



DOI: 10.25178/nit.2025.2.19

Статья

## Историко-поэтический подкорпус Национального корпуса казахского языка

**Айнур А. Сейтбекова, Анар М. Фазылжан, Асел К. СейдаMAT,  
Мадина К. Абаева, Айкерим Мурсал**

*Институт языкознания имени А. Байтурсынулы, Республика Казахстан*

В статье анализируются ключевые аспекты оцифровки образцов устного народного творчества казахского народа XV–XIX вв., написанных на арабской графике, и их интеграция в Национальный корпус казахского языка (НККЯ). Данная работа составляет первый этап создания историко-поэтического подкорпуса НККЯ. В ходе исследования проведён сравнительный анализ существующих поэтических подкорпусов на других языках (русский, чешский, башкирский и персидский), что позволило выявить наиболее эффективные методы и подходы для создания казахского подкорпуса.

Важным результатом работы стало создание модели метатекстовой разметки, включающей 28 параметров, с учётом специфики казахской поэзии. Были определены ключевые элементы казахского стиха: структура строф, количество слогов, рифмы и стопы. Разработанная разметка позволяет точно отражать поэтические особенности текстов и учитывать влияние восточной литературы и народных жанров на развитие казахской поэтической традиции. Одной из важных инноваций стала семантическая разметка устаревших слов.

Представлена разработка интерфейса подкорпуса, который даёт возможность пользователям исследовать поэтические произведения в арабской графике и их транскрибированные варианты на кириллице. Это делает подкорпус ценным инструментом для лингвистических и литературных исследований.

**Ключевые слова:** поэтический подкорпус; метатекстовая разметка; арабская графика; письменность; текстовая база; казахский язык; казахская поэзия; Национальный корпус казахского языка

---

Исследование проводилось в рамках работ, поддержанных программно-целевым финансированием Министерства науки и высшего образования РК, проект ИРН BR21882249 «Совершенствование Национального корпуса казахского языка (НККЯ) как средства межкультурной коммуникации и расширение его подкорпусов».

---



### Для цитирования:

Сейтбекова А. А., Фазылжан А. М., СейдаMAT А. К., Абаева М. К., Мурсал А. Историко-поэтический подкорпус Национального корпуса казахского языка // Новые исследования Тувы. 2025. № 2. С. 312-338. DOI: <https://doi.org/10.25178/nit.2025.2.19>

**Сейтбекова Айнур Аташбеккызы** — кандидат филологических наук, заведующая отделом история языка и тюркологии Института языкознания имени А. Байтурсынулы. Адрес: Республика Казахстан, г. Алма-Ата, ул. Курмангазы, д. 29. Эл. адрес: ainurseit@mail.ru

**Фазылжан Анар Мураткызы** — кандидат филологических наук, директор Института языкознания имени А. Байтурсынулы. Адрес: Республика Казахстан, г. Алма-Ата, ул. Курмангазы, д. 29. Эл. адрес: nar20@bk.ru

**СейдаMAT Асел Казбеккызы** — докторант, младший научный сотрудник отдела история языка и тюркологии Института языкознания им. А. Байтурсынулы. Адрес: Республика Казахстан, г. Алма-Ата, ул. Курмангазы, д. 29. Эл. адрес: assel.seidamat@gmail.com



**Абаева Мадина Кабылкызы** — кандидат филологических наук, заведующая отделом психолингвистики Института языкознания имени А. Байтурсынулы. Адрес: Республика Казахстан, г. Алма-Ата, ул. Курмангазы, д. 29. Эл. адрес: madina-258@mail.ru

**Мурсал Айкерим** — докторант, младший научный сотрудник отдела история языка и тюркологии Института языкознания имени А. Байтурсынулы. Адрес: Республика Казахстан, г. Алма-Ата, ул. Курмангазы, д. 29. Эл. адрес: aigerimmursal@mail.ru



## Historical and Poetic Subcorpus of the National Kazakh Language Corpus

**Ainur A. Seitbekova, Anar M. Fazylzhan,  
Asel K. Seydamat, Madina K. Abaeva, Aykerim Mursal**

*The Institute of Linguistics named after Akhmet Baitursunuly, Republic of Kazakhstan*

The article analyzes the key aspects of digitizing the samples of Kazakh oral folk literature from the 15th to 19th centuries, originally written in Arabic script, and their integration into the National Corpus of the Kazakh Language (NCKL). This work constitutes the first stage in the creation of a Historical and Poetic Subcorpus of the NCKL. As part of the study, a comparative analysis of existing poetic subcorpora in other languages (Russian, Czech, Bashkir, and Persian) was conducted, allowing for the identification of the most effective methods and approaches for developing the Kazakh subcorpus.

A significant outcome of the project is the development of a metatextual annotation model comprising 28 parameters that consider the specifics of Kazakh poetry. Key elements of Kazakh verse were identified, including stanza structure, syllable count, rhyme schemes, and metrical feet. The annotation system developed enables an accurate representation of the poetic features of the texts while accounting for the influence of Eastern literature and folk genres on the evolution of the Kazakh poetic tradition. One of the important innovations introduced is the semantic annotation of archaic vocabulary.

The article also presents the design of the subcorpus interface, which allows users to explore poetic works in their original Arabic script alongside their transcribed Cyrillic versions. This makes the subcorpus a valuable tool for linguistic and literary research.

**Keywords:** poetic subcorpus; metatextual annotation; Arabic script; writing system; text database; Kazakh language; National Corpus of the Kazakh Language

### Financing

The research paper was obtained within the framework of the program BR 21882249 "Improvement and Expansion of the Subcorpus of the National Corpus of the Kazakh Language as a Means of Intercultural Communication".



### For citation:

Seitbekova A. A., Fazylzhan A. M., Seydamat A. K., Abaeva M. K. and Mursal A. Historical and Poetic Subcorpus of the National Kazakh Language Corpus. *New Research of Tuva*, 2025, no. 2, pp. 312-338. (In Russ.). DOI: <https://doi.org/10.25178/nit.2025.2.19>

**SEITBEKOVA, Ainur Atashbekkyzy**, Candidate of Philology, Head, Department of History of Language and Turkology, The Institute of Linguistics named after Akhmet Baitursunuly. Postal address: 29 Kurmangazy St., Almaty, Republic of Kazakhstan. E-mail: [ainurseit@mail.ru](mailto:ainurseit@mail.ru) **ORCID: 0000-0003-4268-1572**

**FAZYLZHAN, Anar Muratkyzy**, Candidate of Philology, Director, The Institute of Linguistics named after Akhmet Baitursunuly. Postal address: 29 Kurmangazy St., Almaty, Republic of Kazakhstan. E-mail: [nar20@bk.ru](mailto:nar20@bk.ru) **ORCID: 0000-0002-0562-4846**

**SEIDAMAT, Assel Kazbekkyzy**, Doctoral Student, Junior Researcher, Department of History of Language and Turkology, The Institute of Linguistics named after Akhmet Baitursunuly. Postal address: 29 Kurmangazy St., Almaty, Republic of Kazakhstan. E-mail: [assel.seidamat@gmail.com](mailto:assel.seidamat@gmail.com) **ORCID: 0009-0001-0471-3249**

**ABAYEVA, Madina Kabylkyzy**, Candidate of Philology, Head, Department of Psycholinguistics, The Institute of Linguistics named after Akhmet Baitursunuly. Postal address: 29 Kurmangazy St., Almaty, Republic of Kazakhstan. E-mail: [madina\\_abaeva@mail.ru](mailto:madina_abaeva@mail.ru) **ORCID: 0000-0001-6463-9354**

**MURSAL, Aikerim**, Doctoral Student, Junior Researcher, Department of History of Language and Turkology, The Institute of Linguistics named after Akhmet Baitursunuly. Postal address: 29 Kurmangazy St., Almaty, Republic of Kazakhstan. E-mail: [aigerimmursal@mail.ru](mailto:aigerimmursal@mail.ru) **ORCID: 0000-0001-8035-4236**



## Введение

Языковой корпус — крупная лингвистическая база, обеспеченная полной и удобной разметкой — широко используется во многих странах мира как вспомогательный инструмент, проливающий свет на множество вопросов, связанных с развитием языка. В свою очередь, корпус требует постоянных изысканий и кропотливой работы лингвистов и программистов: необходимо пополнять и совершенствовать корпусную базу согласно запросам современных пользователей. Профессор А. Жубанов, основоположник казахской прикладной лингвистики, отмечает следующее о важности и необходимости автоматизации текстов на казахском языке: «Определение стилевых, структурных, семантических, функциональных характеристик казахских текстов — это одна из актуальных проблем. Поэтому создание компьютерной базы автоматизированных текстов на казахском языке — проблема, имеющая и научную, и практическую ценность» (Жубанов, 2010: 191).

С учетом современных возможностей цифровых технологий и направлений и темпа развития науки, с целью повышения стойкости языка и укрепления государственного авторитета казахского языка в 2016 г. в Институте языкознания им. Ахмета Байтурсынова под руководством А. М. Фазылжан была начата разработка<sup>1</sup> Национального корпуса казахского языка (далее — НККЯ). В настоящее время НККЯ<sup>2</sup> — это обширный корпусный ресурс, который доступен для широкой публики (Жаңабекова, 2017; Жаңабекова, Пирманова, Карбозова, 2020; Фазылжан, 2023). Он состоит из нескольких подкорпусов: терминологического, диалектологического, культурно-репрезентативного, параллельного и др. Каждый подкорпус имеет собственные функциональные и лингвистические особенности. Существование обширной, систематизированной, аннотированной, адаптивной базы эмпирических материалов — незаменимая помощь при повышении качества исследований в гуманитарной сфере для специалистов, изучающих историю, культуру и литературу казахского языка. Так, терминологический подкорпус обеспечивает научное исследование и нормализацию специализированной лексики, что особенно важно для стандартизации терминологии в условиях технологического и научного прогресса. Диалектологический подкорпус необходим для изучения географического варьирования казахского языка, позволяя фиксировать диалектные особенности и динамику их изменений. Культурно-репрезентативный подкорпус включает тексты, отражающие национальное языковое сознание, традиции и ментальные особенности, что играет ключевую роль в этнолингвистических исследованиях. Таким образом, расширение числа подкорпусов отвечает потребностям современной лингвистики, социолингвистики и цифровой гуманитаристики, обеспечивая комплексный анализ языка в его различных проявлениях.

Расширение Национального корпуса казахского языка за счет историко-поэтического подкорпуса обусловлено не только общими технологическими возможностями и социальными задачами, но, прежде всего, конкретными научными потребностями, выявленными в процессе использования корпуса: необходимостью изучения эволюции казахского языка (для исследования исторической динамики, архаизмов, диалектизмов и трансформации литературных норм); недостатком данных для корпусной лингвистики (поэтические и исторические тексты демонстрируют особенности синтаксиса, метафоры и стилистики, которые не были адекватно представлены в основном корпусе); определенными запросами исследователей (в ходе работы над корпусом ученые, филологи и лингвисты неоднократно указывали на нехватку специализированных данных для анализа влияния устной традиции, эпоса и классической поэзии на современный язык); необходимостью поддержки инструментов укрепления национальной идентичности казахов.

Историко-поэтический подкорпус, добавленный в состав подкорпусов в 2023 г.<sup>3</sup>, охватывает лингвистическую информацию из письменного наследия казахского языка (на основе арабской графики), особенностей письма, авторских рукописей, лексики, традиционной казахской речи, структуры

<sup>1</sup> Исследование проводилось в рамках работ, поддержанных грантом Министерства науки и высшего образования РК ИРН BR21882249 «Совершенствование Национального корпуса казахского языка (НККЯ) как средства межкультурной коммуникации и расширение его подкорпусов».

<sup>2</sup> Национальный корпус казахского языка. Главная страница [Электронный ресурс] // Национальный корпус казахского языка. URL: <https://qazcorpus.kz/?lang=ru> (дата обращения: 25.02.2025).

<sup>3</sup> Историко-поэтический корпус Национального корпуса казахского языка. Главная страница [Электронный ресурс] // Историко-поэтический корпус Национального корпуса казахского языка. URL: [https://qazcorpus.kz/\\_tarihi-poetikalyq-ishkorpus/index.php?soz=%D1%81%D3%A9%D0%B7](https://qazcorpus.kz/_tarihi-poetikalyq-ishkorpus/index.php?soz=%D1%81%D3%A9%D0%B7) (дата обращения: 25.02.2025).



стихосложения, искусства слова, стиля, устойчивых оборотов. Наличие историко-поэтических корпусов в национальных корпусах позволяет поднять на новый уровень с качественной и статистической точки зрения исследовательские работы, связанные междисциплинарным подходом.

Актуальность данного вида подкорпуса представлена и в обзоре изменений Национального корпуса русского языка, где отмечено, что «пополнение Поэтического корпуса включало в себя тексты поэтов второй половины XX в., а также большую коллекцию русских переводов античной поэзии: “Илиаду” Гомера в переводе Н. И. Гнедича, “Энеиду” Вергилия переводах В. Я. Брюсова и С. М. Соловьева и сатиры Горация в переводе А. А. Фета» (Савчук и др., 2024: 11).

На данный момент, в состав историко-поэтического подкорпуса Национального корпуса казахского языка включены поэтические тексты XV–XIX вв. Выбор данного временного периода обусловлен рядом объективных научных факторов. Во-первых, сохранилось значительное количество письменных источников этого времени, что обеспечивает репрезентативность языкового материала. Во-вторых, рукописи поэтических произведений XV–XIX вв. хорошо документированы, исследованы и доступны для анализа, что позволяет проводить их систематизированное включение в корпус. Кроме того, поэзия этих столетий представляет собой важнейший пласт казахского литературного наследия, оказывая влияние на дальнейшее развитие языка, его поэтические и риторические традиции. Таким образом, на этом этапе формирования подкорпуса приоритет был отдан материалу, который сочетает высокую степень сохранности, лингвистическую и культурную значимость, а также доступность для включения в корпус.

