

УДК 81:004.8

DOI: <https://doi.org/10.5281/zenodo.20179005>

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ЛЕКСИКО-СТИЛИСТИЧЕСКИХ ОСОБЕННОСТЕЙ РУССКИХ НАРОДНЫХ СКАЗОК

Норец М. В.¹, Норец Е. М.², Норец Н. К.¹

¹Крымский федеральный университет им. В. И. Вернадского,
Симферополь, Россия;

²Высшая IT-школа Томского государственного университета,
Томск, Россия
E-mail: mnorets@yandex.ru

Исследование посвящено выявлению лексико-стилистических паттернов русских народных сказок с применением методов машинного обучения. На материале корпуса из 549 сказок проведены частотный анализ, TF-IDF, анализ тональности и стилистическая разметка. Теоретической рамкой выступила морфологическая теория В. Я. Проппа. Установлено коммуникативно-действенное лексическое ядро: доминируют глаголы речи и действия («говорить», «пойти», «взять»), антропонимы («Иван», «царь»), социальные роли («сын», «брат»). Жанровая дифференциация (волшебные, бытовые, о животных, кумулятивные) подтверждена метрикой TF-IDF. Общая тональность нейтрально-позитивна (среднее 0,026), однако сказки о животных имеют выраженный негативный уклон (-0,197), а кумулятивные – позитивный (+0,203). Количественные результаты коррелируют с пропповскими нарративными функциями: частотные маркеры прямо соответствуют функциям «героя», «дарителя», «перемещения», «трансформации». Структурное единство русского сказочного эпоса находит статистическое подтверждение. Предложенный подход демонстрирует эффективность симбиоза классической фольклористики и цифровых методов для объективного, масштабируемого анализа культурного наследия.

Ключевые слова: русские народные сказки, компьютерная лингвистика, машинное обучение, частотный анализ, анализ тональности, лексико-стилистические паттерны, жанровая дифференциация.

ВВЕДЕНИЕ

Представленное исследование посвящено анализу русских народных сказок и выявлению в них лексико-стилистических паттернов, которые формируют уникальный характер национального повествования. В эпоху глобализации и стандартизации обращение к культурному наследию приобретает особую значимость для сохранения национальной идентичности. Народные сказки представляют собой концентрированное выражение коллективного опыта, а применение современных методов исследования способствует его цифровому сохранению и новым интерпретациям фольклора.

Актуальность работы определяется применением компьютерных методов анализа, в частности, методов машинного обучения. Данный подход обеспечивает объективность исследования, возможность обработки больших данных и выявление скрытых лингвистических паттернов в фольклорных текстах.

Объектом исследования выступает корпус русских народных сказок, составленный на основе сборника А. Н. Афанасьева «Народные русские сказки» (издание 1984–1985 гг.) [10] и дополненный авторскими лингвистическими пометами.

Предмет исследования: лексико-статистические закономерности в сказках.

Целью научно-исследовательской работы является выявление и классификация стилистических паттернов, характерных для русских народных сказок.

ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА ИССЛЕДОВАНИЯ

Теоретической основой для данного исследования послужила морфологическая теория сказки В. Я. Проппа [14]. В своей фундаментальной работе «Морфология волшебной сказки» Пропп выделил инвариантную структуру сказочного нарратива, разложив его на последовательность функций действующих лиц [13]. Этот подход позволил перейти от описания сюжетного разнообразия к анализу глубинной повествовательной структуры, что является методологической основой для нашего поиска лексико-статистических инвариантов.

С развитием цифровых технологий методы компьютерной лингвистики и стилометрии стали активно применяться для анализа художественных и фольклорных текстов. В области цифровой лингвистики работы демонстрируют возможности выявления частотных паттернов, коллокаций и ключевых слов в больших массивах текстов [1, 9, 16]. Стилометрия, используя статистические методы для идентификации авторского стиля, доказала свою эффективность в задачах атрибуции текстов и анализа стилистических особенностей.

Применительно к русскому фольклору можно отметить ряд исследований, направленных на количественный анализ сказок [16]. Однако в большинстве случаев подобные исследования либо ограничиваются частотным анализом, либо используют готовые, но неспециализированные корпуса.

Новизна настоящего исследования заключается в симбиозе классической фольклористики и современных методов машинного обучения на основе авторского размеченного массива данных. Также, в данной работе используется целенаправленно созданный и жанрово размеченный корпус русских народных сказок, что позволяет проводить сравнительный анализ не на уровне всего фольклорного массива, а на уровне жанров. Кроме того, был применен комплекс машинных методов: от классического частотного анализа и улучшенного частотного анализа (TF-IDF) для выявления ключевых слов до анализа тональности, что позволяет получить многомерную характеристику текстов, при этом, корректность количественных показателей оценивалась через призму качественной теории В. Я. Проппа.

