

Folklore Discourse and Its Representation in the National Corpus of the Kazakh Language

Zhamila Omirbekova

Department of Kazakh Linguistics, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Karlygash Sarekenova

Department of Kazakh Linguistics, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Gulgaisha Sagidolda

Department of Kazakh Linguistics, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Ardak Beisenbai*

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Bauyrzhan Oserov

Department of Kazakh Linguistics, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Syrym Zhanzhigitov

Department of Kazakh Linguistics, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Abstract—This article examines the concept of *folklore discourse* within the theoretical and methodological foundations of contemporary discourse studies and corpus linguistics. The study identifies the distinctive features of folklore discourse in the Kazakh language, including its genre-related, pragmatic, and linguistic characteristics, and analyzes how these features are represented in the National Corpus of the Kazakh Language. Based on empirical linguistic data, the structural and discursive markers of Kazakh folklore texts are described, and the scholarly and methodological issues that arise when integrating folklore texts into a corpus are specified. In addition, the article highlights the particularities of researching folklore discourse. The aim of the study is to characterize folklore discourse within the framework of corpus linguistics and to determine the scholarly potential of folklore texts included in the national corpus. The findings demonstrate the importance of building a corpus of folklore texts in the context of the digitalization of the Kazakh language. The conclusions of the article contribute both theoretically and practically to future research in folklore studies, corpus linguistics, and discourse studies.

Index Terms—folklore discourse, national corpus, corpus linguistics, Kazakh folklore

I. INTRODUCTION

Folklore occupies a special place in the spiritual and cultural heritage of the Kazakh people. Folklore texts are an important source that preserves the nation's historical memory, worldview, and social experience. This emphasis on culturally marked units is consistent with evidence that ethnocultural names constitute salient elements of Kazakh youth's linguistic consciousness, which supports the need for their systematic identification and corpus-based representation in digitized folklore materials (Akhmetova et al., 2025a). However, the traditional oral transmission of folklore creates certain difficulties for its inclusion in scholarly circulation. For this reason, at the present stage, digitizing folklore texts and integrating them into the national language corpus has become one of the key scientific and practical tasks.

In recent years, corpus linguistics has become an independent and rapidly developing field of language studies. Corpus research makes it possible to analyze linguistic data comprehensively in both quantitative and qualitative terms. This methodology also opens new scientific horizons for the study of folklore discourse, because folklore texts, while serving as linguistic material, also contain a complex layer of cultural, pragmatic, and cognitive information.

In modern linguistics, the notion of discourse is treated as a broader category than text. Alongside linguistic structure, it includes extralinguistic factors, such as *social context*, *communicative situation*, the roles of *speaker* and *listener*, and *cultural codes*. From this perspective, folklore discourse is a discourse system formed within the traditional cultural space of a particular ethnic group and capable of reflecting collective consciousness.

In world linguistics, the directions of *natural language modeling*, developing *linguistic algorithms*, and *automatic text processing* are shaping a new paradigm of scientific research. These processes have underpinned the rapid development of corpus linguistics, which enables the systematization, analysis, and interpretation of *large-scale linguistic data*. These approaches open the way to a comprehensive study of folklore texts not only in a traditional descriptive manner but also

through *discourse-based* and *quantitative-statistical* methods.

In Kazakhstani scholarship as well, the importance of *digitization* and *corpus-based research* is increasing. In particular, bringing folklore texts, one of the largest layers of national cultural heritage, into scholarly circulation within the *National Corpus of the Kazakh Language* is among the most pressing issues. In this regard, analyzing folklore discourse on the basis of corpus data in its *linguistic*, *structural*, and *functional* dimensions contributes to the formation of new research directions in Kazakh linguistics.

The scholarly study of Kazakh folklore has a deep tradition and has been widely considered in the works of Sh. Ualikhanov, A. Margulan, A. Divaev, M. Auezov, S. Qasqabasov, R. Berdibaev, A. Qongyratbaev, Zh. Zhakypov, and other scholars from the perspectives of *literary studies*, *ethnography*, and *text studies*. Although these studies provide a comprehensive description of the *genre nature*, *poetic system*, and *historical-cultural content* of folklore, the possibilities for systematically analyzing large volumes of folklore as linguistic data have remained limited.

In this context, a comprehensive study of folklore discourse using *corpus-linguistic methods* is a new and highly promising direction within the national research landscape. Folklore texts in the *National Corpus of the Kazakh Language* make it possible to identify the structure of folklore discourse and its diverse characteristics.

II. LITERATURE REVIEW

Theoretical Foundations of Folklore Discourse and the Problem of Its Corpus Representation

In contemporary linguistic research, folklore is increasingly viewed not as a collection of texts but as a *discursive system* in which linguistic structures function in relation to *cultural memory*, the *communicative situation*, and *genre regulation*. This perspective foregrounds the *corpus approach* as a tool for verifying and refining discourse features on large-scale datasets. In applied terms, this is reflected in studies where folklore texts are examined through recurring formulas, stable structures, evaluative vocabulary, and *variability*, as well as through translation and interpretive mechanisms that preserve cultural meaning across languages and registers (Hilmiyatun et al., 2022; Yuniari et al., 2023; Mahmud et al., 2022; Li et al., 2024). At the same time, corpus representation of folklore raises questions of metadata standardization, including *genre*, *region*, *performer*, and *time of recording*, the harmonization of orthographic and graphic variants, and the description of *formulaicity* as a marker zone of discourse that makes it possible to operationalize a text's folklore affiliation in a digital environment. Genre-typological perspectives also remain methodologically relevant for corpus modeling because they clarify cross-tradition regularities and genre-related constraints that can be operationalized in annotation and comparative corpus analysis (Sagymzhanova & Sagyndyk, 2024). In this context, solutions are important in which traditional philological tasks, such as typology, motifs, and discourse formulas, are integrated with corpus modeling and measurable distributional and frequency indicators, enabling a shift from descriptive observations to reproducible analytical procedures (Lin et al., 2024; Lo et al., 2020; Nguyen et al., 2020). In addition, studies of folklore-based narrative material (e.g., toponymic legends and myths) demonstrate how folklore discourse encodes cultural memory through stable narrative explanations and culturally motivated naming practices, which can be captured in a corpus via metadata and discourse-level tags (Mamysheva & Seitova, 2022).

The scholarly study of Kazakh folklore has a deep and well-established tradition in national scholarship. In the works of Sh. Ualikhanov, A. Divaev, A. Margulan, M. Auezov, and other researchers, folklore was approached as a historical and cultural source, and later it was systematized theoretically. Qasqabasov (2011) analyzed folklore as an artistic system in terms of genre structure and poetics, demonstrating the multi-layered nature of the folklore text; his research provides a foundation for identifying the genre and stylistic basis of folklore discourse. Qongyratbaev (1991) systematized the historical development of Kazakh folklore and proposed a genre classification that offers theoretical guidance for structuring folklore texts within a national corpus framework.