Включение историко-поэтического подкорпуса в Национальный корпус казахского языка представляется важным шагом в сохранении и изучении как прошлого, так и настоящего казахского языка. Этот подкорпус станет уникальным источником, отражающим богатую духовную культуру, риторику и поэтическое искусство, что представляет собой бесценное наследие для исследователей и будущих поколений. В связи с этим, в январе 2023 г., в рамках программно-целевого финансирования Министерства науки и высшего образования Республики Казахстан (ИРН BR21882249) по проекту «Совершенствование Национального корпуса казахского языка (НККЯ) как средства межкультурной коммуникации и расширение его подкорпусов» на период 2023–2025 гг., началась работа по созданию историко-поэтического подкорпуса казахского языка.

Цель настоящей работы — описать проект создания историко-поэтического подкорпуса Национального корпуса казахского языка, содержащий образцы устного народного творчества XV–XIX вв. В ходе исследования нами решались следующие задачи: 1) составить хронологизацию казахской поэзии и составить характеристики ее этапов; 2) проанализировать опыт разработки поэтических корпусов на таких языках, как русский, башкирский, чешский и персидский; 3) описать трудности, возникшие при лингвотекстологической работе с оцифрованными рукописями, особенно ошибки транскрибирования арабографических текстов на кириллицу; 4) определить параметры разработки сетки метаразметки, максимально отражающей информацию о поэтических текстах данного периода; 5) проанализировать и определить поэтические параметры, отражающие особенности казахского стиха; 6) представить семантическую разметку и разработать интерфейс историко-поэтического подкорпуса.

В качестве материала исследования послужили научные труды, посвящённые созданию поэтических корпусов и сам процесс формирования подкорпуса историко-поэтических текстов НККЯ, ведущими разработчиками которого являлись авторы статьи. В том числе теоретические труды создателей Национального корпуса казахского языка А. Жубанова (Жұбанов, 2016; Жұбанов, Жаңабекова, 2017), А. Жанабековой (Жанабекова, 2013; Жаңабекова, Кожахметова, 2021), А. М. Фазылжан (Фазылжан, 2023), исследователей поэтического корпуса русского языка Е. А. Гришиной (Гришина и др., 2009), К. М. Корчагина (Корчагин, 2015), В. А. Плунгяна (Плунгян, 2014), Д. В. Сичиनावы (Сичинава, 2012) стали основой нашего исследования. Также при определении поэтической метаразметки мы руководствовались трудами А. Байтұрсынова (Байтұрсынов, 2003), Ч. Ч. Валиханова (Валиханов, 1986), Р. Сыздыковой (Сыздыкова, 1970), К. Омирәлиева (Өмірәлиев, 2010), З. Ахметова (Ахметов, 1973), З. Базарбаевой (Базарбаева, 2008), К. Жанабаева (Жанабаев, 2014; Жанабаев, Ислямова, Сейітбекова, 2022), и пр.

В процессе исследования по формированию поэтического корпуса и анализу международного опыта в данной области осуществляется описание, сравнительный анализ исследования текстов наци-



ональной литературы, написанных на кириллице и латинице, с использованием лингвотекстологических и поэтических параметров метатекстовой разметки поэтических текстов, описание историко-сравнительных методов, сравнительный анализ поэтических параметров, контекстный анализ семантических характеристик и формирование базы исторических поэтических текстов и их классификация с использованием метода систематизации.

Источниками для создания историко-поэтического подкорпуса послужили литературное и культурное наследие казахского народа, поэтические образцы устного народного творчества (повествовательные дастаны, религиозные дастаны, героический эпос, лирический эпос, героические дастаны, айтысы, стихи назидания и пр.). Источниками подкорпуса послужил фонд Научно-инновационного центра «Рукопись и фольклористика» при Институте литературы и искусства имени М. О. Ауэзова<sup>1</sup>.

### ***Хронологизация казахской поэзии***

В связи с тем, что данный подкорпус представляет собой историко-поэтический подкорпус, целесообразно остановиться на особенностях и хронологизации казахской поэзии.

Казахская народная поэзия является неотъемлемой частью казахской литературы и представляет собой духовное и культурное наследие народа. Эта традиция восходит к древним временам существования тюркоязычных племен. Основу народного фольклора и поэзии составляют их сказки, легенды и пословицы. Казахская народная поэзия, уходящая корнями в глубь веков, является синкретическим явлением, тесно переплетающимся с устной литературой, музыкой и фольклором. Это наследие передавалось из поколения в поколение, пережив многие исторически значимые периоды, и сохранилось до наших дней в виде рукописей, выполненных в различных формах каллиграфии. Также важно отметить, что лексический состав поэтических произведений, дошедших до наших дней, включает в себя значительное количество древнетюркских, среднетюркских и арабо-персидских слов. Это отражает письменную взаимосвязь различных культур в разные исторические эпохи. В периоды расцвета тюркских цивилизаций в Средней Азии, период существования Тюркского каганата (VI–VIII вв.) общепринятым письменным языком был тюркский (Гумилев, 1999). Однако, начиная с X в., с распространением ислама в этот регион, арабский и персидский языки начали оказывать сильное влияние на литературные и научные тексты (Суворов, 2015). В результате этих изменений лексика произведений обогатилась заимствованиями из арабского и персидского языков, что стало отражением новой культурной и религиозной среды. Этот процесс демонстрирует, как менялись и взаимодействовали языковые традиции через письменные системы, создавая уникальные многослойные тексты, которые сохранились до наших дней.

Поэтому эти рукописи являются важным источником для изучения казахской поэзии, а также истории языка и культуры народа.

Особенно важным является выделение периодов, в которые сохранились рукописи казахской поэзии. Периодизация рукописей отражает эволюцию не только литературных жанров, но и изменений в письменных традициях и графических системах. Это позволяет исследователям более точно установить временные рамки создания поэтических произведений и выявить особенности каждого этапа, начиная с древней казахской поэзии, связанной с устным народным творчеством, и заканчивая поздними письменными образцами. Периодизация казахской поэзии тесно связана с общей периодизацией казахской литературы, поскольку поэзия является одной из ключевых форм литературного выражения.

В периодизации казахской литературы мы основываемся на трудах авторитетных исследователей, таких как Х. Суюншалиев (Суюншалиев, 1983), А. Кыраубаева (Кыраубаева, 1999), Н. Келимбетов (Келимбетов, 1991) и М. Жолдасбеков (Жолдасбеков, 1990), которые считают началом истории казахской литературы древнетюркский период V–VII вв. Данные ученые аргументируют свою позицию тем, что именно в этот период зарождаются основы казахской письменности и поэтической традиции, что отражается в фольклоре и древних текстах, сохранившихся в письменной форме. Далее по

<sup>1</sup> Научно-инновационный центр «Рукопись и фольклористика». Главная страница [Электронный ресурс] // Научно-инновационный центр «Рукопись и фольклористика». URL: <https://litart.kz/> (дата обращения: 13.11.2024).



хронологической периодизации казахской литературы следует Исламский период (X–XII вв.), за ним период Золотой Орды или кыпчакский период (XII–XIV вв.). Затем идет литература эпохи Казахского ханства (XV–XVIII вв.), литература первой половины XIX в., литература второй половины XIX в., и заканчивается период литературой начала XX в. (Жумагулов, 2012).

В связи с этим, полноценную текстовую базу, планируемую для ввода в историко-поэтический подкорпус, можно разделить на несколько хронологических периодов, отражающих этапы развития казахской поэзии. Такой подход позволит систематизировать материалы и обеспечить более глубокое понимание их историко-культурного контекста. Поэтому составители поэтического подкорпуса казахского языка, в соответствии с указанными этапами развития казахской литературы, обозначили эти периоды следующим образом:

1. Древнетюрская поэзия V–IX вв.,
2. Средневековая поэзия X–XIV вв.,
3. Староказахская поэзия XV–XIX вв.,
4. Казахская поэзия XX–XXI вв.

На первом этапе (2023–2025 гг.) разработки базы историко-поэтических текстов были охвачены староказахские поэтические произведения XV–XIX вв. Этот период выбран из-за большого объема сохранившихся текстов и их ключевой роли в развитии казахской поэзии. В этих произведениях ярко проявляется мастерство стихосложения, что делает их особенно важными для анализа поэтической традиции. Кроме того, они хронологически ближе и понятнее современным пользователям корпуса. Фокус на этом этапе позволяет систематизировать материал и заложить основу для последующего расширения подкорпуса с добавлением более древних текстов.

### Обзор корпусов

На сегодняшний день известно несколько примеров поэтических корпусов, доступных широкой публике. К ним относятся поэтические корпуса: башкирского языка<sup>1</sup>, русского языка<sup>2</sup>, чешского языка<sup>3</sup>, персидского языка<sup>4</sup>, и др.

Среди них одним из крупнейших электронных подкорпусов, имеющим большие размеры и имеющее особое поэтическое обозначение, является поэтический подкорпус русского языка, состоящий из 13 миллионов словоупотреблений. Работа по созданию Поэтического подкорпуса русского языка осуществлялась группой Института русского языка Российской академии наук (далее — ИРЯ РАН) под руководством В. А. Плунгяна, корпус был открыт для свободного доступа в декабре 2006 г. (Гришина и др., 2009). Тексты данного поэтического подкорпуса состоят из двух систем разметки. Во-первых, это лингвистическая разметка общего корпуса (грамматическая, морфологическая и пр.), а во-вторых, это поэтическая разметка, обозначающая структурные особенности стихотворений и предназначенная для литературоведов и лингвистов (метр, длительность строки, клаузула, строфа, рифма и т. д.) (Сичинава, 2012).

Поисковая система в интерфейсе русского поэтического подкорпуса позволяет найти не только отдельные слова и словоупотребления, встречающиеся в стихотворениях, но также предоставляет возможность поиска по таким структурным стихотворным параметрам, как *метр*, *строфа*, *клаузула*. Источник текстов в корпусе — тексты, предоставленные П. М. Нерлером, издательством «Академический проект», а также проектами «Русская виртуальная библиотека» и «Фундаментальная

<sup>1</sup> Башкирский поэтический корпус [Электронный ресурс] // Башкирский поэтический корпус. URL: <http://webcorpora.net/bashcorpus/search/> (дата обращения: 21.11.2024).

<sup>2</sup> Поэтический корпус русского языка. Главная страница [Электронный ресурс] // Поэтический корпус русского языка. URL: <https://ruscorpora.ru/corpus/poetic?search=CgQuAggJMAE=> (дата обращения: 19.11.2024).

<sup>3</sup> Корпус чешского стиха. Главная страница [Электронный ресурс] // Корпус чешского стиха. URL: <https://lindat.mff.cuni.cz/services/teitok-live/czechverse/> (дата обращения: 21.11.2024).

<sup>4</sup> Персидский поэтический корпус [Электронный ресурс] // Персидский поэтический корпус. URL: [https://linghub.ru/persian\\_poet\\_corpus/#](https://linghub.ru/persian_poet_corpus/#) (дата обращения: 21.11.2024).

<sup>5</sup> Поэтический корпус русского языка. О корпусе [Электронный ресурс] // Поэтический корпус русского языка. URL: <https://ruscorpora.ru/corpus/poetic?search=ChIqDAoICAAQChgyIApACjICCAk=> (дата обращения: 19.11.2024).



электронная библиотека»<sup>5</sup>. Также составители поэтического подкорпуса русского языка опирались на труды известного ученого, исследователя русской поэтики М. Л. Гаспарова (Гаспаров, 1974, 2002, 2013) при определении условий сбора текстов в корпусную базу (Орехов, 2015: 113–127).

В результате обозначены следующие требования, предъявляемые к текстам:

1. Авторские произведения, вносимые в поэтический подкорпус, должны входить в крупные книжные серии, содержащие издания русской поэзии и фольклора, как «Библиотека поэта»;
2. По авторским произведениям, вносимым в поэтический подкорпус, должны существовать исследования;
3. Для авторских произведений последних 50-ти лет, вносимых в подкорпус, должна существовать литературная критика (Корчагин, 2015).

На основе приведенных параметров, составители проводят тщательный отбор поэтических произведений. Это в первую очередь позволяет наиболее полно охватить произведения XVIII–XXI вв., вносимые в корпусную базу, и в результате составить базу отборных текстов.

Мы полагаем, что наличие чётких критериев отбора текстов является одним из ключевых преимуществ поэтического подкорпуса. По нашему наблюдению, особенностью этого подкорпуса является продвинутая метатекстовая разметка, которая охватывает:

- лексико-грамматическую информацию;
- маркировку ритмических и метрических особенностей стихотворений;
- разметку поэтических фигур, таких как аллитерации, рифмы, строфические формы и другие литературные приёмы (Сейтбекова, Елесбай, 2024).

Важным аспектом также является исторический контекст: в подкорпусе уделяется особое внимание изменениям языка и стиля на протяжении времени. Это позволяет исследователям проследить эволюцию поэтического языка на различных этапах его развития и изучать язык в динамике.

Таким образом, методологический подход русского поэтического подкорпуса направлен на детализированную разметку всех ключевых элементов поэтического текста. Это способствует более глубокому изучению как формальных, так и содержательных аспектов поэзии, а также изменений языка и стиля в историческом контексте.

