

UDC 811.512.122'1

IRSTI 16.31

<https://doi.org/10.48371/PHILS.2025.4.79.004>

## LEXICOGRAPHIZATION OF ETHNO-CULTURAL UNITS OF THE NATIONAL CORPUS

\*Barmenkulova A. S.<sup>1</sup>

\*<sup>1</sup> al-Farabi Kazakh National University, Almaty, Kazakhstan

**Abstract.** The article addresses the issue of lexicographing ethnocultural units in accordance with the national corpus of the Kazakh language. It focuses on the ethnocultural values – worldview-related, spiritual-material, and sociocultural – preserved in the language and shaped by the historical and social development of the Kazakh people, their way of life, traditions, and interaction with their natural surroundings. This research is relevant within the framework of the modern anthropocentric paradigm, which also seeks to reveal the functional significance of the native language. The study aims to systematize ethnocultural units in the Kazakh language and substantiate their representation in the national corpus from a scientific and methodological perspective.

The research employs descriptive, comparative-historical, structural-semantic, and corpus-linguistic methods. The main focus and challenge of the study lie in systematizing ethnocultural units within the structure of the national linguistic corpus, identifying their semantic characteristics, and exploring ways to preserve and utilize these units in digital format.

The scientific and practical significance of the research is that the accumulation of ethnocultural units in the national corpus of the Kazakh language and the elucidation of their semantic structure contribute to expanding the body of linguistic knowledge. In this regard, the cultural and spiritual attributes of traditional Kazakh items such as *kebezhe* (a wooden chest), *qazan* (a large cauldron), and *torsyq* (a leather flask) can be understood through ethnocultural units. The study's findings and conclusions involve the collection and semantic analysis of ethnocultural units, as well as their preparation for integration into the corpus database. Furthermore, these units have been incorporated into the theoretical framework of an ethnocultural subcorpus.

The theoretical and practical significance of the research lies in its contribution to ethnolinguistic and linguocultural studies within Kazakh linguistics, based on the national corpus of the Kazakh language. The findings enhance the cognitive potential of linguistic units and emphasize the role of language in reflecting national culture and identity. This issue is particularly important for the development of the national corpus of the Kazakh language as a comprehensive cognitive information resource and an advanced linguistic technology aligned with contemporary paradigms.

**Keywords:** National corpus of the Kazakh language, ethnocultural units, anthropocentric paradigm, automatization, dictionary, cultural-semantic meaning, lexicography, cultural-semantic subcorpus

## Introduction

The article's central focus is the introduction of ethnocultural units into the National Corpus of the Kazakh language. The classification of ethnocultural units and the analysis of the names of subject cultures are also explored while justifying the need for their inclusion in the corpus. Ethnocultural units are linguistic symbols with a significant cumulative role in preserving the Kazakh national language, traditions, worldview, and spiritual culture, ultimately representing the national identity. Therefore, an essential matter is the transmission of Kazakh images of the world through ethnocultural units to the modern generation. Consequently, it is our responsibility to imbue the present generation with our national values and bring the National Fund of Knowledge to future generations. There is a pressing need to establish national corpus for linguistic and linguodidactic studies. The Kazakh national corpus serves as an effective and informative tool that preserves and transfers national-cultural units across generations. This article focuses on the lexicographic methods for incorporating ethnocultural elements into the Kazakh national corpus.

Nowadays in Russian and foreign linguistics, a direction called corpus linguistics is developing. The aim of corpus linguistics is the study of speech patterns in the material of large volumes of text (corpuses), pre-processed, labelled and systematized in electronic form. The difference between such text corpora and simple electronic libraries is that in them, with the help of special programs, it is possible to search for fragments of necessary texts according to given parameters, and further generalize and analyses the obtained data. The search itself takes only a short time, measured in seconds.