Таким образом, данное исследование предлагает интерпретируемый количественный подход к изучению фольклора.

Материалом исследования выступает корпус из 549 русских народных сказок, составленный и систематизированный авторами настоящей работы. Предполагается проведение классификации текстов по параметрам жанровой принадлежности, основным персонажам и тематике. Несмотря на ограниченный объем, выборка представляется репрезентативной, так как охватывает ключевые жанровые разновидности (волшебные, бытовые, кумулятивные сказки и сказки о животных) и воспроизводит основные сюжетные модели русского фольклорного канона.

Перед осуществлением лингвистического анализа была проведена предобработка текстов, а именно:

1. Приведение к нижнему регистру;
2. Удаление цифр и спецсимволов;
3. Токенизация;
4. Лемматизация;
5. Удаление стоп-слов.

Отбор стоп-слов был осуществлен на основе стандартного набора для русского языка из библиотеки NLTK (набор готовых инструментов, позволяющий компьютеру «понимать» и анализировать человеческий текст).

Для проведения лексического и стилистического анализов был разработан комплекс программных модулей на языке Python с использованием следующих технологий и библиотек:

1. Библиотека pandas для работы со структурированными данными, в том числе с загрузкой, фильтрацией и агрегацией данных.
2. Библиотека nltk (Natural Language Toolkit) для токенизации и использования стоп-слов.
3. Библиотека rymorphy2 для морфологического анализа и лемматизации русскоязычных текстов.
4. Библиотека scikit-learn для реализации метода TF-IDF и векторизации текстов.
5. Библиотеки matplotlib и seaborn для визуализации результатов анализа, wordcloud для генерации облаков слов.

Каждый этап анализа (частотный анализ, анализ ключевых слов, анализ тональности, стилистический анализ) сопровождался разработкой отдельного программного кода, что обеспечило воспроизводимость исследования и возможность его масштабирования. Код и корпус помеченных сказок представлены в открытом доступе на международной облачной платформе для хостинга IT-проектов и совместной разработки «Гитхаб» [8].

Проводимый частотный анализ позволяет выявить ключевые лексические доминанты русского сказочного нарратива. Результаты демонстрируют устойчивый тематический каркас, организованный вокруг нескольких смысловых кластеров.

Так, в процессе исследования можно обратить внимание на коммуникативно-действенную природу сказок. Ядро высокочастотной лексики формируют:

1. Глаголы коммуникации и действия: говорить (3610), стать (2476), пойти (1829), взять (1787), идти (1471), сказать (1411).
2. Антропонимика: Иван (2578), царь (2379), царевич (2340).
3. Социальные и семейные роли: сын (1467), брат (1111), старик (1008), царевна (1003), старуха (782), дочь (780).
4. Бытовая и природная лексика: конь (1053), домой (887), день (828), лес (776), вода (767).

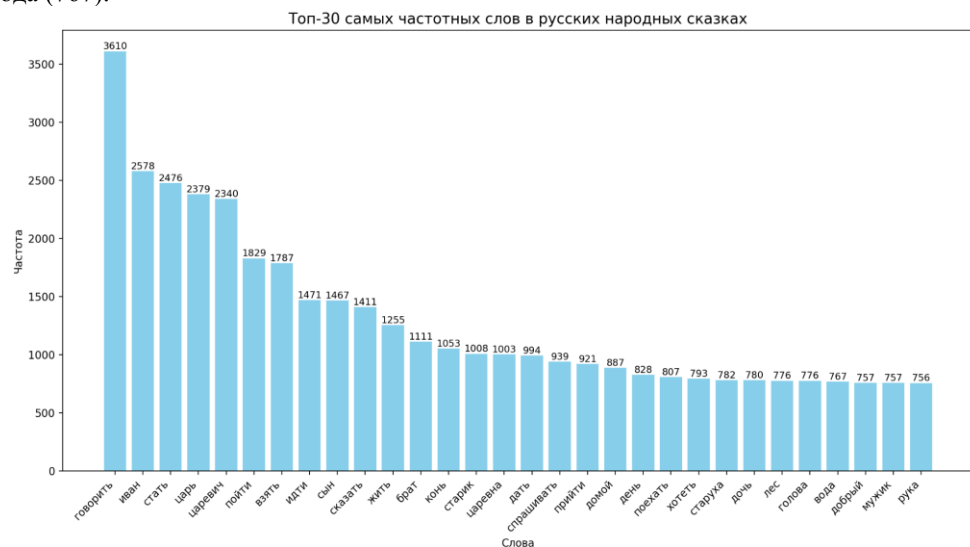


Рис. 1. Распределение наиболее частых слов в корпусе

Частотный анализ частей речи выявил абсолютное доминирование существительных (45.5%) и глагольных форм (36.9%), что составляет в сумме более 82% всего лексикона и подтверждает предметно-действенный характер сказочного повествования. Прилагательные (10.1%) и наречия (4.8%) занимают следующие за существительными и глагольными формами позиции.

