# A. Karshigayeva[1], A. Mursal[2*], G. Tlegenova[3], E. Maralbek[4]

*[1]Turan University, Almaty, Kazakhstan;*
*[2,3,4]A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan*
*(e-mail: aigerimmursal@mail.ru)*

# Study of the lexical composition and peculiarities of language use based on the frequency dictionary of Abay's works

The article analyzes the history of the creation of frequency dictionaries of the Kazakh language, their place in linguistic-statistical studies, as well as their significance for the study of the language system. A linguistic and structural analysis of the frequency dictionaries compiled based on the works of Abai Kunanbayev, with the aim of determining the frequency and patterns of usage of lexical units employed by the poet was conducted. During the study, both frequently and rarely used vocabulary were systematized, and their significance was examined for identifying morphological, syntactic, and lexical features of the Kazakh language. It was found that frequency dictionaries play an important role in the language learning process, particularly in teaching Kazakh as a foreign or second language. The relevance of frequency dictionaries is also emphasized in the context of creating a language corpus, systematizing the frequency of lexical unit usage, developing machine translation systems, creating text editors, and advancing the Kazakh language in the digital space. A comparative analysis of vocabulary with varying degrees of usage frequency in Abai's works was conducted, demonstrating its place in the historical development of the Kazakh language. In conclusion, the significance of frequency dictionaries for identifying the structural, functional, and stylistic features of the Kazakh language is thoroughly examined, along with an analysis of their role in both theoretical and applied linguistics. Frequency dictionaries are considered an important linguistic tool in the research, teaching, and development of the Kazakh language.

*Keywords:* frequency dictionary, linguostatistics, word usage, language corpus, statistical study.

## *Introduction*

Frequency dictionaries of the Kazakh language were developed on the basis of one of the important directions of linguistics — the method of linguostatistics. Although statistical data is often used in scientific research on the Kazakh language, the number of scientists specializing in linguistic statistics is limited. The close connection of linguistics with numerical methods has also contributed to the interest of specialists in the field of mathematics in this direction. In the 70s of the 20th century, the prominent Kazakh mathematician K. Bektayev opened the Department of Applied Linguistics at the Institute of Linguistics and systematized the work on compiling frequency dictionaries. During this period, frequency dictionaries were published in the Kazakh language for the first time, and they became a valuable methodological tool for language teaching and linguistic research.

Scientists such as A.K. Zhubanov, A. Akhabayev, S. Myrzabekov contributed to the compilation of frequency dictionaries. These dictionaries were developed covering various authors and genres. However, in the research process, manual calculation of the frequency of words in texts became difficult. To solve this problem, there was a need to use automated calculation methods. However, despite the insufficient development of computer technologies in Kazakhstan at that time, domestic researchers conducted frequency analysis using electronic calculators.

Numerous facets of the Kazakh language depend on the development of frequency dictionaries and their application in linguostatistical research. First of all, they enhance the process of teaching Kazakh effectively and let language learners choose the most important vocabulary. Second, dictionaries give lexicographers useful information that serves as the foundation for the development of national Kazakh language corpora. Furthermore, the utilization of frequency dictionaries aids in identifying the quantitative features of the language and becomes a crucial instrument in the creation of text processing and machine translation systems. These dictionaries can aid in the optimization of editing work in printing. Furthermore, frequency dictionaries play a significant role in the process of the Kazakh language's conversion to the Latin alphabet, the

---

* Corresponding author's e-mail: *aigerimmursal@mail.ru*

control of keyboard letter layout, the systematization of orthographic standards, and the creation of new orthographic rules.

In general, the results of the study allow to understand more deeply the structure of the Kazakh language, to determine the process of development of its vocabulary composition, to systematize language units by quantitative feature.

*Methods and materials*

One of the main sources for the study of Abai's language is the "Dictionary of Abai's Language". This dictionary was used to systematize lexical units in the poet's works and to determine the frequency of their use. The study of Abai's language helps to analyze the linguistic expression of historical, philosophical and religious concepts, to study the evolution of word meanings, to describe the process of development of the Kazakh literary language.

The investigation was conducted using the linguostatistical analysis method. In order to ascertain the structural and functional nature of the lexicon, this approach enables us to evaluate the lexical and grammatical characteristics of Abai's writings using quantitative indicators [1]. Additionally, it enables us to examine the frequency of word forms in the text in order to determine their semantic load.

The purpose of this study is to analyze the lexical fund of the Kazakh language and its structural features on the basis of linguostatistical methods. In particular, the aim is to study the patterns of Abai Kunanbayev's linguistic word usage by determining the frequency of words used in his works. A. Zhubanov, A. Zhanabekova, B. Karbozova and A. Kozhakhmetova's "Frequency Dictionary of the Kazakh Language" is considered to be the most important methodological aid and is an important linguistic work that allows for a systematic description of the structure of the Kazakh language.

In scientific literature, the compilation of frequency dictionaries is widely used when studying the creative heritage of a particular author. According to scientists, by identifying the most frequently used words in the text, it is possible to gain a deeper understanding of the main themes, style and linguistic features of the author [2]. In this regard, we grouped frequently used words in Abai's works and analyzed their morphological, syntactic and semantic features.