The extensive advancement of contemporary computer technologies and software has sparked the attention of researchers towards corpus linguistics, resulting in remarkable prospects for information acquisition and processing. In the past, scholars were obligated to spend numerous months and years manually retrieving essential data from libraries; yet currently, this information is readily accessible within a few minutes with the click of a button. However, the potential benefits of such information capabilities must be considered alongside significant challenges. The most pressing among these is the prevalence of overwhelming amounts of data. The advent of the Internet has undoubtedly made accessing information easier, faster, and more affordable. However, it has also brought about a new challenge – navigating the vast sea of data to locate credible and relevant information. We believe that V.P. Zakharov's definition provides the most comprehensive understanding of the “corpus” concept. The corpus, as defined by the researcher, is a comprehensive collection of linguistic data that is sizeable, electronically stored, structured, labelled and designed to tackle specific linguistic objectives [1]. In general, this definition can be characterized as functional, as it centers on the linguistic aspects of organized text collections. Thus, the definition of “corpus” varies in the following ways: firstly, a corpus typically comprises numerous texts in electronic form such as those found on the Internet

or discs. Secondly, linguistic data must be labelled to facilitate its analysis for linguistic purposes. Finally, following the analysis, there should be an option to distribute the linguistic material obtained according to its genre, year of creation, and subject matter, but solely on a statistical basis. Kazakhstan's spiritual and cognitive progress highlights the need to establish the National Corpus of the Kazakh language. It is worth noting that the establishment of national corpora in foreign linguistics began in the 1960s, while in Russian linguistics, it started in the 1980s. Only in recent years has the field of Kazakh linguistics begun to place greater emphasis on the use of corpus linguistics. Professor A. Zhubanov highlighted the importance of corpus linguistics in the Kazakh language, making him a pioneer in this field. According to scholarship, corpus linguistics, as a unique area of Kazakh linguistics, enables Kazakh journalists to access ample practical material, competently locate the requisite linguistic data, and suitably process it. This article presents a fresh perspective on empirical techniques that ensure the legitimacy and dependability of research on the Kazakh language. Furthermore, it assists in integrating pertinent linguistic data into academic discussion [2, p. 5-6]. A dictionary is a valuable repository of language resources. The lexicon ought to comprise figurative words that have impacted society and are present in a publicly accessible dictionary. The vocabulary and phraseological wealth presented in dictionaries not only represent linguistic heritage and treasures, but they also hold meaning for all members of society. This information reflects the centuries-old worldview of our people and their relationships with nature, as well as their expertise in mastering natural resources. As such, it can collectively contribute to the material and spiritual wealth of our ethnocultural units. The representation of national identity among diverse ethnocultural groups reveals a unique language attribute that continuously evolves to incorporate present-day values and circumstances. It encompasses the revitalization and modification of concepts, substance, terminology, aesthetics, layout, pedagogical strategies, and ambitions over time. The transfer of ethnocultural units in explanatory dictionaries of the Kazakh language is not consistently implemented, and therefore, the communication of national values to dictionary users requires further refinement. The development of methods and techniques for the lexicography of ethnocultural units in Kazakh is crucial. This issue pertains to the lexicography of ethnocultural units in the National Corpus of the Kazakh Language. It is essential to investigate and distinguish the techniques used to incorporate ethnocultural components into the national corpora worldwide. Furthermore, this exploration will spotlight the methods of lexicography of ethnocultural units within the semantic structure of the Kazakh language by generalizing linguistic units with ethnocultural connotations in the explanatory dictionary.

### **Methods and materials**

The study of lexical units has advanced notably within the field of linguistics. Techniques such as accumulation and grouping (classification) are employed

to gather ethnocultural units in the Kazakh language. Explanatory dictionaries utilise descriptive, synonymic, pictorial, and illustrative methods to ascertain the lexicons of ethnocultural units. The integration of ethnocultural elements into the National Corpus is achieved through the use of coding, modelling, and algorithmic programming procedures.

## **Results**

From this perspective, the current corpus provides information coverage in the cultural studies.

The dictionary-lexical set of language vocabulary, social group dialects, and the works of writers hold significant cultural and spiritual roles in our society.

A dictionary is a reference work that offers alphabetical explanations of word meanings, phrases, and idioms [3]. Since ancient times, dictionaries have been instrumental in interpreting foreign languages. Their use dates back to the 5th century BC in Greece, the 12th century BC in India, and the 13th century in Russia. The first dictionary of Turkic people was authored by Mahmut Kashgari in the 11th century and is known as «Divani lugat-at-Turk» Later, multilingual or bilingual dictionaries were created. For example, when analysing the contrast between Arabic-Persian and Turkic-Mongolian terminology, there are various significant studies to take into account. Initially, Zamakhshari's «Muqaddi mat al-adab» dictionary from the 12th century provides useful insights. Furthermore, Abu Hayyan's «Kitab al-idrak Li Lisan al-adab» grammar and collection of words from the 14th century offer valuable information. Furthermore, the «Bugat al-Mushtak» dictionary by Jamal al-Din from the 15th century, which focuses on the translation between Arabic and Kipchak, is of great significance. This is followed by the work of P. M. Melioransky, entitled «An Arab Philologist on the Turkish Language».

