Научная статья

УДК 81:398
DOI 10.17223/18137083/97/2

Событийная модель адыгского героического эпоса: количественное исследование

Оксана Владимировна Гончарова

Российский университет дружбы народов имени Патриса Лумумбы
Москва, Россия

goncharova-oxvl@rudn.ru, https://orcid.org/0000-0003-1044-6244

Аннотация

Статья посвящена применению методов компьютерной фольклористики к корпусу адыгского нартского эпоса в русском переводе. Цель исследования – показать, как комплекс количественных процедур позволяет формализовать структуру действия и распределение социальных ролей в эпосе. Материалом служит издание «Нарты. Адыгский героический эпос» (СПб.: Вита Нова, 2017). На языке Python выполняются лексико-статистический анализ, тематическое моделирование LDA (Latent Dirichlet Allocation), построение сетей совстречаемости персонажей, событийная разметка и автоматическое выделение мотивов. Результаты показывают преобладание пространственной, военной и хозяйственной лексики, глаголов движения и боя. Сетевой анализ показывает центральное положение Сосруко и Сатаней (среди женских персонажей) при общем дисбалансе в пользу мужских фигур. Распределение типов событий демонстрирует преобладание перемещения, военных и коммуникативных действий. Предлагаемая событийная модель рассматривается как прототип схемы описания фольклорных текстов, ориентированной на интеграцию в более широкий круг гуманитарных исследований.

Ключевые слова

компьютерная фольклористика, нартский эпос, адыгский фольклор, событийная модель, корпусный анализ, гендерные роли, автоматическое выделение мотивов, обработка естественного языка

Благодарности

Исследование выполнено в рамках проекта № 050738-0-000 системы грантовой поддержки научных проектов РУДН

Для цитирования

Гончарова О. В. Событийная модель адыгского героического эпоса: количественное исследование // Сибирский филологический журнал. 2026. № 4. С. 24–42. DOI 10.17223/18137083/97/2

© Гончарова О. В., 2026

Event-based model of the Adyghe heroic epic: a quantitative study

Oxana V. Goncharova

Peoples' Friendship University of Russia named after Patrice Lumumba
Moscow, Russian Federation

goncharova-oxvl@rudn.ru, https://orcid.org/0000-0003-1044-6244

Abstract

This article applies the methods of computational folkloristics to a corpus of the Adyghe Nart epic in Russian translation. The study aims to demonstrate how quantitative procedures can formalize narrative action and the distribution of social roles. The research material is based on the 2017 edition of *Narty. Adygskii geroicheskii epos*, comprising approximately 100,000 word tokens. Automatically segmented into episodes, the text is processed in Python using lexical statistics, LDA topic modeling, character co-occurrence networks, event annotation, and dictionary-based motif detection. Each action yields an event record detailing the subject, object, event type, and local context. The results demonstrate a predominance of spatial, military, and economic terms, together with verbs denoting motion and combat. Network analysis confirms Sosruko as the central character and Sataney as the leading female figure within a heavily male-dominated corpus. Also identified is the quantitative predominance of movement, military, and communicative actions. Risk-related and religious motifs are found to dominate, with family and gender themes remaining relatively underrepresented. A specific analysis of animal vocabulary reveals the functional superiority of the "horse" as a status marker and active narrative participant. The proposed reproducible event model facilitates folklore text description and allows for category refinement as new data is accumulated. Despite limitations arising from translated texts and automated annotation, this model serves as a prototype designed to enable cross-corpus comparisons and integration into broader humanities and social science research.

Keywords

computational folkloristics, Nart epic, Adyghe folklore, event model, corpus-based analysis, gender roles, automatic motif detection, natural language processing

Acknowledgments

This research was conducted as part of project no. 050738-0-000 under the RUDN University grant system for the support of scientific projects

For citation

Goncharova O. V. Sobytiynaya model' adygskogo geroicheskogo eposa: kolichestvennoye issledovaniye [Event-based model of the Adyghe heroic epic: a quantitative study]. *Sibirskii Filologicheskii Zhurnal* [*Siberian Journal of Philology*]. 2026, no. 4, pp. 24–42. (In Russian) DOI 10.17223/18137083/97/2

Введение

Возникновение компьютерной фольклористики (Computational Folkloristics) представляет собой закономерную эволюцию систематизации и количественного анализа устных повествований [Lauer, 2023, p. 1–4]. В начале XX в. Й. Г. Хан, А. Аарне и С. Томпсон разработали приемы анализа, которые стали одной из ранних предпосылок последующего становления алгоритмических подходов. Идея авторов заключалась в декомпозиции сложных сюжетов на элементарные, повторяющиеся компоненты – мотивы и типы сюжетов – для последующего количественного сравнения, подсчета и классификации повествований по географическо-

му и культурному принципам. Предполагалось, что более частотные варианты сюжетов являются либо более древними, либо лучше сохранившимися [Lauer, 2023, p. 1–4]. Существенный вклад в развитие области внесла работа В. Я. Проппа «Морфология сказки» (1968), где была предложена система функциональных ролей, описывающих действия персонажей независимо от их конкретных воплощений. Ключевой идеей данного подхода стало положение о том, что для анализа повествования первостепенное значение имеют не конкретные персонажи, а совершаемые ими действия [Pannach et al., 2021, p. 223].

Данные принципы создали теоретическую основу для современных вычислительных методов, но, несмотря на успехи в данной области [Nguyen et al., 2012; Pompeu, 2019, p. 16–17; 30–33] (табл. 1), задачи машинного обучения и NLP (Natural Language Processing) в фольклористике сопряжены со многими трудностями. В частности, многие повествования сложно отнести к одному жанру, например, сказки без магических элементов часто ошибочно классифицируются как легенды [Nguyen et al., 2012].

Таблица 1

Методы машинного обучения, корпуса и задачи
в исследованиях компьютерной фольклористики

Table 1

Machine learning methods, corpora, and tasks
in computational folkloristics research

Метод / Модель	Корпус	Задача
SVM (Linear Kernel)	14 963 голландских народных сказок	Жанровая классификация
Random Forests	~ 35 000 датских народных сказок	Тематическое моделирование
Hierarchical Attention Network (HAN)	Multilingual Folk Tale Database (MFTD)	Классификация категорий ATU (Aarne – Thompson – Uther)
Random Forests	Корпус португальской литературы (BDCamões)	Классификация литературных жанров
Naïve Bayes	~ 35 000 датских народных сказок	Переклассификация по темам

Для преодоления подобных ограничений перспективным инструментом оказывается тематическое моделирование, в частности метод латентного размещения Дирихле (Latent Dirichlet Allocation, LDA), позволяющий выявлять смысловые связи, не всегда очевидные при традиционном анализе. Так, в исследовании датских народных сказок было зафиксировано, что после нескольких итераций LDA-модель сформировала категоризацию, существенно расходящуюся с экспертной. Например, история о встрече человека с загадочным стариком, которую эксперты относили к рубрикам «подземные жители» или «ведьмы», была отнесена моделью к темам «Наши предки: способы мышления и духовная жизнь» и «Вступление в союз с дьяволом» [Broadwell et al., 2017, p. 15–18].