This work consists of several stages. At **the first stage**, a text corpus of Abai's works was created and data from the texts were prepared for digital processing. At **the second stage**, the frequency of word usage was determined and their statistical indicators were calculated. At **the third stage**, by interpreting the obtained results from the linguistic point of view, the place and use of the ancient lexical stratum in Abai's works in the modern Kazakh language were determined.

*Results and discussions*

The growth of modern technological capabilities has created conditions for bringing linguistic research to a new level. From this point of view, the need to reconstruct frequency dictionaries of the Kazakh language is quite obvious. At present, frequency dictionaries are created on the basis of small text corpuses and databases of national languages. A corpus is a set of data characterizing structural and lexical features of texts written in a certain language [3; 58]. These resources are based on a wide range of textual materials, including works of various authors and linguistic programmes. Once the national corpora of the Kazakh language are formed and begin to function fully, the development of frequency dictionaries of various types using special software will be greatly simplified.

One of the main fields of linguistic statistics is the compilation of frequency dictionaries. However, this field is not limited to the calculation of word frequencies. On the contrary, frequency indicators are used to determine the role of lexical units in the text structure, to assess the semantic load of words, to describe the functional system of a language. For this reason, statistical data play an important role in linguistic research and can be effectively used for lexical and morphological analyses.

One of the topical issues of modern Kazakh lexicology is to determine the boundaries of the modern lexical system of the Kazakh language. In this regard, the importance of studies based on the frequency of word usage is increasing. Compilation of frequency dictionaries contributes not only to applied linguistics, but also to theoretical aspects of Kazakh language study. The significance of such dictionaries can also be assessed by comparing them with other language systems. For example, a comprehensive frequency dictionary including stylistic features of the Russian language has been created. This dictionary not only contributed to the development of the Russian language, but was also widely used in the process of teaching it to foreign citizens. Proof of this is that in her research N.S. Kamarova noted that Abai is a user of the rich value of the

folk language, as well as a poet creating a new word, a new meaning, and also considers the purpose of creating Abai's dictionary. In the course of her research the scientist seeks answers to the questions about how many words used in the writing of the great poet's works, in what sense, analyses foreign words and features of their use, included in Abai's dictionary in the Kazakh language [4; 284].

The Russian scholar M.V. Khokhlova, who studied the compatibility of the Russian language frequency dictionary, text corpus and lexical units, notes that a number of statistical indicators used in lexicographic studies are used to differentiate, represent lexicon on different grounds. In her research, the scientist gives a brief historical overview of the use of quantitative methods in text analysis and discusses issues related to statistical lexicography [5; 88].

The Azerbaijani scholar M.T. Umudova in her research divided the words she studied into linguistic, lexical and semantic, grammatical, morphological, poetic and lexical groups depending on the degree and how they are currently used in the language. The statistical method received its second development due to the development of information technology and is actively engaged in solving applied problems, including the processing and presentation of textual data [6].

Currently, several stylistic directions are included in the development of the "Frequency Dictionary of the Kazakh Language". In particular, frequency dictionaries have been created for scientific style, official-documentary style, journalistic style, artistic-literary style and colloquial style. Thus, it can be assumed that the proposed frequency dictionary fully covers various functional and stylistic layers of the Kazakh language. This dictionary contains 2,102,496 word uses, consisting of 30,515 word combinations, and is presented in three different structures: an alphabetic-frequency dictionary, a reverse alphabetic-frequency dictionary and an frequency-alphabetic dictionary.

Frequency dictionaries are of particular importance in determining the functional and lexical structure of a language. In addition, they allow analyzing the place and scope of words in the general lexical system of the vernacular language. In particular, the ability to determine the level of usage of certain lexical groups with the help of frequency dictionaries, to control the penetration of new words (neologisms) into the Language and to study the regularities of the formation of terminological systems is enhanced. Such databases help to identify lexical richness and stylistic peculiarities. In addition, data from frequency dictionaries can be a valuable source for studying language typology, comparing the Kazakh language with other Turkic languages or other language systems. According to the data from the *Abai Language Dictionary*, the 10 most frequently used verbs in Abai's works are presented below (Fig.).
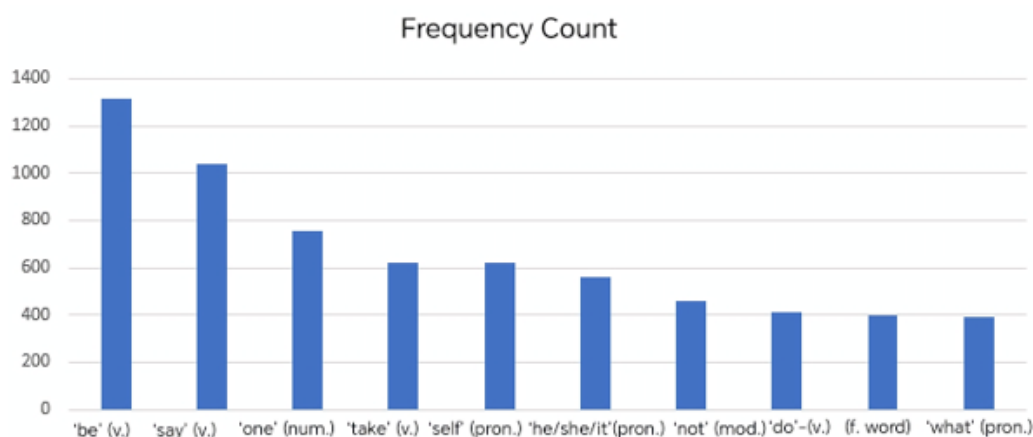