Полные списки частотных слов по общему рейтингу и по жанровым категориям представлены в таблице 1 и 2.

Таблица 1. Частотность слов в русских народных сказках по рейтингу

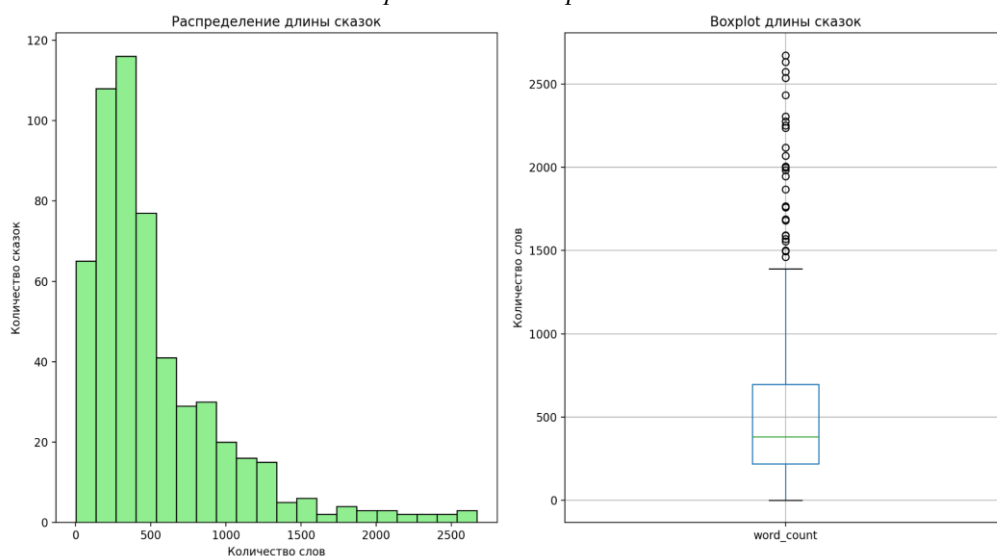
Слово	Абсолютная частота (разы)	Слово	Абсолютная частота (разы)	Слово	Абсолютная частота (разы)
говорить	3610	жить	1255	поехать	807
Иван	2578	брат	1111	хотеть	793
стать	2476	конь	1053	старуха	782
царь	2379	старик	1008	дочь	780
царевич	2340	царевна	1003	лес	776
пойти	1829	дать	994	голова	776
взять	1787	спрашивать	939	вода	767
идти	1471	прийти	921	добрый	757
сын	1467	домой	887	мужик	757
сказать	1411	день	828	рука	756

Таблица 2. Частотность слов в русских народных сказках, группированных по жанровым категориям

О животных		Волшебная		Бытовая		Кумулятивная	
волк	380	Иван	2440	говорить	835	жить	54
лиса	351	говорить	2431	стать	507	дать	51
говорить	302	Царевич	2324	пойти	457	говорить	42
медведь	246	Царь	1943	Царь	436	прийти	39
стать	214	стать	1720	взять	417	медведь	36
пойти	183	взять	1268	мужик	412	стать	35
идти	158	сын	1230	сказать	327	пойти	34
кот	151	пойти	1155	жить	301	брат	34
Уник. слов: 4214		Уник. слов: 14226		Уник. слов: 7974		Уник. слов: 1570	

Жанровая дифференциация в свою очередь позволила выяснить, что в сказках о животных чаще всего фигурируют волк, лиса и медведь; в волшебных сказках используется лексика, связанная с аристократией и властью (Иван, царевич, царь); в бытовых сказках – социально-коммуникативная направленность тематика (говорить, стать, мужик, царь, брат); а в кумулятивных – акцент на базовых действиях (жить, дать, прийти).

Рис. 2. Распределение и Boxplot длины сказок



Количественные параметры текстов показали, что средний объем произведения (с учётом предобработки по стоп-словам) составляет 530 слов, при этом диапазон сказок: от минимальных (20 слов) до развернутых повествований (2757 слов). Результаты анализа длины сказок представлены в таблице 3.

