

DOI: <https://doi.org/10.63332/joph.v5i6.2008>

Using the Poetic Corpus of Abay for the Quality Training of Philologists

Pirmanova K.K.¹, Madiyeva G.B.², Serikbayeva A.D.³, Tokmyrzayev D.O.⁴, Ongalbek A.A.⁵, Ualikhan A.B.⁶, Zadayeveva A.A.⁷

Abstract

In the context of the digital transformation of education and the need to train highly qualified teaching staff capable of integrating humanitarian knowledge with modern technologies, the use of digital corpora of texts in the educational process is particularly relevant. One of the promising tools in this area is a poetic corpus created based on the works of Abai Kunanbayev, a key figure in Kazakh literature and philosophical thought. This article examines the concept, structure and methodological capabilities of a small poetic corpus of Abai's poetry in the context of training future philology teachers. The emphasis is on the interdisciplinary and applied significance of the corpus approach for mastering philological disciplines, forming analytical, interpretive and digital competencies, as well as a deeper immersion in the national literary and cultural heritage. The creation of the corpus includes the stages of collecting, digitizing, linguistic markup and annotating poetic texts. Such a corpus may include original texts in Kazakh, translations into Russian, subject tags, frequency dictionary, morphological and syntactic markups, which provides ample opportunities for complex text analysis. Particular attention is paid to the construction of the corpus taking into account the methodological requirements of corpus linguistics and pedagogical applicability within the framework of university programs. One of the central theses of the article is the assertion that working with the corpus of Abai's poetry allows implementing the principle of "learning through research", involving students in project and research activities. Students learn to work with digital texts, formulate hypotheses, perform quantitative and qualitative analysis of vocabulary, semantics and stylistics. This, in turn, contributes to the development of the most important professional qualities in future teachers: analytical thinking, linguistic intuition, understanding of the cultural context and the ability to work with large volumes of text.

Keywords: Abai Kunanbayev, Research Potential, Lexeme, Methodology, Meter, National Language Corpus, Poetic Corpus, Educational Potential, Poetry, Methodological Opportunities, Tasks.

Introduction

The development of national language corpora is considered to be one of the priority areas of modern linguistics in recent decades. Valuable markups - syntactic, morphological, semantic,

¹ PhD postdoctoral student of Al-Farabi Kazakh National University, Almaty, Kazakhstan, Email: kunsulu.pirmanova@mail.ru, <https://orcid.org/0000-0003-3783-3199>

² Doctor of Philological Sciences, Professor Al-Farabi Kazakh National University, Almaty, Kazakhstan, Email: gmedi@mail.ru, Scopus Author ID 57148340900, <https://orcid.org/0000-0003-3198-0024>

³ PhD postdoctoral student of Al-Farabi Kazakh National University, Almaty, Kazakhstan, Email: sd_aizhan@mail.ru, <https://orcid.org/0000-0002-2585-4892>

⁴ Programmer Al-Farabi Kazakh National University, Almaty, Kazakhstan, Email: dark.han@mail.ru, <https://orcid.org/0000-0002-8679-9698>

⁵ Teacher of Kazakh language and literature, Talent general education school N2, Shymkent, Kazakhstan, Email: aida.kz_1996@mail.ru, <https://orcid.org/0009-0009-3985-7013>

⁶ Junior research assistant Al-Farabi Kazakh National University, Almaty, Kazakhstan, Email: ak.bota@mail.ru, <https://orcid.org/0009-0003-5797-8421>

⁷ Kazakh National Women's Teacher Training University, Almaty, Kazakhstan, <https://orcid.org/0000-0003-0149-2695>



frame (for example, in FrameNet) and others - provide ample opportunities for studying highly objective language material. Currently, national corpora have been developed for American English, British, Russian, Czech, Croatian, Hungarian, Italian, Persian, Bashkir, Tatar and other languages.

Abai is the central figure of Kazakh poetry, a philosopher, reformer and humanist. His poetic legacy is of great importance for national culture, education and self-awareness. However, in educational and research practice, his works are often used fragmentarily or in adapted formats [1].