Следующий поэтический корпус, имеющий лингвистическую (морфологическую) и поэтическую (метрическую) разметку текстов — это поэтический корпус башкирского языка. Данный корпус был создан в Башкирском государственном университете коллективом исследователей под руководством Б. В. Орехова в 2012–2013 гг. (Орехов, 2019а). Поэтический корпус башкирского языка на момент написания статьи охватывает около 17 тысяч поэтических произведений 101 поэта, написанных в XX–XXI вв., в целом содержащий 1,8 млн словоупотреблений<sup>1</sup>. Поиск текстов поэтического корпуса возможен с использованием лексико-грамматических и поэтических обозначений.

На наш взгляд, можно выделить следующие особенности поэтического подкорпуса башкирского языка:

- учёт агглютинативной природы башкирского языка в разметке. Башкирский язык относится к агглютинативным, то есть слова образуются путём последовательного присоединения аффиксов, что существенно влияет на морфологическую и синтаксическую структуру текста. В разметке необходимо учитывать каждую добавленную морфему, так как она может менять значение или грамматическую функцию слова;
- фокус на традиционных формах башкирской поэзии (например, *кыска йыр*, тюркские формы стихосложения). В подкорпусе особое внимание уделено классическим поэтическим формам, таким как *кыска йыр* ‘короткие песни’ и традиционные тюркские стихотворные формы. Это позволяет детально исследовать структуру стихосложения и ритмические особенности башкирской поэзии. Учет специфики метрического строя этих форм имеет ключевое значение для их анализа;
- семантическая и ритмическая разметка, отражающая музыкальные элементы, что особенно важно для эпических произведений. Ритмическая разметка является важным компонентом анализа,

<sup>1</sup> Башкирский поэтический корпус. Главная страница [Электронный ресурс] // Башкирский поэтический корпус. URL: <http://webcorpora.net/bashcorpus/search/> (дата обращения: 21.11.2024).



особенно для произведений, связанных с музыкальной традицией и устной эпической поэзией (например, баиты и эпические песни). Для точного анализа таких текстов необходимо учитывать ритмическую и мелодическую структуру произведений. Семантическая разметка также играет важную роль, помогая отразить культурные и исторические контексты, которые являются неотъемлемой частью содержания поэтических произведений.

Таким образом, поэтический подкорпус башкирского языка предоставляет уникальные возможности для всестороннего исследования башкирской поэтической традиции, её ритмических, лингвистических и культурных особенностей, а также её связи с устным народным творчеством.

Корпус чешского стиха (Corpus of Czech Verse) охватывает поэтические произведения на чешском языке, написанные в XIX–XX вв. В настоящее время корпус содержит почти 1700 поэтических сборников (почти 80 000 стихотворений, более 2,5 миллионов стихотворных строк). В создании корпуса чешского стиха участвовали исследователи Чешского литературного института, и первый период работы был завершён в 2013 г. (Plecháč, Kolár, 2015). В интерфейсе данного корпуса каждой словоформе в базе поэтического корпуса присвоена фонетическая транскрипция, определена поэтическая разметка в соответствии с грамматической категорией и структурными особенностями каждого произведения. На основе базы поэтического корпуса чешского языка создано приложение «Эвфонометр», позволяющее определить фонетическую мелодику чешских стихотворений (там же).

Корпус чешского стиха может быть незаменимым инструментом для исследователей языка поэзии при разработке области поэтики, рассматривающей мелодику звука.

Поэтический корпус персидского языка (опубликован в 2020 г.) — на данный момент один из самых молодых поэтических корпусов по времени создания и по объёму (Орехов, Степина, 2022). Корпус создан Б. В. Ореховым и Д. С. Степиной. Предварительная морфологическая разметка произведена Т. Л. Кононовой. Сегодня база данных корпуса включает 16842 поэтических произведения, охватывающих период с IX–XVII вв. База постоянно пополняется. Общая поэтическая база корпуса на момент написания статьи составляет 4,3 миллиона слов. Авторы корпуса особенно отмечают жанровое разнообразие корпуса, включающее в себя до 15 наименований (*газели, касыды, маснави, рубаи* и т. п.). Тексты персидского поэтического корпуса имеют морфологическую и поэтическую метрическую систему обозначений. По результатам изучения данного корпуса можно сказать, что одним из безусловных преимуществ метаразметки текста является транскрипция и перевод (с персидского на английский) каждого слова в стихах. Он представляет собой незаменимое вспомогательное средство для исследователей персидской поэзии и персидского языка, для востоковедов и тюркологов.

Исследуя опыт создания поэтических корпусов в современной корпусной лингвистике, мы также проанализировали следующие корпусные базы: полный корпус англосаксонской поэзии<sup>1</sup>; корпус древнеанглийской поэзии (по анализу Йорк-Хельсинки)<sup>2</sup>; базу эстонских рунических стихотворений<sup>3</sup>; поэтический корпус тувинского языка<sup>4</sup>, представленный поэтическими текстами тувинских писателей из литературного альманаха «Улуг-Хем» и других источников (Ш. М. Куулар «Собрание стихов», А. А. Даржай «Игил ызы»). Однако в результате сравнения указанных образцов корпусов с поэтическими корпусами русского, башкирского и персидского языков было выявлено, что вышеуказанные образцы не охватывают результаты поэтического анализа и являются не более чем базами данных, представленных в виде коллекции систематизированных поэтических произведений, принадлежащих к определенному временному периоду или жанру.

Исследование опыта создания поэтических корпусов для русского, башкирского, чешского и персидского языков показывает, что подходы к разработке корпуса должны учитывать уникальные лингвистические и культурные особенности каждого языка.

<sup>1</sup> The Complete Corpus of Anglo-Saxon Poetry. Главная страница [Электронный ресурс] // The Complete Corpus of Anglo-Saxon Poetry. URL: <https://sacred-texts.com/neu/ascp/> (дата обращения: 21.11.2024).

<sup>2</sup> The York-Helsinki Parsed Corpus of Old English Poetry. Главная страница [Электронный ресурс] // The York-Helsinki Parsed Corpus of Old English Poetry. URL: <https://www-users.york.ac.uk/~lang18/pcorpus.html> (дата обращения: 23.11.2024).

<sup>3</sup> The database of Estonian runic songs. Главная страница [Электронный ресурс] // The database of Estonian runic songs. URL: <https://www.folklore.ee/regilaul/introduction> (дата обращения: 23.11.2024).

<sup>4</sup> Электронный корпус тувинского языка. Главная страница [Электронный ресурс] // Электронный корпус тувинского языка. URL: <https://tuvancorpus.ru/?q=content/korpusy> (дата обращения: 23.11.2024).



Применяя этот опыт к казахскому языку, были выделены следующие ключевые аспекты, которые позволили создать историко-поэтический подкорпус, способствующий всестороннему исследованию казахской поэзии:

*Агглютинативная природа казахского языка.* Как и в башкирском корпусе, агглютинативная природа казахского языка требует разметки, которая позволит учитывать аффиксацию на морфологическом уровне. Это облегчит анализ синтаксической структуры и грамматических функций слов, создавая условия для точного исследования текстов;

*Метрика и ритмика в традиционных формах.* Важным элементом, аналогичным русскому и башкирскому, чешскому, персидскому корпусам, является акцент на метрических и ритмических особенностях текстов. В казахском корпусе следует обеспечить разметку, отражающую ритмические структуры традиционных форм стихосложения, таких как *айтысы*, *жыры* и *терме*, что позволит исследователям глубже анализировать эти формы с точки зрения поэтической структуры;

*Жанровое и стилистическое разнообразие.* Опыт персидского, чешского корпуса подчеркивает важность жанровой классификации, особенно при включении перевода текстов. Казахский поэтический корпус может включать разнообразные жанры, такие как *жыр*, *терме*, *дастан* и эпические произведения. Это позволит выделить уникальные особенности казахской поэзии и её взаимосвязь с традиционной литературной и устной культурой;

*Семантическая разметка и исторический контекст.* Как и в русском корпусе, семантическая разметка с учётом исторического контекста является необходимой для анализа культурных и исторических элементов. В казахском корпусе это позволит не только исследовать язык и стиль в динамике, но и проследить изменения, связанные с историческим развитием, создавая условия для исследований в области культурологии и исторической лингвистики.

Созданный на основе выделенных рекомендаций историко-поэтический подкорпус казахского языка даёт возможность исследователям доступ к инструменту, способствующему не только изучению формальных и содержательных аспектов казахской поэзии, но и углубленному пониманию её культурного значения. Примеры русского, башкирского, чешского и персидского корпусов демонстрируют, что продвинутая лексико-грамматическая, метрическая и семантическая разметка, дополненная жанровой классификацией, создаёт основу для межкультурных исследований и помогает в сохранении и изучении культурного наследия через призму языка и поэзии.

### ***Лингвотекстологическая работа с оцифрованными рукописями***

подавляющее большинство старо-казахских поэтических произведений XV–XIX вв. было записано арабской графикой и в настоящее время хранится в Научно-инновационном центре «Рукопись и фольклористика» при Институте литературы и искусства имени М. О. Ауэзова<sup>1</sup>. Эти рукописи представляют собой важный источник для изучения казахской поэтической традиции, и их сохранность позволяет провести лингвотекстологический анализ, а также внести их в текстовую базу историко-поэтического подкорпуса для дальнейших исследований. За годы независимости, в период с 2004 по 2013 г., специалисты Института литературы и искусства им. М. Ауэзова в рамках государственной программы «Культурное наследие» подготовили 100 томов фольклорного наследия и представили его широкой публике в серии «Слова предков» («Бабалар сөзі»). Разработчики многотомной серии в предисловии к первому тому указывают на несколько причин, по которым некоторые поэтические образцы устной литературы, сохранившиеся в виде рукописей, не были ранее опубликованы. Во-первых, они не соответствовали общественным и идеологическим интересам. Во-вторых, характер фольклорных произведений не был учтён, и их не публиковали из-за требований, предъявляемых к письменной литературе. Опубликованные тексты часто подвергались сокращению и редактированию из-за строгого политико-идеологического цензурного контроля<sup>2</sup>. Сегодня эти издания, вышедшие многотомной серией, стали богатым исследовательским материалом для лингвистов.

<sup>1</sup> Научно-инновационный центр «Рукопись и фольклористика». Алтын қор цифровые ресурсы [Электронный ресурс] // Научно-инновационный центр «Рукопись и фольклористика». URL: <https://litart.kz/> (дата обращения: 13.11.2024).

<sup>2</sup> Бабалар сөзі [Слова предков] : в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2004. Т. 1. С. 5–6.



По степени обнаружения, публикации и распознавания поэтических образцов устной литературы, сохранившихся в виде рукописей, их можно разделить на три группы.

*Первая группа* включает *жыры* и *дастаны*, которые ранее не были опубликованы. Некоторые из них основаны на сюжетах классической восточной литературы. Например, хисса Асылбека и Гүлжиһан (Асылбек пен Гүлжиханның қиссасы), Мариям сирота (Жетім Мариям), Карга батыр (Қарға батыр), хисса Шеризат Кулшат (Шеризат Құлшат қиссасы), Мырқы батыр (Мырқы батыр), хисса Нугыман-Нагиман (Нұғымен-Нағиман қиссасы), Сырым батыр (Сырым батыр), Этот хисса Зенгир Жухуд (Үшбу хисса Зеңгір Жүһүд), История о Абушахыма (Хикая Әбушаһыма), и т. д.

*Вторая группа* включает распространенные среди народа, опубликованные произведения, такие как «Козы Корпеш — Баян сулу» (Қозы Көрпеш — Баян сұлу)<sup>1</sup>, «Кыз Жибек» (Қыз Жібек)<sup>2</sup>, «Богенбай батыр» (Бөгенбай батыр)<sup>3</sup>. Поскольку тексты эпосов и дастанов претерпели множество веков, у них бывает по два и более варианта. Порой основной сюжет произведения пополняется новыми историями, а во втором варианте такие изменения не сохраняются, тогда как в третьем варианте конфликт и развязка могут быть совершенно иными.

*Третья группа* охватывает малоизвестные произведения с несколькими вариантами. Например, «Мырзаш батыр» (версия Ниязбека Кыдырмоллы, 1–2–3 части), «Мырзаш батыр» (версия Сагидоллы Нуралина), «Мырзаш батыр» (версия Ержана Ахметова)<sup>4</sup> и т. д.

Составителями историко-поэтического подкорпуса казахского языка в рамках Национального корпуса казахского языка были выделены следующие трудности и лингвистические особенности при анализе арабографичных рукописей:

- в арабографичных рукописях встречающиеся среднетюркские лексические и грамматические употребления зафиксированы в неизменном виде, как в оригинале: *түрлік, тірік, дүр, бірлән, илә, сәна, үшбу, анығ, андын, айла, соңыра*, и т. д.;
- также в этих рукописях арабские и персидские слова были написаны на языке оригинала по принципу цитирования. Например, *Аллаһ, нашында* и др.

Следует отметить, что процесс преобразования арабографичных рукописей в кириллическую графику, используемую в настоящее время, осуществлялся различными специалистами, и данный опыт не является уникальным. В ходе этих преобразований значительное количество арабо-персидских слов было адаптировано с учётом фонетических особенностей казахского языка, а не исходных языков. В результате такие заимствованные лексические единицы, сохраняя своё семантическое содержание, получили графическое оформление и произношение, соответствующее казахской фонетической системе. Данная адаптация способствовала упрощению восприятия текстов казахоязычной аудиторией, обеспечивая их доступность при сохранении культурной и исторической значимости оригиналов. Например, хотя в рукописи слово *Аллах* дано как *Аллаһ*, в кириллице сохраняется форма «Алла».

Одно из таких персидских слов, адаптированных к фонетическим особенностям казахского языка, — это *нашін*. Оно прошло процесс фонетической адаптации, сохраняя свое значение, но изменяясь в соответствии с правилами произношения и звуковой системы казахского языка. В арабографичном тексте рукописи слово написано как *нашында* (рис. 1), но при преобразовании в кириллицу записан вариант *нашінде* с мягкими гласными, которая более соответствует казахскому произношению.