The American folklorist Richard Bauman describes folklore through the concept of *verbal art as performance*. In his view, the meaning of a folklore text is revealed not only in its structure but also in the communicative action of performance, that is, in its discursive nature (Bauman, 2004). This theory provides a basis for examining the pragmatic dimensions of folklore discourse, including *addressee*, *impact*, *evaluation*, and *ritual function*. Richard Bauman and Charles Briggs conceptualize genre as a *discursive regulatory system* and demonstrate the role of folklore genres in organizing social experience. From their perspective, genre is not merely a textual form but a discursive model with social and ideological significance (Bauman & Briggs, 1992). This approach makes it possible to describe folklore genres within the national corpus as discursive units rather than as purely formal text types.

In recent years, the development of corpus linguistics has offered new methodological opportunities for traditional text studies and folklore studies. Corpora enable the systematic compilation of large text collections and support their comprehensive quantitative and qualitative analysis (McEnery & Hardie, 2012). This shift has also influenced folklore research and contributed to the emergence of the notion of a *folklore corpus*. In Kazakh linguistics, corpus-related issues represent a relatively new direction. Although the National Corpus of the Kazakh Language has been created and has begun to enter scholarly circulation, the place of folklore texts in the corpus, their discursive description, and the task of creating a specialized folklore *subcorpus* have not yet been investigated systematically. Folklore discourse is a type of discourse grounded in oral tradition and reflecting collective consciousness and cultural experience. It is realized within a specific genre framework, such as fairy tale, legend, epic, and proverb, and is characterized by stable linguistic patterns (Foley, 1995).

The research literature commonly identifies the following core features of folklore discourse: (1) *structural patterning*, that is, the widespread use of fixed expressions and recurring constructions; (2) *repetition and variability*, meaning that a single plot or motif is presented in multiple versions; (3) *evaluation and modality*, expressed in explicit axiological assessments of characters and actions; in addition, research on sacred vocabulary in the linguistic consciousness of Kazakh youth underscores the cultural-semantic weight of sacral lexicon, reinforcing the need to identify, tag, and analyze such units when folklore texts are integrated into a national corpus (Akhmetova et al., 2025b); (4) *mythological time and space*, modeled outside concrete historical coordinates; (5) *oral nature*, meaning that the text originally existed in oral rather than written form; and (6) *collective authorship*, whereby folklore texts are attributed to the people rather than to an individual author. From the perspective of discourse theory, folklore texts function as tools that consolidate social norms and transmit cultural knowledge (Fairclough, 1995). Folklore discourse also has a complex genre organization. Epic narratives, fairy tales, legends, proverbs, riddles, and ritual song traditions differ in discursive structure, pragmatic purpose, and linguistic resources. Therefore, research on folklore discourse requires an integrated and interdisciplinary approach.

Digitizing folklore texts and presenting them in corpus format is a widely established practice in world linguistics. For example, the Finnish SKVR project (*Suomen Kansan Vanhat Runot*) represents a large-scale digital corpus of folk poetry. Harvilahti (2013) evaluates this database as a model for studying folklore with modern methods because the texts are annotated by parameters such as genre, region, performer, and time. Similarly, the Russian National Corpus contains subcorpora dedicated to folk speech and oral texts. Kachinskaja and Malysheva (2014) outline scientific and methodological principles for integrating oral texts close to folklore into a corpus; they argue that corpus inclusion of oral discourse requires annotation that preserves discourse features. The high frequency of stable structures in folklore texts aligns with Propp's (1998) theory of stable compositional models, which emphasizes that folklore texts follow a canonical structure. In the same direction, Qaidar's (2004) ethnolinguistic research interprets folklore structures as condensed representations of ethnocultural information. Corpus data substantiate this view by showing that such structures function as discursive markers within folklore discourse.

The National Corpus of the Kazakh Language is the key source that enables quantitative and text-based research on folklore discourse. Although the corpus does not yet contain a dedicated folklore subcorpus, folklore texts are included within the cultural-representative subcorpus. These texts are compiled across several directions, including folklore materials, authorial oral-literary heritage, ethnographic works, scholarly works, and articles (National Corpus of the Kazakh Language). Aitova and Baktykzy (2025) note that integrating folklore texts into the corpus expands the digital representation of the national language and contributes to the formation of a new direction in linguo-folklore studies. They propose a model of a parallel fairy-tale corpus, demonstrating automated presentation of quantitative and qualitative analysis of linguistic units, including frequency profiles and parallel aligned versions.

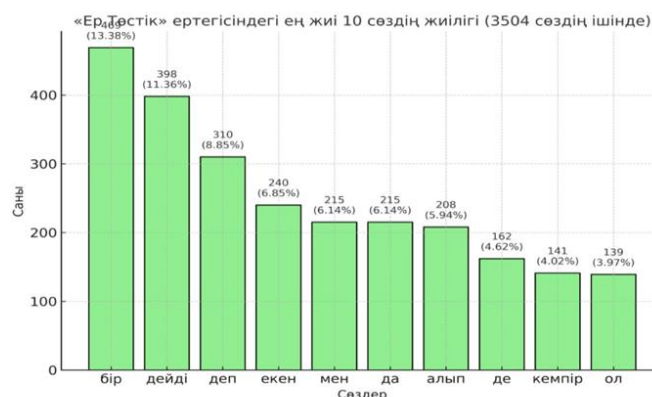


Figure 1. Frequency Distribution of Words in the Fairy Tale Er Tostik (as Proposed by Aitova and Baktykzy)

Note: The chart shows the relative frequency of repetition (*Қайталанудың салыстырмалы жиілігі*) of selected words across text segments (*Мәмін сегменнтері* (*Ер Төстік*)) numbered 1–10. Each line corresponds to one lexical item: *мен* (I), *екен* (it turns out / reportedly), *ден* (saying / that), *дейді* (says), and *бір* (one / a). The values on the y-axis reflect how often each word occurs in each segment relative to the segment's total word count, allowing comparison of distribution patterns across the narrative.

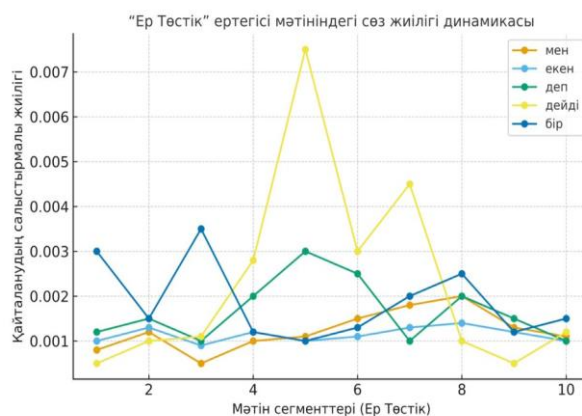


Figure 2. Word Repetition Frequency Across Segments of the Er Tostik Fairy-Tale Text (as Proposed by Aitova and Bakytzyzy)

Note: The vertical bars represent the count (Саны) of each of the ten most frequent words (Сөздер) in the text, with the percentage share of the total (3,504 tokens) shown above each bar. The x-axis includes: бір (one / a), дейді (says), деп (saying / that), екен (it turns out / reportedly), мен (I), да (also / and), алып (taking / having taken), де (particle: also / emphasis), кемпір (old woman), and ол (he / she / that). The height of each bar reflects how frequently the word appears in the overall text.

