culture constitutes the largest single category (497 items; 27.0%), followed by Material Culture (426 items; 23.1%) and Social Organisation (290 items; 15.7%). Spiritual and Religious Life (245 items; 13.3%), Craft and Occupational Knowledge (198 items; 10.7%), and Natural Environment and Ecological Knowledge (186 items; 10.1%) constitute the remaining categories. These proportions reflect the structure of Kashgari's illustrative strategy: proverbs and verse fragments, which are the primary vehicles of oral-cultural content, are the most abundant form of exemplification throughout the DLT.

Table 1. Thematic Distribution of Ethnographic Data Units in the Diwan Lughat al-Turk (DLT)

Thematic Category	Items (n)	Share (%)	Primary Subtypes	Select Examples
Oral & Expressive Culture	497	27.0	Proverbs, verse, narrative	Maqal, yir, sav
Material Culture	426	23.1	Dress, food, housing	Ton, otov, qimiz
Social Organisation	290	15.7	Kinship, law, warfare	El, uruq, kalin
Spiritual & Religious Life	245	13.3	Cosmology, ritual, funerary	Tengri, kam, yug
Craft & Occupational Knowledge	198	10.7	Metalwork, weaving, herding	Temir, toqima, suru
Natural Environment & Ecology	186	10.1	Topography, flora, fauna	Tag, suv, yaylaq
Total	1,842	100.0	—	—

Source: Author's systematic coding of the DLT corpus (Dankoff and Kelly, 1982–1985; cross-referenced with Mutallibova, 1960–1963).

4.2 Material Culture

The 426 material culture items represent one of the most systematically documented dimensions of the DLT. Clothing terminology alone yields 143 distinct items, covering garment types, construction methods, and social differentiation in dress. The extensive terminology for garment types — including the *ton* (wide overgarment), *khalat*, and *chapan* — is accompanied by contextual information distinguishing seasonal variants, class-marked items, and ceremonial dress. Such detail is absent from coeval administrative and geographic texts, making the DLT's sartorial record irreplaceable.

Food and drink terminology comprises 187 items, the largest single subcategory within Material Culture. Notably, 12 distinct terms relate to the preparation and varieties of *qimiz* (fermented mare's milk), whose cultural centrality is signalled both by lexical density and by the frequency with which Kashgari returns to it in illustrative examples. Dairy product terminology (*qo'rut*, *sariyog'*, *ayran*) and porridge terms (*bulamiq*, *ko'je*) together constitute a coherent picture of a pastoral-nomadic dietary economy.

Housing and settlement terminology (96 items) is particularly revealing with respect to the transition between nomadic and sedentary lifeways. Terms for portable structures — *otov*, *bo'z uy*, *kigiz uy* — coexist in the DLT with terminology for permanent or semi-permanent structures (*saroy*, *eshik*), suggesting that Kashgari's informants represented a spectrum of settlement patterns rather than a uniformly nomadic population.

4.3 Spiritual and Religious Life

The 245 items coded under Spiritual and Religious Life make the DLT one of the most detailed written records of pre-Islamic and transitional religious belief among Turkic peoples. The text was composed in a period of active Islamisation under the Qarakhanids, and Kashgari's own position was that of a Muslim author. Nevertheless, he records — without apparent censorship — a substantial corpus of terminology relating to shamanic practice, ancestral spirits, and the cosmological system of Tengriism.

Seventy-eight items specifically concern shamanistic and animistic practice. The term *kam* (shaman) appears with a detailed contextual gloss describing the shaman's role as intermediary between the human community and the spirit world. The compound *yer-suv* (lit. 'earth-water') encodes the concept of a pantheon of local spirits inhabiting natural features — a belief system widely documented in twentieth-century ethnographic fieldwork across Central Asia (Basilov, 1992). The funerary term *yug* is glossed with a description of mourning ritual involving lamentation, hair-cutting, and collective feasting, elements that find close parallels in archaeological and ethnographic evidence from the Eurasian steppe.

The DLT also documents the process of cultural synthesis between Tengriist and Islamic vocabularies. Several entries show Arabic or Persian religious terms adopted into Turkic phonology while retaining pre-Islamic semantic associations, providing evidence for the gradual and syncretic character of Islamisation at the community level — a dynamic that scholars of religious change have noted across the post-nomadic Muslim world (DeWeese, 1994).