<sup>1</sup> Бабалар сөзі [Слова предков] : в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2009. Т. 54. С. 11–54. Вариант: Ш.Уалиханова.

<sup>2</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2006. Т. 26. С. 88–287; Бабалар сөзі [Слова предков] (2009а): в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2009. Т. 53. С. 37–82. Вариант: Мусабай жырау.

<sup>3</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана : Фолиант, 2010. Т. 58. С. 11–138.

<sup>4</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2010. Т. 63. С. 97–349. Версия: Ниязбека Кыдырмоллы, С. 350–362. Версия: Ержана Ахметова, С. 363–410. Версия: Сагидоллы Нуралина.

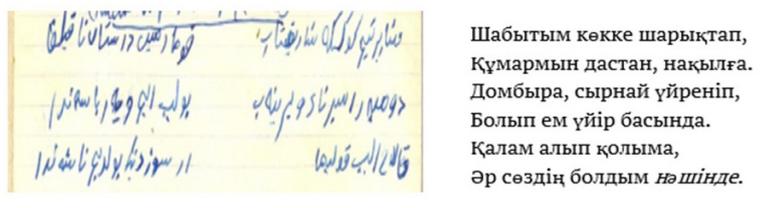


Рисунок 1. Отрывок из второй страницы произведения

«Қисса Шеризат — Құлшат»<sup>1</sup>.

Figure 1. Excerpt from the second page of the work “Kissa Sherizat — Kulshat”.

Согласно содержанию стихотворения, это слово обозначает «по ритму речи». В казахском языке по сей день встречаются словосочетания *нашына/нәшіне/машына келтірді*. Это согласовано с мнением Р. Сыздықовой о том, что «слова в казахском языке гармонируют по мягкости-твердости в соответствии с сингармоническим тембром. Несмотря на обозначение слова, несмотря на присутствие и мягких, и твердых гласных, в целом оно должно обозначаться либо в мягком, либо в твердом варианте» (Сыздықова, 2000: 87).

Кроме того, необходимо выделить несколько ключевых трудностей, с которыми сталкиваются специалисты при преобразовании арабографичных текстов рукописей в кириллическую графику:

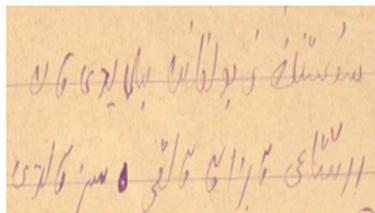
- отсутствие гласных и знаков препинания в арабской вязи. Арабская графика не предусматривает использование заглавных букв и знаков препинания, а также содержит скрытые гласные. Эти особенности существенно осложняют точное воспроизведение и интерпретацию текста при переводе его на кириллицу;
- фонетическая несовместимость языковых систем. Казахский и арабский языки принадлежат к различным языковым системам, что делает арабскую письменность недостаточной для точной передачи звуков казахского языка;
- отсутствие единых стандартов для использования арабской вязи. В XIX в. не существовало общепринятых норм для применения арабской графики в казахских текстах, что приводило к вариативности написания одних и тех же слов;
- проблемы с устаревшими лексическими единицами. Распознавание и корректное преобразование лексических архаизмов и устаревших слов также представляют значительные трудности. При переводе этих слов на современную графику нередко возникают орфографические ошибки, а их форма и содержание изменяются до неузнаваемости, что затрудняет их правильный анализ и интерпретацию (Сыздықова, 2009: 92).

Каждая из указанных проблем усложняет процесс перевода арабографичных рукописей в кириллическую графику и требует системного и внимательного подхода. В связи с этим специалистами и составителями корпуса была проведена лингвотекстологическая работа с рукописями и их найденными кириллическими вариантами, включающая коррекцию и адаптацию текста, установление точного смысла, внесение редакционных и корректорских правок, а также выравнивание нумерации и согласование арабской графики с кириллическими текстами. Например, при сверке арабографичной рукописи с её найденным кириллическим вариантом выражение *жеткізер бізді димешке* было исправлено на *жеткізер бізді бейішке* ‘доставят нас в рай’ в соответствии с контекстом. Исправление было необходимо, так как в казахском языке отсутствует слово «димеш», что указывает на допущенную грамматическую ошибку.

Далее некоторые выдержки из исправлений, внесенных в некоторые тексты.

В эпосе «Как Галдан-Церен взял Абылая в плен» (Қалдан Серен Абылайды тұтқынға алғаны) (рис. 2) нами была произведена следующая замена: *Арттағы қазақ халқы иесіз қалды* ‘И остался народ казахский без владыки’. Причина замены словосочетания *иесіз қалды* ‘и остался народ казахский в смятении’ на *иесіз қалды* объясняется контекстом. В предшествующих и последующих строфах опи-

<sup>1</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2005. Т. 18. С. 158–199.



Аң іздеп Абылай хан ұйықтап қапты,  
Мың батыр қыз алатын мұны тапты.  
Мың батыр жолы болып Абылайдан,  
Танымай сол арада жөн сұрапты.  
Сол жерде хан Абылай ойға қалды,  
Жан-жаққа артын ойлап көзін салды.  
*Соғыстың, не болғанын білмейді хан,  
Артағы қазақ халқы иесіз қалды.*

Рисунок 2. Отрывок из третьей страницы рукописи под названием «Как Галдан-Церен взял Абылая в плен» (Қалдан Серен Абылайды тұтқынға алғаны)<sup>1</sup>.

Figure 2. Excerpt from the third page of the manuscript entitled “How Galdan-Tseren captured Abylai” (Kaldan Seren Abylaydy tutkynga algany).

сывається, как далада ұйықтап жатқан Абылайды, қалың қалмақ қапыда байлап алғаны ‘когда спал Абылай в степи, калмыки внезапно захватили его’. Абылай был казахским ханом, и, таким образом, естественно, что народ, лишившийся своего хана, остался без владыки, что логически соответствует контексту повествования.

Также при сверке арабографичной рукописи с её найденным преобразованным кириллическим вариантом нами было обнаружено, что следующее предложение из рукописи вовсе отсутствовало:

*«Қазақ халқы басқа халыққа бағынады, күндердің күнінде менің бұл сөзім тарихта қалмақ, қош Олжабай, — деп қоштасты. Олжабай қайтып кетеді. Хан Абылай алған Тобыш сұлудан Қасым хан туғанда қан ұстап жалғыз туады, онан кейін Сары Наурызбай туады»*

‘Казахский народ будет повиноваться другому народу, однажды эти мои слова останутся в истории, прощай, Олжабай, — попрощался он. Олжабай уезжает. От красавицы Тобыш, на которой женился хан Абылай, рождается сын Касым хан, рождается один, в крови; затем рождается Сары Наурызбай’.

Изменение отдельных слов или фраз в произведениях устного народного творчества при преобразовании с арабографичного текста на кириллицу может привести к ряду негативных последствий, затрагивающих как содержание, так и форму произведений:

- искажение устной традиции. Изменение текста нарушает аутентичность устного народного творчества, что приводит к утрате культурных и этнолингвистических элементов;
- нарушение ритмико-мелодической структуры. Модификация слов может разрушить ритм и мелодику, важные для запоминания и передачи произведений;
- искажение смыслов. Лексические изменения могут исказить или утратить исходные значения, затрудняя правильную интерпретацию произведений;
- нарушение фонетической целостности. Преобразование может нарушить сингармонизм и фонетическую гармонию, что ухудшит звучание текста;
- потеря эмоциональной и эстетической ценности. Лексические изменения могут снизить эмоциональный и художественный эффект произведений.

Вышеизложенное позволяет сделать вывод, что ошибки, допущенные в процессе преобразования произведений устного народного творчества с арабографичных текстов на кириллицу, не позволяют сохранять их культурную, историческую и лингвистическую ценности. Эти недостатки исправляются составителями и специалистами подкорпуса путём сверки оригинальных рукописей с кириллическими версиями, найденными в процессе исследования.

<sup>1</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2005. Т. 57. С. 257-272.



### **Историко-поэтические метатекстовые разметки**

Параметры метаразметки текста отличаются особыми метатекстовыми параметрами в связи с содержанием подкорпуса. Например, в метаразметке поэтического подкорпуса русского языка указано 17 параметров, 6 из них — общие, 11 параметров взяты в соответствии со структурными особенностями русских стихотворений: число строк, строфика, графическая строфика, метр, число стоп/иктов, слогов, клаузула, рифма, дополнительные параметры, метрическая формула, предложение, словоформа.

Поэтическая метатекстовая разметка историко-поэтического подкорпуса образцов устного народного творчества XV–XIX вв., разрабатываемого в рамках НККЯ, включает 28 параметров. Эти параметры можно разделить на три группы по видам информации о тексте, которую они предоставляют:

1. Базовые параметры, используемые во всех корпусах: название поэтического произведения, тема текста, стиль текста, жанр текста, возраст аудитории, вид подкорпуса, количество словоупотреблений, век и год, количество страниц в тексте, текст введен в корпус (кем), время введения текста;

2. Параметры, обозначающие поиск, публикацию, известность или неизвестность авторов рукописей, принадлежащих к древним эпохам: краткая аннотация текста, повествователь текста, рукопись предоставлена (кем), графика текста, первая публикация текста, варианты текста, место хранения рукописи, текст перевел, форма текста, текст пронумерован (кем), примечания;

3. Необходимые параметры, связанные со структурой и размером стихотворений в поэтическом тексте: количество строф, количество строк, количество стоп, количество слогов, вид рифмы.

Остановимся на принципах использования некоторых параметров, которые считаются необходимыми среди поэтических обозначений текстов, входящих в текстовую базу историко-поэтического корпуса:

*Краткая аннотация текста.* Краткая информация об основном содержании произведения;

*Повествователь текста.* У некоторых произведений, вносимых в корпус, отсутствует автор (сказитель). В данном случае, при наличии, указывается имя-отчество сказителя;

*Рукопись предоставлена (кем).* Рукописное наследие, не опубликованное в советское время из-за жесткой социально-идеологической цензуры, исчезнувшее или хранимое в домах по сей день, порой сдается в центры редких рукописей. Поэтому необходимо указывать имя-отчество человека, предоставившего рукопись;

*Век и год/период — ясно или неясно.* У некоторых рукописей отсутствует год публикации или его сложно разобрать. Поэтому ставится пометка «ясно» или «неясно»;

*Графика текста.* Поскольку образцы устной литературы XV–XIX вв. были написаны разной графикой (арабско-кадимской, арабско-джадидской, арабской тоте), то приводятся сведения о том, какой графикой был написан текст;

*Первая публикация текста.* Первые публикации ранних дастанов и эпосов не сохранились. Поэтому читателю важно показать первую публикацию;

*Место хранения рукописи.* Ценные сведения о том, где хранятся оригиналы устной литературы, написанной арабской и латинской графикой;

*Варианты текста.* Ценные сведения о том, что существуют две и более версии ранних устных дастанов и эпосов;

*Примечания.* Большинство рукописных текстов редкой рукописи арабской графикой были записаны в зашифрованных, пронумерованных связках (папках) тетрадей, а на страницах тетрадей стихи записывались по-разному. Стихи, написанные крупным шрифтом у некоторых авторов, занимают одну страницу, а несколько строк стихотворений, написанных некоторыми авторами мелким шрифтом, умещаются на одной странице тетради. В связи с этим параметр «Примечания» был взят согласно порядку расположения стихов на бумаге.

На основе вышеизложенных параметров метатекстовой разметки составителями, а также исследователями историко-поэтического подкорпуса казахского языка была предпринята попытка составить метатекстовую характеристику по образцам имеющихся у нас рукописей. Образец одного из таких текстов представлен в (табл. 1).



Таблица 1. Образец «Историко-поэтической метатекстовой разметки» жыра (эпоса) «Айдос батыр».  
Table 1. Sample of the “Historical and poetic metatext markup” of the zhyr (epic) “Aidos Batyr”.

Параметр разметки	Содержание
Название поэтического произведения	Айдос батыр
Краткая аннотация текста	<p>Жырдың оқиғасы қызықты, тартымды болып келеді. Жыршы Айдостың шайқас кезіндегі ерлігін сипаттауға ерекше мән берген. Батырдың соғыс кезіндегі күш-қайраты, ерлігі айрықша мадақталып суреттеледі. Сол сияқты “Айдос батыр” жырының сюжеттік құрылымында әлем халықтарының фольклорлық шығармаларында жиі ұшырасатын түс көру, түс жору, бал ашу, жалбарыну, бата беру, қоштасу, атымен арыздасу, жоқтау сияқты тұрақты мотивтер ұшырасады.</p> <p>‘История эпоса интересна и привлекательна. Сказитель уделяет особое внимание описанию героизма Айдоса во время сражения. Восхваляется, ярко описывается сила батыра во время войны, его подвиги. Также в сюжете жыра “Айдос батыр” встречаются такие мотивы, характерные для фольклорных произведений народов мира, как увидеть сон, гадание, мольба, бата, прощание, ссора с конем, оплакивание и пр.’</p>
Тема текста	Об Айдос батыре
Сказитель текста	Айдарбеков Жүмаділ
Век или год/период «ясно» или «неясно»	*
Графика текста	Араб-тоте
Первая публикация текста	Некоторые отрывки из жыра «Айдос батыр» были опубликованы в 1883 году в Ташкенте в «Киргизской хрестоматии» Я. Лутша.
Стиль текста	Художественный
Жанр текста	Исторический эпос
Место хранения рукописи	Рукопись эпоса хранится в Фонде Рукописей ОГК (Ш. 765, тетр. 3).
Варианты текста	*
Форма текста	Устная
Количество строк	8 строф
Количество строк	30 строф
Количество стоп	3
Количество слогов	11
Вид рифмы	Рифма «қара өлең» ( а а б а )
Возраст аудитории	Нейтральный
Текст перевел	Қарашаш Алпысбаева
Вид подкорпуса	Историко-поэтический подкорпус
Количество словоупотреблений	3911
Количество страниц в тексте	25
Текст пронумеровал(а)	Әйгерім Мурсал
Текст введен в корпус	Гүлжиһан Көбденова
Период введения в корпус	17.11.2023
Примечание:	В два ряда на каждой странице

Прим.: звёздочкой (\*) в графах «Век или год/период» и «Варианты текста» обозначены случаи отсутствия конкретных данных. В первом случае это связано с устной природой произведения, из-за чего невозможно установить точный период его возникновения. Во втором случае звёздочка указывает на отсутствие других известных вариантов данного текста.