Таблица 3. Анализ длины русских народных сказок

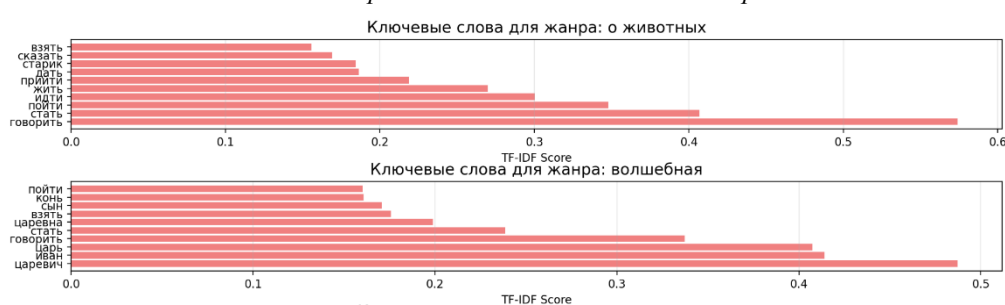
Длина сказки	count	mean	min	max
короткая	86	111.3	20	294
средняя	385	455.7	90	2432
длинная	78	1357.6	632	2672

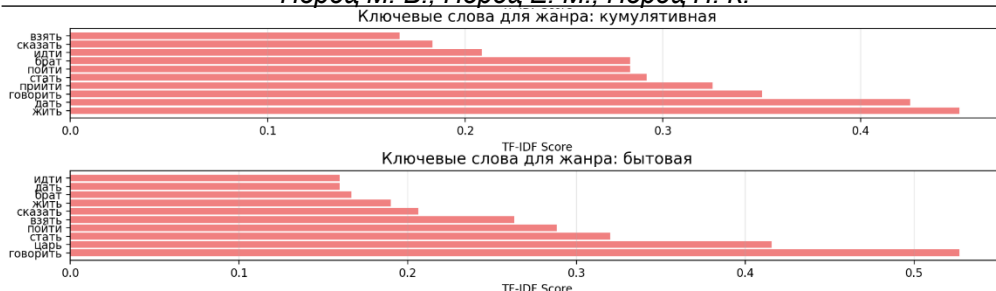
Общее количество слов (после предобработки): 290918, средняя длина: 529.9 слов.

Полученные данные создают статистически репрезентативную основу для последующего стилистического и ключевого анализа.

Проведенный анализ с применением метрики TF-IDF выявил жанровую специфику лексического состава русских народных сказок.

Рис. 3. Распределение ключевых слов по жанрам





В результате ключевого анализа было отмечено, что:

1. Волшебные сказки сопровождаются атрибутами власти и статуса: царевич (0.4873), Иван (0.4142), царь (0.4074), царевна (0.1990).
2. Сказки о животных акцентируют действие и взаимодействие: говорить (0.5741), стать (0.4068), пойти (0.3479).
3. Кумулятивные сказки отличаются социальной направленностью: жить (0.4501), дать (0.4251), говорить (0.3501).
4. Бытовые сказки совмещают элементы волшебного и социального: говорить (0.5270), царь (0.4157), стать (0.3200).

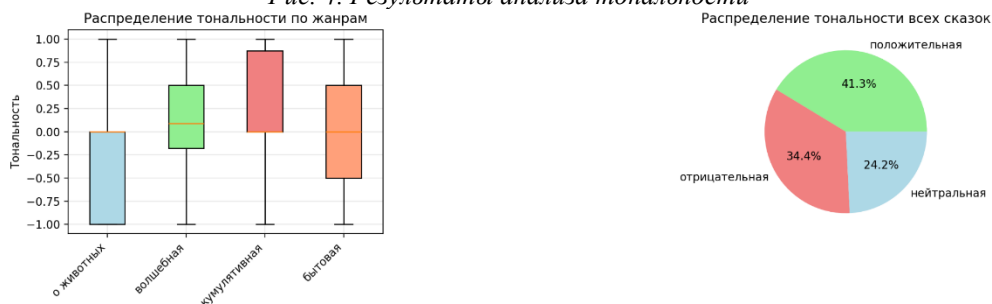
Полученные данные свидетельствуют о системном характере лексической организации сказочных жанров, где каждый жанр формирует собственное тематическое ядро, отражающее его специфику и культурные функции.

Для анализа тональности был составлен словарь условно положительных и условно отрицательных слов и оценивалась их доля вхождения в текст каждой сказки.