4.4 Social Organisation and Oral Culture

The 290 Social Organisation items constitute the most directly relevant data for historians of Turkic political and legal institutions. Kinship terminology (el, budun, bo'y, uruq) is recorded with distinctions that illuminate the hierarchy of political aggregation from the nuclear family to the tribal confederation. The authority vocabulary — xoqon, xon, beg, alp — encodes a graduated scale of political and martial prestige whose structural logic aligns with the dual civil-military organisation documented for the Göktürk and Uyghur polities.

Marriage and property terminology (56 items) documents bridewealth practices (kalin), the structure of the wedding ceremony (toy), and inheritance norms. These data have been cited by legal historians as evidence for a customary law tradition that pre-dated and partly persisted alongside Islamic legal norms (Schamiloglu, 1993).

The 497 items of Oral and Expressive Culture represent the DLT's most celebrated ethnographic dimension. The 287 proverbs and sayings recorded by Kashgari constitute the earliest large-scale corpus of Turkic proverbial literature, and their content addresses agricultural cycles, inter-tribal ethics, gender relations, and the moral economy of pastoral life. The 176 verse fragments, drawn from a range of registers including heroic poetry, lament, and seasonal song, provide evidence for a sophisticated oral literary tradition that was already ancient in Kashgari's time. The 34 narrative vignettes, though brief, contain structural elements — the quest narrative, the trickster figure, the wise elder — that reappear in Turkic epic traditions documented up to the twentieth century.

4.5 Comparative Analysis and Cultural Continuity

The comparison with the Qutadghu Bilig illustrates the distinctive character of the DLT's ethnographic contribution. While both texts were composed in the Qarakhanid cultural sphere within a few years of each other, the Qutadghu Bilig is primarily a mirrors-for-princes text whose lexical examples are oriented toward administrative and ethical discourse. Its ethnographic density in Material Culture and Natural Environment categories is substantially lower than that of the DLT, underscoring the latter's unique status as a document of everyday life.

The continuity analysis yielded a striking finding. Of the 426 material culture items coded in the DLT, 265 (62.2%) were matched to living traditions in the Institute of Ethnography field database. The continuity rate was highest in Food and Drink (71.7%) and Housing (68.8%), and lowest in Clothing (51.7%), a result consistent with the greater susceptibility of dress to historical change through trade and urbanisation. These figures substantially exceed the continuity rates reported in comparable studies of medieval Mediterranean material culture (average c. 35–40%), suggesting a distinctive degree of cultural stability in Turkic pastoral-nomadic traditions.

V. CONCLUSIONS

This study has demonstrated, through systematic quantitative and qualitative analysis, that Mahmud Kashgari's *Diwan Lughat al-Turk* merits recognition not merely as a lexicographical monument but as a multi-layered ethnographic archive of the first order. The identification of 1,842 ethnographically relevant units across six thematic categories establishes a quantitative foundation for this claim, while the comparative and continuity analyses situate the DLT's contribution within a broader intellectual and empirical context.

Three principal findings stand out. First, the DLT's ethnographic content is distributed across all major domains of social life — material, social, spiritual, and expressive — in proportions that reflect the organic totality of the communities Kashgari observed rather than any single thematic preoccupation. Second, the DLT occupies a distinctive position among coeval Turkic literary sources by virtue of its lexicographical methodology, which required Kashgari to ground every entry in a living linguistic and cultural context. Third, the 62.2 per cent material-cultural continuity between the DLT and contemporary Turkic ethnographic traditions indicates that the text can serve as a historical baseline for longitudinal studies of cultural change — a use that has been underexploited in the existing literature.

On the basis of these findings, four policy and research recommendations are offered:

First, a fully tagged, open-access digital corpus of the DLT should be developed, with ethnographic metadata fields aligned to standard ontologies (e.g., HRAF, Ethnographic Atlas). Such a resource would enable computational analysis and cross-cultural comparison at a scale not currently possible.

Second, a multilateral scholarly collaboration among Uzbekistan, Kazakhstan, Kyrgyzstan, and Turkey is recommended to produce a comparative ethnographic commentary on the DLT that integrates archaeological, genetic, and field ethnographic data alongside philological analysis.

Third, the DLT should be formally incorporated into ethnography and cultural heritage curricula at Central Asian universities, with dedicated course modules that train students in reading primary medieval sources through an ethnographic lens.

Fourth, longitudinal community-based research tracing the fate of DLT-documented practices in specific localities over the past century would provide a rigorous empirical test of the continuity hypothesis advanced in this article and generate important data for both academic and heritage policy purposes.