Figure. The 10 most frequently used verbs by Abai according to the data from the "Abai Language Dictionary"

In his works, Abai Kunanbayev paid special attention to expanding the synonymy of the Kazakh language. The poet used some words in several meanings, expanding their semantic possibilities. Researcher R. Syzdyk noted: *"The fact that some simple words acquired an evaluative connotation and became species names, contributed to both the growth of synonymy and ambiguity of words. For example, the words pysyk* (nimble)*, ku* (sly)*, tentek* (naughty)*, ury* (thief) *often occur in Abai's works as names characterizing a certain social group (tentek* (naughty) *— 17 times, ury* (thief) *— 21 times, pysyk* (nimble) *— 20 times, ku* (sly) *— 37 times)"* [7; 28]. The growing interest in Kazakh culture worldwide is evidenced by the increasing number of publications and studies on Abai's legacy in China, Turkey, Iran, and other countries. Thus, despite some

"distortions", Abai's works continue to captivate readers due to their philosophical depth and artistic mastery [8].

The creative heritage of Abai Kunanbayuly is one of the valuable topics worthy of study not only in Kazakh literature, but also at the world level. In this regard, since the 20th century, the poet's works have been considered in various scientific directions. As a result of these studies, the **"Dictionary of Abai's Language"** was compiled, which contains statistical data on the words used by the poet [9]. According to these data, Abai's personal vocabulary is **6030** words. In addition, according to the analysis conducted by the scientist K. Bektayev, who founded Kazakh linguistic statistics, using electronic computers, a total of **46850** word usages were recorded in the poet's works, and the number of unique words is estimated at **6017** [9; 87].

The frequency of use of the above-mentioned words "tentek" (naughty), "ury" (thief), "pysyk" (nimble), "ku" (sly) in Abai's works is as follows (Table 1):

T a b l e   1

**Statistical analysis of the use of the words "tentek" (naughty), "ury" (thief), "pysyk" (nimble), "ku" (sly) in Abai's works**

| Word | Frequency of use: | %: |
|---|---|---|
| Tentek (naughty) | 17 | 0,28 |
| Ury (thief) | 21 | 0,35 |
| Pysyk (nimble) | 20 | 0,33 |
| Ku (sly) | 37 | 0,61 |

These data allow a deeper understanding of the expansion of word meanings in Abai's linguistic use, the development of synonymic series, and the structure of the poet's vocabulary. Based on the information from the "Abai" encyclopedia, scientists from the A. Baitursynuly Institute of Linguistics compiled the "Language Frequency Dictionary" based on Abai's collection, published in 1977. This dictionary contains text material consisting of a total of 46,596 words, and the words are systematized according to their lexical, morphological, and syntactic features [10]. In addition, the words in the dictionary are divided into word classes, and statistical indicators of each category are determined.

According to dictionary data, 15,758 nouns are used in Abai's works. The number of verbs is 15,266, and the total use of pronouns is 4,473 times. In addition, it has been determined that 4,021 adjectives, 2,046 interjections, 1,686 adverbs, and 1,074 numerals occur in the poet's works [10; 35]. Abai's creative heritage covers a variety of topics. In particular, the themes of nature, love, philosophical reflections, and homeland are widely reflected in the poet's poetry. These themes significantly influenced his lexical use and formed a system of frequently used units in his vocabulary (Table 2).

T a b l e   2

**Main themes in Abai's work and the most frequently used words within the themes**

| Love | Nature | Philosophical reflections | Native land |
|---|---|---|---|
| My blessing | Mountain | Wisdom | Kazakh people |
| Love | Water | Understanding | People |
| Be loved | Stone | Thoughts | My land |
| My favorite | Flower | World | My native land |
| Heart | Summer | Life | Mother land |
| My ray | Winter | Depth | Kazakh |
| My sweet | Grass | Perfection | Land |

Abai Kunanbayuly's creative path is divided into several chronological periods, and his literary heritage has passed through various stages of development. Researcher A. Zhanabekova says in this regard: "The evolution of vocabulary in Abai's works clearly demonstrates his personal development as a poet and thinker. At first he used simple, understandable language, but over time the words in his works began to acquire a deeper meaning and to be characterized by emotional intensity. These changes indicate the philosophical maturity of the poet and the expansion of his views on the world around him. Therefore, this evolution is not limited only to the increase in vocabulary, but also reflects the inner spiritual quest of Abai" [11] — she emphasizes.