In practical terms, the poetic corpus can be used in studying such disciplines as "Introduction to Literary Studies", "Modern Linguistics", "Methodology of Teaching Language and Literature", "Digital Technologies in Philology". It can become the basis for seminars, workshops, term papers and final qualifying works. Through the analysis of specific poems by Abai in the corpus environment, students acquire the skill of interpreting the text not only from the position of the reader, but also from the position of the researcher, teacher and mediator of cultural knowledge.

The use of the corpus also contributes to the development of students' digital competence, including basic skills in working with text databases, using linguistic analysis programs (AntConc, SketchEngine, etc.), interpreting statistical data, and visualizing results. This makes future teachers more flexible and ready to use digital resources in their own teaching practice.

The Abai Poetry Corpus can also be used in a bilingual or even trilingual format (Kazakh - Russian - English), which is especially important for the multilingual educational context of Kazakhstan. This allows students not only to compare the original and the translation, but also to identify the features of the poetic form, national and cultural connotations, and translation difficulties, thereby expanding the scope of linguacultural and translation competence.

Based on the analysis of corpus data, it is possible to create electronic teaching aids, methodological recommendations, interactive assignments, and even online platforms, which meets the challenges of distance and blended learning. In addition, the corpus can be useful in the context of advanced training of existing teachers - as a source of methodological renewal and innovative lesson content.

Thus, a small poetic corpus based on the works of Abai is not just a digital resource, but a pedagogical innovation that provides a link between the classical literary heritage and current educational practices. Its use in the training of philology teachers allows not only to improve the content of education, but also to form a new type of teacher - culturally sensitive, methodologically trained, technologically equipped and capable of independent research and teaching of native literature and language.

In the future, such corpus projects can be expanded and adapted for other classics of Kazakh and world literature, which will make philological education more modern, open and based on the principles of digital humanities. The Abai Corpus can become a model for integrating digital resources into the educational system, while maintaining the depth of the humanitarian approach and national-cultural identity.

Material and Methods.

The article is based on the resources of the Russian national corpus, scientific articles on the problems of using materials from national language corpora, as well as the poetic heritage of Abai Kunanbayev, which is subject to digital accumulation in the space of the poetic subcorpus

of the Kazakh national language corpus.

Within the framework of this study, the hypothetical-inductive method, analysis of scientific research in the field of corpus linguistics, synthesis of scientific ideas in the context of the idea of using national language resources in research and education were used.

In order to determine the research and educational potential of national corpora, the following actions were carried out:

- a review of sources containing information on the purpose of national language corpora was conducted;
- features of language corpora that are of research, educational and methodological value were determined;

Results.

The issue of using corpora in training specialists is an important area of modern linguistics and methodology. This the fact is indisputable. The corpus contains educational potential. Let us turn to the studies that reveal these points.

Tony McEnery and Andrew Hardie, considering the use of the concordance method in the study of poetic texts, draw attention to the conceptual approach to the study of English, which helps to form a "linguistic picture of the world" of a particular author. Their book "Corpus Linguistics: Method, Theory and Practice" (2012) is a basic work in this area, devoted to the study of the use of corpus linguistics in teaching English as a foreign language (EFL/ESL) [2].