The extensive range of dictionaries is broadly classified into encyclopaedic dictionaries of a linguistic, philological, and reference nature. Dictionaries function as socially significant tools that provide normative guidance on the meaning and usage of words. Their purview includes theoretical and practical applications of vocabulary methodology and language learning technology in lexicography. Encyclopaedic dictionaries contain concepts, sources, or subjects related to a specific field, while alphabetical linguistic dictionaries aim to clarify particular linguistic vocabulary and phraseology. These dictionaries are produced when the language is fully established, the language culture is mature, and there is a requirement for a standard definition of the vocabulary.

The scholars listed above – I. Kenesbaev, M. Balakaev, G. Musabaev, A. Iskakov, K. Akhtanov, R. Syzdyk, A. Kaidar, Sh. Sarybaev, U. Aitbaevich, B. Kaliev, K. Husain and M. Malbakov – have significantly advanced the field of Kazakh linguistics, particularly in the area of lexicography.

The cultural and spiritual heritage of a nation is its invaluable wealth. Because of the concepts of human cognition, the real world is best expressed through collaboration. The lexical and cultural semantics of words arise from

people's memory, encompassing their diverse and figurative meanings that facilitate language communication. It is undeniable that the quantitative nature and diverse semantic possibilities of a word, transforming it into a cultural subject, entail a noble duty for every representative of a nation. The Kazakh language, along with other global languages, is in no way inferior in terms of vocabulary and dictionary composition. For the first time, a two-volume Explanatory Dictionary of the Kazakh language has been published. In 1999, the initial edition of the Explanatory Dictionary of the Kazakh language was released as a single «Dictionary of the Kazakh language», which summarised the people's vocabulary. The National Dictionary functions as a cultural and linguistic reference to standardise the language by presenting a comprehensive lexicon. The socio-cultural role of spoken language has been strengthened. The «Dictionary of the Kazakh Literary Language», recently published, is the most comprehensive explanatory dictionary of the Turkic language. The dictionary is of substantial quality, as it covers spiritual, cultural, and educational content, with pertinent updates from the latest academic research [4]. Based on the fifteen-volume «dictionary of the Kazakh literary language», a one-volume «Kazakh dictionary» has been published [5]. It is imperative to review and expand the lexicon to improve the vocabulary within the dictionary. Linguistic dictionaries come in various types, such as monolingual, multilingual, and translation dictionaries. Translation dictionaries play a critical role in second language education. Language learners need to learn not only individual vocabulary but also culturally significant terms and their corresponding translations.

An explanatory dictionary records individual words in a register, explaining their formation and revealing their cultural connotations with relevant examples. Therefore, learners of the Kazakh language can fully appreciate the cultural significance of each word. To enable this opportunity for language learners, the first step should be to create an explanatory dictionary of cultural words in this language. This should take priority over translation dictionaries. To achieve our objective, we will develop a comprehensive glossary of cultural terms. Furthermore, we intend to incorporate ethnocultural concepts into the cultural database at a future date.

Lexicographic research is necessary firstly for dictionary compilation and secondly for both discriminative and applied linguistics. For example, before conducting a study, it is required to find out what information lexicographers need. For example, the authors' main request when compiling the Academic Interpretative Dictionary of the Russian Language necessitated the search for the following information:

- a new word that appeared in specific periods;
- an initial form of the word;
- examples (quotations) revealing words with known meanings;
- to find an example of meanings not given in the dictionary;
- new examples that are given to a meaning;
- new types of lexical and syntactic combinations;

- new phraseological phrases;
- new scientific explanation of special terms [6].

Lexicographic patterns exhibit systemic interrelations. For instance, conventional methods identify clusters of similar words, while corpus-based lexicographic analyses illustrate the usage of related words in multiple situations and contexts.

In the field of traditional lexicography, the distribution of lexicographic information based on lexico-semantic and cultural-semantic categories of ethnocultural units possesses distinctive features in automated lexicography. Crucial to these matters is an examination of algorithmic and logical factors in programme creation, along with the exclusive use of data stored in computer memory. Automating the Kazakh language in lexicography can improve vocabulary technology. The lexicons found in the corpus enable lexicographers to enhance their provision of essential language materials and information for users conducting language searches.

