SRSTI: 16.21.23

### A.A. Seitbekova

A. Baitursynov Institute of Linguistics, 050010, Republic of Kazakhstan, Almaty, Shevchenko str., 28 Orcid: 0009-0004-9137-8638 e-mail: ainurseit@mail.ru

# PECULIARITIES OF REPRESENTING ARABIC AND PERSIAN LOANWORDS IN DICTIONARIES

Abstract. This article provides a comprehensive examination of the place and role of Arabic and Persian elements in the historical lexicography of the Kazakh language. The study, based on the analysis of historical, etymological, and dialectological dictionaries, considers in detail the phonetic, morphological, and semantic characteristics of borrowed vocabulary. Particular attention is paid to the works of L.Z. Budagov (Comparative Dictionary of Turkic-Tatar Dialects) and R. Syzdykova, which reveal the peculiarities of the functioning of loanwords in ancient written monuments.

The article demonstrates the historical significance of Arabic-Persian elements, the ways of their penetration into the Kazakh language, as well as the process of de-etymologization that they underwent over time. Through specific examples (Jamiʻ al-Tawarikh, Türik Shajara, and epic works of the batyr cycle), the semantic dynamics of borrowed words are traced along with their contemporary meanings in the Kazakh language. Issues of graphic variability and orthographic instability in the medieval writing tradition are also discussed. The findings can serve as a theoretical basis for compiling a historical dictionary of the Kazakh language, for identifying the intrinsic nature of borrowed elements, and for determining their place within the system of lexicography.

**Keywords:** historical lexicography, borrowed vocabulary, Arabic-Persian elements, etymology, deetymologization, historical dictionary, written monuments, semantic dynamics, phonetic principle, morphological analysis.

#### Introduction

Arabic and Persian borrowings occupy a significant place in the lexical composition of the Kazakh language. To determine their historical formation and to describe their semantic, phonetic, and morphological features, it is essential to rely on historical, etymological, and dialectological dictionaries. In this regard, historical lexicography makes it possible to explain the meanings of loanwords used in different periods of language development and to reveal the particularities of their usage.

Among the most notable studies are L.Z. Budagov's Comparative Dictionary of Turkic-Tatar Dialects and R. Syzdykova's research on historical lexicography. These works clarify the use of Arabic-Persian elements found in ancient written monuments and epic heritage and trace their semantic continuity in modern Kazakh. Consequently, the place of Arabic and Persian borrowings in historical lexicography is important not only for understanding the history of the language but also for illuminating the cultural and civilizational links of the Kazakh language.

### Materials and methods

The main materials of the study are the historical, etymological, and dialectological dictionaries of the Kazakh language, as well as ancient written monuments and epic texts. Specifically: Budagov's Comparative Dictionary of Turkic-Tatar Dialects served as a basis for comparing Arabic and Persian borrowings with their original languages and for identifying their semantic fields; the Concise Etymological Dictionary of the Kazakh Language was used as a source for describing the origins and historical usage of loanwords; Qadirgali Jalairi's Jamiğ-at tawarikh and Abilgazi Khan's Türki shezhire were treated as primary materials demonstrating the use of loan units in medieval manuscripts. Epic poems of the fifteenth to eighteenth centuries (Alpamys, Kambar, Edige) provided key evidence for identifying the historical-cultural functions of borrowings in oral literature.

The following methods were employed: the comparative-historical method to compare the usage of Arabic and Persian words in Turkic languages and in Kazakh; etymological analysis to identify the original meanings, initial forms, and historical development of loanwords; semantic analysis to examine meanings in different periods and the process of de-etymologization; phonetic-

morphological analysis to systematize phonological and structural changes; and a phraseological-syntactic approach to elucidate the meanings of fixed expressions containing Arabic-Persian elements. Overall, the materials allowed a comprehensive representation of the place and role of borrowed elements in Kazakh. Methodologically, systematic comparison of historical data, multifaceted analysis, and reliance on established scholarly conclusions ensured the robustness of the findings.