Statistical study of Abai's language and vocabulary allowed us to identify 100 words most used by him (Table 3). This information is compiled on the basis of the "Dictionary of Abai's language", the list was compiled by B.N. Nurzhanova suggested [12].

T a b l e   3

**A list of the 100 most frequently used words based on the Abai language dictionary by B.N. Nurzhanova**

| Word list | Frequency count | Word list | Frequency count |
|---|---|---|---|
| Bol — 'be' (v.) | 1317 | qudai — 'God' (n.) | 160 |
| de — 'say' (v.) | 1037 | is — 'work' (n.) | 160 |
| bir — 'one' (num.) | 757 | zhurek — 'a heart' (n) | 158 |
| al — 'take' (v.) | 618 | osy — 'this' (pron.) | 151 |
| oz — 'self' (pron.) | 618 | dunie — 'world' (n.) | 145 |
| ol — 'he/she/it'(pron.) | 562 | kerek — 'need' (mod) | 145 |
| zhoq — 'not' (mod.) | 458 | Alla — 'Allakh' (n.) | 134 |
| Qyl — 'do'-(v.) | 409 | narse — 'a thing' (n.) | 133 |
| da (f. word) | 398 | zhaqsy — 'good' (adj.) | 130 |
| Ne — 'what' (pron.) | 391 | bai — 'rich' (adj) | 129 |
| sez — 'feel' (n.) | 376 | Qazaq — 'Kazakh' (people) | 127 |
| sol — 'that' (pron.) | 352 | qara — 'look' (v.) | 126 |
| kor — 'see' (v.) | 338 | oila — 'think' (v.) | 126 |
| ber — 'give' (v.) | 337 | ish (aux.pron) | 125 |
| de — 'say' (f. word.) | 334 | Yet — 'od' (v.) | 123 |
| bil — 'know' (v.) | 317 | qol — 'hand' (n.) | 122 |
| I (per. pron) | 315 | oi — 'thought' (n.) | 122 |
| bul — 'this' (pron.) | 309 | biz — 'we' (pron.) | 121 |
| kel — 'come' (v.) | 304 | omir — 'life' (n.) | 113 |
| zhur — 'go'(v.) | 295 | yeki — 'two' (num) | 111 |
| bar — 'have' (mod.) | 285 | gylym — 'science' (n.) | 111 |
| ait — 'say' (v.) | 271 | zhet — 'reach' (v.) | 110 |
| adam — 'a man' (n.) | 270 | zhas — 'young' (adj.) | 109 |
| eken — 'would'(ax.v.) | 243 | zhol — 'way, path' (n.) | 107 |
| sen — 'you' (pron) | 240 | ushin — 'for' (f. word.) | 107 |
| e (v.) | 233 | ol — 'die' (v.) | 106 |
| tur (v.) | 230 | bar — 'go' (v.) | 104 |
| zhan — 'soul'(зат) | 220 | qoi –'stop' (v.) | 103 |
| kop — 'many' (adv.) | 217 | qait — 'be back' (v.) | 96 |
| konil — 'mood' (n.) | 216 | izde — 'search' (v.) | 94 |
| mal — 'cattle' (n.) | 214 | bai — 'reach' (adj) | 91 |
| tap — 'find' (v.) | 210 | endi — 'else'(adv.) | 90 |
| ket — 'go away' (v.) | 209 | ui — 'house' (n.) | 89 |
| qal — 'stay' (v.) | 204 | paida — 'gain' (n.) | 84 |
| bala — 'a boy, a baby' (n.) | 204 | men –'and' (f.word) | 83 |
| koz — 'eye' (n.) | 200 | til — 'tongue, language' (n.) | 81 |
| aqyl — 'mind' (n.) | 199 | ot — 'pass' (v.) | 81 |
| kisi — 'a man' (n.) | 195 | tus — 'get off' (v.) | 77 |
| zher — 'land' (n.) | 195 | zhe — 'eat' (v.) | 74 |
| bari — 'al' (pron) | 193 | at — 'shoot' (v.) | 73 |
| Yel — 'people' (n.) | 183 | qyzyq — funny/interesting (n.) | 72 |
| yemes — 'is not' (v.) | 182 | zhane — 'and' (f. word) | 69 |
| ma — 'Here!' (pron) | 174 | ozge — 'other' (adj) | 69 |
| sal — 'put'(v.) | 173 | qaida — 'where' (pron) | 68 |
| kun — 'day/sun'(n.) | 171 | qaigy — 'sadness' (n.) | 66 |
| birey –'somebody' (pron.) | 168 | Goi — (f. word) | 66 |
| shyq — 'go out' (v.) | 166 | Zhat — 'lie down' (v.) | 66 |
| son — 'after' (f. word) | 165 | me (q.pron.) | 64 |
| bas — 'head' (n.) | 163 | Yer — 'brave' (adj) | 64 |
| kim — 'who' (pron.) | 162 | sui — 'love' (v.) | 63 |

Statistical studies of the works of Abai Kunanbayuly allow us to draw certain conclusions. Of course, we cannot say that the obtained figures fully characterize the quality of the poet's language and the Kazakh language in general, but their significance cannot be ignored. The results of comparing Abai's frequently used words with words frequently used in the Kazakh language in the second half of the 19th century prove that his language is a reflection of the Kazakh literary language of that era. In this regard, frequency dictionaries and semantic dictionaries of Abai's works are a valuable source in the process of learning the Kazakh language. The use of frequency dictionaries at the initial stage of language learning is considered one of the most effective methods. Using frequency dictionaries based on Abai's works as a teaching tool helps Kazakh language learners understand the frequency of use of words, master the language structure, and enrich their vocabulary.