D. Baibera, S. Conrad, R. Reppen's Book «Corpus Linguistics. Investigation language structure and use» [7]. (further – Corpus Linguistics). The researcher, working through a corpus approach, separately examines the six main problems facing lexicographers:

1. What meanings are associated with a specific word?
2. What is the frequency of use of a particular word compared to the frequency of other words adjacent to it?
3. What are the non-linguistic models of a given word (about case, historical periods, dialects, etc.)
4. What words are usually found together with a given word, and how to use their combination in different registers is coordinated in the sequence.
5. How are the meanings and types of word usage coordinated?
6. How are synonymous words used and how are they used in their place in different contexts? [8].

The objective side of lexicography research using a corpus approach lies in the derivation of a set of concordances that broadly illustrate a given word's usage in the corpus. From these concordance lists (KWICs), it is possible to extract various meanings associated with the word in different contexts.

The primary objective of corpus creation is to simplify the research work of technicians in such fields. For instance, a corpus represents not only a complex field of inquiry but also a potentially consequential discovery.

The scale of the concordance in the corpus will be considerable and may thus generate a wealth of data. In extensive corpuses, the number of concordances could run into several thousand pages, with one instance possibly accommodating hundreds of concordances [9].

## Discussion

Ethnocultural units are preserved much more along phrasal word



combinations than in individual words. So, since the corpus is one of the sources of information that represents our culture, it must be understood how important it turns out to include such phrases in abundance in the Corpus. These updates in the Corps are revealed in the practice of lexicography, and they are transmitted in the Corps from beginning to end in a single system. Thus, a stable model of the lexicography of ethno-cultural units was developed. In the corps, this model is useful for distinguishing the lexicography of ethno-cultural units. For example, in the Corps, ethnocultural units are given the following interpretation:

1. The National Corps of the Kazakh Language is compiling a database of source texts in Excel to transfer to the Ethno-Cultural Units Corpus.

2. The texts of the ethno-cultural units collected in Excel are then processed for inclusion into the main database.

3. The ethnographic units collected through the Excel program have been integrated into the main database and linked to the cultural and semantic corpus.

4. Ethnographic units are designated in the corpus with the label (ethn.).

5. The lexical meaning of each ethnographic unit is explained in the corpus.

6. The corpus reveals and categorizes the culture and semantics of each ethnographic unit.

7. The Corpus includes examples of ethno-cultural units taken from works of fiction.

8. Ethnocultural units are discussed in proverbs and sayings, which are recorded in the Corpus.

9. The Corpus provides phraseological phrases relating to ethno-cultural units.

10. The Corpus offers illustrative materials that depict ethno-cultural units.

11. The culture and semantics of buildings are explored through the Corpus.

12. Search work for each designation is improved as part of this study.

We will provide a corpus of ethno-cultural units, offering the examples as follows.

In the Kazakh perspective, a plethora of ethnocultural units encompasses a wide range of proverbs, sayings, tales and rituals. It is stated that the cauldron holds significant cultural importance for the Kazakh people. For instance, the

*Cauldron* ethn., which is cast from bronze, copper, or iron, is utilized for cooking purposes.

The Kazakhs have long revered the black cauldron. Each member of the family, whose pot was always close to the fire, nailed it to the source of energy. A boiler house was built at the place of residence. In nomadic circles, it was specially designed for a cauldron made of dense linen felt so that it would not break. A worthy vessel that survived the country's famine. In times of war, many children survived by hiding under the kettle. This pot is used to cook almost all the dishes on the table. Cooking kurt, boursaki (pastry), making irimshik (cheese), meat soup and many others are cooked in a black pot. The pot is placed in a three-legged hearth or hearthstone. There is a saying «Black pot, yellow boy condition»

that goes back to ancient times. When they say, «When the cauldron shakes, the blessing hums», many thoughts come to mind. The Cauldron was considered the source of life, the blessing of the homeland. On a difficult day, the Black Cauldron did not throw the old man out. He believed that «prosperity is our happiness». When the girl married, her parents gave her all their possessions, but not the Black Cauldron. The Kazakh never gave the Black Cauldron as a gift. The Cauldron could be used temporarily when the neighbors gave a feast, but after using the Cauldron they put food in it and the container was not left empty [10, p. 454].

