

УДК 519.2:801.81

DOI: 10.22213/2410-9304-2024-4-90-97

## Теоретико-графовый подход к анализу вариативности русских сказок

Н. Д. Москин, доктор технических наук, доцент,

Петрозаводский государственный университет, Петрозаводск, Россия

А. А. Лебедев, кандидат филологических наук,

Петрозаводский государственный университет, Петрозаводск, Россия

*В работе рассматривается теоретико-графовый подход, который применяется для анализа русских народных волшебных сказок. Подобное формальное описание структуры анализируемых фольклорных текстов опирается на закономерности, изложенные В. Я. Проппом в ряде своих работ и, в частности, в его известной книге «Морфология сказки». Это позволяет систематизировать и анализировать ключевые события, происходящие в текстах. Вершины теоретико-графовой модели представляют персонажей сказки, которые классифицируются на определенные фиксированные группы, а ребра обозначают встречи между персонажами, сопровождающиеся значимыми для сказочного сюжета действиями. На примере текста «Терешечка» из сборника А. М. Афанасьева показано, как строится модель и вычисляются числовые характеристики графов. Для этого в том числе используются инструменты информационной системы «Фольклор». С помощью этой программы можно не только хранить тексты коллекции и их теоретико-графовые модели, но также в автоматизированном режиме создавать формальные структуры, выполнять их визуализацию, агрегацию и сравнение. Дискриминантный анализ показал, что по формальным параметрам можно идентифицировать сюжет сказки о молодце-удальце, молодильных яблоках и живой воде. Значение коэффициента Лямбда Уилкса (Wilks' Lambda) равно 0,0275781 (меньше 0,05), что говорит о хорошем качестве классификатора. Исходя из этого следует заключить, что для систематизации и сравнения сюжетов русских народных сказок можно применять математические модели, основанные на графах. При этом открываются перспективы изучения как вариативности сказок, так и выделения наиболее значимых и типичных для определенной географической местности или временного периода сюжетов.*

**Ключевые слова:** русская сказка, теоретико-графовая модель, вариативность сюжета, дискриминантный анализ, информационная система «Фольклор».

### Введение

Поиск аналогий и обнаружение различий между отдельными фольклорными текстами и группами текстов, вопросы структурирования и формального анализа сюжетов волшебных русских сказок – тема, интересующая специалистов как из сферы фольклористики, так и собственно лингвистики [1, 2]. Одной из основных задач в данном контексте является классификация и систематизация типов сказок, базирующаяся на материале русскоязычных текстов. Заметными работами в этой области являются труды [3, 4]. Следует отметить, что анализ сюжетов сказок можно проводить не только применительно к текстам, написанным на русском языке, но и включать в исследование сопоставление сюжетов из разных культур. В частности, проблема межкультурного анализа в аспекте фольклорной традиции рассматривается в статье [5]. Этот подход позволяет исследовать особенности передачи сюжетов между народами и оценивать разные варианты методологии изучения сказок. Также важно формальное описание структуры анализируемых фольклорных текстов, что подробно описал В. Я. Пропп в своей работе «Морфология сказки» [6].

Для более эффективного решения обозначенных нами лингвистических и собственно фольклорных задач полезно применять разнообразные математические методы. Такой комбинированный подход обеспечивает более точный анализ текстов, улучшает систематизацию собранного материала и обеспечивает более продуктивное сравнение текстов с учетом разнообразных критериев. Рассмотренная графовая модель позволяет систематизировать и анализировать ключевые события, происходящие в волшебных сказках. Известно, что у сказок есть определенная структура, которую можно формализовать и сравнивать между собой, а также выделять ключевые элементы в каждом сюжете. В основе предлагаемой графовой модели – персонажи сказок и встречи между ними, формирующие сюжет.

Вершинами графа выступают персонажи сказки, распределенные по группам согласно способу, предложенному в работе [7]. Выделяются следующие персонажи: Герой – Н, Антигерой – А, Прорицатель – Р, Даритель – Д, Помощник – П, Антипомощник – V, Глупец – G, Антидаритель – W, Родственник – Т. Дополнительно включены Награда – N и Препятствие – R, которые, хотя и

не являются в буквальном смысле персонажами, важны для развития сюжета.

Персонажи одной категории нумеруются по мере их появления в сказке: например, первый помощник обозначается как П1, второй – как П2 и т. д. Иногда под одним персонажем понимается сразу несколько людей – это возможно, если они выполняют единую функцию в тексте сказки и не разделены. К примеру, сорок одна дочь бабы-яги из сказки «Баба-яга и Заморышек» рассматривается в модели как единый персонаж, так как все действия происходят одновременно с каждой из таких дочерей (их всех одновременно выводят из терема, сватают с Заморышком и его братьями, а затем им одновременно рубят головы).