In Abai's works, one can notice that certain words are repeated many times. In addition, a number of words used by the poet appear only once in his works. These words reflect the poet's style, the richness of his vocabulary, and certain terminological features. Below are some examples of words that appear only once in Abai's works (Table 4).

T a b l e  4

**Words used only once by Abai in his works**

| Words used by Abai once in his works | Meanings of words: |
|---|---|
| 'Tegindik' (for free (gratis)) | A neologism used as a translation of the word "ne darom" (for free) in the translation of M. Lermontov's work |
| 'Zhersogar' (tiller) | A word from the noun class. The heel part of a horse's hoof. |
| 'Tezdik'(speed) | Not in vain, used in the sense of not being a descendant. |
| 'Ustirtteu' (lightly) | Taking things lightly, not taking the matter seriously. |
| 'Ulyqtau' (glory) | To respect, to appreciate |
| 'Turaqtau' (to settle in) | To settle, to live |
| 'Tiriltu' (revitalise) | To revive, to restore |

This data describes Abai's unique use of language. It also shows that the poet chose to use certain words to convey certain shades of meaning. Such studies contribute to a better understanding of the development of the Kazakh literary language and to the improvement of language learning methodologies.

"The Dictionary of Abai's Language", which systematically describes the vocabulary in the works of Abai Kunanbayuly, is one of the unique works that, apart from explanatory dictionaries, also contains the frequency of words and word forms. Another work that complements the Abai dictionary and is similar in content is the frequency dictionary in the novel «Abai Zholy» (Abai's Path) by M. Auezov. This dictionary allows us to study not only the words used in Abai's works, but also the structure of the Kazakh language in his era [13].

According to statistical analysis, the 135 most common words in Abai's language account for 50 % of his entire text. This data helps to identify the features of the poet's vocabulary and to understand his style more deeply. In addition, the share of words used by Abai only once or twice is 64 %. However, these words account for only 8 % of the entire text. This indicator proves not only the richness of the Kazakh language vocabulary, but also the lexical scope of Abai's work.

As for the research of foreign researchers, Russian scholar E.A. Morozkina, V.V. Vorobyev, Z.M. Safina, while studying the translation of the work of the American writer William Gilmore Sims "Grayling; or Murder Will Out" written in 1841, provide information on the 10 most common words in the original and translated texts. Scientists conclude that this is due to the increased interest of researchers in the possibilities of applied methods of computational linguistics in studying the quantitative component of the text corpus and some linguistic phenomena of the originals and their translations [14; 135].

The scientific article "The Practice of Creating Dictionaries of the Language of M. Auezov", which studies the proportion of the most frequently used words in Abai's vocabulary, provides the following data:

– 75 most frequently used words account for 40.8 % of all word usage;
– 150 words account for 50.8 %;
– 1000 words account for 80 % of the text [15].

For comparison, according to the frequency dictionary of modern Kazakh newspapers compiled by A. Akhabayev, the 57 most frequently used words cover 25 % of the entire text, 301 words — 50 %, 1168 words — 75 %, and 1602 words — 80 %. These data allow us to compare the structural features of

Abai's language with the Kazakh press language [16; 102]. As for the rare words in Abai's works, 2975 words used only once by the poet constitute 49.4 % of his entire vocabulary. The frequency of these words covers 6.35 % of all text units. Meanwhile, in studies taken from the texts of Kazakh newspapers, 5720 words used once constitute 46.04 % of the total words and 3.81 % of all text units.

The above data show that frequency dictionaries have great theoretical and applied significance. They are an effective tool for language teaching, creating teaching aids, identifying language minima, and conducting lexicographic research [17]. In addition, frequency dictionaries are widely used in improving information search systems, developing machine translation, and text processing programs.

## Conclusion

It can be seen that dictionaries based on the works of Abai Kunanbayuly occupy an important place in Kazakh linguistics. The words used by the poet reveal linguistic features not only of the era in which he lived but also of the present day. Therefore, these dictionaries are among the most valuable lexicographic works in the Kazakh language.

This article presents a statistical analysis of the most frequently used words in Abai's works, as well as a list of rare words. The conducted research demonstrates the vastness of the poet's vocabulary, proving that it reflects the richness and quality of the Kazakh language as a whole. These dictionaries are an extremely important tool for studying the lexical structure of the Kazakh language.