1. on most of the earthly hearths a cauldron is boiling, white foam is boiling, and abundant meat is boiling [11, p. 454].

2. in every hearth a hundred small earthly hearths are boiling, a cauldron is boiling, and the cauldron is boiling abundant flesh [12, p. 454].

3. The old woman alone washed the cauldron, filled it with water and put the meat in [13, p. 25-26]. Examples of fiction that use the cauldron are given.

Cauldron is the source of life, a symbol of the beginning of creation. With such symbolic concepts as Cauldron, there are regular turns of words in our language that refer to Cauldron, which shows that proverbs and sayings have a place in people's lives. For example, «a cauldron of satiety - no worries». Fire «cauldron other sadness separate» fire another homeland other; black cauldron-symbol of unity, peace of the country: one cauldron drank; stirring up enmity, two cauldrons drank into one house. These symbols and signs are cultural and linguistic representations of Cauldron in the consciousness of the nation. The country is the well-being, the prosperity, the world of people.

1. If you say I will put on Allah,

Let it be Azan.

If you say that you will put on man,

Let it be a cauldron.

3. A spoon clatters out of the broken cauldron.

Since ancient times, the concept of Cauldron, the sacred, has had great significance. It is a superstition of our people, which is only good.

Chest ethn. A box for storing utensils. Ethnographic works note that in the description of each of them, its shape was square, with the only difference being that the height of the four thresholds was increased by 10-15 cm from the main post [10, p. 454]. Ethnographic data about the chest play an important role in the everyday life of this product. These characteristics of the product are reflected in phraseological expressions. For example: Chest belly, wide belly – 1) the most common alternative to courts in oral literature; 2) used in fiction as an equivalent of rich people who came thick, stocky.

Chest, belly, striped chest – in the oral literature of the twentieth century and works on modern historical themes were used judge, mayor. In the rich language of our country there are a lot of proverbs and sayings that indicate the special place of the breast, emphasise it and accurately convey the knot of ninety words. There is a group of proverbs about daughters-in-law, such as «a



good daughter-in-law is a daughter-in-law, a bad daughter-in-law is death», «a daughter-in-law is not a bad person, but a bad place», «a good daughter-in-law is like your daughter, a bad groom is like your son». He went into the kitchen, put the chest down and the earth began to lay the table: – Shaltik has been sighing all day, snoring, snuffling, can't sleep! – Amina giggles [14, p. 197].

Waterskin is an ethnomedical drinking vessel known as «kymyz», typically crafted from the neck leather of camels, deer, and bulls. Waterskins are typically made from the hides of camels, horses, and cows, as well as camel necks. The waterskin is crafted by sewing together hump, horn, and camel neck skin.

Since shepherds always have their drinks with them, they are referred to as soul waterskin carriers. The horn exterior is adorned with an array of decorations. The vessel, adorned with intricate carvings, is referred to as a leather container for liquids. The bevelled hides are then stacked on top of one another, and their edges are sutured with copper or thick thread. Both sides of the vessel are intricately drawn upon using the «perch» (button) tool. A small «ear» was extracted from the neck, affixed with a rivet, and punctured with a small round hole where a complete ribbon was attached. She adorned herself with a braid of ribbon. After completing the waterskin, it undergoes a process of drying, sanding, washing, and cleaning. The leather water container was subsequently dried, coated with oil, treated with ghee smoke, and utilized for producing liquor.

When the chicken, which had been placed in a container, was boiled and cooked, it became entangled and packed together, like people gathering at twilight in a crowded space. The container was intentionally carried out of the village in a vessel tied to the shoulder, and the individuals poured honey kymyz as they chatted [15, p. 454].

«On a bad horse, to fatten its withers, one does not tie a waterskin», «A bad man runs out of cattle, next to his neighbour he will not pull up» He who has happiness settled in his head and his reputation raised, is in the habit of looking at others with contempt, keeping away from them, and taking to the skies. Such a person, for example, gets very fat and does not tie a waterskin to the horse that uses it, but, judging from the shock, he runs out of cattle, begins to get caught, is jealous of the bad man, and keeps others out of the neighbourhood.