Ребра графа представляют собой встречи, осуществляемые между персонажами, которые сопровождаются действиями, играющими важную роль в развертывании сказочного сюжета. При этом, как отмечается в [8, с. 84], «лексическое обрамление сказки относительно небогато, да и сами языковые конструкции представляют собой, в большинстве случаев, застывшие штампы».

В тексте персонажи могут многократно встречаться между собой, и каждая подобная встреча обозначается отдельным ребром графа. Обычно эти взаимодействия выражаются глаголами: *На третий день посылает царь меньшего своего сына, Ивана-царевича, «Здравствуй, старая баба-яга!» – говорит Иван-царевич и т. п.* Здесь и далее тексты сказок выделяются курсивом и цитируются по источнику [9].

Целью работы является проверка гипотезы о возможности использования теоретико-графовых моделей для решения задачи классификации сказочных сюжетов.

### Основные сложности, связанные с формализацией связей в фольклорном тексте

#### 2.1. Подразумеваемое действие, не выраженное глаголом в тексте

В ряде случаев действие в сказке может быть только подразумеваемым (частое явление при встрече главного героя и его помощника – глаголы говорения (*заговорил, рассказал, сказал, молвил*), могут отсутствовать, но восстанавливаются читателем из контекста: «К счастью, на дороге <стояла> печка»). Если последовательность действий повторяется, но упоминается в тексте не полностью, то подразумевается, что все действия были повторены в точности (что отражается в формировании модели). К примеру, в сказке «172. Сказка о молодце-удальце, молодильных яблоках и живой воде» [9] подробно описывается, как главный герой беседует с первой бабой-ягой, а

затем с ней же беседует Царь-девица (*только успела она <баба-яга> отпустить царевича, как прилетает царь-девица и застаёт бабу-ягу всю растрепанную*), а встречи с двумя другими бабами-ягами предельно сжаты в тексте: *Две другие бабы-яги то же сказали*.

#### 2.2. Встречи трех и более лиц

Как правило, встречи между персонажами являются бинарными (*Макарка говорит царю, Василиса зарыла череп в землю и т. п.*). Но иногда во взаимодействиях принимают участие три персонажа. Это бывает, когда действия двух из них направлены на третьего (например, *отец и мать создают Терешечку* – в этом случае устанавливается только две связи), или в процессе дарения предмета (см. пункт 2.5.)

#### 2.3. Сохранение роли персонажа после смерти

В ходе развития сюжета сказки роль персонажа сохраняется даже после его смерти. Например, рассматривая «другую дочь», убитую бабой-ягой, мы условно сохраняем ее как персонажа и после ее смерти: *Приехал отец и повёз одни косточки <другой дочери>*. Аналогичная ситуация – смерть главного героя и последующее его воскрешение второстепенным персонажем при помощи живой воды: *Конь Ивана-царевича собрал его косточки в одно место, вспырынул живою водой*.

#### 2.4. Невозможность однозначно установить взаимосвязи между объектами

Редка, но возможна ситуация, когда установить взаимосвязь между объектами однозначно не удастся. К примеру, в сказке «140. Зорька, Вечорка и Полуночка» мы встречаем следующую фразу: *Король <...> тотчас же обвенчал Зорьку, Вечорку и Полуночку на своих дочерях*. Мы знаем, что у короля есть три дочери (младшая, средняя и старшая), и исходим из рационального допущения, в соответствии с которым каждый богатырь получил ровно одну дочь в жены, но кому именно какая дочь досталась – понять из текста произведения однозначно нельзя.

#### 2.5. Ситуации дарения и обмена

Особо следует обратить внимание на описание в ходе построения моделей ситуаций дарения и обмена. Дарение – часто встречающееся действие в сказках (например, *Баба-яга дает Василисе череп, Царевна дает Ивану-царевичу золотое кольцо и т. п.*). В таких случаях в модели фиксируются три связи (одинаково нумеруемые), по одной между каждым из трех объектов.

Реже встречается вариант обмена предметами. Например, в сказке «176. Сказка о молодце-

удальце, молодильных яблоках и живой воде» [9] Иван-царевич оставляет девушке своего коня, а получает при этом летучего сокола.

В тексте сказки это выражено в виде одного предложения: «Садись же ты, добрый молодец, на моего летучего сокола; а своего коня у меня оставь», в модели же включает в себя шесть связей (первые три из них указывают на то, что Иван-царевич берет летучего сокола у первой прекрасной девушки, а еще три – что Иван-царевич оставляет коня первой прекрасной девушке).