These figures indicate that integrating folklore texts into a national corpus is a complex multi-stage process. First, texts are collected from diverse sources, including manuscripts, archival materials, printed collections, and audio recordings. They are then converted into digital format and edited. At this stage, orthographic and graphic normalization is performed while preserving the authenticity of the original text. The next crucial stage is metadata annotation. For folklore texts, information such as *genre*, *narrator or performer*, *place and time of recording*, and *collector* is essential. This information makes it possible to study folklore discourse in spatial and temporal dimensions. Linguistic annotation also plays a special role in corpus inclusion. Alongside morphological, lexical, and syntactic tagging, there is a need to introduce additional markers specific to folklore. For example, marking fixed structural expressions, archaisms, dialect elements, or mythological names through specialized tags enables deeper analysis of folklore discourse. However, the specific properties of folklore discourse create challenges for automatic processing. *Variability*, flexible syntactic structures, and non-standard linguistic units may reduce the accuracy of automated analysis systems. Therefore, in developing a folklore corpus, it is effective to combine automated procedures with manual annotation.

International Corpus-Based and Computational Approaches to Folklore: Annotated Collections, Digital Repositories, and Quantitative Modeling of Variability

International research traditions show that the digitization of folklore is developing along two complementary trajectories: (a) the creation and publication of annotated collections and corpora as research infrastructure, and (b) the development of quantitative models that explain *variability* and the diffusion of plots. In the first trajectory, representative studies provide folklore texts with deep structural annotation, including functions, motif sequences, and event schemas, which makes it possible to examine a “plot grammar” and to train computational models to recognize narrative regularities (Finlayson, 2017; Hagedorn & Darányi, 2022; Nguyen et al., 2020). In the second trajectory, folklore is treated as “cultural data”: researchers analyze character and event networks and invariants of plot relations, identify patterns in the evolution of stories and their network organization (Karsdorp & van den Bosch, 2016; de Lima et al., 2016), and discuss mechanisms of cultural transmission, ranging from local stability of formulas to interregional trajectories of dissemination (Ross & Atkinson, 2016; Acerbi et al., 2017). A substantial body of work applies phylogenetic and comparative-evolutionary methods to reconstruct the “relatedness” of variants and to estimate the depth of plot roots (da Silva & Tehrani, 2016; Tehrani et al., 2016), as well as studies that link the spread of folklore motifs to evidence on migrations and contact zones (Bortolini et al., 2017). A distinct line of research focuses on automated extraction of semantic and thematic patterns in legends and fairy tales using modern language models and machine-learning approaches, expanding the analytical toolkit from frequency-based descriptions to context-sensitive identification of motifs and discursive roles (Tangherlini & Chen, 2024; Lo et al., 2020). Finally, in a number of studies folklore is used as a source of ethnocultural knowledge, for example, concerning human–animal relations, which makes the corpus not only a “linguistic” but also a cognitive-cultural resource for interdisciplinary interpretation (Nakawake & Sato, 2019; Wiysahnyuy & Valentine, 2023). Taken together, these approaches establish a methodological framework in which a corpus of folklore texts is understood as a data system requiring mandatory metadata and multi-level annotation, and folklore discourse is treated as an object amenable to reproducible description through frequency-based, network-based, and evolutionary models.

A Bibliometric Snapshot of “Folklore + Discourse” Research in Web of Science: Dynamics, Geography, and Publication Venues

In addition to the theoretical foundations of folklore discourse and corpus-based digital approaches, it is methodologically appropriate to include a bibliometric perspective that empirically demonstrates how research interest in the combination of the keywords “folklore” and “discourse” is distributed in *Web of Science*. Such an analysis does not

replace the substantive interpretation of individual studies; rather, it provides an important context by revealing temporal “waves” of activity, leading countries, and key publication venues where findings are accumulated. Below, analytics are presented for a *Web of Science* sample retrieved by the keywords “folklore” and “discourse” (total N = 15 publications).

The distribution of publications by year demonstrates an uneven dynamic and the presence of pronounced peaks in research activity (see *Figure 1*). The earliest isolated work appears in 2009 (1 publication), followed by single outputs in 2017 and 2018 (1 publication each), and a moderate level in 2020 (1 publication). Against this background, a qualitative shift is evident at the beginning of the 2020s: in 2021 the number of publications rises to 4, indicating a strengthening of attention to the topic and likely reflecting the expansion of discourse-based and corpus-digital methods for describing folklore material. Subsequent dynamics confirm the consolidation of interest without steady linear growth: after the 2021 increase, a decline occurs in 2022 (2 publications), followed by a new peak in 2024 (4 publications), comparable to the 2021 maximum. The appearance of a publication in 2025 (1 work) indicates the continuation of the research line; however, more precise conclusions about the trend require the accumulation of data for subsequent years and comparison with an expanded set of search queries and synonymous keyword forms (see *Figure 3*).

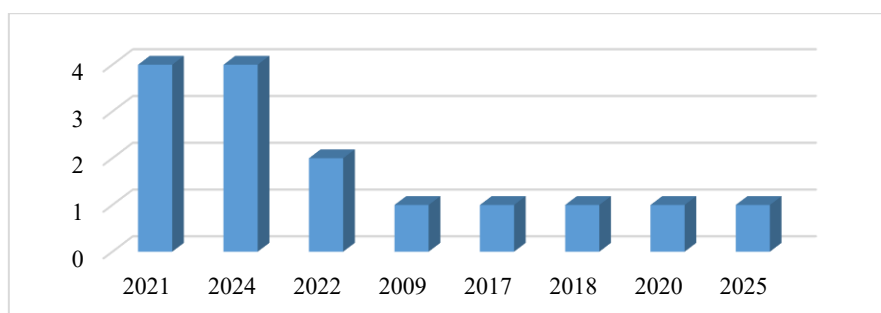


Figure 3. Dynamics of Publication Activity for the Keywords “Folklore” and “Discourse” in *Web of Science* (2021–2025; N = 15) (Source: Authors’ Calculations Based on *Web of Science* Data)

The country distribution reveals a noticeable asymmetry in research activity and the dominance of certain national centers (see *Figure 2*). The largest contribution is made by Russia (6 publications), which indicates a stable tradition of studying folklore material in a discourse perspective and an active use of linguistic approaches to describing the genre and pragmatic features of texts. Ukraine follows with 2 publications, whereas other countries are represented by single works: Australia, India, Malaysia, Portugal, Slovenia, Turkey, and Turkiye (1 publication each). This structure forms a core–periphery pattern and suggests that the current agenda is largely shaped by an Eastern European scholarly context, while contributions from other countries are episodic and likely associated with individual projects, case studies, or interdisciplinary topics. The simultaneous presence of two naming variants for one country (Turkey and Turkiye) reflects indexing and metadata normalization practices; therefore, subsequent refinements of the sample should merge such duplicates to obtain a more accurate picture of the distribution (see *Figure 4*).