The frequency dictionary of Abai's works allows you to determine the frequency of use of words in the Kazakh language. In turn, these indicators help to understand the general patterns of the Kazakh language, its development direction and structural features. In addition, dictionaries created on the basis of Abai's works contribute to the study of historical changes, thematic features, grammatical structure and lexical layers of the Kazakh language.

By comparing Abai's most frequently used words with common words in the Kazakh language, it was determined that they can be used for teaching the Kazakh language. These dictionaries are considered valuable material for Kazakh language learners.

## Acknowledgements

## References

1 Мурсал А. Абай шығармаларындағы ескі кітаби лексика және оның лингвостатистикалық ерекшеліктері / А. Мурсал, А.Ә. Жаңабекова, Г.Б. Тлегенова // Tiltanym. — 2024. — № 4 (96). — Б. 187–200. https://doi.org/10.55491/2411-6076-2024-4-187-200

2 Жұбанов А.Қ. Қазақ тілінің жиілік сөздігі / А.Қ. Жұбанов, А.Ә. Жаңабекова, Б.Д. Карбозова, А.Қ. Қожахметова. — Алматы: «Қазақ тілі» баспасы, 2016. — 792 б.

3 Жұбанов А. Корпустық лингвистика / А. Жұбанов, А. Жаңабекова. — Алматы: «Қазақ тілі» баспасы, 2017. — 336 б.

4 Камарова Н.С. Иностранные слова, включенные в словарь Абая на казахском языке: особенности использования / Н.С. Камарова // Сборник материалов Международного симпозиума Томского государственного университета. Секция 6. Словари образных средств языка и художественной речи. — 2021. — С. 280–286. http://dx.doi.org/10.17223/978-5-907442-19-1-2021-93

5 Хохлова М.В. Статистические методы в лексикографических исследованиях: представление частотной лексики / М.В. Хохлова // Terra Linguistica. — 2023. — Т. 14. — № 3. — С. 80–93. https://doi.org/10.18721/JHSS.1430

6 Umudova M.T. Types of Obsolete Words (Archaisms and historicisms) / M.T. Umudova // International Journal of Social Science And Human Research. — 2022. — Volume 05. — Issue 12. — P. 5897–5900. https://doi.org/10.47191/ijsshr/v5-i12-74

7 Сыздық Р. Абай шығармаларының тілі / Р. Сыздық. — Алматы, 1968. — 172 б.

8 Ахметова Ш. Абай тілі сөздігі / Ш. Ахметова, Ә. Болғанбаев, Т. Жанұзақов. — Алматы, «Өнер», 2011. — 616 б.

9 Бектаев Қ.Б. М.О. Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері / Қ.Б. Бектаев, А.Қ. Жұбанов, С. Мырзабеков, А.Б. Белботаев. — Алматы-Түркістан, 1995. — 346 б.

10 Белботаев А. Абай тілінің жиілік сөздіктері және оның маңызы мен мәні туралы / А. Белботаев, Ж. Бұланбаева. — Алматы, 2014. — 213 б.

11 Жаңабекова А. Абай Құнанбайұлы тілінің жиілік сөздігі / А. Жаңабекова. — Алматы. 2023. — 667 б.

12 Нуржанова Б.Н. Жазушы тілі сөздігін түзудің ғылыми-теориялық негіздері / Б.Н. Нуржанова. — Алматы, 2018. — 112 б.

13 Бектаев Қ.Б. М. Әуезовтің «Абай жолы» романының жиілік сөздігі / Қ.Б. Бектаев, А.Қ. Жұбанов, С. Мырзабеков, А.Б. Белботаев. — Алматы, 1979. — 334 б.

14 Морозкина Е.А. Статистические методы исследования в художественном переводе. / Е.А. Морозкина, В.В. Воробьев, З.М. Сафина // Доклады Башкирского университета. — 2023. — Том 8. — № 3. — С. 130–138. https://doi.org/10.33184/dokbsu-2023.3.15

15 Жұбанов А. М. Әуезов тілінің сөздіктерін жасау тәжірибесі / А. Жұбанов, С. Мырзабеков. — Алматы: А. Байтұрсынұлы атындағы Тіл білімі институты, 2022. — 89 б.

16 Ахабаев А. Қазіргі қазақ газеттері тілінің алфавитті-жиілік сөздігі / А. Ахабаев // Қазақ тексінің статистикасы. — Алматы, 1973. — 731 б.