## Literature review

Some dictionaries aim to describe the historical development of a language's vocabulary across several periods; these are termed historical dictionaries. By contrast, other dictionaries set out to capture words frequently used in contemporary language and analyze them; these are modern-language dictionaries [1]. Accordingly, we consider it appropriate to examine the usage features of the loanword stratum-which showcases each nation's linguistic wealth-through diachronic historical, etymological, and dialectological dictionaries. As G. Änes aptly notes: «The historical background is an extralinguistic factor, and it serves as a primary support in studying loanwords and the ethnographic vocabulary of a language» [2].

Historical dictionaries have traditionally explained the semantics of loan units and collocations found in early written monuments and oral literature and were often appended to books. An example offering extensive analyses of the origins of numerous loanwords is L.Z. Budagov's Comparative Dictionary of Turkic-Tatar Dialects. The dictionary also attends to Arabic, Persian, and Mongolian words in Turkic written heritage that are obscure to modern readers. It contains many items borrowed from Arabic-Persian, frequently used across Turkic languages, and indicates the donor language and the intermediary routes by which items entered particular languages. For instance, mir is explained as a shortening of Arabic amīr, denoting various titles such as «chief administrator, leader, prince», while mirab (Kazakh mūrap) is glossed as «an official overseeing water management». Special attention is paid to the meanings of Arabic and Persian words in collocation, their usage and orthography across Turkic languages. For example, Mongolian akhte is presented with its variants (akhte, aqta, aqte; Uigh. aqda; Turk. ağdağ etc.) and semantic range. In Kazakh written monuments regarded as early models (Qadirgali bi's Jami' attawarikh and Abilgazi Khan's Türki shezhire), ahta at denotes «war-horse, riding horse», and related forms such as akhtači/akhtači function as titles attached to proper names; akhtekhana denotes «stable.»

Thus, by presenting forms alongside their Arabic originals and drawing comparative parallels with cognate Turkic languages, Budagov's dictionary is a uniquely valuable resource for establishing etymologies [3].

The Concise Etymological Dictionary of the Kazakh Language likewise provides examples. On bazar, one view derives it from bay «abundant» + -sar (a suffix of abundance/plurality): bay + sar > basar > bazar. Parallels are adduced with -sar/sär in other items (e.g., Tatar qamsar «sandy area», qulmaq-sar «place where hops grow», Chuvash lapsar «lowland», Cheboksary «place abundant in chub fish») [4]. R. Syzdykova argues that bazar has been in Kazakh since very early times, originally meaning «a gathering, a festive/crowded place,» later narrowing to «marketplace,» and suggests that baz may relate to Persian dar-bāza «gate» (with dar/där «small door» + baz «gate»). Scholars such as A. Ysqaqov, Ä. Qaidar, R. Syzdyk, L. Rüstemov, N. Qarasheva, and Ä. Quryshjanov present their principles for compiling etymological dictionaries.

The results of the study reveal that Arabic and Persian loanwords constitute a significant and multi-layered stratum of the Kazakh lexicon, reflecting both cultural contact and historical processes of integration. Their presence is not limited to isolated lexical units but extends into semantic fields associated with governance, social organization, material culture, and military terminology. Analysis of historical dictionaries, manuscripts, and epic texts allows us to trace their phonetic, morphological, and semantic evolution across several centuries.

One of the most illustrative sources is L.Z. Budagov's Comparative Dictionary of Turkic-Tatar Dialects, which provides detailed explanations of Arabic and Persian words preserved in early written monuments. For example, the dictionary records the shortened form mir derived from Arabic amīr,

denoting high-ranking officials such as «chief administrator, leader, prince». A related form, mirab (Kazakh mūrap), is glossed as «an official overseeing water management», which highlights how administrative structures and social hierarchies influenced the borrowing process. Budagov also devotes attention to words of Mongolian origin with Arabic-Persian connections, such as akhte («war horse, riding horse»), attested in Qadirgali bi's Jamiʿ at-tawarikh and Abilgazi Khan's Türki shajara. Titles like akhteči were attached to personal names (e.g., Buqatai akhtači, Hurun akhtači), while akhtekhana denoted «stable». Such evidence underscores that borrowings penetrated not only vocabulary but also institutional and cultural practices of the medieval Kazakh steppe.

