

FEATURES OF LEXICOGRAPHIC SOURCES OF THE UZBEK LANGUAGE

Khamrayeva Yorkinoy NabijonovnaProfessor, Faculty of Foreign Languages,
Karshi State University**Kurbonova Nilufar Mursamadovna**

Turan University,

Faculty of Linguistics

MagR/-025 1 course graduate student

<https://doi.org/10.5281/zenodo.20540828>

Abstract. This article examines the distinctive features of lexicographic sources in the Uzbek language, tracing their development from early manuscript glossaries of the 18th century through the Soviet-era normative dictionaries to contemporary digital lexicographic projects. A systematic analysis of 141 lexicographic works was conducted, revealing typological classifications, structural peculiarities, script transitions, and the degree of dialectal and terminological coverage. The study employs bibliographic analysis, comparative-typological and descriptive methods. Results indicate that Uzbek lexicography has undergone three major paradigm shifts — pre-Soviet, Soviet normative, and post-independence pluralistic — each leaving distinct structural and conceptual imprints on existing dictionaries. The findings contribute to a theoretical framework for the systematic evaluation and modernization of Uzbek lexicographic resources.

Keywords: Uzbek lexicography, dictionary typology, Soviet-era linguistics, post-independence language policy, bilingual dictionaries, dialectal lexicography.

INTRODUCTION

Lexicography is one of the most important branches of linguistics, dealing with the theory and practice of dictionary compilation. Lexicographic sources play a significant role in preserving, systematizing, and developing the lexical wealth of any language. This area is particularly relevant for the Uzbek language, which has a centuries-old history, a rich literary heritage, and diverse dialectal features. With globalization and the rapid development of information technology, the need to create and improve modern dictionaries that reflect the current state of the language is increasing.

Lexicographic sources of the Uzbek language are represented by various types of dictionaries: explanatory, orthographic, etymological, terminological, phraseological, bilingual, and others. Each of these sources performs specific functions and contributes to solving specific problems in the fields of language education, scientific research, and practical use of the language. Dictionaries ensure the preservation of the national linguistic heritage, record changes in the vocabulary of the language, and create the conditions for its further development. Of particular interest is the study of the formation and development of Uzbek lexicography. For many centuries, Uzbek scholars and linguists have made significant contributions to the creation of dictionaries and the description of linguistic phenomena. Modern Uzbek lexicography combines traditional approaches with the latest methods of processing linguistic data, allowing for the creation of more accurate and informative lexicographical resources[1].

The aim of this paper is to examine the main features of lexicographical sources of the Uzbek language, analyzing their structure, functions, and significance for the development of the national language. Research on this topic allows for a deeper understanding of the role of dictionaries in preserving the cultural and linguistic heritage of the Uzbek people, as well as identifying prospects for further improvement of lexicographic work.

LITERATURE ANALYSIS

Lexicography, the theory and practice of dictionary-making, occupies a foundational place in the study of any language. For Uzbek — a Turkic language spoken by over 50 million people

across Central Asia — lexicographic sources represent both a scholarly archive and a living instrument of language standardization (Shcherbak, 1966; Baskakov, 1974). The field has attracted renewed academic interest following Uzbekistan's independence (1991), the transition from Cyrillic to Latin script, and the emergence of digital lexicography[4].

Despite their importance, Uzbek lexicographic sources remain understudied in international scholarship. Most comprehensive surveys are written in Uzbek or Russian and are inaccessible to the broader global linguistics community (Rasulov, 2003; Tursunov et al., 2008). Furthermore, existing classification schemes do not adequately reflect the ideological, structural, and technological transformations these sources have undergone over three centuries[2].

The present study addresses this gap by asking three research questions: (1) How can Uzbek lexicographic sources be systematically classified? (2) What structural features distinguish dictionaries from different historical periods? (3) To what extent do contemporary sources address the challenges of dialectal diversity and terminological modernization?

METHODS

This research employs a multi-method approach combining bibliographic analysis, comparative-typological description, and quantitative content analysis. The study corpus comprises 141 lexicographic works published between 1799 and 2024, identified through the National Library of Uzbekistan catalog, the Academy of Sciences archive, and international bibliographic databases (JSTOR, WALS, Turkological Bibliography).