17 Mursal A. Archaic Bookish Words in Abai Kunanbayuly's Works: A Linguostatistical Analysis / A. Mursal, G. Tlegenova, A. Zhanabekova, A. Cherepanova // Forum for Linguistic Studies. — 2025. — 7(3). — 195–208. https://doi.org/10.30564/fls.v7i3.8506

А.А. Қаршығаева, А. Мурсал, Г. Тлегенова, Е. Маралбек

## Абай шығармаларының жиілік сөздігіндегі лексикалық құрылым және тілдік қолданыс ерекшеліктері

Мақалада қазақ тіліндегі жиілік сөздіктердің құрылу тарихы, олардың лингвостатистикалық зерттеулердегі орны және тілдік жүйені зерттеудегі маңыздылығы талданған. Абай Құнанбайұлының шығармалары негізінде жасалған жиілік сөздіктерге лингвостатистикалық және құрылымдық талдау жүргізіліп, ақын қолданған сөздердің жиілігі мен таралу заңдылықтары анықталған. Зерттеу барысында Абай шығармаларында жиі және сирек қолданылатын сөздер жүйеленіп, олардың қазақ тілінің морфологиялық, синтаксистік және лексикалық ерекшеліктерін айқындаудағы маңызы қарастырылды. Зерттеу нәтижесінде жиілік сөздіктердің тіл үйрету үдерісіндегі рөлі, оның ішінде қазақ тілін шет тілі немесе екінші тіл ретінде оқытудағы тиімділігі айқындалды. Сондай-ақ тілдік корпус құру, лексикалық бірліктердің қолданыс жиілігін жүйелеу, машиналық аударма жүйелерін әзірлеу, мәтіндік редакторлар жасау және қазақ тілін цифрлық кеңістікте дамыту мәселелерінде жиілік сөздіктердің өзектілігі негізделді. Абай шығармаларындағы жиі және сирек қолданылатын сөздерге салыстырмалы талдау жасалып, олардың қазақ тілінің тарихи дамуындағы орны көрсетілді. Қорытындыда жиілік сөздіктердің қазақ тілінің құрылымдық, функционалдық және стильдік ерекшеліктерін анықтаудағы маңызы жан-жақты зерделеніп, олардың теориялық және қолданбалы тіл біліміндегі рөлі сараланған. Жиілік сөздіктер қазақ тілін зерттеу, оқыту және дамыту салаларында маңызды лингвистикалық құрал ретінде қарастырылған.

*Кілт сөздер*: жиілік сөздік, лингвостатистика, сөз қолданыс, тілдік корпус, статистикалық зерттеу.

А.А. Каршигаева, А. Мурсал, Г. Тлегенова, Е. Маралбек

## Исследование лексического состава и особенностей языкового употребления на основе частотного словаря произведений Абая

В статье анализируется история создания частотных словарей казахского языка, их место в лингвостатистических исследованиях, а также значение для изучения языковой системы. Проведён лингвистический и структурный анализ частотных словарей, составленных на основе произведений Абая Кунанбаева, с целью определения частотности и закономерностей употребления лексических единиц, использованных поэтом. В ходе исследования систематизирована как часто, так и редко употребляемая лексика, а также рассмотрено её значение для выявления морфологических, синтаксических и лексических особенностей казахского языка. Установлено, что частотные словари играют важную роль в процессе обучения языку, в частности — при преподавании казахского языка как иностранного или второго. Актуальность частотных словарей подчёркивается в контексте создания языкового корпуса, систематизации частотности употребления лексических единиц, разработки систем машинного перевода, совершенствования текстовых редакторов и развития казахского языка в цифровом пространстве. Проведён сравнительный анализ лексики с различной степенью употребительности в произведениях Абая, показано её место в историческом развитии казахского языка. В заключение всесторонне рассматривается значение частотных словарей для выявления структурных, функциональных и стилистических особенностей казахского языка, а также анализируется их роль в теоретической и прикладной лингвистике. Таким образом, частотные словари выступают важным лингвистическим инструментом в исследовании, преподавании и развитии казахского языка.

*Ключевые слова:* частотный словарь, лингвостатистика, словоупотребление, языковой корпус, статистическое исследование.

## Reference

1 Mursal, A., Zhanabekova, A.A., & Tlegenova, G.B. (2024). Abai shygarmalaryndagy eski kitabi leksika zhane onyn lingvostatistikalyq erekshelikteri [Old book vocabulary in Abai's works and its linguistic and statistical features]. *Tiltanym — Linguistic*s, 4 (96), 187–200. https://doi.org/10.55491/2411-6076-2024-4-187-200 [in Kazakh]

2 Zhubanov, A.K., Zhanabekova, A.A., Karbozova, B.D., & Kozhakhmetova, A.K. (2016). *Qazaq tilinin zhiilik sozdigi* [Frequency dictionary of the Kazakh language]. Almaty: «Qazaq tili» baspasy [in Kazakh].

3 Zhubanov, A., & Zhanabekova, A. (2017). *Korpustyq lingvistika* [Corpus linguistics]. Almaty: «Qazaq tili» baspasy [in Kazakh].

4 Kamarova, N.S. (2021). Inostrannye slova, vkliuchennye v slovar Abaia na kazakhskom yazyke: osobennosti ispolzovaniia [Foreign words included in the words of Abai in the Kazakh language: usage features]. *Sbornik materialov Mezhdunarodnogo simpoziuma Tomskogo gosudarstvennogo universiteta. Sektsiia 6. Slovari obraznykh sredstv yazyka i khudozhestvennoi rechi — Collection of materials of the International Symposium of Tomsk State University. Section 6. Dictionaries of figurative means of language and artistic speech* (pp. 280–286). http://dx.doi.org/10.17223/978-5-907442-19-1-2021-93 [in Russian].