Another widely attested example is the word bazar. Competing etymological interpretations have been offered in the literature. One hypothesis explains it as a compound of the Turkic root bay («abundant, rich») and the suffix -sar («abundance, plurality»), yielding basar > bazar. This explanation is supported by parallels in other Turkic languages, such as Tatar qamsar («sandy area») and Chuvash lapsar («lowland»). R. Syzdykova, however, argues that bazar has been part of Kazakh vocabulary since very early times, originally denoting «a gathering or festive place», later narrowing to «marketplace». She also points to a Persian connection: dar-bāza («gate»), where dar means «small door» and baz means «gate». This interpretation highlights the polygenetic pathways through which loanwords entered Kazakh, often blending Turkic word formation with Persian semantic influence.

The semantic field of material culture and defense provides further striking examples. The Persian word qandaq (originally «ditch, trench») entered Kazakh with the meaning «specially dug pit» or «treasure pit». In epic texts such as Alpamys, it appears in verses describing a storage pit for valuables. Later contexts, however, broadened the meaning: in some manuscripts, qandaq denotes an imprisonment pit or, alternatively, a defensive trench. In Türki shajara, qandaq is used in the sense of a trench behind which warriors took cover. The term was widely attested in fourteenth-century Golden Horde texts, demonstrating its deep historical roots. The semantic dynamics of qandaq illustrate how borrowings adapted to the sociocultural realities of the steppe, moving from economic functions (storage) to military applications (defense).

Similarly, the word top («cannon, artillery») illustrates semantic shifts within the military domain. Epic poems such as Alpamys and Kambar show that Turkic and Mongol armies used both firearms and cannons, with top referring specifically to artillery. Scholars connect this term to the Persian tufang («rifle, firearm»). The diminutive derivative tapancha («pistol"»), formed from tufang + -cha, entered Kazakh in the late nineteenth or early twentieth century. This process demonstrates not only borrowing but also productive word formation based on Persian roots, which continued to generate new lexemes within Kazakh.

The study also identifies numerous cases of de-etymologization, a process by which the original form and meaning of borrowed words gradually became obscure. For example, the Old Persian word dej («hostile, rude») survives in modern Kazakh only as part of derivatives such as dushman («enemy») and dezhkhim («executioner»), where the original meaning is entirely lost. Another case is pod («to protect»), once productive in Persian but now recognizable only in fossilized compounds. As L.Z. Rustemov notes, such processes blur the semantic boundaries of borrowed items, leading to delinking from their original roots. De-etymologization thus represents a universal phenomenon whereby loanwords are absorbed into the internal lexical system of Kazakh, gradually losing transparency.

Another important result concerns orthographic and graphic instability in medieval texts. Because most manuscripts were written in Arabic script, which allowed multiple ways of rendering certain sounds, the same loanword often appears in variant forms. For example, in Türki shajara the word mudir («administrator») is written as mudar, while tagallaq is rendered as tagalliq. These orthographic shifts were partly due to the absence of stable spelling conventions in the medieval period, as well as the adaptation of Arabic script to Turkic phonology. Such variation complicates lexicographic analysis, making it necessary to reconstruct etymologies by comparing across multiple sources and languages.

From a methodological perspective, the results confirm that several analytical approaches must be combined:

Phonetic analysis is essential for identifying sound correspondences and reconstructing the original form of loanwords.

Morphological analysis enables the recognition of word-formation processes based on Arabic-Persian roots, especially when these borrowings served as bases for new Turkic derivatives.

Semantic analysis highlights the dynamic meanings of loanwords across different periods, as in the case of bazar or qandaq.

Phraseological and syntactic analysis is crucial for understanding fixed expressions and idioms, many of which included Arabic-Persian components in medieval texts.

Finally, epic literature such as Alpamys, Kambar, and Edige confirms the cultural integration of borrowings. Words denoting governance (mir), social titles (akhteči), weaponry (top, tufang), and imprisonment (qandaq, zindan) reveal how deeply Arabic and Persian elements became embedded in the oral tradition. Their continued presence in epic poetry underscores not only their lexical assimilation but also their symbolic role in representing power, warfare, and cultural values of the Kazakh people.