В результате проверки выяснилось, что общая эмоциональная картина характеризуется нейтральной медианной тональностью (0.000) со смещением в позитивную область (среднее значение 0.026) и значительным разбросом показателей (стандартное отклонение 0.604), что указывает на общую нейтральность сказочного нарратива. Однако жанровая дифференциация выявляет заметные различия.

В сказках о животных прослеживается выраженная негативная тенденция (-0.197), что отражает дидактическую функцию жанра, основанную на демонстрации последствий неправильного поведения. Волшебные сказки имеют умеренную позитивную окраску (0.121), что может соответствовать ориентации жанра на преодоление испытаний. Бытовые сказки имеют нейтрально-негативный характер (-0.019), который обуславливается социальной критикой и иронией, а кумулятивные сказки – позитивную направленность (0.203), связанную с ритмической организацией и комическими элементами.

Рис. 4. Результаты анализа тональности



ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ...



Подборки максимально положительных и отрицательных сказок, определенных по экстремальным значениям, демонстрирует, что даже между жанрами существует значительный разброс (таб. 5.). Список результатов анализа тональности представлен в таблице 4.

Таблица 4. Анализ результатов тональности русских народных сказок

жанр	mean	std	count
Бытовая	-0.019	0.643	174
волшебная	0.121	0.501	260
Кумулятивная	0.203	0.606	22
О животных	-0.197	0.718	93

Определено, что средняя тональность: 0.026, медиана: 0.000, стандартное отклонение: 0.604.

Таблица 5. Анализ тональности русских народных сказок

Топ-3 положительных сказок (жанр)	Топ-3 отрицательных сказок (жанр)
Теремок (кумулятивная)	Колобок (о животных)
Маша и медведь (волшебная)	Курочка Ряба (бытовая)
Каша из топора (бытовая)	Петушок золотой гребешок (о животных)

Важно отметить, что использованный словарный метод имеет ограничения, особенно для архаичных и метафорических текстов. В будущих исследованиях целесообразно применить более сложные модели, учитывающие контекст, а также способные отслеживать изменение эмоционального окраса текста от начала к концу сказки.

ВЫВОДЫ

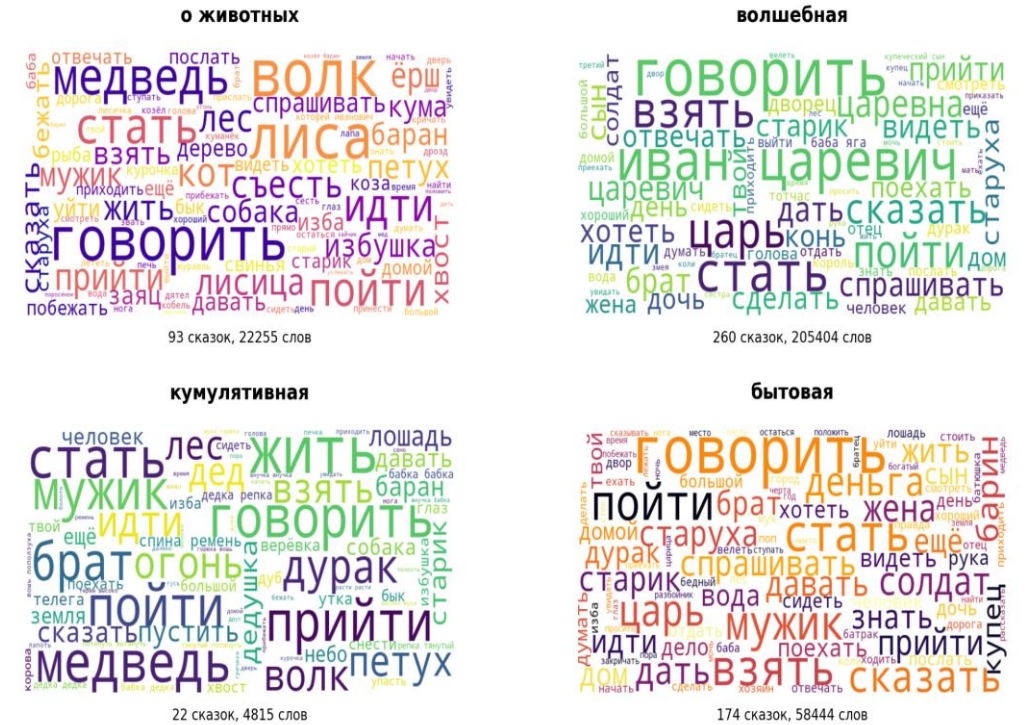
Полученные в результате частотного, ключевого и тонального анализа данные коррелируют с морфологической теорией сказки, разработанной В. Я. Проппом.