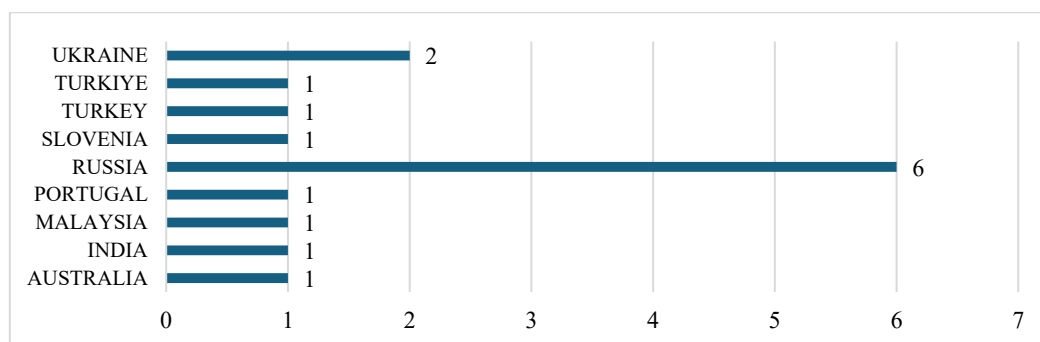


Figure 4. Distribution of Publications by Country in the *Web of Science* Sample for the Keywords “Folklore” and “Discourse” (N = 15) (Source: Authors’ Calculations Based on *Web of Science* Data)

To make the bibliometric block not only descriptive (based on distributions) but also substantively transparent, we present a structured overview of the cited publications identified in the *Web of Science* sample retrieved by the keywords “folklore” and “discourse” (full set N = 15). Due to space constraints, the table reports only items with at least one citation in the WoS export (citations > 0), while records with zero citations are omitted. This overview serves two functions: (1) it provides a transparent view of the most visible works in the dataset and enables rapid verification of the bibliometric “core”; and (2) it supports a shift from general distributions (Figures 1–3) to a more content-focused interpretation by indicating which thematic clusters are represented among the cited literature.

TABLE 1
CITED WEB OF SCIENCE PUBLICATIONS FOR THE KEYWORDS “FOLKLORE” AND “DISCOURSE” (FROM THE FULL SET N = 15): AUTHORS, TITLES, THEMATIC CLASSIFICATION, YEAR, AND CITATION COUNTS (WOS EXPORT; CITATIONS > 0; SORTED BY DECREASING CITATIONS)

<i>Authors</i>	<i>Title of the work</i>	<i>Classification (field → subfield)</i>	<i>Year</i>	<i>Citations (WoS export)</i>
Yesypovych, K.; Pavlova, A.; Siliutina, I. (et al.)	Folklore as an Object of Philological Research: Preservation, Analysis and Interpretation of Folk Creativity	Folklore linguistics → cognitive / communicative / linguocultural approaches	2024	25
Sarmiento, C.	Folk Culture and Political Power: Practices and Representations of Moliceiro Culture in Portugal	Discourse-oriented cultural studies → cultural sociology / representation	2009	2
Sreehari, K. R.; Kasi, E.	Continuity and Change of Folklore among Irulas of Kerala: Discourse Analysis of Tribal Folk Songs from the South Indian State of India	Discourse analysis → ethnolinguistics / oral tradition studies	2024	2
Aleksandrov, O. A.; Bogoslovskaya, Z. M.	Metalanguage Reflexion in the Text Culture of the Russian Germans: Variation of Forms and Meanings	Sociolinguistics → perceptual dialectology / discourse and field linguistics	2021	2
Osman, S. A.	Addressing Rape Culture through Folktale Adaptation in Malaysian Young Adult Literature	Discourse and cultural analysis → gender discourse / adaptation studies	2021	2
Turan, T.	Folklore and Stylistics	Stylistics → folklore discourse stylistics (foregrounding, repetition, parallelism)	2024	1
Mamonova, N. V.; Yukhmina, E. A.	Instagram Network Folklore in the Light of the Linguosynergetics	Digital discourse studies → social media folklore / linguosynergetics	2018	1
Pribac, T. B.; Kaucic, M. G.	Zoofolkloristics: Imagination as a Critical Component	Discourse and culture → eco or zoo folkloristics / methodological reflection	2024	1

The data in Table 1 refine the overall picture obtained from the distributions (Figures 1–3) by focusing on the *cited segment* of the dataset. Although records with zero citations from the full set (N = 15) are not displayed, the cited items already reveal a clear structure of visibility: the highest-impact work is associated with an integrative problem statement on folklore as an object of philological research, whereas the remaining cited publications largely represent sociocultural and discourse-focused case studies in which folklore is examined as a mechanism for representing power, identity, norms, and values. The cited subset also captures the expansion of folklore beyond traditional oral culture (e.g., digital/network folklore) and highlights interdisciplinary perspectives (including eco-/zoofolkloristics), indicating that contemporary “folklore + discourse” research develops across multiple analytical traditions rather than within a single monojournal core. Overall, restricting the table to cited items reduces redundancy while preserving the evidence needed to interpret bibliometric trends and to motivate subsequent comparisons of corpus-based and discourse-oriented approaches to folklore analysis.

III. MATERIAL AND METHODS

Materials of the Study

The material of the study consists of Kazakh folklore texts presented in the *National Corpus of the Kazakh Language*, as well as corpus sub-collections or segments in which folklore is included in a *cultural-representative fund*. The analysis covers genres that most clearly represent folklore discourse in the corpus: *epic narratives and batyr epics, fairy tales, legends and traditions, proverbs and sayings, ritual and everyday-life texts, and song genres*. The choice of a corpus-based source is justified because it provides systematic storage and the verifiability of linguistic observations, which is especially important for texts that historically existed in an oral tradition and are characterized by *variability*. The corpus format makes it possible to view folklore texts not only as individual works but also as an integrated *discursive mass* in which recurring formulas, clichés, typical motifs, and structural patterns can be identified, for example *opening formulas, temporal formulas, stable epithets, and symbolic numerals*. In addition, as an external analytical layer, the study uses the results of a bibliometric sample from *Web of Science* based on the keywords *folklore* and *discourse* (N = 15 publications). This makes it possible to describe a broader “map” of research attention to the topic, including dynamics by year, distribution by country, and distribution by publication venue, and to relate the article’s corpus-discursive objectives to current research trends.

Methods and Instruments

The study is conducted in an interdisciplinary framework at the intersection of *discourse studies, corpus linguistics, folklore studies, textology, and information technologies*, because digitizing folklore texts and integrating them into a corpus is inherently a complex multi-level process. The following methods are applied: (1) *qualitative and quantitative methods* aimed at identifying the features of folklore discourse in the corpus, including the analysis of linguistic units in terms of *frequency, collocation, and structural and functional characteristics*, such as formulas, fixed expressions, evaluative units, recurring patterns, archaisms, symbolic numerals, and related phenomena; (2) *information processing methods* that include collecting and selecting folklore texts, analyzing them, and preparing the material for description both in terms of content and linguistic organization; (3) the method of *generalization and systematization*, which allows the unification and grouping of texts by genre and discourse parameters and their description as a structured corpus-based

dataset; and (4) *bibliometric analysis* based on *Web of Science* data, including a quantitative description of publication activity for the specified keywords and the analysis of distributions by year, country, and publication venue (based on the N = 15 sample) as an additional tool for contextualizing the research field and clarifying its interdisciplinary environment.