The National Corpus provides extensive opportunities, particularly for scholars in correspondence, to conduct a thorough investigation into the cultural semantics of this language. Thanks to the significant advancement in computer technology, manual work is now greatly facilitated, and data can be effortlessly entered and retrieved. Thus, a language that aims to decipher the evolution of language and its direction can provide valuable insights to researchers.

The National Corpus allows for diverse scientific investigations.

It enables a comprehensive analysis of the lively expressions of Kazakh material culture, which are adeptly utilised in daily life, based on research conducted within the National Corpus system.

## Conclusion

Language Corpus has a great potential for application in all aspects of linguistics, including education, compiling diverse dictionaries (such as frequency, explanatory, regional, phraseological, ethnographic, etc.), and producing educational materials and translations.

Since ethnocultural linguistic units are embedded in a computer database within language corpus, operating under a control system, it is feasible to acquire knowledge of such units from corpus materials. The transference of ethnocultural linguistic units in the corpus holds significant importance. One of the most intricate projects encompassing various texts is the creation of a corpus of ethnocultural linguistic units.

There are currently no entries for ethnocultural units within any corpus. In our future research, it is essential to consider methods for integrating ethnocultural vocabulary into the Kazakh National Corpus. This study aims to introduce ethnocultural terminology into the National Corpus of the Kazakh language. Theoretical issues regarding lexicography and corpus research in the Kazakh language have been explored extensively, with particular emphasis on the lexicography of ethnocultural terms.

## REFERENCES

- [1] Zakharov V.P. Korpusnaja lingvistika: Uchebno-metod, posobie [Corpus linguistics: Educational method, manual]. – SPb., 2005. – 48 s. [in Rus.]
- [2] Zhubanov A.K. Principy avtomatizacii morfologijacheskoj razmetki tekstov Nacional'nogo korpusa kazahskogo jazyka (NKKJa) [Principles of automation of morphological markup of texts of the National Corpus of the Kazakh Language] // Materialy Mezhdunarodnyj nauchno-prakticheskoy konferencii «kontrastivnye issledovanija i prikladnaja lingvistika» 29-30 oktjabrja 2014 goda v Minskom gosudarstvennom lingvisticheskom universitete. – Minsk. – 4 komp.s. [in Rus.]
- [3] Malbaqov M. Bır tıldı tüsındırme sözdıktıñ qırylymdyq negızderı [Structural foundations of a monolingual Explanatory Dictionary]. – Almaty, 2002. – 368 b. [in Kaz.]
- [4] Qazaq ädebi tılınıñ sözdıgı [Dictionary of the Kazakh literary language]. On bes tomdyq. 1-tom. – Almaty, 2011. – 752 b. [in Kaz.]
- [5] Qazaq sözdıgı [Kazakh dictionary]. (Qazaq tılınıñ birtomdyq ülken tüsındırme sözdıgı) – Almaty: «Däur» baspasy, 2013. – 1488 b. [in Kaz.]
- [6] Grishina E.A, Savchuk S.O. Nacional'nyj korpus ruskogo jazyka kak instrument dlja izuchenija variativnosti grammaticheskikh norm [National corpus of the Russian language as a tool for studying the variability of grammatical norms] // Trudy mezhdunarodnyj konferencii «Korpusnaja lingvistika – 2008» 6-10 oktjabrja 2008 g. – SPb., 2008. – S. 161-19-69. [in Rus.]
- [7] Zhubanov A. O koncepcii po sozdaniju nacional'nogo korpusa kazahskogo jazyka (NKKJa) [On the concept of creating a national corpus of

the Kazakh language] // «Kontrastivnye issledovaniya i prikladnaya lingvistika: materialy mezhdunar. Nauch. Konf. – Minsk: MGLU, 2015. – 227 s. [in Rus.]

[8] Zhanabekova A. Rol' lingvisticheskoy annotatsii na opyte sozdaniya nacional'nogo korpusa kazahskogo yazyka [The role of linguistic annotation on the experience of creating a national corpus of the Kazakh language] // Sbornik: «Problemy sovremennoj prikladnoj lingvistiki» (otvetstvennyj – A.V. Zubov). – Minsk: MGLU, 2013. dekabrja. – S. 2012-216. [in Rus.]