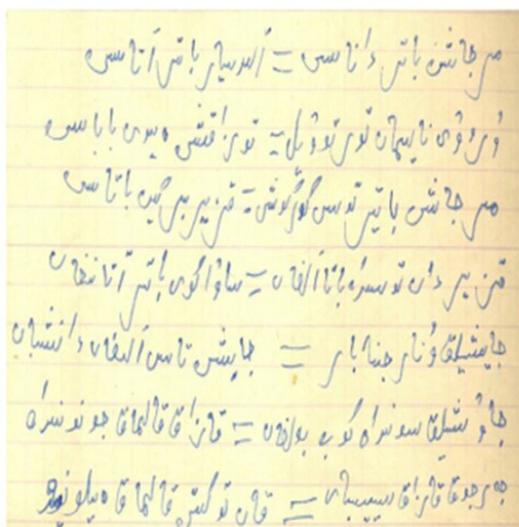
И так по составленным образцам метатекстовой разметки сделаны следующие выводы. Во-первых, перед созданием модели метатекстовой разметки историко-поэтического подкорпуса образцов устного народного творчества XV–XIX вв., разрабатываемого в рамках НККЯ, был проведён всесторонний анализ и сравнительное исследование аналогичных параметров в ведущих поэтических подкорпусах, таких как поэтический подкорпус русского языка, поэтический подкорпус башкирского языка, а также поэтический подкорпус персидского языка. Во-вторых, с учётом особенностей сохранившихся рукописей устной литературной традиции казахского языка, специалистами и составителями данного историко-поэтического подкорпуса была разработана более точная и детализированная модель метатекстовой разметки, включающая 28 параметров. Данная разметка учитывает специфические особенности анализируемых текстов, что обеспечивает её высокую точность и расширенную применимость для описания материалов устной поэтической традиции.

### Поэтическая разметка

Структура казахских стихотворений является категорией теории казахской литературы и считается основным условием и закономерностями системы казахских стихотворений. Первым, кто изучал структуру казахской поэзии как категорию теории литературы, был А. Байтұрсынов. Он разделил параметры стихотворения на следующие: строфа, строка, стопа, рифма (Байтұрсынұлы, 1991: 82).

К. Омиралиев, исследователь языка казахской поэзии, отмечает основные признаки стихотворений XVIII–XIX вв.: 1) повторение вслед за тем, кто говорит первым; 2) нестабильная рифма: а/б/а/б, а/б/в/ю; 3) вид ритма и интонации; 4) отсутствие стабильности в количестве слогов; 5) ударение на первую стопу стихотворной строки (Өмірәлиев, 1976: 82). Мы выявили ряд особенностей рассматриваемых поэтических текстов и сочли целесообразным остановиться на каждой из них.

Если в рукописи стихотворение не делится на строфы и является астрофическим, при переводе на кириллицу строки, рифмующиеся друг с другом и обозначающие единую мысль, объединяются в строфу, то есть строки не делятся по четыре, по шесть или по восемь согласно правилам, а формируют многострочные строфы с разным количеством строк — по 4, по 5, иной раз по 20 или более (рис. 3).



#### Вариант на кириллице

Мырзаш батыр данасы,  
Алдияр батыр атасы.  
Руы найман төртуыл,  
Тұзақшы еді бабасы.  
Мырзаш батыр түс көргіш,  
Қызыр берген батасы.

Қызырдан түсте бата алған,  
Сәуегәй батыр атанған.  
Жайшылық ұнар және бар,  
Жайшын тас алған данышпан.  
Жаушылық сонан көп болған  
Қазақ, қалмақ жонынан.  
Жері жоқ қазақ сиыспай,  
Қан төкті қалмақ елінде

Рисунок 3. Отрывок из жыра Мырзаш (по варианту Ержана Ахметова)<sup>1</sup>.  
Figure 3. Excerpt from Myrzash fat (according to Yerzhan Akhmetov's version).

Многие исследователи отмечали, что объединение множества строк одной рифмой — распространённое явление в эпосах (Сыздықова, 1970: 22). В поэтических произведениях башкирского языка

Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2010. Т. 63. С. 350.



в изосиллабических и гетеросиллабических стихотворных формах 7–8-строчные строфы называются *кыска-күй*, а 9–10-строчные — *узун-күй* (Орехов, 2019b: 41). В связи с этим в эпохах многострочные стихотворения, обозначающие единую мысль, объединены в одну строфу.

*Количество слогов.* Основа ритма казахской поэзии — слог. Такая силлабическая система стихосложения, основанная на количестве слогов, характерна также для узбекского, татарского, каракалпакского, киргизского, уйгурского, азербайджанского, туркменского языков (Ахметов, 1973). Исследователь уйгурской поэзии А. Хамраев отмечает: «семи-восьми-слоговые, также десяти-слоговые стихотворные размеры являются древними образцами тюркской поэзии»<sup>1</sup>. К примеру, в каракалпакском дастане «Шарьяр» количество слогов — 7–8. Приведем отрывок из этого дастана:

Бурынғы айем заманда,  
Ол заманнын кадиминде.  
Коп ноғайдынг ишинде,  
Уллы ноғай журтында (Сапарниязов, Хожаниязов, 1959: 11).

Сюжет этого дастана схож с сюжетами таких эпосов, распространенных среди узбекского, каракалпакского, татарского, башкирского и алтайского народов, как «Алпамыс» и «Кобыланды». Возможно поэтому и в казахской поэзии встречаются 7–8, 11 слогов.

*Количество стоп.* Стихотворный размер зависит не только от количества слогов в строке, но и от количества стоп. Стопа (*икт*) — это колебания звуковой волны, заметные при декламировании стихотворения. Стопа — наименьшая единица казахского стихотворения, а в лингвистике это универсальная семантически-синтаксическая, интонационная наименьшая синтагма (Базарбаева, 2022: 337). З. Ахметов, описывая структуру стихотворения, говорит, что первые стопы казахских стихотворений состоят из трех слогов, срединные из четырех, а заключительные — снова из 4 (Ахметов, 1973: 81). К примеру, в «Хиссе Асылбека и Гүлжихан»<sup>2</sup> (рис. 4).

3	4	4
Патшаның / оң көзінде / бұлар болып, /		
3	4	4
Сыйласып / зор сияпат, / құрмет көріп. /		
3	4	4
Байларды / құзырына / қабыл етті, /		
3	4	4
Падиша / Ғабдүмәлік / көңіл бөліп. /		

Рисунок 4. Стихотворная структура «Хиссы Асылбек и Гүлжихан»<sup>3</sup>.  
Figure 4. The poetic structure of “Khissyasyylbek and Gulzhikhan”.

Б. Кенжебаев объясняет перемену мест стоп по количеству слогов следующим образом: «Есть три причины, по которым стопы стихотворных строк и слоги в составе стоп могут поменяться: по словесному ударению, по тому, какую мысль хочет высказать поэт, по ударению и ритму стихотворения» (Кенжебаев, 2004: 81). В корпусных текстах стопа обозначается косой линией. В основном такие стихотворения состоят из 2 или 3 стоп.

*Вид рифмы.* Одно из главных требований поэзии — рифма. Есть несколько видов рифмы. В рассматриваемых нами поэтических произведениях наиболее часто встречаются следующие виды: рифма *қара өлең* (А-А-Б-А); сплошная рифма (А-А-А-А); вольная рифма (А-Б-В-Г-Б); попеременная рифма (А-Б-В-Б); перекрестная рифма (А-Б-А-А); парная рифма (А-А-Б-Б); перекрестная рифма (А-А-Б-В-Б-В-Б-Б-Б); современные виды рифмы (А-Б-А-Б-В-Г-Д). Мы опираемся на высказывание А. Лорда о том, что «Формула обозначает важную, основную идею произведения» (Lord, 1991: 107). Такие формулы позволяют определить виды рифмы в стихотворении.

<sup>1</sup> Хамраев А. Т. Рифма и звуковая организация уйгурского стиха (в сопоставительном плане с русским стихом) : дисс. ... кандидата филол. н. Алматы, 1994. С. 8.

<sup>2</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2005. Т. 21. Указать страницу. С. 222.

<sup>3</sup> Там же.



Ниже рассмотрим примеры для каждого из видов рифм:

«Аңшыбай батыр» (вариант 1) — *сплошная рифма (А-А-А-А)*:

— *Қайсы болса да, біл, — деді. (А)*

*Орақшы мінген кер еді. (А)*

*Ол да мықты ер еді. (А)*

*“Алысалық, кел”, — деді. (А)*

В строках стихотворения по 7–8 слогов. Каждая строка четко обозначает мысль и показывает идею сказителя. В основном стихотворения, написанные сплошной рифмой, динамично и логично демонстрируют высказываемую мысль.

Стихотворение Майлыкожи Султанкожаулы «Көштің көркі түйе еді» написано с рифмой (А-Б-В-Б):

*Көштің көркі түйе еді, (А)*

*Шыңырауға қауға байлап бақ. (Б)*

*Жетекшіл келер жылқы мал, (В)*

*Семізін мініп, айдап бақ. (Б)*

В этом стихотворении также по 7–8 слогов. Попеременная рифма, объединяя по две строки, украшает стихотворение. Если верблюды — украшение коша (кочевья), то лошади — это ценный скот, который можно вести на поводу: проводится сравнение. Этот вид рифмы распространен в казахской поэзии со времен Абая. В каждой строфе рифмуются вторая и четвертая строка, и количество строк не превышает 4.

Эпос Мурына Сенгирбаева «О Аншибай батыре» (Аңшыбай батыр туралы) (11-й вариант) написан парной рифмой (А-А-Б-Б).

*Ақылың жоқ баласың, (А)*

*Қарғам, қайда барасың? (А)*

*Менде бір киер киім жоқ, (Б)*

*Сататұғын тиын жоқ. (Б)*

Попарно рифмуются строки стихотворения. Первая строка и вторая строка, третья и четвертая. Автор приводит свои примечания и замечания в стихотворении, а также информацию и о своем состоянии.

Хисса Ниязбека Кыдырмоллы «Мырзаш батыр» написана рифмой *қара өлең (А-А-Б-А)*:

*Бастайын бір-екі істі келсе шамам, (А)*

*Баянсыз сұм дүнияға болмай алаң. (А)*

*Құдайым өзі оңдаса сапарымды, (Б)*

*Ұзартып алға қарай создым қалам. (А).*

В таком виде рифмы четко прослеживается мысль каждой строки, и рифмуются первая, вторая и четвертая строки. Автор умещает в одну строфу мысль о том, как он, не отвлекаясь ни на что, берет перо в руки и пытается завершить пару дел по мере возможности.

Исторический эпос «Сырым батыр» Азамата Тайыра Ауатанулы, собирателя западно-казахстанского устного народного творчества, написан вольной рифмой (А-Б-В-Г-Б):

*Екі ғасыр бұрынғы, (А)*

*Сонау атам заманда. (Б)*

*Айбат шегіп ысқырып, (В)*

*Екі басты айдаһар, (Г)*

*Аранын ашып адамға. (Б).*

В стихотворении 7–8 слогов. В каждой строфе рифмуются вторая и пятая строки, остальные строки не повинуются общему правилу. Используются разные виды рифмы. Однако все пять строк обращают внимание на одну проблему и выражают одну мысль.



Хисса Асылбека и Гүлжихан (Асылбек пен Гүлжихан қиссасы), Мариям сирота (Жетім Мариям), Карга батыр (Қарға батыр), Хисса Шеризат Кулшат (Шеризат Кулшат қиссасы), Мырқы батыр (Мырқы батыр), Хисса Нұғыман-Нағиман (Нұғыман-Нағиман қиссасы), Сырым батыр (Сырым батыр), Этот хисса Зенгир Жухуд (Үшбу хисса Зеңгір Жүһүд), История о Абушахыма (Хикая Әбушаһыма) и прочие — основаны на классических сюжетах восточной литературы.

В процессе внесения материалов в корпус было замечено, что часто встречается тип рифмы А-А-Б-А. К примеру: эпос Айдарбекова Жумадила «Айдос батыр», стихотворение Майлыкожи Султанкожаулы «Стихотворение хана Худияра его сыну» (Қудияр ханның ұлына айтқан өлеңі), Посвящение Майлыкожи Султанкожаулы «Мулле Есенбаю, отправляющемуся в хадж» (Қажылық сапарына аттанып бара жатқан Есенбай молдаға), «Стихотворение-жалоба Кадиру» (Қадірге айтқан арыз өлеңі) Майлыкожи Султанкожаулы, «Как поэт Майлы восстанавливал прошлую справедливость» (Майлы ақын өткен датқаларды қылған сипаты) Майлыкожи Султанкожаулы, «Второе слово поэта Майлы торе Сейтжаппару» (Сейітжаппар төреге екінші айтылған Майлы ақын сөзі) Майлыкожи Султанкожаулы, «Хисса батыра Мырзаша» Ниязбека Кыдырмоллаулы, «Мырзаш батыр» Сагидоллы Нуралина, эпос Ашима Нурлыбаева «Мырқы батыр» (Мырқы батыр), «Асанқайғы» (Асанқайғы) Ильяса Ыскакулы, эпос «Богенбай батыр» (Бөгенбай батыр), эпос «Мырзаш батыр» Ержана Ахметова, «Этот хисса Зенгир Жухуд» (Үшбу хисса Зеңгір Жүһүд) Кошенова Ташмухаммеда Баймухаммедулы, «Хисса Асылбека и Гүлжихан» (Асылбек пен Гүлжихан қиссасы) Жусупбекулы Нура, «Қарга батыр» (Қарға батыр) Абиқая Кусайынұлы, «Мариям сирота» (Жетім Мариям) Машырапа Ақылбекова и пр. Этот вид стихосложения был замечен в середине девятнадцатого века.