5 Khokhlova, M.V. (2023). Statisticheskie metody v leksikograficheskikh issledovaniiakh: predstavlenie chastotnoi leksiki [Statistical methods in lexicographic research: the representation of partial vocabulary]. *Terra Linguistica, 14*(3), 80–93. https://doi.org/10.18721/JHSS.1430 [in Russian].

6 Umudova, M.T. (2022). Types of Obsolete Words (Archaisms and historicisms). *International Journal of Social Science and Human Research, 5*(12), 5897–5900. https://doi.org/10.47191/ijsshr/v5-i12-74.

7 Syzdyk, R. (1968). *Abai shygarmalarynyn tili* [The language of Abai's works]. Almaty [in Kazakh].

8 Akhmetova, Sh., Bolganbaev, A., & Zhanuzakov, T. (2011). *Abai tili sozdigi* [Dictionary of the Abai language]. Almaty: «Oner» baspasy [in Kazakh].

9 Bektaev, K.B., Zhubanov, A.K., Myrzabekov, S., & Belbotaev, A.B. (1995). *M.O. Auezovtin 20 tomdyq shygarmalar teksterinin zhiilik sozdikteri* [Frequency dictionaries of texts of M.O. Auezov's works in 20 volumes]. Almaty-Turkistan [in Kazakh].

10 Belbotaev, A., & Bulanbaeva, Zh. (2014). *Abai tilinin zhiilik sozdikteri zhane onyn manyzy men mani turaly* [On frequency dictionaries of the Abai language and its meaning and significance]. Almaty [in Kazakh].

11 Zhanabekova, A. (2023). *Abai Qunanbaiuly tilinin zhiilik sozdigi* [Frequency dictionary of the language of Abai Kunanbaevich]. Almaty [in Kazakh].

12 Nurzhanova, B.N. (2018). *Zhazushy tili sozdigin tuzudin gylymi-teoriialyq negizderi* [Scientific and theoretical foundations of the construction of a Dictionary of the writer's language]. Almaty [in Kazakh].

13 Bektaev, K.B., Zhubanov, A.K., Myrzabekov, S., & Belbotaev, A.B. (1979). *M. Auezovtin «Abai zholy» romaninyn zhiilik sozdigi* [Frequency dictionary of B.M. Auezov's novel "The Way of Abai"]. Almaty [in Kazakh].

14 Morozkina, E.A., Vorobiev, V.V., & Safina, Z.M. (2023). Statisticheskie metody issledovaniia v khudozhestvennom perevode [Statistical research methods in literary translation]. *Doklady Bashkirskogo universiteta — Report of Bashkir University, 8*, 3, 130–138. https://doi.org/10.33184/dokbsu-2023.3.15 [in Russian].

15 Zhubanov, A., & Myrzabekov, S. (2022). *M. Auezov tilinin sozdikterin zhasau täzhiribesi* [The experience of creating dictionaries of the Auezov language]. Almaty: A. Baitursynuly atyndagy Til bilimi instituty [in Kazakh].

16 Akhabaev, A. (1973). Qazirgi qazaq gazetteri tilinin alfavitti zhiilik sozdigi [Alphabetical frequency dictionary of the language of modern Kazakh newspapers]. *Qazaq teksinin statistikasy — Statistics of the Kazakh text.* Almaty [in Kazakh].

17 Mursal, A., Tlegenova, G., Zhanabekova, A., & Cherepanova, A. (2025). Archaic Bookish Words in Abai Kunanbayuly's Works: A Linguostatistical Analysis. *Forum for Linguistic Studies, 7*(3), 195–208. https://doi.org/10.30564/fls.v7i3.8506.

## Information about the authors

**Karshigayeva, Ainur Aralbekovna** — Candidate of Philological Sciences, Associate Professor, Department of Media Communications and Intercultural Communication at the Higher School of Media Communications and Cultural Studies, Turan University, Almaty, Kazakhstan. E-mail: kar_ainur@mail.ru; ORCID: https://orcid.org/0000-0002-8721-5308

**Mursal, Aikerim** (corresponding author) — Doctoral student, Akhmet Baitursynuly Institute of Linguistics, Almaty, Kazakhstan; E-mail: aigerimmursal@mail.ru; ORCID: https://orcid.org/0000-0001-8035-4236

**Tlegenova, Gulden Bakytkazyevna** — Doctoral student, Akhmet Baitursynuly Institute of Linguistics, Almaty, Kazakhstan; E-mail: gulden_20.88@list.ru; ORCID: https://orcid.org/0000-0001-9842-3324

**Maralbek, Ermukhamet** — Doctor of Philosophy (PhD), Head of the Department of Akhmet Studies at the Akhmet Baitursynula Institute of Linguistics, Leading Researcher, Almaty, Kazakhstan. E-mail: muhamet.maralbek@gmail.com; ORCID: https://orcid.org/0000-0001-5954-8669