Помимо решения задач классификации методы NLP и машинного обучения открывают возможности для анализа структуры и содержания фольклорных текстов, в частности, для автоматического выделения мотивов и сюжетных линий

[Lauer, 2023, p. 1–4], используя семантический анализ ролей персонажей, распознавание их взаимодействий и последующее восстановление последовательностей событий, соответствующих функциональным ролям, описанным В. Я. Проппом [Toro Isaza et al., 2023, p. 6509–6531; Lauer, 2023, p. 1–4]. Еще одно направление – исследование гендерных стереотипов. Так, анализ сказок из Project Gutenberg выявил, что мужские персонажи значительно чаще вовлечены в события, связанные с битвами, властными полномочиями, предпринимательством и получением знаний, в то время как женские персонажи преимущественно связаны с домашними обязанностями и эмоциональной сферой [Toro Isaza et al., 2023, p. 6509–6531].

Машинное обучение и NLP предоставляют инструменты для изучения эволюции и распространения фольклора, заимствуя методы из биологии для реконструкции дерева родства схожих сюжетов [Tehrani, 2023; Lauer, 2023, p. 1–4]. Данный подход позволяет делать выводы о том, какие версии сюжета являются более древними, как они изменялись под влиянием других культур и какие факторы способствовали их стабильности или трансформации. Также NLP помогает анализировать процесс трансформации, который описывается тремя факторами: преемственностью (continuity), вариацией (variation) и выбором (selection) [Bleier, Winslow, 2019, p. 61]. Например, проект «The Rake Cycle» прослеживает эволюцию английской народной песни «Unfortunate Rake» от древнейшей формы до адаптации в американский блюз «St. James Infirmary Blues», показывая, как один и тот же сюжет (смерть молодого человека из-за болезни, похороны, назидательный финал) сохранялся на протяжении двух столетий, в то время как конкретные детали (профессия героя, место действия, причина болезни) менялись в соответствии с местными контекстами [Ibid., p. 58–63].

Выбор подходящего набора инструментов является ключевым для успешного применения машинного обучения в фольклористике. Широкое распространение сохраняют такие методы машинного обучения, как метод опорных векторов, метод случайного леса, наивный байесовский классификатор и логистическая регрессия [Li, 2025]. В качестве признаков для этих моделей обычно используются формулы TF-IDF, частотность *n*-грамм (последовательностей из *n* подряд идущих слов в тексте: униграмма ($n = 1$) – отдельное слово, биграмма ($n = 2$) – пара соседних слов и т. д. [van Gompel, van den Bosch, 2016, p. 2–3]) и различные стилистические характеристики (например, длина предложений или частота употребления знаков препинания [Li, 2025; Lô et al., 2021]).

В свою очередь, современные архитектуры глубокого обучения позволяют работать с более сложными лингвистическими конструкциями. Модели, использующие только кодирующий блок (encoder-only), например BERT и RoBERTa, хорошо подходят для задач, требующих понимания контекста, таких как классификация или извлечение признаков [Debelak et al., 2025, p. 3–4, 6]. Особенно эффективной для анализа длинных текстов, какими являются полные сказки, зарекомендовала себя иерархическая сеть с механизмом внимания (Hierarchical Attention Network, HAN) [Pompeu, 2019, p. 31–32; Hagedorn, Darányi, 2022, p. 4].

Несмотря на значительный потенциал исследований в данной области [1], проблемой остается качество и доступность данных [Lauer, 2023, p. 1–4]. Многие

[1] Следует отметить также отечественные разработки в данной области, в частности исследования, ориентированные на сопоставительный анализ мотивов и сюжетов [Березкин, 2022], работы по моделированию структуры фольклорных текстов (теоретико-графовые модели) [Москин, Лебедев, 2024], а также опыт квантитативного описания сюжетно-композиционных и метрических характеристик фольклорных текстов [Козьмин, 2005; 2009].

оцифрованные коллекции содержат неточные транскрипции, лишены полных метаданных, используют противоречивые системы классификации и т. д. Серьезную трудность представляет также и то, что существующие корпусы и, как следствие, обученные на них модели в значительной степени отражают западные культурные традиции [Toro Isaza et al., 2023; Lauer, 2023, p. 1–4]. Для работы с материалом иных культур необходимо учитывать контекст, избегая методов, которые абстрагируют текст от породившей его среды [Haring, 1994, p. 3–5]. Кроме того, необходимо проверять источники на предмет адаптаций для западной аудитории, поскольку такие версии зачастую содержат упрощенные или искаженные трактовки оригинала [Tsutsumi, Jinnai 2025].

Материалы и методы

В качестве материала исследования использован текст адыгского нартского эпоса из русскоязычного издания «Нарты. Адыгский героический эпос» (СПб.: Вита-Нова, 2017) [2]. Общий объем корпуса составляет около 100 тыс. словоформ. Текст был взят в оригинальном виде, представленном в издании, с сохранением авторской пунктуации, разметки прямой речи и системы заголовков.

Для автоматизированной обработки текста применялись инструменты NLP и статистического анализа на языке Python. Исходный текст в формате .docx был загружен, разбит на абзацы и сегментирован на фрагменты, далее именуемые эпизодами. Эпизод определялся как блок текста между заголовками и абзацами, оформленными как заголовки (например, «Сказание о нарте Сосруко», «Рождение Сосруко»). Данные эпизоды использовались в качестве минимальных единиц анализа.

Был сформирован список именованных персонажей и ключевых сущностей. На основе текста составлен словарь CHAR_INFO, в котором для каждого имени фиксировались тип сущности (человек, божество, коллектив, институт, географический объект) и гендерная принадлежность (мужской, женский, коллективный, неизвестный). Для каждого эпизода методом строкового поиска фиксировался набор присутствующих имен. На основе этих данных были построены матрица совстречаемости и сеть совстречаемости персонажей, где вершины соответствовали персонажам, а ребра – совместному появлению имен в пределах одного предложения.

Текст был обработан морфологическим анализатором и синтаксическим парсером зависимостей для русского языка. В каждом предложении глаголы-предикаты рассматривались как кандидаты на события. Для каждого глагола фиксировались лемма, грамматические характеристики и контекст предложения. На основе синтаксических зависимостей определялся субъект действия; если субъект совпадал с одним из имен из словаря персонажей, событие атрибутиро-

---

Важным направлением является онтологическое описание и формализация жанровых классификаций для задач корпусного представления и поиска фольклорных данных [Гриневич, 2021].

[2] В качестве исходного текста использовано издание «Нарты. Адыгский героический эпос» (СПб.: Вита Нова, 2017), поскольку издание доступно автору в качественной цифровой копии, пригодной для автоматизированной обработки. Использование более ранних изданий, представленных преимущественно в виде сканов, требует масштабной ручной корректуры, что увеличивает долю технических ошибок распознавания и нормализации текста и, как следствие, снижает воспроизводимость количественных результатов.