The instrumental foundation of the study is the *National Corpus of the Kazakh Language* platform as the source of empirical data and the means of verification. Within corpus work, standard procedures are used, including *corpus search and concordance sampling*, identifying the recurrence of formulas and constructions, observing the distribution of units across genres and corpus segments (where annotation is available), and basic techniques of *corpus statistics*, such as frequency counts, detecting stable combinations, and identifying recurring schematic patterns. Methodologically, the corpus enables the analysis of folklore discourse in both a *quantitative* dimension, including frequency, distribution, and recurrence, and a *qualitative* dimension, including meaning, function, pragmatics, and cultural code.

IV. RESULTS AND DISCUSSION

The corpus-based manifestation of folklore discourse is most clearly observed through epic poetry, fairy tales, proverbs and sayings, and ritual and everyday songs. Although these genres differ in their discursive structure, they share common cultural codes and linguistic patterns. One of the key features of folklore discourse is the stable use of formulaic expressions. Drawing on materials from the National Corpus of the Kazakh Language, we can examine the frequency of recurring structures in discursive texts. For example:

Ерте, ерте, ертеде,

Ешкі жүні бөртеде... (“Long, long ago, when the goat’s wool was bumpy...”)

This structure functions as a discourse marker signaling the beginning of a fairy-tale narrative. Corpus evidence shows that this formula occurs in dozens of variants, while the word «*ертеде*» appears thousands of times. Such a formula introduces the listener into a mythological space, foregrounds the folklore nature of the text, and conveys the conventional (non-literal) character of time and space.

Folklore discourse also frequently contains words that are rarely used in contemporary standard literary language. For example:

Алдасаным жарқ етті,

Ақ сауытын кигізіп... (“My broadsword flashed, having donned the white armor...”)

Here, *алдасан* and *сауым* are not merely weapon-related terms; they are recurrent lexemes typical of heroic epic discourse. Corpus-based observation indicates that these lexemes are used predominantly in heroic-epic contexts alongside semantics of valor and warfare. The high frequency of archaic lexemes in heroic epics (*алдасан*, *сауым*, *найза*) aligns with Potebnya’s (2006) ideas about the cumulative function of language: archaisms preserved in folklore are linguistic traces of earlier historical and cultural periods. In Kazakh linguistics, this perspective was further developed by Zhanpeisov (2013), who specifically examined the historical and cognitive content of epic vocabulary.

Proverbs and sayings constitute the most compact and conceptually dense genre of folklore discourse. Corpus data clearly demonstrate their role in representing cultural concepts. For example: *Қонақ келсе – құт.* (“When a guest comes, it is a blessing.”) In this text, the concept *қонақ* is not only nominative but also evaluative. Corpus materials show that this word frequently co-occurs with lexemes such as *құт*, *береке*, and *ырыс* (“blessing,” “prosperity,” “abundance”). In this way, folklore discourse linguistically reinforces one of the key values of Kazakh culture, namely hospitality. Frequency-based patterns of cultural concepts in proverbs relate directly to the concept theory developed by Karasik and Stepanov (2016), who treat the concept as a point of intersection between language and culture.

Another essential feature of folklore discourse is variability. The National Corpus of the Kazakh Language makes it possible to compare multiple versions of the same text side by side. For example:

Version 1: Жеті күн, жеті түн жол жүріпті. (“He traveled for seven days and seven nights.”)

Version 2: Қырық күн, қырық түн жол кешіпті. (“He journeyed for forty days and forty nights.”)

This difference shows that folklore discourse typically expresses not literal time but conventional time, while symbolic numerals (*жеті*, *қырық*) carry cultural meaning. Corpus comparison can also help identify how such variants depend on geography and the period of collection. The frequency of symbolic numerals (*үш*, *жеті*, *қырық*) resonates with Lévi-Strauss’s (2017) account of mythological thinking, where numerals acquire symbolic value.

Evaluative constructions are also clearly visible in folklore texts, reflecting the linguistic image of a national worldview. For example:

Жаманнан жирен,

Жақсыдан үйрен. (“Shun the bad; learn from the good.”)

Corpus data indicate that the opposition *жақсы* vs. *жаман* (“good” vs. “bad”) is highly frequent in folklore discourse. This binary model reinforces moral norms, regulates social behavior, and performs an educational function. Evaluative oppositions such as *жақсы* – *жаман* and *адал* – *арам* (“honest/pure” vs. “dishonest/impure”) point to the normative character of folklore discourse.

Taken together, the examples above demonstrate that folklore discourse in the National Corpus of the Kazakh Language can serve as a fully valid linguistic data source. The corpus enables the study of folklore texts from quantitative perspectives (frequency, distribution), qualitative perspectives (meaning, function), and comparative perspectives (genre, region, period). Corpus linguistics opens new research opportunities for folklore discourse: by relying on large-scale

textual data, it becomes possible to determine the frequency, distributional range, and usage patterns of specific linguistic phenomena. For instance, the frequency of stable epithets in epic poetry or the use of particular syntactic models in fairy tales can be supported by explicit quantitative evidence. In addition, comparing folklore discourse with contemporary standard language makes it possible to investigate the historical development of Kazakh, linguistic continuity, and change, thereby contributing to a deeper understanding of language evolution. Accordingly, the main linguistic features of folklore discourse can be summarized in *Table 2* below.

TABLE 2
MAIN LINGUISTIC FEATURES OF FOLKLORE DISCOURSE

No.	Feature	Example
1	Structures	<i>Ертеде</i> (“long ago”), <i>жеті күн, жеті түн</i> (“seven days and seven nights”), <i>бай болыпты</i> (“there once lived a rich man”), <i>төрт-мүлігі сай болыпты</i> (“he had all livestock in abundance”), etc.
2	Variability	<i>Ұрысат алып Танаиштан, / Шоң батыр сонда аттанды, / Жаудан кекті алмаққа / Қаруын алып, саптанды</i> (“Having received permission from Tanash, Shon Batyr set off to take revenge on the enemy; he took up his weapons and formed ranks.”) \n <i>Алтайға әкеп жабады. / Қырғызға бірге барам, – деп, / Сұрағанда рұқсат, / Барақ батыр бермеген</i> (“They drive [him/it] toward Altai. Saying, ‘I will go together to the Kyrgyz,’ when he asked for permission, Batyr Barak did not grant it.”)
3	Evaluativity	<i>батыр – оң, жау – теріс</i> (“batyr/hero = positive; enemy = negative”)

Taking these features into account, let us analyze how they are manifested in folklore discourse texts within the National Corpus of the Kazakh Language. We begin with *structures*.