I.Vorobey, considering the use of the concordance method in the study of poetic texts by R.M.Rilke, draws attention to the conceptual approach to the study of the text, which helps to form a "linguistic picture of the world" of a particular author. Considering the features, or rather, the shades of meaning of the terms "concept", "content", "meaning", "notion", "significance", the researcher emphasizes that the main difference of the concept as a linguistic phenomenon is the emotional component, the determination of the cultural context, the "translingual character" and acts as a "conditional mental unit that does not exist outside of thinking and is materialized with the help of language, by means of a word" [3]. Mentioning the division of concepts into universal, ethnic, group and individual, including author's, the researcher dwells in detail on the phenomenon of crystallization of meaning, that is, the formation of a word with an unambiguous meaning. Using the method of conceptual analysis, I. Vorobey divides it into concordance and interpretation. Sylvian Granger, the founder of the Learner Corpus Research direction, studies the written and oral speech of language learners based on specially collected corpora of texts. It has influenced the teaching of academic writing and stimulated the creation of pedagogical approaches and corpora for the high-quality training of teachers of philology. This corpus contains texts of academic essays written by students with different native languages, including philologists studying English as a foreign language. Analysis of these texts allows us to identify typical errors, interlingual interferences and features of non-local academic discourse. It is important that ICLE is structured by countries and types of tasks, which makes it a powerful tool for contrastive analysis and the development of targeted teaching methods [4].

The poetic corpus of Abay will allow, in particular, to determine the gravitation of lexical units to poetic meters. B.V. Orekhov writes about this [5]. For this, the already traditional collostructural analysis of A. Stefanovich and S. Gris can be used, which is based on measuring the degree of "attraction" or "repulsion" of a construction by words. To determine these two

phenomena, the Fisher method is used, in which the frequency of occurrence of a lexeme in a construction is calculated throughout the corpus. Thus, if the frequency exceeds the expected one, we can talk about attraction and, conversely, if the expected level decreases, about repulsion of the word by the construction. We plan to apply for this purpose a collexemic analysis, which involves the following stages: fixing the number of occurrences in the corpus for each lexeme and, separately, the statistics of occurrence in the construction. This will allow us to calculate the necessary ratio and determine the moments of "attraction" or "repulsion". In particular, this concerns the counting of lexemes and their appearance in subcorpora of meters. In this way, it is possible to determine the gravitation of a particular word form to a particular metrical construction or, according to M.L. Gasparov, to determine the "semantic aura of the meter" [6]. N.R. Dobrushina notes the possibilities of the Russian language corpus in compiling assignments: "...the existence of the National Corpus significantly simplifies the process of compiling exercises of various kinds" [7]. S. Anokhina [8] draws attention to the possibilities of using the NKRY in the professional training of primary school teachers. The author suggests such forms of work as studying the category of polysemy and homonymy using examples provided by the corpus in the "children's literature" subcorpus. According to the author of the article, students go through a number of stages of working with textual material: familiarization with texts from the NKRY subcorpus, analysis of word meanings in different texts, determination of the primary-secondary nature of LSV, as well as studying homonymy, synonymy and antonymy by searching for them in the texts of children's literature of the same or different historical periods. In this case, the analysis is based on the search for authorial motives in the selection of synonyms.

E. A. Grishina highlights in her article interesting approaches to the analysis of markers of Russian colloquial speech based on the material of film transcripts in the corresponding subcorpus and finds the following signs of oral speech: the introduction of special interphrase particles-connectors, ellipsis of conjunctions, explication of the performative, introduction of metatext, explication of roles, intensification of phatics, anocopa, rematization and emphatization of the text, implementation of annotational excess, discrepancy between the transcript and subtitles, etc. [9].

Methodologically, the poetic corpus of Abai for the high-quality training of teachers of philology provides the following opportunities:

1. Lexical-semantic exercises

Search for synonyms/antonyms in context (comparison of usage)

Identification of lexical collocations: students look for stable phrases with a given word

Analysis of polysemy: distinguishing the meanings of one word based on context

Example: find the word "vesti" in the corpus and classify its meanings.

2. Grammatical-structural exercises

Analysis and transformation of constructions: for example, passive and active forms

Comparison of syntactic models: how the infinitive, participle, etc. are used

Working with variability: which grammatical forms are used more often and in what contexts

Example: compare the use of the constructions "need to do" and "should do".

3. Discursive and pragmatic exercises

Study of speech acts: how requests, apologies, gratitude, etc. are expressed in the corpus

Analysis of text coherence: working with discourse markers, conjunctions, introductory words

Speech characteristics of a character/author through corpus examples

Example: find typical ways of starting a letter or address in the corpus.