валось конкретному герою. При наличии прямого дополнения в виде существительного фиксировалась лемма объекта. В результате был сформирован корпус событий со следующими полями: «эпизод», «предложение», «герой-субъект», «тип сущности», «гендер», «глагол», «объект», «контекст». Каждое событие было отнесено к одному из содержательных типов на основе словаря EVENT_TYPE_MAP. Леммы глаголов были распределены по нескольким укрупненным категориям: военные действия, хозяйственно-экономические, социально-политические, религиозно-ритуальные, семейно-воспитательные, перемещения и коммуникативные акты.

Для выявления сфер жизни, представленных в эпосе, применялась комбинация частотного анализа и тематического моделирования. На уровне всего корпуса подсчитывалась частотность существительных и глаголов с предварительным удалением стоп-слов. Использовался стандартный список стоп-слов для русского языка, дополненный набором высокочастотной, но содержательно слабой лексики [3]. Из анализа также исключались типичные дискурсивные маркеры и лексемы с низкой дифференцирующей способностью. Тематическое моделирование методом LDA проводилось на уровне эпизодов. Перед построением модели текст каждого эпизода был лемматизирован; в анализ включались только знаменательные части речи (существительные, глаголы, прилагательные). Матрица «эпизод × лемма» формировалась с помощью инструмента CountVectorizer с ограничениями по минимальной и максимальной частоте для отсева слишком редких и избыточно частотных лексем. На основе полученной модели были интерпретированы тематические блоки (например, военные действия, семейно-родовые отношения).

Для связи количественных данных с традиционными представлениями о мотивах реализован экспериментальный метод их автоматического выделения. Для ряда ключевых мотивов («чудесное рождение Сосруко» и др.) были заданы словарные описания, включающие характерные лексемы, пороговое значение совпадений и список обязательных персонажей. Эпизод считался релевантным мотиву при выполнении этих условий. Для анализа репрезентации фауны изучалась совстречаемость зоонимов в пределах одного предложения. На основе парной совстречаемости была построена взвешенная сеть «животное – животное».

В целом в настоящем исследовании были сознательно комбинированы уже описанные в литературе методы – частотный анализ лексики, тематическое моделирование LDA на уровне эпизодов, построение сетей совстречаемости персонажей и других сущностей, а также подсчет распределения типов событий между героями – с рядом специально разработанных процедур. К последним относятся корпусная сегментация текста на эпизоды по структуре заголовков, правилоориентированное выделение акторов с присвоением им типа сущности и гендера, построение словарей событийных и мотивных категорий (военные, хозяйственные, религиозные и др.), а также экспериментальная схема автоматического отнесения эпизодов к мотивам на основе сочетания лексических индикаторов и набора задействованных персонажей. Все количественные процедуры были объединены общей моделью «событийной записи», в которой для каждого случая действия фиксируются субъект, объект, тип события и контекст, что позволяет сопоставлять результаты различных уровней анализа. Необходимо также подчеркнуть, что целью работы является не интерпретация смысла эпоса, а демонстрация того, каким образом подобный комплекс методов позволяет описывать крупные массивы

[3] Полный список исключений см.: https://github.com/brainteaser-ov/epos.

фольклорных текстов в единой системе интерпретируемых категорий, пригодных для дальнейшего сравнения. Описанный метод реализован в виде анализа на языке Python [4], обеспечивающего воспроизводимость результатов.

Результаты и обсуждение

Лексико-статистический анализ показывает, что в корпусе доминируют существительные, связанные с пространством («земля», «дом»), участниками военных и хозяйственных действий («нарты», «конь», «меч», «всадник») и семейными отношениями («сын», «мать», «отец»). Среди глаголов наиболее частотны обозначения движения, боя и физического воздействия (рис. 1).

Распределение персонажей по эпизодам и совместная встречаемость показали, что с точки зрения распределения типов на индивидуальных персонажей приходится 778 упоминаний (77,6 % всех фиксаций), на богов – 80 (8,0 %). Если рассматривать только индивидуальные персонажи (люди и боги, 29 имен), суммарно они дают 858 упоминаний. Среди них 23 мужских персонажа, 5 женских и один персонаж с неустановленным полом. По числу эпизодов распределение следующее: мужские персонажи – 716 эпизодных упоминаний (83,4 %), женские – 132 (15,4 %), на персонажа с неизвестным полом приходится 10 упоминаний (1,2 %). Внутри человеческих персонажей соотношение близко: 81,7 % упоминаний приходятся на мужские и 17,0 % – на женские.

Ранжирование по числу эпизодов показывает наличие круга персонажей с очень высокой частотностью. Сосруко фигурирует в 214 эпизодах (24,9 % всех упоминаний индивидуальных персонажей), Бадыноко упоминается в 100 эпизодах (11,7 %), Сатаней – в 80 эпизодах (9,3 %). Далее следуют Шужей (51 эпизод, 5,9 %), Шауей (50; 5,8 %), Малечипх (48; 5,6 %), Тлепш (39; 4,5 %), Куйцук (36; 4,2 %), Батараз (32; 3,7 %) и Насрен (25; 2,9 %). Таким образом, Сосруко выделяется как центральный герой, на которого приходится примерно четверть всех эпизодов с участием индивидуальных персонажей, а Бадыноко и Сатаней образуют ближайшее окружение по частоте участия в повествовании (табл. 2).

В качестве отдельного блока анализа была построена сеть совместной встречаемости персонажей: вершинами сети выступали именованные сущности (герои, боги, коллективы), ребро между двумя вершинами добавлялось в том случае, если соответствующие имена встречались в одном и том же предложении (рис. 2). Вес ребра отражал число совместных появлений. Для каждой вершины были рассчитаны два показателя: степень (число различных соседей) и взвешенная степень (сумма весов всех ребер, связанных с вершиной).