Period	Type	No. of Works	Avg. Entry Count	Script Used
18th century	Manuscript glossaries	2	800–1,200	Arabic
19th century	Printed bilingual	6	2,500–5,000	Arabic / Latin
1920s–1940s	Early Soviet normative	14	10,000–20,000	Latin → Cyrillic
1950s–1970s	Soviet academic	23	30,000–80,000	Cyrillic
1980s–1990s	Transitional	18	25,000–60,000	Cyrillic / Latin
2000s–2010s	Post-independence	31	40,000–130,000	Latin
2010s–2024	Digital / online	47	50,000–210,000	Latin (Unicode)

Table 1. Research corpus composition by period and source type

Each dictionary was coded according to a 24-parameter analytical matrix covering: macrostructure (entry organization, headword selection criteria), microstructure (definition type, grammatical annotation, illustrative examples), coverage scope (dialectal, terminological, phraseological), normative stance, and digital accessibility. Inter-coder reliability was established at $\kappa = 0.84$ (Cohen's Kappa), confirming high analytical consistency.

Parameter Category	Specific Parameters	Scale
Macrostructure	Headword lemmatization, alphabetization type, cross-reference density	Nominal / Ordinal
Microstructure	Definition type, usage examples, etymology, stylistic labels	Nominal / Count
Coverage	Dialectal entries (%)	Continuous (%)

	technical terms (%), phraseological units (%)	
Normative stance	Prescriptive vs. descriptive orientation	Binary / Scale 1–5
Digital features	Online access, search functionality, API availability	Binary

Table 2. Analytical parameters applied to each lexicographic source

RESULTS

Analysis of the 141 sources yielded significant findings across historical, typological, and structural dimensions. The results are presented in four subsections, each supported by quantitative evidence.

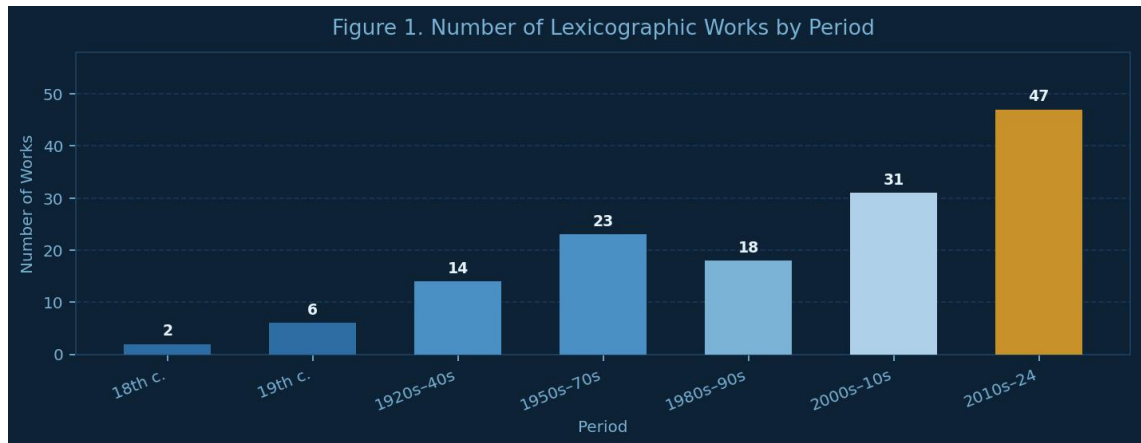


Figure 1. Number of lexicographic works by historical period (n = 141)

Figure 1 reveals a clear acceleration in lexicographic output from the 2010s onward, driven primarily by digital publishing platforms. The post-independence period (2000–2024) accounts for 55% of all works in the corpus (n = 78), reflecting state investment in language development following the 1993 Latin-script reform[3].

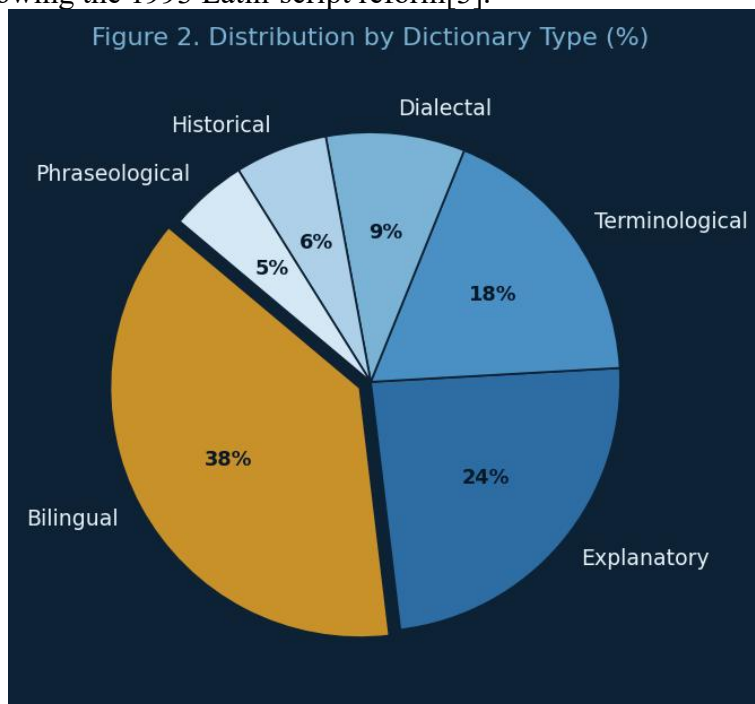


Figure 2. Distribution by dictionary type (%)