[9] Jūbanov A. Qazaq leksikografiasynyñ («Tıl – qazyna» atty elektrondyq bazasy jaiynda [About the electronic base of Kazakh lexicography (“Tıl – kazyna”)] //Elektrondy-aqparattyq keñistiktegi qazaq tılı: terminologia jäne tıl мәдениеті. QR Bılım jäne Ğylym ministrlyğı Ğylym komitetitınıñ A.Baitürsynūly atyndaǵy Tıl bılımı instituty üymdastyrǵan respublikalyq ğylymi-teorialyq konferensia materialdary. – Astana 2009. [in Kaz.]

[10] Qazaq etnografialyq kategorialar, űǵymdar men ataularynyñ дәstürlı jüiesı. Ensiklopedia [Traditional system of Kazakh ethnographic categories, concepts and names]. 1-tom. – Almaty: 2011. – 738 b. [in Kaz.]

[11] Tolybaev Q. «Äset» [Aset].– Almaty: 2011. – 197 b. [in Kaz.]

[12] Jünisov S. «Şyǵarmalary» [Works]. – Almaty, 2011. – 33 b. [in Kaz.]

[13] Mailin B. «Ait künderi» [Ait festivities]. – Almaty, 2011. – 17-18 b. [in Kaz.]

[14] Müstafin Ğ. «Köz körgen» [Koz korgen]. – Almaty, 2011. – 10-12 b. [in Kaz.]

[15] Tarazi Ä. «Qııryqty jüldyz» [Comet]. – Almaty, 2011. – 25-26 b. [in Kaz.]

## ЭТНОМӘДЕНИ БІРЛІКТЕРДІ ҰЛТТЫҚ КОРПУСТА ЛЕКСИКОГРАФИЯЛАУ

\*Барменкулова А.С.<sup>1</sup>

<sup>1</sup>Әл-Фараби атындағы Қазақ ұлттық университеті  
Алматы, Қазақстан

**Аңдатпа.** Мақала этномәдени бірліктерді қазақ тілінің ұлттық корпусына сай лексикографиялау мәселесіне арналған. Қазақ халқының тұрмыс кешкен ортасына, қоршаған табиғатына сай қалыптасып, тілде сақталған өзіндік тарихи-әлеуметтік даму ерекшеліктері мен өмір-тіршілігіне, тұрмыс-салты мен дәстүрлеріне байланысты дүниетанымдық, рухани-материалдық, этномәдени құндылықтарын зерттеу заманауи антропоэлектік парадигма аясында өзектеледі. Сондай-ақ осы парадигма аясында ана тілінің қолданыстық функцияларының мәнін ашу көзделеді. Зерттеудің мақсаты – қазақ тіліндегі этномәдени бірліктерді жүйелеп, олардың ұлттық корпуста көрініс табуын ғылыми-әдіснамалық тұрғыдан негіздеу. Зерттеу әдістері – сипаттамалық, салыстырмалы-тарихи, құрылымдық-семантикалық, корпус-лингвистикалық әдістер қолданылды. Зерттеудің басты бағыттары және проблемасы – ұлттық тілдік корпусның құрылымында этномәдени бірліктерді жүйелеу, олардың семантикалық

ерекшеліктерін анықтау және осы бірліктерді цифрлық форматта сақтау мен пайдалану жолдарын қарастыру.

Зерттеудің ғылыми және практикалық маңызы – этномәдени бірліктердің қазақ тілінің ұлттық корпусында жинақталып, олардың мағыналық құрылымын ашу арқылы тілдік білім қорының кеңеюі. Осыған орай жинақталған этномәдени, рухани сипаттағы кебеже, қазан, торсық сияқты мәдени бұйымдардың ұлттық сипаты мен менталды ерекшеліктерін, мәдениеті мен руханиятын этномәдени бірліктер арқылы тани аламыз. Зерттеудің нәтижелері мен қорытындысында этномәдени бірліктер жинақталып, олардың мәдени-семантикалық мағыналары ашылады, сондай-ақ корпусық базаға енгізуге дайындалды. Сонымен қатар, этномәдени бірліктер этномәдени ішкорпустың теориялық мәселелері ретінде толықтырылды. Зерттеу нәтижелерінің теориялық және практикалық маңызы – алынған нәтижелер қазақ тілінің ұлттық корпусы негізінде қазақ тіл біліміндегі этнолингвистикалық, лингвомәденитанымдық зерттеулерді толықтыруға және тілдік бірліктердің танымдық әлеуетін арттыруға бағытталады. Бұл мәселе тіл арқылы ұлт мәдениетін (болмысын) қазіргі жаңа парадигмаға сай лингвистиканың кешенді танымдық ақпарат көзі, озық технологиясы ретінде қазақ тілінің ұлттық корпусын дамытуға өзінің үлесін қосуымен де маңызды.