Ранжирование по взвешенной степени показывает, что наиболее связанным персонажем является Сосруко (взвешенная степень 86; степень 13), на его долю приходится 19,3 % суммарной взвешенной степени сети. Далее следуют Бадыноко (47; 10,5 %), Сатаней (36; 8,1 %) и Малечипх (29; 6,5 %). Коллектив «Нарты» имеет взвешенную степень 28, что сопоставимо с Малечипх, а коллектив «Хасу» – 15. В сумме первые десять узлов по величине взвешенной степени (Сосруко,

[4] Файл с анализируемым текстом и код для выполнения доступен на https://github.com/brainteaser-ov/epos/tree/main

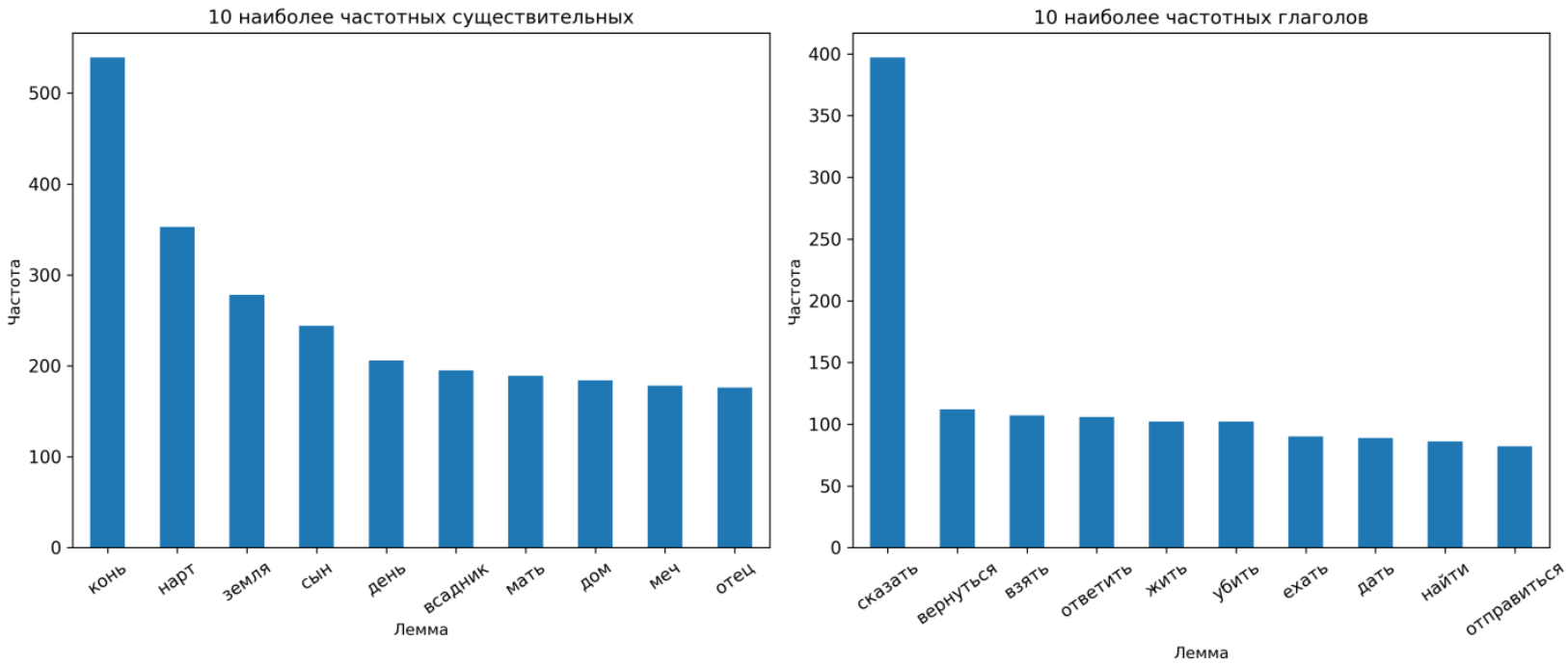


Рис. 1. Наиболее частотные существительные и глаголы в корпусе адыгского нартского эпоса
Fig. 1. Most frequent nouns and verbs in the corpus of the Adyghe Nart epic

Таблица 2

Наиболее частотные персонажи нартского эпоса:
тип сущности, пол и частота упоминаний по эпизодам

Table 2

Most frequent characters of the Nart epic:
entity type, gender, and frequency of mentions per episode

Name	Type	Gender	Frequency in episodes
Сосруко	human	male	214
Бадыноко	human	male	100
Нарты	collective	collective	95
Сатаней	human	female	80
Шужей	human	male	51
Шауей	human	male	50
Малечипх	human	male	48
Хасу	collective	collective	45
Тлепш	god	male	39
Куйцук	human	male	36
Батараз	human	male	32
Насрен	human	female	25
Унаджоко	human	male	23
Пануко	human	male	15
Химиш	human	male	15
Емынежа	human	female	13
Барымбух	human	female	13
Дабеч	human	male	11
Маруко	human	male	11
Япанес	human	male	11
Тха	god	male	10
Хижа	human	unknown	10
Тхаголедж	god	male	9
Пшая	human	male	7
Мазитха	god	male	7
Псатха	god	male	7
Канжоко	human	male	7
Амыш	god	male	6
Гогуаж	human	male	5
Хаса Нартов	collective	collective	3
Страна Нартов	place	unknown	2
Созреш	god	male	2
Нашепко	human	female	1

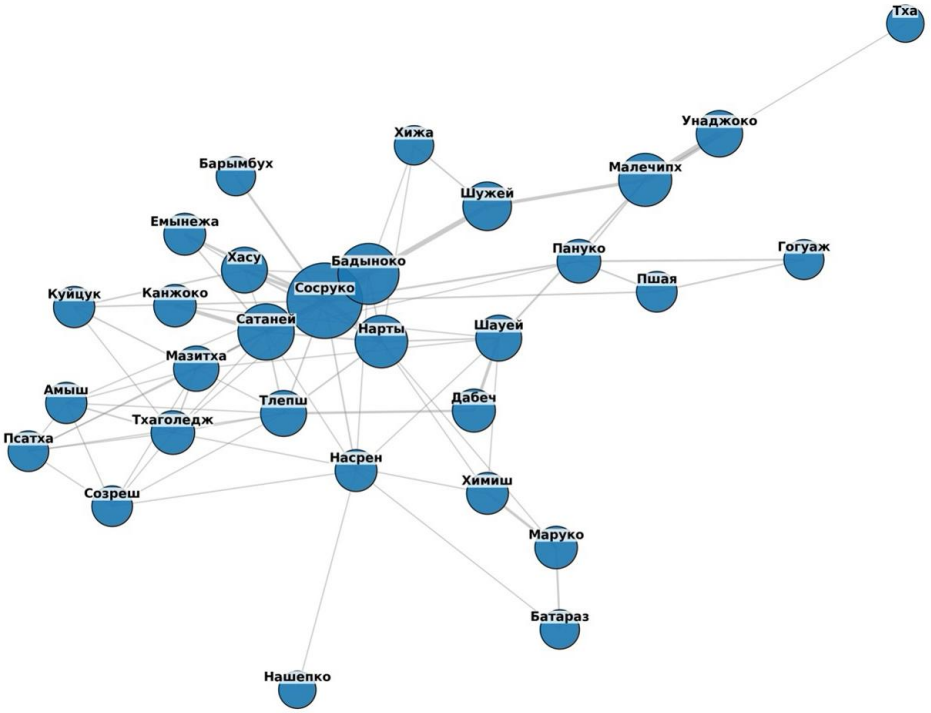


Рис. 2. Сеть совместной встречаемости персонажей адыгского нартского эпоса
Fig. 2. Co-occurrence network of characters in the Adyghe Nart epic

Бадыноко, Сатаней, Малечипх, Нарты, Шужей, Унаджоко, Хасу, Тлепш, Шауей) составляют 68,8 % общей взвешенной степени сети, т. е. основная часть совместных появлений сосредоточена вокруг ограниченного круга центральных персонажей и коллективов (табл. 3).