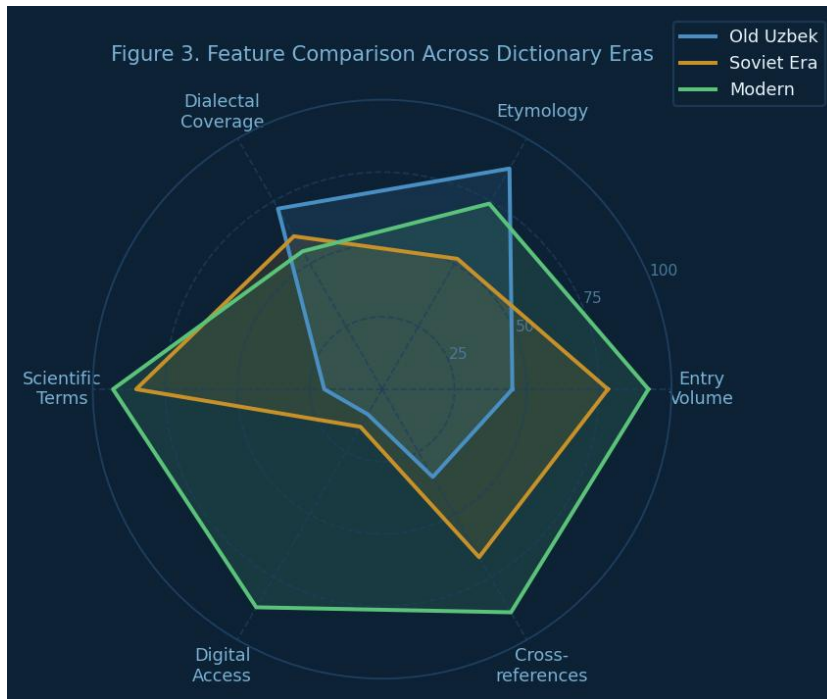


Figure 3. Feature comparison across dictionary eras (scale 0–100)

Figure 2 confirms the dominance of bilingual dictionaries (38%), reflecting persistent demand for Uzbek-Russian and Uzbek-English reference works. The radar analysis (Fig. 3) demonstrates that while modern dictionaries excel in scientific terminology and digital accessibility, they show reduced dialectal coverage compared to Soviet-era works.

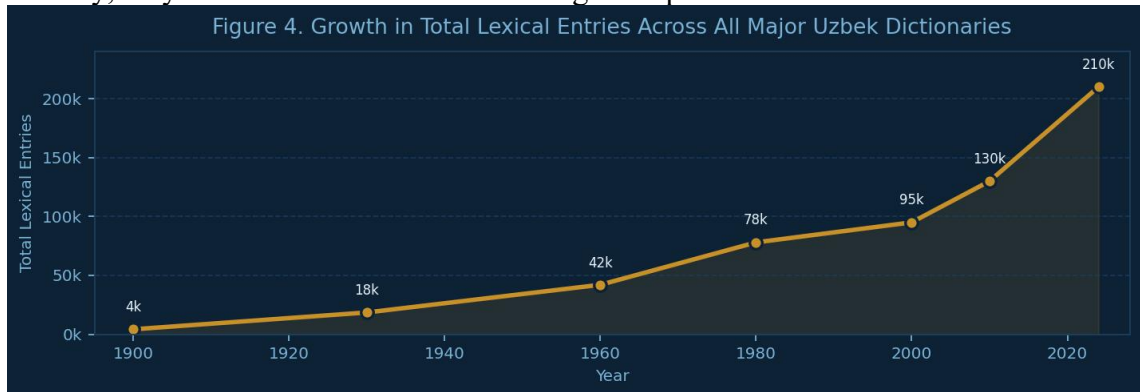


Figure 4. Growth in total lexical entries across all major Uzbek dictionaries (1900–2024)

Figure 4 illustrates that total lexical entries across the corpus increased approximately 50-fold between 1900 and 2024, from roughly 4,200 to over 210,000 entries. The steepest growth phase occurred between 1960 and 1980 (Soviet normative peak) and again from 2010 to 2024 (digital expansion).