Выявленное частотное ядро непосредственно отражает ключевые функции пропповской морфологии:

1. Глаголы коммуникации (говорить, сказать) вербализуют функции «оповещения», «обмана», «клеветы».

- Общий позитивный сдвиг может быть коррелятом компенсаторной природы сказки по концепции В. Я. Проппа, однако для подтверждения этой гипотезы требуются методы, анализирующие динамику тональности внутри текста.

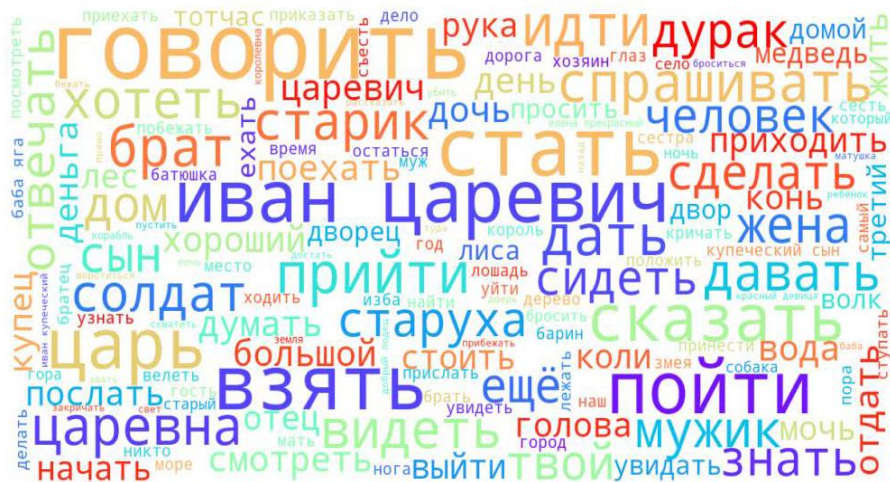
Рис. 5. Облака слов на основе частотного анализа



Помимо частотного ядра необходимо отметить и жанровую специфику, которая в свете пропповской типологии приобретает конкретные интерпретации:

1. Волшебные сказки демонстрируют полную реализацию 31 функции с акцентом на аристократию и власть.
2. Сказки о животных редуцируют персонажную парадигму, сохраняя базовые функции «вредительства» и «ликвидации недостачи».
3. Бытовые сказки трансформируют функции, замещая магические элементы социальными взаимодействиями.

Экстремальные значения тональности в конкретных сказках (Колобок, Теремок) маркируют ключевые сюжетные точки «нарушения запрета» и «наказания».



Выявленные закономерности демонстрируют, что структурное единство русского сказочного эпоса, описанное Проппом на уровне нарративных функций, находит прямое отражение в лексико-статистических параметрах текстов.

Проведенное исследование позволило достичь поставленной цели и выявить устойчивые лексико-стилистические паттерны, характерные для русских народных сказок, с помощью методов машинного обучения.

Осуществлена верификация теории В. Я. Проппа количественными методами. Показано, что ключевые нарративные функции («герой», «даритель», «помощник», «перемещение») имеют прямые корреляты в виде лексико-статистических маркеров, что подтверждает структурное единство сказочного эпоса на глубинном уровне.

Авторами настоящей работы планируются дальнейшие исследования, а именно – сравнительный частотный анализ, использующий сопоставление с корпусом литературных произведений и современного языка; углубленный анализ тональности на основе более сложных языковых моделей; разработка модели машинного обучения для автоматического определения жанра неизвестной сказки.

Таким образом, комбинация традиционных гуманитарных знаний и современных компьютерных методов открывает новые возможности для глубокого и масштабируемого изучения культурного наследия.