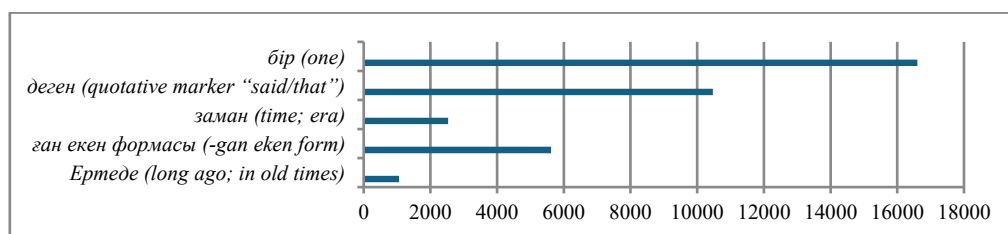


Figure 5. Frequency of Folklore Discourse Structures in the National Corpus of the Kazakh Language

Based on this figure, we can observe that verb-based constructions, as well as the numeral *бір* (“one”), occur with high frequency. This suggests that constructions built around these elements are widely distributed, for example *бір бай болған екен* (“there once lived a rich man”), *бір заманда* (“once upon a time”), and so forth.

TABLE 3
EXAMPLES OF VARIABILITY IN FOLKLORE DISCOURSE TEXTS IN THE NATIONAL CORPUS OF THE KAZAKH LANGUAGE

No.	Type of variability	Examples
1	One suffix realized in two forms: the -менен form	<i>Асыл Алтай сәніменен / Тастан суын қайнатып. / Алтайда бейғам ел жатыр, / Балығы шоршып көл жатыр, / Жер қайысқан шерікпен / Келе жатыр он батыр</i> (“Noble Altai, with its splendor, makes water boil from stone. In Altai, a carefree people live; fish splash in the lake; with an army that bends the earth, ten batyrs are coming.”) \n <i>Алтынды-күміс сәнменен / Сылдыраған әнменен / Көз жанарын алады</i> (“With golden and silver adornment, with a tinkling song, it captivates the eyes.”)
2	Variability of numerals	<i>Жеті күн, жеті түн той жасалды. / Елдің басы қосылды...</i> (“For seven days and seven nights, a feast was held. The people gathered...”) \n <i>Қырық күн, қырық түн той жасалды. / Елдің басы қосылды...</i> (“For forty days and forty nights, a feast was held. The people gathered...”)
3	Omitting letters in use	<i>Ұлықпан хакім пайғамбар, / Жәрдем бер, – деп сұрады. / Мертіккендер орнынан / Төрт-бес күнде тұрады.</i> (“Ulyqpan Hakim, the prophet, asked: ‘Grant help.’ The wounded rise from their place in four or five days.”) \n <i>Луқпан хакім аузында / Шеріктердің жадында / Жәрдем ет, хакім, – деген ед</i> (“On Luqpan Hakim’s lips, in the soldiers’ memory, was the phrase: ‘Help us, Hakim,’ it was said.”) \n <i>Дұшмандар естен танады. / Ер Танаишты іздеген / Абылай, Барақ елшісі / Неше күндей таба алмай</i> (“The enemies lose consciousness. Seeking Er Tanash, Ablai, Barak’s envoy, could not find him for many days.”)

These patterns of variability are directly related to the oral transmission of folklore discourse texts.

Taking the above features into account, we propose a model of a corpus-based representation of folklore discourse. We believe that it should include the following characteristics.



Figure 6. Corpus Model of Folklore Discourse

When folklore texts are incorporated into a corpus, they should be treated not merely as literary or ethnographic evidence, but as a specific discourse system. In this view, folklore discourse is understood as a stable communicative model that consolidates social memory, a collective worldview, and cultural norms through language. An analysis of folklore texts in the National Corpus of the Kazakh Language has identified a number of discourse-specific features. In

particular, *structures* and *repetition* function as the key organizing mechanisms of folklore discourse. For example, traditional opening formulas in fairy tales such as «*Ерме, ерме, ермеде...*» (“Long, long ago...”) and conventionalized time expressions such as «*жети күн, жети түн*» (“seven days and seven nights”) not only signal the genre identity of the text but also introduce the audience into a discourse space. At the corpus level, such formulas can be detected via frequency patterns and treated as markers that enable the automated recognition of folklore texts. In addition, corpus evidence allows the evaluative nature of folklore discourse to be demonstrated more precisely. In heroic epics and legends, the linguistic representation of positive versus negative characters is typically organized through clear binary oppositions. This feature reveals the axiological system of folklore discourse and highlights its educational and ideological functions. Corpus-based analysis makes it possible to compare the frequency, contextual usage, and cross-genre variation of such evaluative units in a systematic way.

In Finnish, Russian, and English folklore corpora, texts are typically annotated with specialized tagging at the genre, motif, and pragmatic levels. By contrast, although the National Corpus of the Kazakh Language contains folklore texts, they have not yet been systematized as a fully developed folklore subcorpus. Therefore, to study Kazakh folklore discourse effectively in a digital format, several scientific and methodological issues must be addressed. First, folklore texts should be identified and separated as a distinct internal subcorpus within the national corpus and tagged with genre and discourse features. Second, annotation models should be developed to mark formulaic structures and recurrent motifs. Third, metadata reflecting the oral nature of folklore texts, including the performer, the time of recording, and the region, should be treated as mandatory components of the corpus structure. This approach integrates traditional qualitative analysis with quantitative evidence and enables a rethinking of folklore within the framework of contemporary digital humanities.

V. CONCLUSION

Folklore discourse within the National Corpus of the Kazakh Language has distinctive scholarly and cultural significance. It enables the preservation of the linguistic and cultural heritage of the Kazakh people in digital space and supports its systematic investigation. Integrating folklore texts into the national corpus is an essential component of the broader digitization of the Kazakh language. The results demonstrate the high scientific potential of examining folklore discourse within a corpus-linguistic framework: folklore materials in the national corpus can serve as an important tool for preserving cultural heritage in a digital format. However, to fully realize this potential, it is necessary to systematize folklore texts as a specialized subcorpus and to deepen genre and discourse annotation. International experience in folklore corpora can provide concrete models in this direction, but the goal should not be mechanical transfer; rather, these solutions should be adapted with careful consideration of the distinctive features of Kazakh folklore.

The study substantiates the scientific and practical importance of building a folklore-text corpus within the digitization of the Kazakh language. A corpus-based investigation of folklore discourse can contribute to the emergence of new directions in Kazakh linguistics and to the effective integration of national cultural heritage into contemporary digital environments. The conceptual and methodological proposals presented in this article can serve as a theoretical foundation for future research in folkloristics, corpus linguistics, and discourse studies.

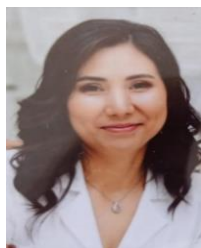
ACKNOWLEDGEMENTS

This article was prepared under the grant financing of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan for 2025-2027 within the framework of the project AP26103909 "Corpus of folklore texts: linguistic and IT solutions in the digitization of the Kazakh language".