4. Contextual exercises

Inserting a missing word based on the context (several options)

Determining style/genre based on context

Identifying errors and illogicalities in the proposed examples against the background of authentic corpus

Example: choose the most appropriate word for the context from several corpus options.

5. Frequency analysis

Comparison of the frequency of synonymous expressions

Selecting the most natural option based on corpus statistics

Studying language trends (e.g., changes in word frequency over time)

Example: compare the frequency of the words “Internet” and “World Wide Web” in different time periods.

6. Project tasks

Creating a mini-corpus on a topic

Written work using corpus analytics (essay, article)

Studying the author’s style (e.g., based on a subcorpus of works by one writer)

Example: create a table of the most frequent epithets of Uabay based on the corpus.

Abay often reflects on the value of reason, knowledge, and thinking. There are different words in poetry that denote intellectual activity. Here are possible synonyms and contextual options: Akyl (reason), Parasat (prudence, common sense), Zerde (consciousness, cognition), Sana (reason, mentality), Oy (thought), Gylym (knowledge, science), Bilym (knowledge) Tuysik (intuition, feeling)

"Akyl, kayrat, zhurekti birdey usta..."

— Here "akyl" is correlated with "kayrat" and "zhurek" - this is a category of moral and intellectual balance.

"Gylym tappay maktanba..."

— "Gylym" as a synonym for the value of reason and the true meaning of life.

Abay uses synonymous words not as equivalent, but to highlight different aspects of human thought: logical, emotional, moral and cognitive.

For students of philology working with the poetic corpus of Abay, we have compiled an exercise

based on the analysis of verb forms. Use the following verb forms from the poetic corpus of Abay "to freeze", "to get tired" Analyze them in terms of tense, aspect and syntactic function.

The review and research position presented in the material can cause a scientific controversy today in the sense that the linguistic community is divided into supporters and opponents of corpus development and their use in teaching and research. For example, Apresyan believes that the reckless use of corpus resources can lead to falsification due to the incorrect use of quantitative data [10].

This position certainly deserves attention, but does not exclude the resource use of corpus materials. Pirmanova K. also opposes blind trust in corpus materials [11.]

Agreeing with the authors of these ideas that corpus texts may contain quantitative, qualitative and other distortions, we nevertheless consider it necessary to defend the position of the value of corpora for research and methodological purposes. In this case, a critical approach to the use of resources is important [12].

Worthy of attention is their position, which considers the possibility of creating corpora of texts with violations of language and speech norms for use in various purposes, including methodological ones. At the same time, according to linguists, the accumulation of texts with errors, for example, in the speech of bilinguals, can become valuable material and a kind of "anti-model" for everyone who is familiar with standardized oral and written speech [13]. Being in a position of accepting different points of view, we will allow ourselves to express our position regarding the fact that the creation of the poetic corpus of Abai is still beyond the mentioned risks, since it should represent examples of classical speech and texts taken from practically canonized sources.

Conclusion

Modern education in the field of philology requires new approaches that combine a deep understanding of the humanitarian content with innovative tools of digital analysis. In this context, the use of the poetic corpus of Abai Kunanbaev is of particular importance as a powerful educational, scientific and cultural resource that contributes to the comprehensive development of future teachers of language and literature. In the context of the digitalization of education and the rapid development of digital humanities, corpus-based training is becoming not just a trend, but a necessity, especially in the system of training highly qualified teachers of the new generation [14].

First of all, it should be emphasized that the corpus created on the basis of the poetic heritage of Abai acts as a multifunctional tool. It simultaneously serves as a database, an analysis tool, a platform for learning and a source of motivation for studying native literature. This combination of functions provides a comprehensive approach to teaching philology students, allowing them not only to master the content of texts, but also to develop critical thinking, research skills, translation skills, interpretation and analysis of texts at different levels - from lexical to semantic-cultural [15].