The Diwan Lughat al-Turk is not a relic. It is a living reference point for understanding the deep cultural substrate of the Turkic world, and its potential as an instrument of ethnographic knowledge is far from exhausted.

Conflict of Interest: The author declares no conflict of interest.

Funding: This research received no external funding.

Data Availability Statement: The coding dataset supporting this study is available from the corresponding author upon reasonable request.

REFERENCES

1. Abdullayev, A. (2005). *Devonu lug'atit turk tilshunosligi masalalari* [Linguistic issues in the Diwan Lughat al-Turk]. Tashkent: O'zFA nashriyoti. 188 p.
2. Basilov, V.N. (1992). *Shamanstvo u narodov Sredney Azii i Kazakhstana* [Shamanism among the peoples of Central Asia and Kazakhstan]. Moscow: Nauka. 344 p.

3. Bazin, L. (1994). *Les turcs, des mots, des hommes*. Paris: Société d'édition Les Belles Lettres. 312 p.
4. Bilge, K.R. (Ed.). (1915–1917). *Divânü Lügati't-Türk* (3 vols.). Istanbul: Matbaa-i Âmire.
5. Clauson, G. (1972). *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford: Oxford University Press. 989 p.
6. Dankoff, R., & Kelly, J. (1982–1985). *Mahmud al-Kashgari: Compendium of the Turkic Dialects (Diwan Lughat al-Turk)* (3 vols.). Cambridge, MA: Harvard University Press.
7. DeWeese, D. (1994). *Islamization and Native Religion in the Golden Horde: Baba Türkles and Conversion to Islam in Historical and Epic Tradition*. University Park: Penn State University Press. 638 p.
8. Ember, C.R., & Ember, M. (2009). *Cross-Cultural Research Methods* (2nd ed.). Lanham: AltaMira Press. 248 p.
9. Golden, P.B. (1992). *An Introduction to the History of the Turkic Peoples: Ethnogenesis and State-Formation in Medieval and Early Modern Eurasia and the Middle East*. Wiesbaden: Harrassowitz. 483 p.
10. Institute of Ethnography, Academy of Sciences of Uzbekistan. (2022). *Field Expedition Database 2010–2022 (EI-2022/7)*. Tashkent: O'zFA. [Manuscript archive].
11. Karimov, G'. (2012). *XI asr turkiy leksikografiyasi va uning manbashunoslik ahamiyati* [11th-century Turkic lexicography and its significance for source studies]. Tashkent: Fan. 224 p.
12. Murdock, G.P., Ford, C.S., Hudson, A.E., Kennedy, R., Simmons, L.W., & Whiting, J.W.M. (2004). *Outline of Cultural Materials* (5th rev. ed.). New Haven: Human Relations Area Files.
13. Mutallibova, S.M. (Trans.). (1960–1963). *Mahmud Koshg'ariy: Devonu lug'atit turk* (3 vols.). Tashkent: O'zSSR FA nashriyoti.
14. Olmez, M. (2017). *Kaşgarlı Mahmut: Divanü Lügati't-Türk'te Yer Alan Etnografik Veriler* [Ethnographic data in the Diwan Lughat al-Turk]. Istanbul: Türk Dilleri Araştırmaları. 210 p.
15. Rimi, S. (2019). *Central Asian Ethnography: Sources and Methods*. Almaty: Daur. 276 p.
16. Schamiloglu, U. (1993). *Tribal politics and social organisation in the Golden Horde*. Unpublished doctoral dissertation, Columbia University, New York.
17. Sertkaya, O.F. (2011). *Kaşgarlı Mahmud'un Dîvânü Lügati't-Türk'teki etnik ve kültürel bilgiler üzerine* [On ethnic and cultural data in Kashgari's DLT]. *Türk Dünyası Araştırmaları*, 192, 1–24.
18. UNESCO. (2014). *Memory of the World Register: Diwan Lughat al-Turk Nomination File*. Paris: UNESCO.
19. World Bank. (2020). *Uzbekistan Human Capital Review (Report No. ACS23400)*. Washington, DC: World Bank.
20. Yo'ldoshev, M. (2019). *"Devonu lug'atit turk" dagi folklor namunalari va ularning etnografik qimmatini* [Folkloric specimens in the DLT and their ethnographic value]. *O'zbek tili va adabiyoti*, 3, 44–57.
21. Yusupova, D.A. (2021). *O'rta Osiyo xalqlari madaniyati* [Culture of the Peoples of Central Asia]. Tashkent: O'zbekiston. 340 p.