Список литературы

1. Андреев Д. С., Фадеева Е. А., Кузнецова Е. А. Как измерить культурный код? Количественный анализ семантики в сказках А. С. Пушкина и в ценностях российской молодежи // Государственное управление. Электронный вестник. – 2023. – № 98. – С. 151–161.
2. Балко Е. С. Анализ лексико-стилистических особенностей русских народных сказок с помощью методов машинного обучения // Лингвистические и культурологические аспекты современного инженерного образования: Сб. тезисов VI Международной научно-практической конференции (Томск, 12–14 ноября 2025 г.). – Томск: Национальный исследовательский Томский политехнический университет, 2025. – С. 10–15.
3. Велилаева Л. Р., Абдулжемилева Ф. И. Культурный код в художественной литературе // Мир науки, культуры, образования. – 2021. – № 6(91). – С. 527–529.
4. Герасимова Н. М. Прагматика текста: фольклор, литература, культура. – СПб.: РИИИ, СПбГУ, 2012. – 364 с.
5. Жикюлина К. П. Алисины сказки: трансформация структуры, сказочных формул и контекстов у голосового помощника в навыке «Давай придумаем» // Litera. – 2024. – № 2. – С. 45–63.
6. Зворыгина О. И. Речевая организация хронотопа русской литературной сказки // Слово. Текст. Контекст. – 2020. – № 1. – С. 38–44.
7. Матвеев В. Э. Использование русских народных сказок в процессе обучения студентов-филологов РКИ с учетом современных технологий // Русистика. 2013. – № 4. – С. 72–80.
8. Международная облачная платформа для хостинга IT-проектов и совместной разработки «Гитхаб». – Режим доступа: <https://github.com/monalenska/russian-folklore-ml/>. – (Дата обращения: 25.04.2026).
9. Москин Н. Д., Лебедев А. А. Теоретико-графовый подход к анализу вариативности русских сказок // Интеллектуальные системы в производстве. – 2024. – Т. 2. – № 4. – С. 90–98.
10. Народные русские сказки А. Н. Афанасьева / Изд. Подгот. Л. Г. Бараг и Н. В. Новиков: в 3 т. – М., 1984–1985.
11. Норец М. В., Элькан О. Б., Норец Н. К. Взаимодействие искусственного интеллекта и цифровой лингвистики // Litera. – 2025. – № 3. – С. 221–230.
12. Поверенная А. А. Лингвокультурные особенности сказочного дискурса: аксиологический аспект (на материале немецкого и русского языков): Дисс. ... канд. филол. наук: 10.02.19. – Майкоп, 2022. – 191 с.
13. Пропт В. Я. Морфология волшебной сказки; Исторические корни волшебной сказки. – М.: КоЛибри, Азбука-Аттикус, 2021. – 511 с.
14. Пропт В. Я. Морфология сказки / Гос. ин-т истории искусств. – Л.: Academia, 1928. – 152 с.
15. Разумова И. А. Стилистическая обрядность русской волшебной сказки: монография. – Петрозаводск: Карелия, 1991. – 163 с.
16. Рафекова Д. Р. Лингвоаксиологический анализ русской народной сказки и языковые механизмы репрезентации ценностей и оценок по корпусу А. Н. Афанасьева // Филологический вестник. – 2026. – Т. 5. – № 1. – С. 40–45.

References

1. Andreyuk D. S., Fadeeva E. A., Kuznecova E. A. *Kak izmerit' kul'turnyj kod? Kolichestvennyj analiz semantiki v skazkah A. S. Pushkina i v cennostyah rossijskoj molodezhi* [How to Measure Cultural Code? A Quantitative Analysis of Semantics in A.S. Pushkin's Fairy Tales and the Values of Russian Youth]. *Gosudarstvennoe upravlenie. Elektronnyj vestnik*, 2023, no 98, pp. 151–161.
2. Balko E. S. *Analiz leksiko-stilisticheskikh osobennostej russkikh narodnyh skazok s pomoshch'yu metodov mashinnogo obucheniya* [Analysis of lexical and stylistic features of Russian folk tales using machine learning methods]. *Lingvisticheskie i kul'turologicheskie aspekty sovremennogo inzhenerenogo obrazovaniya: Sbornik tezisev VI Mezhdunarodnoj nauchno-prakticheskoy konferencii (Tomsk, 12–14 noyabrya 2025 goda)*. Tomsk, Tomskij politekhnicheskij universitet, 2025, pp. 10–15.
3. Velilaeva L. R., Abdulzhemileva F. I. *Kul'turnyj kod v hudozhestvennoj literature* [Cultural code in fiction]. *Mir nauki, kul'tury, obrazovaniya*. 2021, no. 6(91), pp. 527–529.
4. Gerasimova N. M. *Pragmatika teksta: fol'klor, literatura, kul'tura* [Pragmatics of the text: folklore, literature, culture]. Sankt-Peterburg: RIII, SPbGU, 2012. 364 p.