Overall, the analysis demonstrates that Arabic and Persian borrowings are not marginal phenomena but key elements of the Kazakh lexical system. They functioned as carriers of cultural knowledge, reflected technological and institutional change, and were reshaped through processes of phonological adaptation, morphological productivity, semantic shift, and de-etymologization. These findings provide a strong foundation for compiling a comprehensive historical dictionary of the Kazakh language that accurately represents the origins, development, and contemporary relevance of its borrowed vocabulary.

### **Results and discussion**

Etymology involves not only identifying a word's origin but also how it entered the language and when it came into use. In many cases, derivational ties weaken and break over time; the original meaning is lost, and de-etymologization occurs a universal phenomenon. L.Z. Rüstemov illustrates this with dej in Old Persian («hostile, rude»), whose meaning has been forgotten but survives within compounds such as dezhman «enemy,» dezhkhim «executioner» (khim «temper, character»). Similarly, pod (formerly «to protect») has lost its meaning; Rüstemov notes occurrences in compounds like Middle Persian posukh «answer (defense of speech)», describing such cases as delexicalization/de-etymologization the loss of earlier etymological and semantic links [5].

Closer to Kazakh, extensive work has been done on Qadirgali Jalairi's Jamiğ-at tawarikh and Abilgazi Khan's Türki shezhire, yet comprehensive dictionaries assembling their full lexical inventory are still lacking. In Archaisms and Innovations in the Kazakh Language (Arys, 2009), R. Syzdykova provides a concise glossary of the Jamiğ-at tawarikh items relevant to modern Kazakh and examines archaic and innovative units more broadly. Crucially, she compares early loan elements with units in XV–XVIII-century epic texts. For example, qandaq (from Persian khandaq) originally denoted a «pit dug to store valuables,» contrasted with zyndān for a pit to confine people; the word appears in the Golden Horde texts (Säyf-i Sarāy's Gulistān bi-'t-Turki and Nahj al-Farādis) and later in Abilgazi's Türki shezhire. Likewise, top «cannon» is traced to Persian tūfang «gun», with medieval attestations showing both «gun» and «projectile» senses; from tūfang/tofang + diminutive -sha/-cha derives tapan-sha «pistol», now fully integrated into Kazakh. Historical research also notes the spread of top/zeñbirek as artillery terms in Central Asia.

Overall, many borrowings were highly productive in the medieval period and, over time, underwent internal restructuring, with only the root retaining tenuous semantic ties. Given the scarcity of early Kazakh-related manuscripts, the method proposed by R. Syzdykova helps bridge gaps-yet, as she cautions, «without written evidence, however sound our arguments, they often remain conjectural» [6], [7]. The history of lexicography studies dictionaries produced in earlier periods, while historical lexicography systematizes word histories in dictionary form on the basis of written monuments, using contemporary methods. Because the data of lexicographic history are themselves written monuments, they are a primary source for historical lexicography. Any

dictionary requires sources: (a) prior dictionaries and (b) a card index; works cited below belong to the first group and fall within the scope of lexicographic history [8].

It follows that the Kazakh lexicon contains numerous loan elements attested in early manuscripts, later poetic works, and other historical sources. Some have faded from active consciousness and fallen out of use; their reasons and pathways are sought in historical dictionaries. In such dictionaries, preserving the external (graphic) form of Arabic-Persian elements is vital: erroneous spellings obscure meanings and complicate analysis. Since medieval works were written in Arabic script, some letters had two or three graphic variants (e.g., two forms of t, three of s), producing orthographic variability across copies. In Türki shezhire, for instance, certain borrowings deviate from the Arabic/Persian originals (e.g., mudar for Arabic mudir; tağalliq for Arabic tağallaq); later copies often modernize spellings to contemporary norms.

If the phonetic shape of Arabic and Persian words is not accurately preserved, interpretation in context becomes difficult. Hence the need to observe the language's stable phonetic system as a key etymological principle. A morphological perspective also aids in determining the root forms, structure, and etymological nature of borrowings by analyzing word-formation (analytic, synthetic) patterns in Turkic derivatives. Because many Arabic and Persian words are polysemous and were used with multiple senses in medieval texts, attention must be paid not only to phonetic and morphological aspects but also to semantic features, with comparative evidence from dialects and related languages. Finally, identifying the phraseological-syntactic patterns of Arabic-Persian elements in collocations remains a relevant task, as such idioms are abundant in manuscript language; this helps trace the semantic dynamics of obscure units whose meanings are opaque to the modern reader.