Если рассматривать распределение по типам сущностей, человеческие персонажи дают 76,9 % суммарной взвешенной степени (343 из 446), боги – 13,5 % (60), коллективные акторы – 9,6 % (43). Внутри группы богов наибольшие значения взвешенной степени имеют Мазитха (14), Тлепш (15) и Тхаголедж (11), занимая промежуточное положение между наиболее частотными людьми и персонажами второго плана.

Распределение по полу также характеризуется дисбалансом. На мужских персонажей приходится 76,7 % общей взвешенной степени (342), на женских – 12,8 % (57), на коллективы – 9,6 %, на персонажей с неустановленным полом – менее 1 %. Внутри женской подгруппы Сатаней занимает доминирующее положение: взвешенная степень (36) составляет 63,2 % от суммарной взвешенной степени всех женских персонажей. Остальные женские персонажи (Емынежа, Насрен, Барымбух, Нашепко) имеют существенно меньшие значения и располагаются ближе к периферии сети.

Таблица 3

Показатели степени и взвешенной степени
узлов сети совместной встречаемости персонажей

Table 3

Degree and weighted degree
measures of nodes in the character co-occurrence network

Персонаж	Тип	Пол	Степень	Взвешенная степень
Сосруко	человек	мужской	13	86
Бадыноко	человек	мужской	9	47
Сатаней	человек	женский	11	36
Малечипх	человек	мужской	5	29
Нарты	коллектив	коллектив	9	28
Шужей	человек	мужской	3	20
Унаджоко	человек	мужской	1	16
Хасу	коллектив	коллектив	6	15
Тлепш	бог	мужской	9	15
Шауей	человек	мужской	7	15

Моделирование событий показало, что глаголы можно сгруппировать по содержательным категориям. Распределение типов событий в целом по корпусу демонстрирует количественное преобладание военных и хозяйственных действий над религиозными, семейными и коммуникативными. Обобщенная статистика по корпусу (рис. 3) показывает, что наибольшую долю составляют события перемещения. События военного и коммуникативного типа занимают вторую и третью позиции и вместе с перемещением формируют основное ядро корпуса. Семейные события представлены существенно меньшим числом случаев, хозяйственные, социально-политические и религиозные типы образуют сравнительно небольшую долю распределения.

Диаграмма распределения типов событий по полу субъекта показывает, что во всех типах действий доминируют мужские персонажи. Особенно велика их доля в военных, коммуникативных и событиях перемещения. Женские персонажи заметнее всего представлены в военных, семейных и коммуникативных событиях, в хозяйственных, социально-политических и религиозных типах их участие минимально (рис. 4).

Анализ распределения мотивов по тематическим группам (рис. 5), показал, что всего было учтено 276 срабатываний мотивов, каждый из которых мог быть отнесен к одной или нескольким группам. Наибольшее число мотивов связано с риском (89 случаев, около трети всех фиксаций). На втором месте – религиозные мотивы (46 случаев, около 17 %): вмешательство богов, санкции за нарушение их воли и получение / утрата божественных даров.

Далее следуют мотивы, связанные с продвижением героя в иерархии (40 случаев; 14,5 %), хозяйством и ресурсами (38; 13,8 %), и властью (35; 12,7 %). Данные три группы охватывают сюжеты карьерного роста и признания, распределения благ (прежде всего семян проса, скота, оружия) и организации политического порядка. Значительно менее представлены мотивы, отнесенные к семье и гендер-

ным отношениям (по 13 случаев, около 4,7 % каждый). Мотивы, в которых названы обычай и запрет, фиксируются эпизодически (2 случая, менее 1 %) (табл. 4).

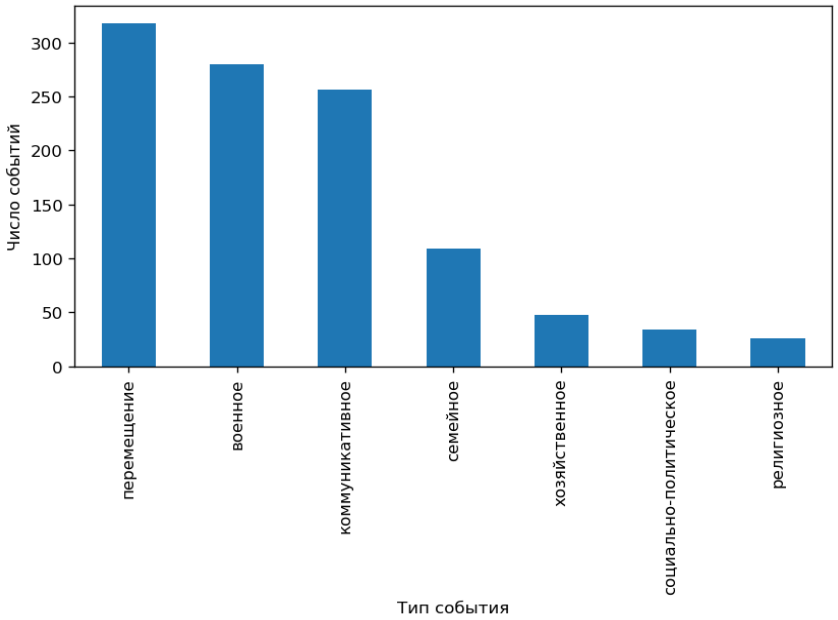


Рис. 3. Распределение типов событий в корпусе адыгского нартского эпоса
Fig. 3. Distribution of event types in the corpus of the Adyghe Nart epic

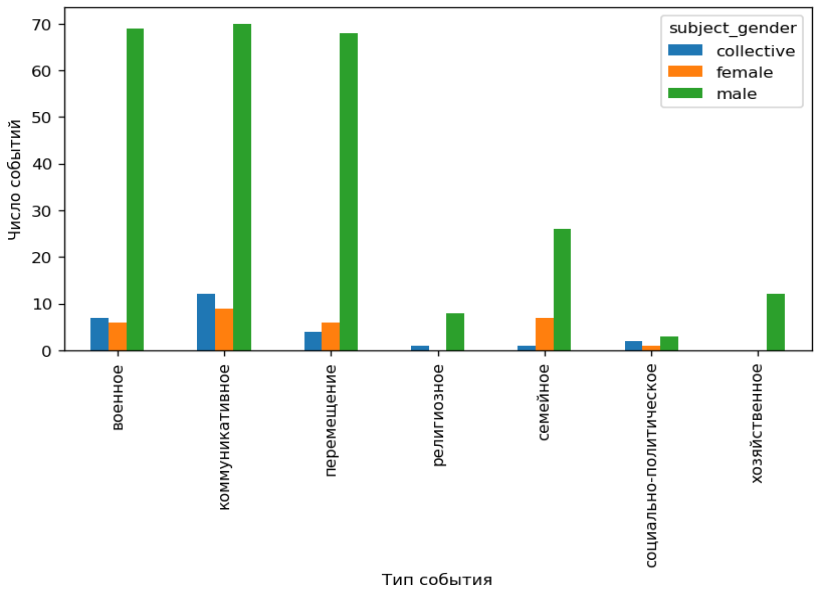


Рис. 4. Распределение типов событий по полу субъекта действия
Fig. 4. Distribution of event types by the gender of the acting subject