Название *қара өлең* было впервые задокументировано и исследовано Чоканом Валихановым (Валиханов, 1986: 55). Причина, по которой этот вид рифмы наиболее часто встречается в рукописях — связь этой рифмы с устным народным творчеством казахского народа. А. Сейдимбек отмечает: «В сравнении с другими видами рифмы, “қара өлең” удобна для передачи отдельной мысли в отдельной строфе. Это — поэтическая галерея образа жизни, морально-этических норм, понятий о красоте, чувств и состояний»<sup>1</sup>.

Поэтическая разметка историко-поэтического подкорпуса казахского языка разработана после детального анализа текстовой базы, а также основывается на выводах, полученных в результате изучения трудов ведущих исследователей казахского стиха, таких как А. Байтурсынов и К. Омиралиев. Их исследования позволили выделить ключевые особенности казахской поэзии, такие как структура строф, рифм и ритмическая организация, которые будут отражены в разметке подкорпуса:

- структура строф: разметка будет фиксировать количество строк в строфах, включая многострочные строфы, характерные для эпических произведений. Это позволит определить единство мысли и ритмическую организацию текста;
- количество слогов: основой ритмической разметки станет количество слогов в строке, что важно для силлабической системы казахской поэзии. Будет фиксироваться диапазон слогов от 7–8 до 11, что характерно для тюркской поэзии;
- стопы: разметка будет включать информацию о количестве и типах стоп, так как это влияет на ритм и интонацию стихотворения. Особое внимание уделяется изменению стоп в зависимости от словесного ударения и ритмической структуры текста;
- типы рифмы: в разметке будут отмечены виды рифм, такие как «қара өлең» (А-А-Б-А), сплошная рифма (А-А-А-А), перекрёстная и попеременная рифма. Это поможет в изучении закономерностей рифмования в казахской поэзии;
- эволюция и традиции: поэтическая разметка будет отражать преемственность традиций устного народного творчества и их интеграцию в письменные формы, фиксируя влияние восточной литературы и народных жанров на казахскую поэзию.

Таким образом, разметка будет охватывать ключевые элементы казахского стихосложения и позволит систематизировать структурные и ритмические особенности поэтических текстов в подкорпусе.

<sup>1</sup> Сейдімбек А. Қазақтың дәстүрлі халық лирикасындағы қара өлең жанры [Жанр «қара өлең» в традиционной казахской народной лирике]: автореф. дис. ... канд. филол. н. Алматы, 1989. С. 34.



### Семантическая разметка

В текстах устной литературы казахского народа XV–XIX вв. встречается множество слов, которые устарели и в настоящее время могут быть незнакомы современным носителям казахского языка. Эти архаизмы отражают культурные и языковые особенности эпохи, а также традиции и обычаи, которые изменились со временем. Их изучение позволяет глубже понять исторические контексты и развитие казахского языка и культуры. Поскольку база подкорпуса состоит из древних текстов с устаревшей лексикой, его семантическая разметка сложна. В связи с этим раскрытие устаревших значений, разработка их семантической разметки повышает ценность историко-поэтического подкорпуса.

Например, словосочетания с неясным значением *irgeden шығарып жіберу* буквально — ‘прогнать с порога: отец не дает сыну часть наследства в связи со ссорой’, *ту түбінде тұр* буквально — ‘стой у флага: стать воином, солдатом, частью армии’, *қосалықтап отау тігу* ‘возводит юрты попарно, объединяя две соседние’, *қазақ шығу, қазақ шығып кету* ‘становиться казахом’, *ұран отын жағу* ‘дать знак о том, что приближается вражеское войско’, *қала шығу* ‘поехать на базар за покупками’ не только обозначают определенные понятия, но и ярко демонстрирует словесное искусство, талант к образной речи и точное описательское мастерство казахского народа (Сыздықова, 2009: 95).

Поэтому составителями историко-поэтического подкорпуса казахского языка было принято решение ввести семантическую разметку для устаревших слов. Это позволит не только сохранить их значение для современных исследователей и читателей, но и прояснить контексты их использования в текстах, что способствует более глубокому пониманию культурных и исторических особенностей устной литературы казахского народа.

Предлагаем образец семантической разметки устаревших слов в тексте. Разметка включает следующие параметры: *слово*; *часть речи*; *тематическая группа*; *контекстуальное значение*. Этот подход обеспечивает структурированное описание лексических единиц, позволяя точно отразить их место и роль в устной литературе казахского народа. В будущем планируем расширить семантическую разметку, вносимую в корпус.

Семантическая разметка присваивается лексическим единицам из образцов устного народного творчества. К примеру, слово *аға*<sup>1</sup> имеет несколько значений в современном казахском языке 1. Старший из мужчин-родственников; 2. Обращение к любому старшему мужчине; 3. Путеводитель, предводитель, лидер; 4. Человек на высокой должности, имеющий власть; 5. Глава предприятия, общества, собрания. В историко-поэтических произведениях<sup>2</sup> это слово употребляется в значениях «военачальник» (*қол басы*), «предводитель ста воинов» (*жүз басы*), «предводитель тысячи воинов» (*мың басы*), «глава тумена — десяти тысяч воинов» (*түмен басы*), «герой, воин» (*ер*) согласно контексту.

Семантическая разметка слова *аға* из указанных стихотворений продемонстрирована в конкордансе следующим образом (рис. 5).

В следующих образцах устного народного творчества слово *ауыр* ‘тяжелый’ используется в значении «многочисленный, бесчисленный народ» (рис. 6).

Слово: АҒА  
 Часть речи: существительное.  
 Тематическая группа: военная должность  
 Семантика в контексте: қол басы (военачальник), жүз басы (глава ста воинов), мың басы (глава тысячи воинов), түмен басы (глава тумена), әскери башсы, қолбасы, ер (герой, воин).  
 Сочетаемость:  
 Есік аға (ишик аға) – в средние века – главный страж порога ханского дворца.  
 Қол ағалық – военачальник.  
 Ел ағасы – глава какого-либо сообщества.  
 От ағасы, ауыл ағасы, жігіт ағасы – главный человек, владелец, глава аула, глава семьи.

Батыр жігіт жау бастар,  
**Аға** жігіт қолч бастар  
 (Шалкиіз жырау.  
 «Бес ғасыр жырлайды». 1. т., 56 с.)

Рисунок 5. Семантическая разметка слова «Аға»<sup>3</sup>.

Figure 5. Semantic markup of the word “Aga”.

<sup>1</sup> Қазақ әдеби тілінің сөздігі [Толковый словарь казахского литературного языка] словарь: в 15 т. / сост. Т. Жанұзақ, С. Омарбеков, Ә. Жүнісбек и др. Алматы : Арыс баспасы, 2011. Т. 1. С. 59.

<sup>2</sup> Бабалар сөзі [Слова предков]: в 100 т. / под ред. С. Қасқабасова, С. Қирабаева. Астана: Фолиант, 2006. Т. 26. С. 21; Бабалар сөзі [Слова предков]: ... 2009. Т. 53. С. 54; Бабалар сөзі [Слова предков]: .... 2009. Т. 54. С. 66; Бабалар сөзі [Слова предков]: ... 2009. Т. 55. С. 46; Бабалар сөзі [Слова предков]: ... 2010. Т. 58. С. 13.

<sup>3</sup> Бес ғасыр жырлайды [Пять веков поют]: сборник стихов : в 2 т. / сост. М. Мағауин, М. Байділдаев. Алматы: Жазушы, 1989 Т. 1. С. 56.



Сере, сере, сере қар,  
Асты кілең, үсті мұз,  
Күн-түн қатса жібімес.  
Мен көлікке қосымды артқанмын,  
Көмбідей ару жаларға  
Күректей мұзды тондырып,  
Кірмембес **ауыр** қолға бас болып,  
Күңіреніп күн түбіне жортқанмын!  
(Доспамбет жырау. «Бес ғасыр жырлайды». 1. т., 52 с.)

**Слово:** АУЫР  
**Часть речи:** прилагательное.  
**Тематическая группа:** военная  
**Семантика в контексте:** многочисленный, бесчисленный народ.  
**Сочетаемость:**  
**Ауыр әскер** – большое войско  
**Ауыр қылыш** – острый меч

Рисунок 6. Семантическая разметка слова «Ауыр»<sup>1</sup>.

Figure 6. Semantic markup of the word “Auыр”.

Асыра қыпшақ баласы,  
Жатқаннан соң байлауда..  
Арыстан туған екен деп,  
**Асыра жұрты** қуанды  
(«Қобыланды батыр» жыры, 8 с.)

**Словосочетание:** Асыра жұрт  
**Тематическая группа:** общественная лексика  
**Семантика в контексте:** бағынышты (подчиненные), қол астындағы жұрт (подданные), бас иген (повиновавшие), мойынсұнған (завоеванные).

Рисунок 7. Семантическая разметка слова «Асыра жұрт»<sup>2</sup>.

Figure 7. Semantic markup of the word “Assira zhurt”.

Казахи весьма искусны в метафоризации, сравнении, преувеличении, восхвалении либо преуменьшении какого-либо действия, в данном случае в историко-поэтических текстах встречаются фразеологизмы с неясным значением, вышедшие из употребления: «бесплодная лошадь» (*бедеу ат*), «мерин» (*ақта ат*), «военный конь» (*алаша ат*) и пр. В связи с этим разрабатывается семантическая разметка для словосочетаний в конкордансе. Например, в эпосе Кобыланды батыр сочетание *асыра жұрт* использовано в значении «подданные, народ, повиновавшие, завоеванные» (рис. 7).

Слово «народ» (*жұрт*) имело большое значение в языковом сознании казахского народа. Это общественно-политическая лексическая единица, обозначающая единение страны, ханства, воссоединение распавшегося на части народа. Слово синонимично с такими единицами, как *ел* ‘страна’, *халық* ‘народ’, также встречается в жырах и дастанах в составе словосочетаний *астана жұрт*, *аса жұрт*, *асыра жұрт*, *ауыр жұрт*, *байтақ жұрт*, *Еділ жұрт*, *алаң жұрт*. В настоящее время сочетаемость данного слова сузилась, и оно употребляется только в ограниченных контекстах.

Р. Сыздыкова, изучавшая устаревшие слова в языке поэзии, отмечает:

«Все сказанное нами в адрес тех, кто ставит перед собой цель собрать, изучить, проанализировать и продемонстрировать древние, устаревшие или начинающие выходить из употребления лексико-фразеологические единицы такого богатого, с долгой историей, гибкого и довольно чистого (сохранившего свой архетип) языка, как казахский, не подчеркивает сложность лексикологических и этимологосемасиологических исследований, а наоборот — указывает на то, насколько интересна, актуальна и значима эта тема» (Сыздыкова, 2009: 95).

<sup>1</sup> Бес ғасыр жырлайды [Пять веков поют]: сборник стихов : в 2 т. / сост. М. Мағауин, М. Байділдаев. Алматы: Жазушы, 1989 Т. 1. С. 32.

<sup>2</sup> Байганин Н. Алпамыс батыр. Қобыланды батыр [Алпамыс батыр. Кобыланды батыр] : в 3 т. Алматы: Қазығұрт, 2012. Т. 2. (На каз. яз.). С. 8.



Сбор текстов в корпус сам по себе является важным этапом для сохранения и изучения языка, однако этого недостаточно для полноценного анализа языковых данных. Важно также пополнять корпус семантической разметкой устаревших и древних слов. Семантическая разметка позволяет не только фиксировать эти слова, но и отслеживать их смысловые изменения, исследовать развитие языка и восстанавливать утраченные культурные и исторические контексты. На наш взгляд, данная разметка будет способствовать углубленному пониманию эволюции языка и его роли в сохранении культурного наследия.

Таким образом, при разработке базы историко-поэтических текстов XV–XIX вв. в рамках НККЯ были достигнуты важные научные результаты, которые не только отражают существующий опыт других стран, но и предлагают новые перспективы по изучению устного и письменного народного творчества. В отличие от ранее проведенных исследований, данное исследование объясняет причины, по которым определенные образцы устного народного творчества XV–XIX вв. не были опубликованы, что вносит важный вклад в понимание публикационной истории этих текстов. Этот аспект редко затрагивается в существующей литературе, и наши выводы предлагают новый способ анализа данного вопроса.

Во-первых, опыт других стран, включая анализ поэтических корпусов башкирского, русского и чешского языков, был полезен для определения ключевых преимуществ подходов к созданию аналогичных корпусов. В существующей литературе отмечается важность создания таких баз данных для лингвистических и литературных исследований. Наше исследование подтверждает эти выводы, но также предлагает новый взгляд на необходимость адаптации таких методик к специфике тюркской поэтической традиции, особенно в отношении арабографической графики.