#### Conclusion

Considering Arabic and Persian elements from the perspective of historical lexicography makes it possible to determine their place and function within the linguistic system. Based on historical, etymological, and dialectological sources, the study yields the following conclusions:

Arabic and Persian borrowings occur frequently in medieval manuscripts, oral literature, and epic poetry, enriching the Kazakh lexicon.

Their semantic fields have shifted over time; as a result of de-etymologization, links to original roots have often become obscured.

Works such as Budagov's Comparative Dictionary of Turkic-Tatar Dialects are invaluable for establishing etymologies through comparative evidence.

Studies by R. Syzdykova, L. Rüstemov, and others provide a solid methodological basis for explaining the archaic lexical layer, including Arabic-Persian borrowings.

Orthographic and graphic features of the medieval writing tradition have led to variant spellings, which complicate semantic interpretation.

In sum, the historical lexicography of Arabic and Persian borrowings contributes to a deeper understanding of the development of the Kazakh language and clarifies its cultural-civilizational connections. This line of inquiry lays theoretical and practical foundations for compiling a historical dictionary of the Kazakh language.

#### References

- 1. Akhanov K. Introduction to Linguistics. Almaty: Mektep, 1965. 240 p. (In Kazakh)
- 2. Änesov G. Main Directions of Etymological Research in Kazakh Linguistics // Issues of Historical Lexicology of the Kazakh Language. Alma-Ata: Gylym, 1988. P. 100–120.
- 3. Amangaziyeva M. The Place of L.Z. Budagov's Comparative Dictionary of Turkic-Tatar Dialects in Kazakh Lexicography // Bulletin of L.N. Gumilyov ENU. Philology Series. 2018. No. 123(2). P. 38–43. DOI: 10.32523/2616-678X-2018-123-2-38-43.
- 4. Concise Etymological Dictionary of the Kazakh Language. Almaty: Gylym, 1966. 240 p.
- 5. Rüstemov L.Z. Pathfinder of the Persian Language. Tehran, 1995. 198 p.
- 6. Syzdykova R. Words Speak. Almaty: Arys, 2004. 232 p. (In Kazakh)

- 7. Köbdenova G. Historical Lexicography of the Kazakh Language: Structure and Principles. Dissertation (PhD) Almaty, 2008. 152 p. (In Kazakh)
- 8. Syzdykova R. Archaisms and Innovations in the Kazakh Language. Almaty: Arys, 2009. 256 p. (In Kazakh)
- 9. Baizhanov T. Military Vocabulary in the Kazakh Language. Almaty: Rauan, 1991. 174 p.
- 10. Seitbekova A., Kairat A., Nazarzoda S. Old Book Vocabulary: Potential for the Revival of Arabic and Persian Words // Tiltanym. 2025. Vol. 97, No. 1. P. 58–67. DOI: 10.55491/2411-6076-2025-1-58-67.
- 11. Malikov G., Alshayeva A. Linguistic Features of Borrowed Words in the Modern Kazakh Language // Comparative Turkish Dialects and Literatures. 2023. Vol. 1, No. 3. P. 33–51.
- 12. Shajara-i Turk. Kazan ed. Kazan, n.d. 183 p. (In Kazakh)

### А.А. Сейтбекова

А.Байтұрсынұлы атындағы Тіл білімі институты, 050010, Қазақстан Республикасы, Алматы қ., Шевченко к-сі, 28 Orcid: 0009-0004-9137-8638 e-mail: ainurseit@mail.ru