The use of the poetic corpus helps to expand the range of educational activity formats: from traditional text analysis to interactive assignments, project and research work, as well as cross-disciplinary projects. Students can not only read and interpret Abai's poems, but also, for example, study the frequency of lexemes, the construction of metaphorical structures, the thematic distribution of key concepts, and identify stylistic and grammatical features of the text.

This is especially valuable in the context of training teachers, who must be able not only to explain, but also to demonstrate the text in development, in dynamics, in different contexts - linguistic, cultural, philosophical [16].

It is also worth noting that the Abai corpus can serve as a resource for bilingual and multilingual education, which is especially relevant for Kazakhstan, where education is carried out in several languages - Kazakh, Russian and English. A parallel corpus (in the original and translation) allows students to analyze not only the semantic and stylistic structure of the original, but also the features of the translation process, which develops their linguacultural competence, sensitivity to nationally marked elements, phraseological units, idioms and poetic imagery. Such training significantly increases the awareness of future teachers in matters of intercultural communication and language policy [17].

In addition, the corpus can be actively used in the methodological training of future teachers. It allows you to develop educational tasks, tests, worksheets, conduct lesson simulations and even create electronic teaching aids. This is especially important in the context of the transition to blended and distance learning, when digital resources are becoming not just auxiliary, but central components of the educational process. Working with the corpus forms pedagogical flexibility in students and a willingness to integrate ICT into teaching practice. It should be emphasized that turning to Abai's work through the prism of digital humanities gives students the opportunity not only to study the poet as a classic, but also to perceive him as a relevant thinker and cultural mediator, whose ideas about morality, reason, love of knowledge and homeland remain significant in the 21st century. Thus, not only the professional but also the personal identity of the future teacher is formed - a bearer of cultural values, connecting the past and the future, traditions and innovations. It is also important that the corpus stimulates students' research activities. With its help, you can complete term papers and theses, conduct mini-research, participate in conferences and projects. The corpus becomes a platform for applied philology, where students not only learn theory, but also apply it in practice. This contributes to the development of academic independence, initiative and the ability to self-education - the key qualities of a modern teacher.

In the long term, the use of the Abai poetic corpus in the educational environment can become a model for creating similar resources on other authors of Kazakh, Russian and world literature. This will open up opportunities for creating national corpus collections that will be both scientific and educational platforms that contribute to the development of digital humanities in Kazakhstan.

Of course, it is necessary to take into account that the introduction of corpus technologies requires certain training from both teachers and students. Here, it is important to develop methodological recommendations, conduct training seminars, and integrate corpus linguistics into the curricula of philological areas. But the experience of foreign and domestic educational institutions shows that this area is successfully developing with due support from universities and the scientific community.

Thus, the development and use of the Abai Kunanbayev poetic corpus in the system of training philology teachers is not just a private methodological innovation, but a strategically important area that promotes the integration of digital technologies into humanitarian education, the preservation and understanding of cultural heritage, and the improvement of the quality of pedagogical education. This is the path to the formation of a new generation of teachers capable of not only teaching language and literature, but also inspiring, exploring, and transforming

The Abai Corpus becomes the link that connects the richness of the national poetic tradition with the possibilities of the digital age, unites artistic text and technology, educational tasks and cultural meanings. Its use in university training is not only a step forward in the methodology of teaching philological disciplines, but also a contribution to the preservation and popularization of the legacy of the great poet, whose poetry continues to inspire, educate, and unite generations.