Во-вторых, лингвотекстологический анализ выявил типичные ошибки и проблемы с транскрибированием арабографических текстов на кириллицу. Эти проблемы обсуждались в существующей литературе, но наше исследование предлагает более глубокую классификацию и объяснение причин возникновения ошибок, что позволяет лучше понять трудности работы с древними текстами.

В-третьих, наше исследование углубило понимание параметров сетки метатекстовой разметки, что соответствует современным тенденциям в литературоведении и корпусной лингвистике. Однако, наш подход ориентирован на специфические особенности тюркской поэтической традиции, что вносит новые идеи в существующую литературу, посвященную корпусной разметке поэтических текстов.

В-четвертых, наша работа подтверждает и развивает идеи о важности анализа поэтических размеров и структур, отражающие особенности казахского стиха. В то время как существующие исследования уже подчеркивали значение рифмы и метра, наше исследование впервые применило сравнительный анализ стихотворных структур в тюркских языках, что позволило выявить закономерности, ранее не описанные в литературе.

В пятых, предложенная нами модель семантической разметки архаических и устаревших слов вносит значительный вклад в теорию корпусной лингвистики. В существующей литературе мало уделяется внимания таким аспектам, и наши результаты открывают новые возможности для дальнейших исследований в этом направлении.

Создание интерфейса для «Историко-поэтического подкорпуса казахского языка» также важно не только для исследователей, но и для более широкой аудитории, включая студентов и преподавателей. Интерфейс будет состоять из двух частей (рис. 8) и (рис. 9).



Рисунок 8. Первое — теги, посвященные параметрам поэтических стихотворений.  
Figure 8. The first is the tags dedicated to the parameters of poetic poems.



Рисунок 9. Второе — параметры стихотворного размера.  
Figure 9. The second is the parameters of the verse size.

Тағы да келді кеңес Қараменнен, Басына бақыт-дәулет бәрін берген. Есептеп аяқты мал айтып болмас, Қара жер қайысқандай төңкерілген. Отарлап қырға көшіп малы сисмай, 50 Исімен мұрын жарып жас көкорай. Айдынды қайда болса бәрін біліп, Қанаттас қонышы еді көл жағалай. Жайласып мал өріске, ел қонысқа, Самалы салқын көлдің көңілге жай. Ордасы Қараменнің тоғыз канат, Жайылған жібек кілем сән-салтанат. Екі үйдің арасына керме керіп, Ойнатып түр-түрімен тізілген ат. Алты ауыл алты ағайын аймақтары, 60 Жылып ертеңді кеш соның бәрі. Кезекпен алты ауылға қонақтасып, Шарада сапырулы қымыздары. Базарбай бөлек ауыл кейін қонып, Дейтұғын келеді ауыл аңыздары. Бір Нияз бір енесі аты шыққан, Жауапсыз жан кіргізбес үйге сырттан. Бүркіттей бүрген адам онбайтұғын, Есіне дұшпан түгіл досы бүкқан. Айбыны Қараменнің сенетұғын, 70 Бұған сай адам жоқ деп жөнетұғын. Екі ұлы Қараменнің Егіз, Жатақ, Ынжық деп елі шулап күлетұғын. Келеді қатар өсіп Егіз, Жатақ, Егізге берген қалың малды матап. Екі ұлы біреуінен бірі өтеді, Бола ма жайылмасқа жаман атак. Жайылды момын аты ел ішіне, Туысы риза болмай жүрісіне. Қарамен құдасымен ат салысып, 80 Қуанып мақтанғандай өресіне. Егіз де көріп тәуір кісі оны, Мырза деп сыйлайтұғын елі мұны. Сен десе қымырайып кететұғын, Болғанмен аузы қисық байдың ұлы. Ағалары ызаланып ұрысса да, Ешкімді тындамайды келсе жыны.



Рисунок 10. Интерфейс оригинальных и транскрибированных текстов в историко-поэтическом подкорпусе.  
Figure 10. The interface of original and transcribed texts in the historical and poetic subcorpus.

литературе о цифровых корпусах часто подчеркивается значимость удобного интерфейса, но наше исследование предлагает новый подход к визуализации текстов на двух графиках (арабографической и кириллической) (рис. 10), что способствует углубленному анализу текстов и истории их передачи.

Если пользователю необходимо произведение «Кыз Жибек» (*Қыз Жібек*), его можно найти по тегу «тема произведения» (*шығарма тақырыбы*). Если пользователю нужно собрать стихотворения с шестью строфами, он нажимает тег «строфа» (*шумақ*). В результате корпус показывает список стихотворений, состоящих из шести строф. Таким образом, поэтический корпус может использоваться для самых разнообразных задач на стыке литературоведения и лингвистики, где каждый пользователь может получить необходимые ему материалы и информацию в интерфейсе поэтического корпуса с помощью указанных тегов.

Историко-поэтический подкорпус в рамках НККЯ предоставляет следующие преимущества.

Пользователь может ознакомиться с текстами поэтических рукописей и глубже исследовать язык того времени, проводя статистический мониторинг языковых средств. Это расширяет эмпирическую базу исследований в лингвистике и литературоведении, предлагая новые возможности для систематизации и анализа данных.

Подкорпус позволяет не только изучить стихотворную форму и историю развития поэзии, но и выявить семасиологические, прагматические и стилистические изменения лексем, а также поэтические размеры. Это помогает лучше понять изменения в тюркской поэтической традиции.

Также пользователь может находить слова в контексте, что улучшает понимание текста и освоение национальной культуры, истории и поэтического мастерства.

Следовательно, подкорпус предлагает новые перспективы для анализа и изучения поэтических текстов и является важным инструментом для дальнейших исследований.

## Заключение

Итак, мы описали проект создания историко-поэтического корпуса, который находится на стадии своего формирования. Историко-поэтический подкорпус казахского языка в рамках НККЯ — это электронный фонд, выполняющий функцию информационного инструмента и состоящий из базы историко-поэтических текстов, сохранившихся со времен древности, а также из специальной поэтической и лингвистической разметки, описывающей особенности поэтического вида устной речи.

Цель исследования, направленного на изучение опыта разработки поэтических подкорпусов других языков и внедрение эффективных практик для создания историко-поэтического подкорпуса казахского языка, была успешно достигнута. Все поставленные задачи выполнены. В частности, был проведён сравнительный анализ разработки поэтических корпусов на русском, башкирском, чешском и персидском языках, что позволило выявить и адаптировать ценные подходы для казахского подкорпуса. Описаны основные трудности, возникшие в процессе лингвотекстологической работы с оцифрованными рукописями, включая проблемы транскрибирования арабографических текстов на



кириллицу. Определены параметры метаразметки, отражающие поэтические особенности текстов, и проведён детальный анализ поэтических параметров казахского стиха. Важным результатом исследования стало создание семантической разметки и разработка интерфейса для историко-поэтического подкорпуса.

Тем не менее, для полной реализации целей проекта необходимо продолжение исследований, особенно в направлении включения в подкорпус произведений более ранних периодов казахской поэзии, таких как средневековая и древнетюркская поэзия. Это расширит текстовую базу и углубит анализ эволюции казахской поэтической традиции. Перспективы дальнейших исследований связаны с продолжением работы по включению текстов разных эпох и углублённым анализом их языковых и культурных особенностей.

В дальнейшем, при продолжении исследований по разработке Историко-поэтического подкорпуса, планируется включение произведений более ранних периодов казахской поэзии. Данный этап является необходимым для расширения и углубления текстовой базы подкорпуса, что позволит охватить все важные этапы развития казахской поэтической традиции.

Развитие подкорпуса зависит от продолжения финансирования в рамках новых исследовательских программ, когда можно расширить корпус за счет текстов более ранних эпох. Это позволит создать более полную и репрезентативную базу данных для изучения исторической эволюции казахского языка и его поэтической традиции. Среди этих произведений особое место займут тексты средневековой поэзии X–XIV вв., которые отражают культурные и литературные процессы эпохи расцвета исламского влияния в регионе. Также будут включены произведения древнетюркской поэзии (V–IX вв.), которые служат ключевым источником для понимания раннего этапа зарождения казахской поэтической традиции и ее связи с древними тюркскими племенами. Эти тексты важны не только для лингвистических и культурных исследований, но и для понимания корней казахского языка и поэзии в контексте общего развития тюркского мира (Сейтбекова и др., 2024)

Работа над обеспечением объемности, многоаспектности и гибкости Национального корпуса казахского языка будет продолжаться с опорой на мировые достижения корпусной лингвистики и на опыт создания и использования национальных корпусов (Фазылжан, 2023: 18).

#### **Вклад авторов:**

*Сейтбекова А. А.* — разработка концепции исследования, написание текста статьи;

*Фазылжан А. М.* — выполнение анализа поэтических текстов;

*Сейдамаат А. К.* — анализ поэтических корпусов на других языках;

*Абаева М. К.* — перевод статьи с казахского на русский язык, оформление статьи; корреспондирующий автор;

*Мурсал А.* — написание текста статьи.

#### **СПИСОК ЛИТЕРАТУРЫ**

- Ахметов, З. (1973) *Өлең сөздің теориясы* [Теория стихотворного слова]. Алматы : Мектеп. 212 с. -
- Базарбаева, З. М. (2008) *Казахская интонация*. Алматы : Дайк-Пресс, 281 с.
- Базарбаева, З. (2022) *Интонология* : в 5 т. Алматы : Everest. Т. 1. 440 с.
- Байтұрсынов, А. (2003) *Әдебиет танытқыш* [Литературовед]. Алматы : Атамұра. 208 с. (На каз. яз.).
- Байтұрсынұлы, А. (1991) *Ақ жол* [Светлый путь]. Алматы : Жалын. 494 с. (На каз. яз.).
- Валиханов, Ч. Ч. (1986) *О формах казахской народной поэзии*. М. : Наука. 416 с.
- Гаспаров, М. Л. (1974) *Современный русский стих. Метрика и ритмика*. М. : Наука. 487 с.
- Гаспаров, М. Л. (2002) *Очерк истории русского стиха. Метрика. Ритмика. Рифма. Строфика*. 2-е изд., доп. М. : Фортуна Лимитед. 319 с.
- Гаспаров, М. Л. (2013) *Метр и смысл. Об одном из механизмов культурной памяти*. М. : Фортуна ЭЛ. 414 с.



Гришина, Е. А., Корчагин, К. М., Плунгян, В. А., Сичинава, Д. В. (2009) Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы / отв. ред. В. А. Плунгян. СПб. : Нестор-История. 502 с. С. 71–113.

Гумилев, Л. Н. (1999) Древние тюрки / сост. А. И. Куркчи. М. : Институт ДИ-ДИК. 480 с.

Жанабаев, К. (2014) Поэтическая система произведений жырау XV–XVIII веков. К начальным основаниям художественного перевода. Алматы : Қазақ университеті. 260 с.

Жанабаев, К., Ислямова, У., Сейітбекова, А. А. (2022) Частотный поэтический словарь языка жырау XV–XVIII вв. // Tiltanym. № 86 (2). С. 28–38. DOI: <https://doi.org/10.55491/2411-6076-2022-2-26-36>

Жанабекова, А. (2013) Роль лингвистической аннотации на опыте создания национального корпуса казахского языка // Проблемы современной прикладной лингвистики: сборник статей / отв. ред. А. В. Зубов. Минск : МГЛУ. 531 с. С. 212–216.

Жаңабекова, А. (2017) Қазақ тілінің ұлттық корпусына енгізілетін метабелгіленімдер әзірлемесі туралы [О разработке метадаанных для включения в национальный корпус казахского языка] // «Ахмет Байтұрсынұлы мұрасы: зерттеу, жүйелеу және насихаттау» атты халықаралық ғылыми-теориялық конференция материалдары [Материалы международной научно-теоретической конференции «Наследие Ахмета Байтұрсыновича: исследование, систематизация и популяризация»] / отв. ред. М. Малбақов. Алматы : Елтаным. 396 с. С. 229–234. (На каз. яз.).

Жаңабекова, А., Пірманова, К. К., Карбозова, Б. Д. (2020) Разработка лексико-семантической разметки в национальном корпусе казахского языка // Тюркология. № 4. С. 201–216.

Жаңабекова, А., Кожамметова, А. К. (2021) Обработка текстов, включенных в метаразметку, на специальном компьютерном программном обеспечении // Tiltanym. № 3. С. 37–52. DOI: <http://doi.org/10.55491/2411-6076-2021-3-37-52>

Жолдасбеков, М. (1990) Асыл арналар [Драгоценные истоки]. Алматы : Жазушы. 352 с. (На каз. яз.).

Жұбанов, А. (2010) Қазақ әдеби тілінің электрондық корпусын түзудің теориялық бастаулары [Теоретические истоки построения электронного корпуса казахского литературного языка] // Тіл және мәдениет: тілдің антропоэктік парадигмасы. Профессор Ж. А. Манкееваның 60 жылдығына арналған республикалық ғылыми-теориялық конференцияның материалдары [Язык и культура: антропоцентрическая парадигма языка. Материалы республиканской научно-теоретической конференции, посвященной 60-летию профессора Ж. А. Манкеевой] / отв. ред. А. Хабиева, Г. Нұрымқызы. Алматы : А. Байтұрсынұлы атындағы Тіл білімі институты. 385 с. С. 191–197. (На каз. яз.).

Жұбанов, А. (2016) Самая ценная часть поискового аппарата в национальном корпусе казахского языка — метаразметка, которая характеризует текст в целом // Tiltanym. № 2. С. 3–9.

Жұбанов, А., Жаңабекова, А. (2017) Корпустық лингвистика [Корпусная лингвистика]. Алматы : «Қазақ тілі» баспасы. 336 с. (На каз. яз.).

Жумагулов, А. Б. (2012) Қазақ әдебиетін дәуірлеу тарихы [История периодизации казахской литературы]. Қарағанды : Академик Е. А. Букетов ат. ҚарМУ. 143 с. (На каз. яз.).