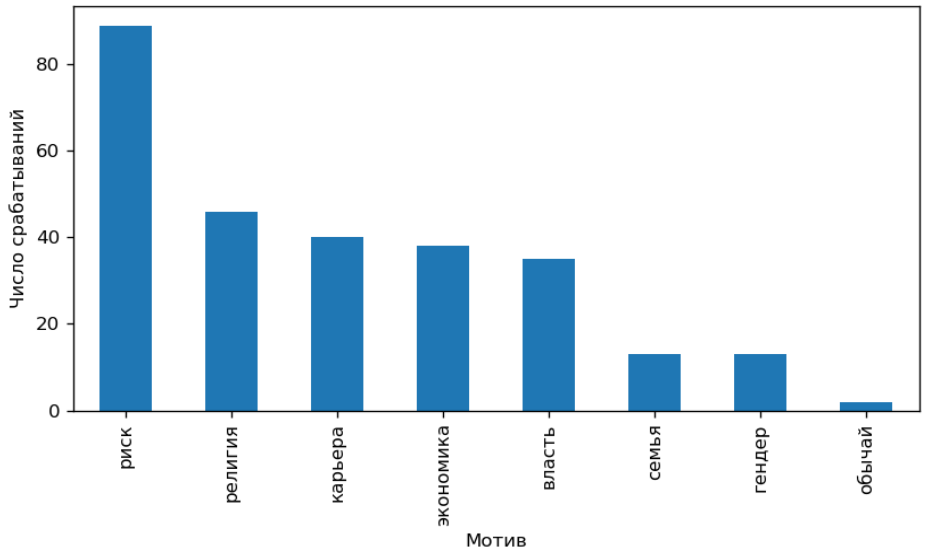


Рис. 5. Распределение мотивов по тематическим группам
Fig. 5. Distribution of motifs by thematic groups

Таблица 4

Распределение типов мотивов
в корпусе адыгского нартского эпоса

Table 4

Distribution of motif types
in the corpus of the Adyghe Nart epic

Тип мотивов	Число фиксаций	Доля от общего числа, %
Риск	89	32,2
Религия	46	16,7
Продвижение, статус (карьера)	40	14,5
Хозяйство и ресурсы	38	13,8
Власть, политический порядок	35	12,7
Семья	13	4,7
Гендерные отношения	13	4,7
Обычай, запрет	2	0,7
Всего	276	100,0

Частотное распределение лексем, связанных с животными, показывает, что в корпусе доминирует лексема «конь»: 523 упоминания (58,4 % всех фиксаций названий животных). Следующие по частоте лексемы имеют меньшие значения: «бык» (53; 5,9 %), «орёл» (48; 5,4 %), «собака» (37; 4,1 %), «птица» (29; 3,2 %), «овца» (28; 3,1 %), «коза», «баран» и «лиса» (по 23; 2,6 %). Остальные лексемы встречаются реже 2,6 % от общего числа упоминаний каждая (табл. 5).

Таблица 5

Частотность лексем, обозначающих животных,
в корпусе адыгского нартского эпоса

Table 5

Frequencies of animal-denoting lexemes
in the corpus of the Adyghe Nart epic

Животное	Число упоминаний	Доля от всех упоминаний животных, %
Конь	523	58,4
Бык	53	5,9
Орёл	48	5,4
Собака	37	4,1
Птица	29	3,2
Овца	28	3,1
Коза	23	2,6
Баран	23	2,6
Лиса	23	2,6
Козёл	22	2,5
Волк	14	1,6
Олень	13	1,5
Змея	11	1,2
Корова	9	1,0
Лошадь	8	0,9
Голубь	7	0,8
Жеребец	6	0,7
Ворон	6	0,7
Медведь	5	0,6
Ягнёнок	2	0,2
Щенок	2	0,2
Змей	1	0,1
Телёнок	1	0,1
Ящерица	1	0,1

Анализ совместной встречаемости связи персонажей с животными показал, что на графе «персонаж – животное» центральным узлом выступает конь (рис. 6), связанный с целой группой ключевых героев: Сосруко, Бадыноко, Батаразом, Шужей, Шауеем, Малечипхом, Сатаней, Канжоко и др.

Конь оказывается не атрибутом одного-двух конкретных героев, а общим маркером статуса и участия в действиях, прежде всего связанных с перемещением и войной.

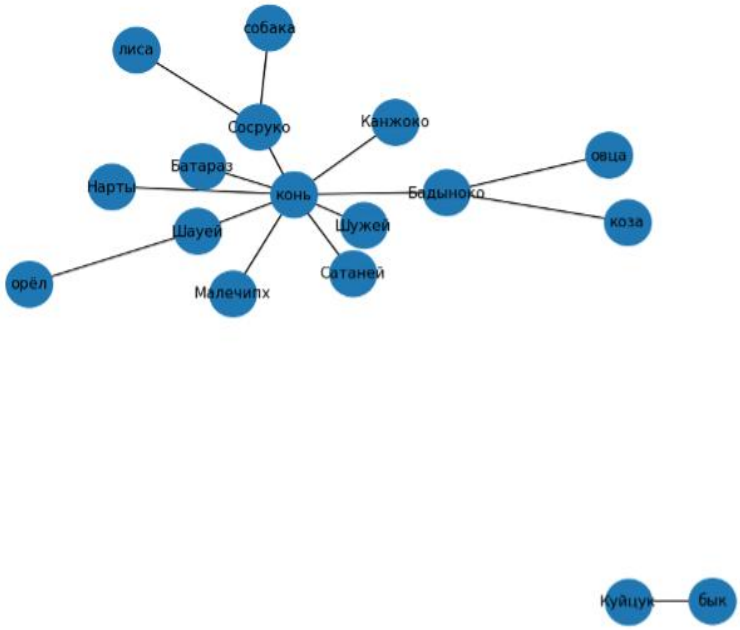


Рис. 6. Сеть связей между персонажами и лексемами, обозначающими животных
Fig. 6. Network of links between characters and animal-denoting lexemes

Заключение

Проведенный анализ показал, что методы компьютерной фольклористики могут быть практически применены к корпусу нартских текстов и дают сопоставимые между собой количественные результаты при условии явного задания единицы анализа и набора категорий. Модель «событийной записи» позволяет описывать один и тот же материал в разных проекциях: через распределение персонажей и типов действий, через тематические профили эпизодов и через схемы мотива. Это показывает, что для эпоса возможно формализованное описание структуры действия, не сводящееся к отдельным сюжетным примерам.