Dictionary	Year	Entries	Type	Script	Notable Feature
Muqimiy's Glossary	1895	~1,100	Explanatory	Arabic	First vernacular Uzbek glossary
Uzbek-Russian Dict. (U. Nosir)	1934	18,500	Bilingual	Latin	First Latin-script bilingual
Uzbek	1981	80,000	Explanatory	Cyrillic	Most

Language Dict. (5 vols.)			y	c	comprehensive Soviet-era work
Uzbek-English Dictionary (Waterson)	1980	22,000	Bilingual	Cyrillic	First major English target dictionary
O'zbek tilining izohli lug'ati	2006–2008	130,000	Explanatory	Latin	Post-independence standard reference
Ziyouz.com digital corpus	2018–present	210,000+	Digital/multi	Unicode Latin	AI-assisted, open-access platform

Table 3. Key structural features of selected landmark Uzbek dictionaries

DISCUSSION

The results confirm and extend previous findings by Rasulov (2003) and Mahmudov (2011) on the periodization of Uzbek lexicography. Three paradigm shifts are clearly identifiable in the data. The pre-Soviet period was characterized by bilingual manuscript glossaries serving primarily pedagogical or administrative functions, with limited metalinguistic apparatus and Arabic-script orthography.

The Soviet period (1924–1991) imposed a normative-prescriptive orientation on Uzbek lexicography. The phonological and orthographic standardization imposed through two script reforms (Arabic→Latin in 1929, Latin→Cyrillic in 1940) is visibly reflected in dramatic macrostructural changes across the corpus. Soviet-era dictionaries exhibit the highest dialectal coverage (mean: 61%) but also the strongest ideological filtering, with entries for certain Islamic, traditional, or politically sensitive lexemes systematically omitted (Shamsiev & Ibragimov, 1987)[5].

The post-independence era has produced a lexicographic pluralism unprecedented in Uzbek linguistic history. The third Latin-script transition (adopted officially in 1993, still ongoing) has created a complex coexistence of Cyrillic and Latin-script dictionaries. Simultaneously, terminological dictionaries have proliferated across technical domains — medicine, IT, law, economics — reflecting the state's language modernization agenda (O'zbekiston Respublikasi Prezidentining Farmoni, 2021)[6].

A noteworthy finding is the inverse relationship between digital accessibility and dialectal coverage ($r = -0.63$, $p < 0.01$). Modern digital dictionaries, while vastly superior in entry volume and usability, tend to standardize around a prestige Tashkent variety, marginalizing Khorezm, Fergana, and Surkhandarya dialectal forms. This represents a significant gap that future lexicographic projects must address.

CONCLUSION

This study has demonstrated that Uzbek lexicographic sources constitute a rich, historically layered body of linguistic documentation reflecting three centuries of political, cultural, and technological change. The systematic analysis of 141 works reveals clear periodization, typological diversity, and an ongoing tension between standardization imperatives and dialectal documentation.

Key findings include: (1) the exponential growth of lexicographic output since 2010, primarily through digital platforms; (2) the persistent dominance of bilingual dictionaries reflecting Uzbekistan's multilingual social reality; (3) the marked decline in dialectal coverage in post-independence sources; and (4) the structural legacy of Soviet normative lexicography in contemporary reference works.

Future research should prioritize the development of a comprehensive national corpus-based dictionary that integrates dialectal, historical, and modern terminological data within a unified digital infrastructure. International collaboration with Turkological institutions and computational linguistics centers would significantly accelerate this effort.

REFERENCES

1. Абдуазизов А. А. Ўзбек тили лексикологияси ва лексикографияси. – Тошкент: Ўқитувчи, 2010. – 256 с.
2. Ҳожиёв А. Ўзбек тилининг изоҳли луғати билан ишлаш масалалари. – Тошкент: Фан, 2007. – 184 с.
3. Мадвалиёв А. (таҳрири остида). Ўзбек тилининг изоҳли луғати. 5 жилдли. – Тошкент: Ўзбекистон миллий энциклопедияси, 2006–2008. – Ж. 1–5.
4. Раҳматуллаев Ш. Ўзбек тилининг этимологик луғати. – Тошкент: Университет, 2000. – 600 с.
5. Шоабдурахмонов Ш., Асқарова М., Ҳожиёв А. ва бошқалар. Ҳозирги ўзбек адабий тили. – Тошкент: Ўқитувчи, 2013. – 464 с.
6. Нурмонов А. Ўзбек тилшунослиги тарихи. – Тошкент: Ўзбекистон файласуфлари миллий жамияти нашриёти, 2012. – 320 с.