Келимбетов, Н. (1991) Ежелгі дәуір әдебиеті [Литература древности]. Алматы : Мектеп. 264 с. (На каз. яз.).

Кенжебаев, Б. (2004) Түрік қағанатынан бүгінге дейін [От турецкого каганата до наших дней]. Алматы : Ана тілі. 344 с. (На каз. яз.).

Корчагин, К. М. (2015) Поэзия XX века в поэтическом подкорпусе Национального корпуса русского языка: проблема репрезентативности // Труды Института русского языка им. В. В. Виноградова. № 3 (6). С. 235–256.

Кыраубаева, А. (1999) Ежелгі әдебиет [Древняя литература]. Алматы : Қазақ университеті. 138 с. (На каз. яз.).

Орехов, Б. В. (2015) Еще раз об исследовательском потенциале поэтического корпуса: метр, лексика, формула // Труды Института русского языка им. В. В. Виноградова. Вып. 6. С. 449–463.

Орехов, Б. В. (2019a) Башкирский стих XX века. Корпусное исследование. СПб. : Алетейя. 344 с.

Орехов, Б. В. (2019b) Метр отрезков длиннее строки в башкирском силлабическом стихе // Известия РАН. Серия Литературы и Языка. Т. 78. № 2. С. 41–50.

Орехов, Б. В., Степина, Д. С. (2022) Персидский поэтический корпус // Труды Института русского языка им. В. В. Виноградова. № 1. С. 65–72.

Өміралиев, Қ. (1976) XV–XIX ғасырлардағы қазақ поэзисының тілі [Язык казахской поэзии XV–XIX веков]. Алматы : Ғылым. 269 с. (На каз. яз.).



Өмірәлиев, Қ. (2010) Көне түркі әдеби ескерткіштері туралы зерттеулер [Исследования древнетюркских литературных памятников]. Алматы : Арыс. 650 с. (На каз. яз.).

Плунгян, В. А. (2014) Неклассический стих Лермонтова: некоторые детали // Ученые записки Петрозаводского государственного университета. Общественные и гуманитарные науки. № 7 (144). С. 40–51.

Савчук, С. О., Архангельский, Т. А., Бонч-Осмоловская, А. А., Доница, О. В., Кузнецова, Ю. Н., Ляшевская, О. Н., Орехов, Б. В., Подрядчикова, М. В. (2024) Национальный корпус русского языка 2.0: новые возможности и перспективы развития // Вопросы языкознания. № 2. С. 7-34. DOI: <http://doi.org/10.31857/0373-658X.2024.2.7-34>

Сапарниязов, Н., Хожаниязов, У. (1959) Шарьяр [Шарьяр]. Нокис : Қарақалпақ мемлекет баспасы, 107 с.

Сейтбекова, А., Елесбай, Н. (2024) Исторический поэтический подкорпус: база староказахских поэтических текстов // Tiltanym. № 3. С. 140–150. DOI: <http://doi.org/10.55491/2411-6076-2024-3-140-150>

Сичинава, Д. В. (2012) Поэтический подкорпус Национального корпуса русского языка: несколько примеров поиска стиховедческой информации // Славянский стих. Т. 9. С. 482–491.

Суворов, М. Н. (2015) Средневековая литература мусульманского мира. СПб. : Президентская библиотека. 151 с.

Суюншалиев, Х. (1983) Қазақ әдебиеті. XVII–XIX ғ.ғ. [Казахская литература. XVII–XIX вв.]. Алматы : Мектеп. 168 с. (На каз. яз.).

Сыздықова, Р. (1970) Абай өлеңдерінің синтаксистік құрылысы [Синтаксическое построение стихов Абая]. Алматы : Ғылым. 173 с.

Сыздықова, Р. (2000) Қазақ тілінің анықтағышы (емле, тыныс белгілері, сөз сазы) [Справочник казахского языка (правописание, знаки препинания, орфоэпия)]. Астана : Елорда. 480 с. На каз. яз.).

Сыздықова, Р. (2009) Қазақ тіліндегі ескіліктер мен жаңалықтар [Архаизмы и неологизмы в казахском языке]. Алматы : Арыс. 182 с. На каз. яз.).

Фазылжан, А. М. (2023) Қазақ тілінің ұлттық корпусын әзірлеу тәжірибесі [Опыт разработки Национального корпуса казахского языка]. Алматы : ЖК Асыл. 446 с. На каз. яз.).

Lord, A. B. (1991). *Epic Singers and Oral Tradition*. Ithaca and London: Cornell University Press. 280 p.

Plecháč, P., Kolár, R. (2015) The corpus of Czech verse // *Studia Metrica et Poetica*. V. 2. № 1. P. 107–118.

Дата поступления: 30.08.2024 г.

Дата принятия: 02.04.2025 г.

#### REFERENCES

- Akhmetov, Z. (1973) *Theory of the poetic word*. Almaty, Mektep. 212 p. (In Kazakh)
- Bazarbayeva, Z. M. (2008) *Kazakh intonation*. Almaty, Daik-Press. 281 p. (In Russ.)
- Bazarbayeva, Z. (2022) *Intonology*: in 5 vols. Almaty, Everest. Vol. 1. 440 p. (In Russ.)
- Baitursynov, A. (2003) *The study of literature*. Almaty, Atamura. 208 p. (In Kazakh)
- Baitursynuly, A. (1991) *Bright path*. Almaty, Zhalyyn. 494 p. (In Kazakh)
- Valikhanov, Ch. Ch. (1986) *On the forms of Kazakh folk poetry*. Moscow Nauka 416 p. (In Russ.)
- Gasparov, M. L. (1974) *Modern Russian verse Metrics and rhythmic*. Moscow, Nauka. 487 p. (In Russ.)
- Gasparov, M. L. (2002) *Essay on the history of Russian verse Metrics Rhythm Rhyme Stanza*. 2nd ed. ext. Moscow, Fortuna Limited. 319 p. (In Russ.)
- Gasparov, M. L. (2013) *Meter and meaning On one of the mechanisms of cultural memory*. Moscow, Fortuna EL. 414 p. (In Russ.)
- Grishina, E. A., Korchagin, K. M., Plungian, V. A. and Sichinava, D. V. (2009) The poetic subcorpus within the Russian National Corpus: general structure and application prospects. *Natsional'nyi korpus russkogo yazyka 2006–2008*. In: *Novye rezul'taty i perspektivy*, ed. by V. A. Plungian. St. Petersburg, Nestor-Istoriya. 502 p. Pp. 71–113. (In Russ.)
- Gumilev, L. N. (1999) *Ancient Turks*, comp. by A. I. Kurkchi Moscow, Institute DI-DIK. 480 p. (In Russ.)
- Zhanabayev, K. (2014) *The poetic system of the zhyrau works of the 15th–18th centuries Towards the foundations of literary translation*. Almaty, Kazakh University. 260 p. (In Russ.)
- Zhanabayev, K., Islyamova, U. and Seitbekova, A. A. (2022) Frequency poetic dictionary of the zhyrau language of the 15th–18th centuries. *Tiltanym*, no. 86(2), pp. 28–38. (In Russ.) DOI: <https://doi.org/10.55491/2411-6076-2022-2-26-36>



Zhanabekova, A. (2013) The role of linguistic annotation based on the experience of creating the National Corpus of the Kazakh Language. In: Zubov, A. V. (ed.) *Problems of modern applied linguistics*. Minsk, MSLU. 531 p. Pp. 212–216. (In Russ.)

Zhanabekova, A. (2017) On the development of metadata for inclusion in the National Corpus of the Kazakh language. In: Malbakov, M. (ed.) *Materials of the international scientific-theoretical conference "The legacy of Akhmet Baitursynuly: research, systematization and popularization"*. Almaty, Eltany. 396 p. Pp. 229–234. (In Kazakh)

Zhanabekova, A., Pirmanova, K. K. and Karbozova, B. D. (2020) Development of lexico-semantic annotation in the National Corpus of the Kazakh language. *Tyurkologiya*, no. 4, pp. 201–216. (In Russ.)

Zhanabekova, A. and Kozhakhmetova, A.K. (2021) Processing of texts included in the metatagging using special computer software. *Tiltany*, no. 3, pp. 37–52. (In Russ.) DOI: <http://doi.org/10.55491/2411-6076-2021-3-37-52>

Zholdasbekov, M. (1990) *Precious sources*. Almaty, Zhazushy. 352 p. (In Kazakh)

Zhubanov, A. (2010) Theoretical foundations of building the electronic corpus of the Kazakh literary language. In: Khabieva, A. and Nuryumkyzy, G. (eds.) *Language and culture: the anthropocentric paradigm of language*. Materials of the republican scientific-theoretical conference dedicated to the 60th anniversary of Professor Zh. A. Mankeyeva. Almaty, Institute of Linguistics named after A. Baitursynuly. 385 p. Pp. 191–197 (In Kazakh)

Zhubanov, A. (2016) The most valuable part of the search engine in the National Corpus of the Kazakh language is the metatagging characterizing the text as a whole. *Tiltany*, no. 2, pp. 3–9 (In Russ.)

Zhubanov, A. and Zhanabekova, A. (2017) *Corpus linguistics*. Almaty, Kazakh language publishing house 336 p. (In Kazakh)

Zhumagulov, A. B. (2012) *History of the periodization of Kazakh literature*. Karaganda, Buketov KarSU. 143 p. (In Kazakh)

Kelimbetov, N. (1991) *Ancient literature*. Almaty, Mektep. 264 p. (In Kazakh)

Kenzhebayev, B. (2004) *From the Turkic Khaganate to the present*. Almaty, Ana tili. 344 p. (In Kazakh)

Korchagin, K. M. (2015) 20th century poetry in the poetic subcorpus of the Russian National Corpus: the problem of representativeness. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova*, no. 3(6), pp. 235–256 (In Russ.)

Kyraubayeva, A. (1999) *Ancient literature*. Almaty, Kazakh University. 138 p. (In Kazakh)

Orekhov, B. V. (2015) Once again about the research potential of the poetic corpus: meter, lexis, formula. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova*, no. 6, pp. 449–463 (In Russ.)

Orekhov, B. V. (2019a) *Bashkir verse of the 20th century. A corpus-based study*. St Petersburg, Aleteya. 344 p. (In Russ.)

Orekhov, B. V. (2019b) Meter of segments longer than a line in Bashkir syllabic verse. *Izvestiya RAN. Seriya literaturny i yazyka*, vol. 78, no. 2, pp. 41–50 (In Russ.)

Orekhov, B. V. and Stepina, D. S. (2022) The Persian poetic corpus. *Trudy Instituta russkogo yazyka im. V.V. Vinogradova*, no. 1, pp. 65–72 (In Russ.)

Ömiraliev, Q. (1976) *Language of Kazakh poetry of the 15th–19th centuries*. Almaty, Gylym. 269 p. (In Kazakh)

Ömiraliev, Q. (2010) *Studies on Old Turkic literary monuments*. Almaty, Arys. 650 p. (In Kazakh)

Plungyan, V.A. (2014) Non-classical verse of Lermontov: some details. *Uchyonye zapiski Petrozavodskogo gosudarstvennogo universiteta Seriya obshchestvennye i gumanitarnye nauki*, no. 7 (144), pp. 40–51 (In Russ.)

Savchuk, S. O., Arkhangelsky, T. A., Bonch-Osmolovskaya, A. A., Donina, O. V., Kuznetsova, Y. N., Lyashevskaya, O. N., Orekhov, B. V. and Podryadchikova, M. V. (2024) The Russian National Corpus 2.0: new possibilities and development prospects. *Voprosy yazykoznaniiya*, no. 2, pp. 7–34. (In Russ.) DOI: <http://doi.org/10.31857/0373-658X.2024.2.7-34>

Saparnyyazov, N. and Khozhaniyazov, U. (1959) *Sharyar*. Nukus, Karakalpak State Publishing House. 107 p. (In Russ.)

Seitbekova, A. and Elesbay, N. (2024) Historical poetic subcorpus: a database of Old Kazakh poetic texts. *Tiltany*, no. 3, pp. 140–150. (In Russ.) DOI: <http://doi.org/10.55491/2411-6076-2024-3-140-150>

Sichinava, D.V. (2012) Poetic subcorpus of the Russian National Corpus: some examples of searching for verse metrics. *Slavyanskiy stikh*, vol. 9, pp. 482–491 (In Russ.)

Suvorov, M. N. (2015) *Medieval literature of the Muslim world*. St Petersburg, Presidential Library. 151 p. (In Russ.)

Suyunshaliev, Kh. (1983) *Kazakh literature. 17th–19th centuries*. Almaty, Mektep. 168 p. (In Kazakh)

Syzdykova, R. (1970) *Syntactic structure of Abai's poems*. Almaty, Gylym. 173 p. (In Kazakh)

Syzdykova, R. (2000) *Kazakh language reference book (spelling, punctuation, orthoepy)*. Astana, Elorda. 480 p. (In Kazakh)



- Syzdykova, R. (2009) *Archaisms and neologisms in the Kazakh language*. Almaty, Arys. 182 p. (In Kazakh)
- Fazylzhan, A. M. (2023) *Experience of developing the National Corpus of the Kazakh language*. Almaty, ZhK Asyl. 446 p. (In Kazakh)
- Lord, A. B. (1991) *Epic Singers and Oral Tradition*. Ithaca, London, Cornell University Press. 280 p.
- Plecháč, P. and Kolár, R. (2015) The corpus of Czech verse. *Studia Metrica et Poetica*, vol. 2, no. 1, pp. 107–118.

*Submission date: 30.08.2024.*

*Acceptance date: 02.04.2025.*