5. Zhikulina K. P. *Alisiny skazki: transformaciya struktury, skazochnyh formul i kontekstov u golosovogo pomoshchnika v navyke «Davaj pridumaem»* [Alice's Fairy Tales: Transforming Structure, Fairytale Formulas, and Contexts in the Voice Assistant's "Let's Think of It" Skill]. *Litera*, 2024, no. 2, pp. 45–63.
6. Zvorygina O. I. *Rechevaya organizaciya hronotopa russkoj literaturnoj skazki* [Speech organization of the chronotope of the Russian literary fairy tale]. *Slovo. Tekst. Kontekst*, 2020, no. 1, pp. 38–44.
7. Matveenko V. E. *Ispol'zovanie russkih narodnyh skazok v processe obucheniya studentov-filologov RKI s uchetom sovremennyh tekhnologij* [Using Russian folk tales in teaching Russian as a foreign language to students of philology, taking into account modern technologies]. *Rusistika*, 2013, no 4, pp. 72–80.
8. *Mezhdunarodnaya oblachnaya platforma dlya hostinga IT-proektov i sovmestnoj razrabotki «Github»*. Rezhim dostupa: <https://github.com/monalenka/russian-folklore-ml> (accessed 25 April 2026).
9. Moskin N. D., Lebedev A. A. *Teoretiko-grafovyj podhod k analizu variativnosti russkih skazok* [A graph-theoretic approach to the analysis of variability in Russian fairy tales]. *Intellektual'nye sistemy v proizvodstve*, 2024, T. 2, no 4, pp. 90–98.
10. *Narodnye russkie skazki* A. N. Afanas'eva [Russian folk tales by A. N. Afanasyev]. Izd. Podgot. L. G. Barag i N. V. Novikov: v 3 t. Moskva, 1984–1985.
11. Norets M. V., El'kan O. B., Norets N. K. *Vzaimodejstvie iskusstvennogo intellekta i cifrovoj lingvistiki* [The Interaction of Artificial Intelligence and Digital Linguistics]. *Litera*, 2025, no. 3, pp. 221–230.
12. Poverennaya A. A. *Lingvokul'turnye osobennosti skazochnogo diskursa: aksiologicheskij aspekt (na materiale nemeckogo i russkogo yazykov): Diss. ... kand. filol. nauk*: [Linguocultural features of fairy-tale discourse: axiological aspect (based on the German and Russian languages). Thesis]. Majkop, 2022. 191 p.
13. Propp V. Ya. *Morfologiya volshebnnoj skazki; Istoricheskie korni volshebnnoj skazki* [Morphology of the fairy tale; Historical roots of the fairy tale]. Moscow, KoLibri, Azbuka-Attikus, 2021. 511 p.
14. Propp V. Ya. *Morfologiya skazki* [Morphology of a fairy tale]. Gos. in-t istorii iskusstv, Leningrad, Academia, 1928. 152 p.
15. Razumova I. A. *Stilisticheskaya obryadnost' russkoj volshebnnoj skazki: monografiya* [Stylistic Ritualism of Russian Fairy Tales: A Monograph]. Petrozovodsk, Kareliya, 1991. 163 p.
16. Rafekova D. R. *Lingvoaksiologicheskij analiz russkoj narodnoj skazki i yazykovye mekhanizmy reprezentacii cennostej i ocenok po korpusu A. N. Afanas'eva* [Linguistic and axiological analysis of Russian folk tales and linguistic mechanisms of representing values and assessments based on the corpus of A. N. Afanasyev]. *Filologicheskij vestnik*, 2026, T. 5, no 1, pp. 40–45.

APPLICATION OF MACHINE LEARNING METHODS TO ANALYZE LEXICAL AND STYLISTIC FEATURES OF RUSSIAN FOLK TALES

Norets M. V., Norets E. M., Norets N. K.

This study aims to identify lexical and stylistic patterns in Russian folk tales using machine learning methods. A corpus of 549 tales (based on A.N. Afanasyev's collection) was used for frequency analysis, TF-IDF, sentiment analysis, and stylistic tagging. Propp's morphological theory served as the theoretical framework. A communicative-active lexical core was identified: dominant verbs of speech and action ("speak," "go," "take"), anthroponyms ("Ivan," "tsar"), and social roles ("son," "brother"). Genre differentiation (magical, everyday, animal, cumulative) is confirmed by the TF-IDF metric. The overall sentiment is neutral-positive (mean 0.026), but animal tales have a pronounced negative bias (-0.197), while cumulative tales have a positive bias (+0.203). The quantitative results correlate with Propp's narrative functions: frequency markers directly correspond to the functions "hero," "giver," "transition," and "transformation." The structural unity of the Russian folktale epic is statistically confirmed. The proposed approach demonstrates the effectiveness of combining classical folklore studies and digital methods for an objective, scalable analysis of cultural heritage.

Keywords: Russian folk tales, computational linguistics, machine learning, frequency analysis, sentiment analysis, lexical and stylistic patterns, genre differentiation.