Результаты демонстрируют, что даже при относительно небольшом по современным меркам корпусе (около ста тысяч словоформ) статистические методы позволяют зафиксировать особенности повествовательной организации и распределения социальных ролей. При этом числовые показатели (распределение событий по типам и полу субъекта, частоты мотивных групп, параметры сетей совстречаемости) оказываются интерпретируемыми в терминах классических фольклористических и антропологических описаний. Важным итогом является не столько уточнение отдельных интерпретаций (например, роли коня или структуры мужских и женских ролей), сколько демонстрация того, что традиционные наблюдения могут быть воспроизведены и проверены средствами формализованного анализа.

Работа выявила и ряд методологических ограничений. Автоматическое извлечение событий и ролей опирается на качество морфологического и синтаксического разбора; словарно заданные категории событий и мотивов заведомо неполны. Использование русского перевода вместо оригинальных адыгских текстов задает дополнительные искажения: часть культурно значимой лексики и грамматических оппозиций нивелируется. Тем не менее даже при данных ограничениях построенная система разметки воспроизводима, а шаги анализа прозрачны и могут быть уточнены при расширении словарей или замене языковой модели.

Перспективой развития представленного подхода является перенос описанной схемы на другие варианты эпоса и иные фольклорные традиции, установление межжанровых связей внутри одной этнической фольклорной традиции, а также интеграция событийной разметки с более сложными моделями представления текста, включая современные языковые модели. Важной задачей представляется создание общедоступных, документированных корпусов с аналогичными уровнями аннотации, что позволит проводить сопоставительные исследования и проверять гипотезы о гендерных ролях, структуре мотивов, распределении хозяйственных и сакральных практик в разных традициях. В этом смысле статья предлагает не только частный анализ одного корпуса, но и прототип методики, ориентированной на масштабируемую обработку фольклорных данных и их включение в более широкие междисциплинарные исследования.

Список литературы

Березкин Ю. Е. Мотив как идеальный тип, или Трикстерские эпизоды в евразийском фольклоре // Фольклор: структура, типология, семиотика. 2022. Т. 5, № 1. С. 10–29. DOI 10.28995/2658-5294-2022-5-1-10-29

Гриневич А. А. Онтология фольклорных жанров // Языки и фольклор коренных народов Сибири. 2021. № 2. С. 255–272. DOI 10.25205/2307-1737-2021-2-255-272

Козьмин А. В. Метрический репертуар «блатной» песни 1920–1930-х гг. и проблема ее происхождения // Живая старина. 2005. № 2. С. 40–42.

Козьмин А. В. Сюжетный фонд сказок: структура и система. М.: РГГУ, 2009. 109 с.

Москин Н. Д., Лебедев А. А. Теоретико-графовый подход к анализу вариативности русских сказок // Интеллектуальные системы в производстве. 2024. Т. 22, № 4. С. 90–97. DOI 10.22213/2410-9304-2024-4-90-97

Bleier R., Winslow S. M. (eds.). Versioning Cultural Objects: Digital Approaches. Norderstedt: BoD, 2019. 195 p. (Schriften des Instituts für Dokumentologie und Editorik; Bd. 13)

Broadwell P., Mimno D., Tangherlini T. The Tell-Tale Hat: Surfacing the Uncertainty in Folklore Classification // Journal of Cultural Analytics. 2017. Vol. 2, no. 1. P. 1–27. DOI 10.22148/16.012

Debelak R., Koch T. K., Aßenmacher M., Stachl C. From Embeddings to Explainability: A Tutorial on Large-Language-Model-Based Text Analysis for Behavioral Scientists // Advances in Methods and Practices in Psychological Science. 2025. Vol. 8, no. 3. P. 1–25. DOI 10.1177/25152459251351285

Gompel M. van, Bosch A. van den. Efficient n-gram, Skipgram and Flexgram Modelling with Colibri Core // Journal of Open Research Software. 2016. Vol. 4. Art. e30. P. 1–10. DOI 10.5334/jors.105

Hagedorn J., Darányi S. Bearing a Bag-of-Tales: An Open Corpus of Annotated Folktales for Reproducible Research // Journal of Open Humanities Data. 2022. Vol. 8, no. 16. P. 1–10. DOI 10.5334/johd.78

Haring L. Introduction: The Search for Grounds in African Oral Tradition // Oral Tradition. 1994. Vol. 9, no. 1. P. 3–22.

Lauer G. Computational folktale studies. A very brief history // Fabula. 2023. Vol. 64, no. 1–2. P. 1–6. DOI 10.1515/fabula-2023-0001

Li Y. Folk Tales from Diverse Cultures: Digital Analysis of Content using Natural Language Processing // Applied Mathematics and Nonlinear Sciences. 2025. Vol. 10, no. 1. P. 1–17. DOI 10.2478/amns-2025-0529

Lô G., Boer V. de, Aart C. J. van. Exploring West African Folk Narrative Texts Using Machine Learning // Semantic Web. 2021. Vol. 12, iss. 2. P. 219–239. DOI 10.3390/info11050236

Nguyen D., Trieschnigg D., Meder T., Theune M. Automatic classification of folk narrative genres // Proceedings of KONVENS 2012. LThist 2012 workshop. Vienna, 2012. P. 378–382.

Pannach F., Sporleder C., May W., Krishnan A., Sewchurran A. Of lions and Yakshis: Ontology-based narrative structure modelling for culturally diverse folktales // Semantic Web. 2021. Vol. 12, no. 2. P. 219–239. DOI 10.3233/SW-200417

Pompeu D. Interpretable Deep Learning Methods for Classifying Folktales According to the Aarne – Thompson – Uther Scheme: Master's Thesis. Lisboa: Instituto Superior Técnico, Universidade de Lisboa, 2019. 60 p.

Tehrani J. J. The Cultural Transmission and Evolution of Folk Narratives // The Oxford Handbook of Cultural Evolution / Eds. J. J. Tehrani et al. Oxford: Oxford Uni. Press, 2023. P. 487–498. DOI 10.1093/oxfordhb/9780198869252.013.39

Toro Isaza P., Xu G., Oloko T., Hou Y., Peng N., Wang D. Are Fairy Tales Fair? Analyzing Gender Bias in Temporal Narrative Event Chains of Children's Fairy Tales // Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Vol. 1: Long Papers. Toronto, Canada: Association for Computational Linguistics, 2023. P. 6509–6531. DOI 10.18653/v1/2023.acl-long.359

Tsutsumi A., Jinnai Y. Do Large Language Models Know Folktales? A Case Study of Yokai in Japanese Folktales // Findings of the Association for Computational Linguistics: ACL 2025. Vienna, Austria: Association for Computational Linguistics, 2025. P. 16124–16146. DOI 10.18653/v1/2025.findings-acl.829

Список источников

Нарты. Адыгский героический эпос / Пер. с адыг. С. Липкина, В. Звягинцевой, М. Петровых [и др.]; ред. А. Дмитренко, Е. Петрова; худож. Х. Савкуев. СПб.: Вита Нова, 2017. 632 с.