REFERENCES

- [1] Acerbi, A., Kendal, J., & Tehrani, J. J. (2017). Cultural complexity and demography: The case of folktales. *Evolution and Human Behavior*, 38(4), 474–480. <https://doi.org/10.1016/j.evolhumbehav.2017.03.005>
- [2] Aitova, N. N., & Bakytqyzy, M. (2025). Qazaq folklorlyq mätinderiniñ korpussy: Ğylymi taldaı jáne zertteu mümkindikteri [Corpus of Kazakh folklore texts: Scientific analysis and research opportunities]. *L. N. Gumilyov atyndagy Euraziia ultyq universitetiniñ habarshysy* [Bulletin of L. N. Gumilyov Eurasian National University], 152(3), 11–27. <https://doi.org/10.32523/2616-678X-2025-152-3-11-27>
- [3] Akhmetova, M., Abilmazhinova, A., Sarekenova, K., Omirbekova, Z., Zhanzhigitov, S., Kamarova, N. S., & Kurmanbekova, Z. (2025a). Sacred vocabulary in linguistic consciousness of the Kazakh youth. *Eurasian Journal of Applied Linguistics*, 11(2), 205–217. <https://doi.org/10.32601/ejal.11217>
- [4] Akhmetova, M. K., Zhanzhigitov, S. Kh., Alpysbayeva, A. Y., Sarekenova, K. K., Bibekov, K., & Abitzhanova, Z. (2025b). The representation of ethnocultural names in the linguistic consciousness of Kazakh youth. *International Journal of Society, Culture & Language*, 13(2), 75–87. <https://doi.org/10.22034/ijsc.2025.2056961.3990>
- [5] Bauman, R. (2004). *A world of others' words: Cross-cultural perspectives on intertextuality*. Oxford: Blackwell Publishing.
- [6] Bauman, R., & Briggs, C. L. (1992). Genre, intertextuality, and social power. *Journal of Linguistic Anthropology*, 2(2), 131–172. <https://doi.org/10.1525/jlin.1992.2.2.131>
- [7] Bortolini, E., Pagani, L., Crema, E. R., Sarno, S., Barbieri, C., Boattini, A., Sazzini, M., da Silva, S. G., Martini, G., Metspalu, M., Pettener, D., Luiselli, D., & Tehrani, J. J. (2017). Inferring patterns of folktale diffusion using genomic data. *Proceedings of*

- the National Academy of Sciences of the United States of America*, 114(34), 9140–9145. <https://doi.org/10.1073/pnas.1614395114>
- [8] da Silva, S. G., & Tehrani, J. J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society Open Science*, 3(1), 150645. <https://doi.org/10.1098/rsos.150645>
- [9] de Lima, E. S., Feijó, B., Casanova, M. A., & Furtado, A. L. (2016). Storytelling variants based on semiotic relations. *Entertainment Computing*, 17, 31–44. <https://doi.org/10.1016/j.entcom.2016.08.003>
- [10] Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. London: Longman.
- [11] Finlayson, M. A. (2017). ProppLearner: Deeply annotating a corpus of Russian folktales to enable the machine learning of a Russian formalist theory. *Digital Scholarship in the Humanities*, 32(2), 284–300. <https://doi.org/10.1093/llc/fqv067>
- [12] Foley, J. M. (1995). *The singer of tales in performance*. Bloomington: Indiana University Press.
- [13] Hagedorn, J., & Darányi, S. (2022). Bearing a Bag-of-Tales: An Open Corpus of Annotated Folktales for Reproducible Research. *Journal of Open Humanities Data*, 8, 16. <https://doi.org/10.5334/johd.78>
- [14] Harvilahti, L. (2013). The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala. *Oral Tradition*, 28(2), 223–232.
- [15] Hilmiyatun, Suwandi, S., Waluyo, H. J., & Wardani, N. E. (2022). Between Ecology and Economics: A Critical Discourse Analysis of Putri Mandalika Folklore. *Theory and Practice in Language Studies*, 12(11), 2388–2396. <https://doi.org/10.17507/tpls.1211.20>
- [16] Kachinskaja, I. B., & Malysheva, A. V. (2014). Ustnaia rech' i fol'klor v korpuse: Printsipy annotirovaniia [Oral speech and folklore in a corpus: Principles of annotation]. In *Trudy Instituta russkogo iazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Institute of the Russian Language] (pp. 214–229). Moscow.
- [17] Karasik, V. I., & Stepanov, Iu. S. (2016). *Iazyk, kul'tura, kontsept: Teoriia i praktika lingvokul'turologii* [Language, culture, concept: Theory and practice of linguocultural studies]. Moscow: Akademicheskii proekt.
- [18] Karsdorp, F., & van den Bosch, A. (2016). The structure and evolution of story networks. *Royal Society Open Science*, 3(6), 160071. <https://doi.org/10.1098/rsos.160071>
- [19] Levi-Strauss, C. (2017). *Mifologicheskoe myshlenie* [Mythological thinking]. Moscow: Akademicheskii proekt.
- [20] Li, X., Sim, N. B., & Zin, Z. B. M. (2024). Translation of Hmong Folklore Terms in Hmong Oral Epic “Creating Heaven and Earth” From the Perspective of Eco-Translatology. *Theory and Practice in Language Studies*, 14(10), 3177–3185. <https://doi.org/10.17507/tpls.1410.20>
- [21] Lin, Y., Halim, H. A., & Jalis, F. M. M. (2024). Building a Parallel Corpus for Chinese Folk Songs Translation Studies: A Case Study of Northern Shaanxi and Hua'er Folk Songs. *Theory and Practice in Language Studies*, 14(2), 454–468. <https://doi.org/10.17507/tpls.1402.17>
- [22] Lo, G., de Boer, V., & van Aart, C. J. (2020). Exploring West African folk narrative texts using machine learning. *Information*, 11(5), 236. <https://doi.org/10.3390/info11050236>
- [23] Mahmud, E. Z., Sobarna, C., Hartono, R., & Moriyama, M. (2022). Chesterman's Pragmatic Strategies in Translating Sundanese Folklore Into English. *Theory and Practice in Language Studies*, 14(4), 1015–1023. <https://doi.org/10.17507/tpls.1404.10>
- [24] Mamysheva, M. E., & Seitova, S. B. (2022). Tarbagatai toponimderindegi ataular semantikasi [Semantics of names in Tarbagatai toponyms]. *Bulletin of L.N. Gumilyov Eurasian National University. Philology Series*, 3(140), 95–102. <https://doi.org/10.32523/2616-678X-2022-140-3-95-102>
- [25] McNery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- [26] Nakawake, Y., & Sato, K. (2019). Systematic quantitative analyses reveal the folk-zoological knowledge embedded in folktales. *Humanities and Social Sciences Communications*, 5, 161. <https://doi.org/10.1057/s41599-019-0375-x>
- [27] Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R., & Winters, J. (2020). How we do things with words: Analyzing text as social and cultural data. *Frontiers in Artificial Intelligence*, 3, 62. <https://doi.org/10.3389/frai.2020.00062>
- [28] Potebnya, A. A. (2006). *Mysl' i iazyk* [Thought and language]. Moscow: Labirint.
- [29] Propp, V. Ia. (1998). *Morphology of the folktale*. Austin: University of Texas Press.
- [30] Qaidar, A. (2004). *Qazaq etnolingvistikasy* [Kazakh ethnolinguistics]. Almaty: Eltanym.
- [31] Qasqabasov, S. (2011). *Qazaq fol'klorynyń tipologiiasy* [Typology of Kazakh folklore]. Almaty: Dáuir.
- [32] Qongyratbaev, Á. (1991). *Qazaq fol'klorynyń tarihy* [History of Kazakh folklore]. Almaty: Ana tili.
- [33] Ross, R. M., & Atkinson, Q. D. (2016). Folktale transmission in the Arctic provides evidence for high bandwidth social learning among hunter–gatherer groups. *Evolution and Human Behavior*, 37(1), 47–53. <https://doi.org/10.1016/j.evolhumbehav.2015.08.001>
- [34] Sagymzhanova, T. B., & Sagyndyk, N. B. (2024). Typological similarities in the folklore traditions of Persian and Turkic peoples. *Bulletin of L.N. Gumilyov Eurasian National University. Philology Series*, 4(149), 175–183. <https://doi.org/10.32523/2616-678X-2024-149-4-175-183>
- [35] Tangherlini, T. R., & Chen, R. (2024). Travels with BERT: Surfacing the intertextuality in Hans Christian Andersen's travel writing and fairy tales through the network lens of large language model-based topic modeling. *Oral Literature*, 79(6), 519–562. <https://doi.org/10.1111/oli.12458>
- [36] Tehrani, J., Nguyen, Q., & Roos, T. (2016). Oral fairy tale or literary fake? Investigating the origins of *Little Red Riding Hood* using phylogenetic network analysis. *Digital Scholarship in the Humanities*, 31(3), 611–636. <https://doi.org/10.1093/llc/fqv016>
- [37] Wiysahnyuy, L. F., & Valentine, N. B. (2023). Folktales as indigenous pedagogic tools for educating school children: A mixed methods study among the Nso of Cameroon. *Frontiers in Psychology*, 14, Article 1049691. <https://doi.org/10.3389/fpsyg.2023.1049691>
- [38] Yuniari, I. G. A. M. E., Kasni, N. W., & Sutjaja, I. G. M. (2023). Transitivity and social context in Balinese folklores: A Systemic Functional Perspective. *Theory and Practice in Language Studies*, 13(5), 1295–1303. <https://doi.org/10.17507/tpls.1305.25>
- [39] Zhanpeisov, E. (2013). *Qazaq eposynyń tili: Tarihi-tanymdyq zertteu* [The language of the Kazakh epic: A historical and cognitive study]. Almaty: Arys.