Funding

This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR21882334 "Kazakh poetic corpus development: morphological and poetic designation of Abai's poems")

References

- Li D., Yerzhanova F., Abisheva S., Sabirova D., Tuleubayeva M., Smagulova B., Umarova A. Analysis of Chekhov's Literary Research Hotspots and the Application of Educational Technology Using CiteSpace // Forum for Linguistic Studies | Volume 06 | Issue 06 | December 2024 | 361-370 | DOI : <https://doi.org/10.30564/fls.v6i6.7166>, Q2, 54.
- Tony McEnery, A. Hardie. *Corpus Linguistics: Method, Theory and Practice*. 2012. Cambridge University Press. 295 p.
- Vorobey I.A. Concordance as part of the method of analyzing a poetic concept. <https://www.academia.edu/>
- Ädel, A. 2015. "Variability in Learner Corpora." In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 401–421. Cambridge: Cambridge University Press
- Orekhov B.V. Once again about the research potential of the poetic corpus: meter, vocabulary, formula. <https://publications.hse.ru/articles/178496831>
- Gasparov M.L. Meter and meaning. On one of the mechanisms of cultural memory. — M.: Fortuna EL, 2012. — P.3-16.
- Dobrushina R.M. How to use the national corpus of the Russian language in education. <https://ruscorpora.ru/old/sbornik2005/18dobrushina.pdf>
- Anokhina S.A. Using materials from the national corpus of the Russian language in studying lexical categories // *National corpus of the Russian language and problems of humanitarian education*. Edited by N.R. Dobrushina. – Moscow 2007. P. 150-159.
- E. A. Grishina On the markers of colloquial speech (preliminary study of the cinema subcorpus in the national corpus of the Russian language, www.ruscorpora.ru)
- Realisation of parameters of the bologna process at teaching of philologists in the Republic of Kazakhstan / Belassarova, Z., Muratbek, N., Sultaniyazova, I., Nurbaeva, A., Sabirova, D. *Review of European Studies*, 2015, 7(7), p. 356–364
- Pirmanova K. K., Serikbayeva A. D., Nurlybayev N. M., Sabirova D. A. Potential of Using Abai's Poetic Corpora for Research, Educational and Methodical Purposes // *Journal of Ecohumanism*, 2024, 3(4), p. 618–626. <https://doi.org/10.62754/joe.v3i4.3589>
<https://ecohumanism.co.uk/joe/ecohumanism/article/view/3589>
- Sabirova, D., Moldagali, M., Nurbayeva, A. The Main Images of Internet Poetry of Kazakhstan. *Journal of Ecohumanism.*, 2024, 3(7), p. 5240–5248.
- Kunsulu, Pirmanova, Tokmyrzayev Darkhan, Zaurbekova Gulkhadisha, Madiyeva Gulmira, Serikbayeva Aizhan, and Sabirova Dina. 2025. "Usage of the Statistical Method to the Lexico-Morphological

- Structure of the Language."Forum for Linguistic Studies 7(3):890–905.
<https://doi.org/10.30564/fls.v7i3.8419>
- Sabirova D.A., Nurbaeva A.M., Sametova Z.S., Baibolov A.U., Moldagali M.B. Artistic Features of Lydia Kossutskaya's Prose for Children (Based on The Material of The Kazakh Literary Online Magazine Daktil) // Journal of Ecohumanism, 2024, 3(3), p. 1511–1516.
<https://ecohumanism.co.uk/joe/ecohumanism/article/view/3616>
- Abisheva, S.D., Khavaidarova, M.M., Sabirova, D.A., Ongarbaeva, A., Moldagali, M. Experience of Implementing Trilanguage in Kazakhstan on the Example of Pedagogical Universities. Journal of Ecohumanism This link is disabled., 2024, 3(7), p. 283–290/<https://www.researchgate.net/publication/383203186>
- Zhumagulova V., Kazhigaliyeva G., Onalbayeva A., Sabirova D., Zadayeva A. Representation of modernization of public consciousness in the mass media discourse | Representación de la modernización de la conciencia pública en el discurso de los medios de comunicación // Opcion, 2019, 35(Special Issue 23), pp. 1171–1187.
- Abisheva S., Polyak D., Seidullaeva G., Kermeshova Z., Sabirova D. Meaning of fiction in the formation of students' identity // Opcion, 2018, 34(85-2), pp. 186–204 ISSN 10344-3611587/ISSNe: 2477-9385/
<https://produccioncientificaluz.org/index.php/opcion/article/view/23792>.