**Тірек сөз:** Қазақ тілінің ұлттық корпусы, этномәдени бірліктер, антропоэлектік парадигма, автоматтандыру, сөздік, мәдени-семантикалық мағына, лекикография, мәдени-семантикалық ішкорпус

## ЛЕКСИКОГРАФИРОВАНИЕ ЭТНОКУЛЬТУРНЫХ ЕДИНИЦ В НАЦИОНАЛЬНОМ КОРПУСЕ

\*Барменкулова А.С.<sup>1</sup>

\*<sup>1</sup> Казахский национальный университет имени аль-Фараби  
Алматы, Казахстан

**Аннотация.** Статья посвящена проблеме лексикографизации этнокультурных единиц в соответствии с национальным корпусом казахского языка. В рамках антропоцентрической парадигмы актуализируется исследование сохранившихся в языке своеобразных историко-социальных особенностей развития казахского народа, сформировавшихся в соответствии с местами его проживания и окружающей средой, а также познавательных, духовно-материальных и этнокультурных ценностей, связанных с его жизнедеятельностью, бытом и традициями. Также в рамках этой парадигмы предполагается раскрытие смысла употребительных функций родного языка. Цель исследования – систематизировать этнокультурные единицы казахского языка и обосновать их присутствие в национальном корпусе с научно-методической точки зрения. Методы исследования: описательный, сравнительно-исторический, структурно-семантический, корпус-лингвистический. Основные направления и проблема исследования –

систематизация этнокультурных единиц в составе национального корпуса, определение их семантических особенностей и рассмотрение путей сохранения и использования данных единиц в цифровом формате.

Научная и практическая значимость исследования – расширение языкового фонда знаний путем сбора этнокультурных единиц в составе национального корпуса и раскрытия их смысловой структуры. В связи с этим этнокультурные и духовные особенности таких культурных предметов, как кебеже, казан, торсык, а также их национальная специфика и ментальные характеристики могут быть осмыслены через этнокультурные единицы. В ходе исследования были систематизированы этнокультурные единицы, раскрыты их культурно-семантические значения. Этнокультурные единицы подготовлены для включения в корпусную базу. Помимо этого, были дополнены теоретические аспекты этнокультурного подкорпуса. Теоретическая и практическая значимость результатов исследования заключается в том, что полученные данные направлены на дополнение этнолингвистических и лингвокультурологических исследований в казахском языкознании на основе Национального корпуса казахского языка, а также на усиление познавательного потенциала языковых единиц. Этот вопрос также важен своим вкладом в развитие Национального корпуса казахского языка как комплексного когнитивного источника информации и передовой технологии, соответствующей новой парадигме осмысления национальной культуры (идентичности) через язык.

**Ключевые слова:** Национальный корпус казахского языка, этнокультурные единицы, антропоцентрическая парадигма, автоматизация, словарь, культурно-семантическое значение, лексикография, культурно-семантический подкорпус

*Мақала түсті / Статья поступила / Received: 16.04.2025.*

*Жариялауға қабылданды / Принята к публикации / Accepted: 25.12.2025.*

***Information about the author:***

Barmenkulova Aida Serikhanovna – PhD student of al-Farabi Kazakh National University, Almaty, Kazakhstan, <https://orcid.org/0000-0002-1471-5846>, e-mail: [bs.aida@mail.ru](mailto:bs.aida@mail.ru)

***Автор туралы мәлімет:***

Барменқұлова Аида Серікханқызы (корреспондент автор) – әл-Фараби атындағы Қазақ ұлттық университетінің PhD докторанты, Алматы, Қазақстан, <https://orcid.org/0000-0002-1471-5846>, e-mail: [bs.aida@mail.ru](mailto:bs.aida@mail.ru)

***Информация об авторе:***

Барменкулова Аида Серикхановна – докторант PhD Казахского национального университета имени аль-Фараби, Алматы, Казахстан, <https://orcid.org/0000-0002-1471-5846>, e-mail: [bs.aida@mail.ru](mailto:bs.aida@mail.ru)