References

Berezkin Yu. E. Motiv kak ideal'nyy tip, ili Triksterskiye epizody v yevraziyskom fol'klore [The motif as an ideal type, or Trickster episodes in Eurasian folklore]. *Folklore: Structure, Typology, Semiotics.* 2022, vol. 5, no. 1, pp. 10–29. DOI: 10.28995/2658-5294-2022-5-1-10-29

Broadwell P. M., Mimno D., Tangherlini T. R. The tell-tale hat: surfacing the uncertainty in folklore classification. *Journal of Cultural Analytics*. 2017, vol. 2, no. 1, pp. 1–27. DOI: 10.22148/16.012

Debelak R., Koch T. K., Aßenmacher M., Stachl C. From embeddings to explainability: a tutorial on large-language-model-based text analysis for behavioral scientists. *Advances in Methods and Practices in Psychological Science*. 2025, vol. 8, no. 3, pp. 1–25. DOI: 10.1177/25152459251351285

Grinevich A. A. Ontologiya fol'klornykh zhanrov [Ontology of folklore genres]. *Yazyki i Fol'klor Korennykh Narodov Sibiri* [Languages and Folklore of Indigenous Peoples of Siberia]. 2021, no. 2, pp. 255–272. DOI: 10.25205/2307-1737-2021-2-255-272

Hagedorn J., Darányi S. Bearing a bag-of-tales: an open corpus of annotated folktales for reproducible research. *Journal of Open Humanities Data*. 2022, vol. 8, no. 16, pp. 1–10. DOI: 10.5334/johd.78

Haring L. Introduction: the search for grounds in African oral tradition. *Oral Tradition*. 1994, vol. 9, no. 1, pp. 3–22.

Koz'min A. V. Metricheskiy repertuar "blatnoy" pesni 1920–1930-kh gg. i problema yeye proiskhozhdeniya [The metrical repertoire of the "blatnaya" song of the 1920s–1930s and the problem of its origin]. *Zhivaya starina*. 2005, no. 2, pp. 40–42.

Koz'min A. V. *Syuzhetnyy fond skazok: struktura i sistema* [The plot fund of folktales: structure and system]. Moscow, RGGU, 2009, 109 p.

Lauer G. Computational folktale studies. A very brief history. *Fabula*. 2023, vol. 64, no. 1–2, pp. 1–6. DOI: 10.1515/fabula-2023-0001

Li Y. Folk tales from diverse cultures: digital analysis of content using natural language processing. *Applied Mathematics and Nonlinear Sciences*. 2025, vol. 10, no. 1, pp. 1–17. DOI: 10.2478/amns-2025-0529

Lô G., de Boer V., van Aart C. J. Exploring West African folk narrative texts using machine learning. *Information*. 2020, vol. 11, no. 5, art. 236, pp. 1–23. DOI: 10.3390/info11050236

Moskin N. D., Lebedev A. A. Teoretiko-grafovyy podkhod k analizu variativnosti russkikh skazok [A graph-theoretic approach to the analysis of the variability of Russian folktales]. *Intellektualnye sistemy v proizvodstve*. 2024, vol. 22, no. 4, pp. 90–97. DOI: 10.22213/2410-9304-2024-4-90-97

Nguyen D., Trieschnigg D., Meder T., Theune M. Automatic classification of folk narrative genres. In *Proceedings of KONVENS 2012 (LThist 2012 workshop)*. Vienna, 2012, pp. 378–382.

Pannach F., Sporleder C., May W., Krishnan A., Sewchurran A. Of lions and Yakshis: ontology-based narrative structure modelling for culturally diverse folktales. *Semantic Web*. 2021, vol. 12, no. 2, pp. 219–239. DOI: 10.3233/SW-200417

Pompeu D. P. *Interpretable Deep Learning Methods for Classifying Folktales According to the Aarne – Thompson – Uther Scheme*. Master's thesis. Lisboa, Instituto Superior Técnico, Universidade de Lisboa, 2019, 60 p.

Tehrani J. J. The cultural transmission and evolution of folk narratives. In *The Oxford Handbook of Cultural Evolution*. J. J. Tehrani et al. (Eds.). Oxford, Oxford Uni. Press, 2023, pp. 487–498. DOI: 10.1093/oxfordhb/9780198869252.013.39

Toro Isaza P., Xu G., Oloko T., Hou Y., Peng N., Wang D. Are fairy tales fair? Analyzing gender bias in temporal narrative event chains of children's fairy tales. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*.

Vol. 1: Long Papers. Toronto, Canada, Association for Computational Linguistics, 2023, pp. 6509–6531. DOI: 10.18653/v1/2023.acl-long.359

Tsutsumi A., Jinnai Y. Do large language models know folktales? A case study of Yokai in Japanese folktales. In *Findings of the Association for Computational Linguistics: ACL 2025*. Vienna, Austria, Association for Computational Linguistics, 2025, pp. 16124–16146. DOI: 10.18653/v1/2025.findings-acl.829

van Gompel M., van den Bosch A. Efficient n-gram, skipgram and flexgram modelling with Colibri Core. *Journal of Open Research Software*. 2016, vol. 4, no. 1, art. e30, pp. 1–10. DOI: 10.5334/jors.105

Versioning Cultural Objects: Digital Approaches. Bleier R., Winslow S. M. (Eds.). Norderstedt, BoD, 2019, 195 p. (Schriften des Instituts für Dokumentologie und Editorik; Bd. 13).

List of sources

Narty. Adygskiy geroicheskiy epos [Narts. Adyghe heroic epic]. S. Lipkin, V. Zvyagintseva, M. Petrovykh et al. (Trans.), A. Dmitrenko, E. Petrova (Eds.), Kh. Savkuev (Illus.). St. Petersburg, Vita Nova, 2017, 632 p.

Информация об авторе

Оксана Владимировна Гончарова, кандидат филологических наук, доцент кафедры общего русского языка и методики его преподавания Российского университета дружбы народов имени Патриса Лумумбы (Москва, Россия)
Scopus Author ID 56037850600
WoS Researcher ID C-4671-2017

Information about the author

Oxana V. Goncharova, Candidate of Philology, Associate Professor, Department of the Russian Language and Methods of its Teaching, Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (Moscow, Russian Federation)
Scopus Author ID 56037850600
WoS Researcher ID C-4671-2017

Статья поступила в редакцию 30.11.2025;
одобрена после рецензирования 31.03.2026; *принята к публикации 31.03.2026*
The article was submitted on 30.11.2025;
approved after reviewing on 31.03.2026; *accepted for publication on 31.03.2026*