Omirbekova Zhamila Kaldybekovna is an Associate Professor at the Department of Kazakh Linguistics at L.N. Gumilyov Eurasian National University, where she has been teaching since September 2005. She earned the degree of Candidate of Philological Sciences in 2005. Her research interests include lexicology, cognitive linguistics, and ethnolinguistics, and she teaches courses such as Ethnolinguistics. Her professional experience includes work as a secondary school teacher in Almaty (1999–2000), Senior Inspector of the Scientific Council at Al-Farabi Kazakh National University (2001–2003), and postgraduate study there (2003–2005). ORCID iD: 0000-0002-3162-0515 Email: ozhamila@inbox.ru



Karlygash Kurmangalikyzy Sarekenova (born 3 May 1974 in present-day Zhetisu Region, Kazakhstan) received her BA (1996), MA in Linguistics (1998), and postgraduate training in Kazakh language studies (10.02.02, 2000) from Al-Farabi Kazakh National University. Since 22 May 2001, she has worked at L.N. Gumilyov Eurasian National University. She has published 19 papers in Scopus-indexed journals with non-zero impact factor, 5 papers in Web of Science-indexed journals, and 44 papers in journals recommended by Kazakhstan's national quality assurance committee, and has authored 3 textbooks, 3 teaching manuals, 3 monographs (plus 4 co-authored), 1 e-learning textbook, and 4 methodological guides, alongside numerous conference publications. ORCID iD: 0000-0003-1383-2247 Email: sareke.kk@gmail.com



Gulgaisha Sagidolda (born 16 August 1963 in Mongolia) is a Doctor of Philology and Professor, recognized as “Best University Teacher of Kazakhstan” (2008, 2015). She graduated in 1986 from Al-Farabi Kazakh National University (then S.M. Kirov Kazakh State University) with a degree in Kazakh Language and Literature, and has worked as a senior researcher at Kazakhstan's Academy of Sciences institutes (1987–2005), held professorial and chair roles at L.N. Gumilyov Eurasian National University (2005–2010). She has participated in 18 research projects (2001–2024), authored 6 monographs and 7 textbooks, and published 100+ scholarly articles; her contributions also include national standards and model curricula in Turkology and Philology. ORCID iD: 0000-0002-4811-9743 Email: asem963@mail.ru



Beisenbai Ardak Bakytuly (*correspondent author*) is a member of the Board and Vice-Rector for Academic Affairs at L.N. Gumilyov Eurasian National University. He graduated from the university with a degree in *Kazakh Language and Literature* and has built his academic career there, progressing from Assistant at the Department of Kazakh Linguistics (2005) to Lecturer (2006), Senior Lecturer (2009), Deputy Dean for Educational Work of the Faculty of Philology (2011), Deputy Dean for Research (2012), Head of the Department of Kazakh Linguistics (2016), and Dean of the Faculty of Philology (2019). He is also a recipient of the Bolashak International Scholarship of the Republic of Kazakhstan. ORCID iD: 0000-0001-7785-650X Email: beisenbai_ab@enu.kz



Bauyrzhan Maratuly Oserov (born 25 October 1991 in Kostanay Region, Kazakhstan) earned his BA in Philology (2012) and MA in Kazakh Language and Literature (2014) from L.N. Gumilyov Eurasian National University. Since 20 November 2013, he has worked as a leading specialist at the National Center for Higher Education Development under the Ministry of Science and Higher Education of the Republic of Kazakhstan, and has published two articles in journals recommended by Kazakhstan's national quality assurance committee. ORCID iD: 0009-0009-4106-2930 Email: b.m.91@bk.ru



Syrym Zhanzhigitov was born on September 6, 1996, in Bike village, Mamlyut District, North Kazakhstan Region. In 2023, he defended his dissertation and was awarded a PhD degree in “Kazakh Language and Literature” from L.N. Gumilyov Eurasian National University (Astana, Kazakhstan). He is currently pursuing postdoctoral studies under the project AP25793567 “Pedagogical Model for the Development of Linguoconflictological Competence of Future Legal Professionals” (2025–2027). He is the author of more than 24 articles in national journals, 8 articles in Scopus-indexed publications, and a participant in over 35 international and national conferences. ORCID iD: 0000-0002-7814-1378 Email: syrymphd@gmail.com