# АРАБ ЖӘНЕ ПАРСЫ СӨЗДЕРІНІҢ ТАРИХИ СӨЗДІКТЕРДЕ БЕРІЛУ ЕРЕКШЕЛІКТЕРІ

Аңдатпа. Бұл мақалада қазақ тілінің лексикалық құрамындағы араб және парсы тілі элементтерінің тарихи лексикографиядағы орны мен қызметі жан-жақты қарастырылады. Зерттеу барысында тарихи, этимологиялық және диалектологиялық сөздіктерді саралай отырып, кірме сөздердің фонетикалық, морфологиялық және семантикалық сипатын талданады. Әсіресе Л.З. Будаговтың «Түркі-татар тілдерінің салыстырмалы сөздігі» мен Р. Сыздықованың еңбектері негізінде көне жазба ескерткіштердегі кірме сөздердің қолданылу ерекшеліктері ашып көрсетілді. Мақалада араб-парсы элементтерінің тарихи қолданыстағы мағыналары, олардың қазақ тіліне ену жолдары, уақыт өте келе деэтимологиялану құбылысына ұшырауы нақты мысалдармен дәлелденеді. «Жамиғ-ат тауарих», «Түркі шежіресі», батырлар жырлары сияқты жазба мұралардағы кірме сөздердің семантикалық динамикасы талданып, олардың қазіргі қазақ тіліндегі мағынасы айқындалады. Сондай-ақ графикалық нұсқалардың әртүрлілігі мен ортағасырлық жазба дәстүрдегі орфографиялық тұрақсыздық мәселелері де назарға алынды. Зерттеу нәтижелері қазақ тілінің тарихи сөздігін түзуге, кірме элементтердің түпкі табиғатын тануға және олардың лексикографиядағы орнын анықтауға теориялық негіз бола алады.

**Тірек сөздер:** тарихи лексикография, кірме сөздер, араб-парсы элементтері, этимология, деэтимологиялану, тарихи сөздік, жазба ескерткіштер, семантикалық динамика, фонетикалық принцип, морфологиялық талдау.

## А.А. Сейтбекова

Институт языкознания имени А. Байтурсынова, 050010, Республика Казахстан, г. Алматы, ул. Шевченко, 28. Orcid: 0009-0004-9137-8638 e-mail: ainurseit@mail.ru

# ОСОБЕННОСТИ ОТРАЖЕНИЯ АРАБСКИХ И ПЕРСИДСКИХ ЗАИМСТВОВАНИЙ В СЛОВАРЯХ

Аннотация. В данной статье всесторонне рассматривается место и роль арабских и персидских элементов в исторической лексикографии казахского языка. В ходе исследования, опираясь на анализ исторических, этимологических и диалектологических словарей, проводится комплексное рассмотрение фонетических, морфологических и семантических характеристик заимствованной лексики. Особое внимание уделено трудам Л.З. Будагова («Сравнительный словарь тюркско-татарских наречий») и Р. Сыздыковой, на основе которых раскрываются особенности функционирования заимствованных слов в древних письменных памятниках.

В статье демонстрируется историческое значение арабо-персидских элементов, пути их проникновения в казахский язык, а также процесс деэтимологизации, которому они подвергались со временем. На конкретных примерах («Жамиат-ат тауарих», «Тюркская родословная», эпические произведения батырского цикла) показана семантическая динамика заимствованных слов, а также их современное значение в казахском языке.

Рассматриваются вопросы вариативности графических форм и орфографической нестабильности в средневековой письменной традиции.

Полученные результаты могут служить теоретической основой для составления исторического словаря казахского языка, выявления сущностной природы заимствованных элементов и определения их места в системе лексикографии.

**Ключевые слова:** историческая лексикография, заимствованная лексика, арабо-персидские элементы, этимология, деэтимологизация, исторический словарь, письменные.

# Автор туралы

**Сейтбекова Айнұр Аташбекқызы -** филология ғылымдарының кандидаты, қауымдастырылған профессор, А.Байтұрсынұлы атындағы Тіл білімі институтының аға ғылыми қызметкері, e-mail: ainurseit@mail.ru, Orcid: 0009-0004-9137-8638.

## Сведения об авторе

Сейтбекова Айнур Аташбековна – кандидат филологических наук, ассоциированный профессор, старший научный сотрудник Института языкознания имени А. Байтурсынова, е-mail: ainurseit@mail.ru, Orcid: 0009-0004-9137-8638.

#### Information about author

**Seitbekova Ainur** – candidate of philological sciences, associate professor, Senior research fellow, A. Baitursynov Institute of Linguistics, e-mail: ainurseit@mail.ru, Orcid: 0009-0004-9137-8638.