Таблица 1. Персонажи и события в тексте «Терешечка»

Table 1. Characters and events in the text of «Tereshchka»

	Вершина 1	Вершина 2	Событие в тексте
1	T1	H1	Отец и мать создают Терешечку
1	T2	H1	Отец и мать создают Терешечку
2	T2	H1	Мать зовет Терешечку
3	H1	T2	Терешечка подъезжает к матери
4	T2	H1	Мать рассказывает Терешечке о ведьме
5	A1	H1	Ведьма Чувилиха зовет Терешечку
6	H1	A1	Терешечка распознает обман ведьмы Чувилихи
7	T2	H1	Мать зовет Терешечку
8	H1	T2	Терешечка подъезжает к матери
9	A1	H1	Ведьма Чувилиха зовет Терешечку
10	H1	A1	Терешечка подъезжает к Чувилихе
11	A1	H1	Чувилиха крадет Терешечку
12	A1	A2	Ведьма Чувилиха велит дочери изжарить Терешечку
12	A1	H1	Ведьма Чувилиха велит дочери изжарить Терешечку
12	A2	H1	Ведьма Чувилиха велит дочери изжарить Терешечку
13	H1	A2	Терешечка жарит дочь ведьмы
14	H1	A1	Терешечка кричит ведьме Чувилихе
15	H1	A1	Терешечка отвечает ведьме Чувилихе
16	A1	H1	Ведьма Чувилиха видит Терешечку
17	A1	P1	Ведьма Чувилиха идет к кузнецу
18	H1	P2	Терешечка обращается к первой стае гусей-лебедей
19	P2	H1	Первая стая гусей-лебедей отказывает Терешечке
20	H1	P3	Терешечка обращается ко второй стае гусей-лебедей
21	P3	H1	Вторая стая гусей-лебедей отказывает Терешечке
22	H1	P4	Терешечка обращается к зачипанному гусенку
23	P4	H1	Зачипаный гусенок помогает Терешечке
24	P4	T1	Зачипаный гусенок и Терешечка летят к отцу
24	H1	T1	Зачипаный гусенок и Терешечка летят к отцу
25	T1	H1	Отец обнимает Терешечку
25	T2	H1	Мать обнимает Терешечку
26	T1	P4	Отец и мать кормят зачипанного гусенка
26	T2	P4	Отец и мать кормят зачипанного гусенка

Если поставить в соответствие персонажам сказки вершины графа, а событиям – ребра, то получится теоретико-графовая модель, представленная на рис. 1.

Так как ребра пронумерованы, то речь идет о построении темпорального графа, хотя это понятие в литературе понимается достаточно широко [10–13].

### Теоретико-графовая модель сказки

Рассмотрим пример построения теоретико-графовой модели на основе текста «112. Терешечка» из сборника [9, с. 183–184]. В этой сказке выделяются следующие персонажи: Родственник – Отец (T1) и мать (T2); Герой – Терешечка (H1); Антигерой – Ведьма Чувилиха (A1), Дочь Чувилихи (A2); Помощник – Кузнец (P1), Первая стая гусей-лебедей (P2), Вторая стая гусей-лебедей (P3), Защищенный гусенок (P4). Связи и соответствующие им события в указанном тексте представлены в табл. 1.

### Анализ результатов

На основе сказок из сборника А. Н. Афанасьева была составлена выборка текстов, представленная в табл. 2. Разделим тексты на две группы: в первую войдут сказки с номерами 93, 98, 102–107, 112–113, 140 и 148, а во вторую – варианты сказки о молодце-удальце, молодильных яблоках и живой воде (номера от 171 до 178).

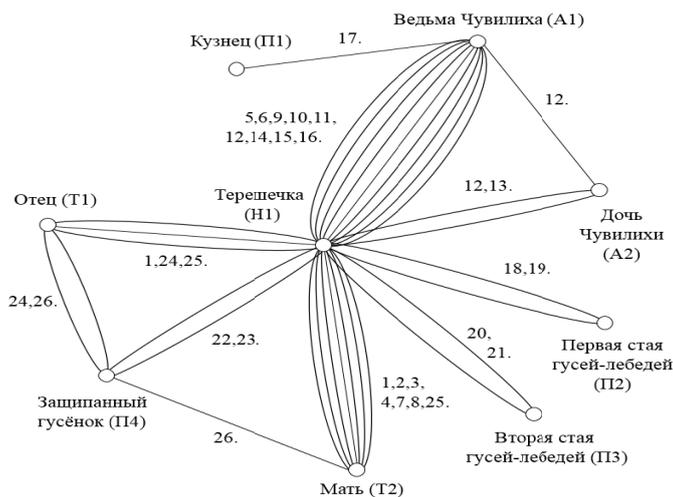


Рис. 1. Теоретико-графовая модель сказки «Терешечка»

Fig. 1. Graph-theoretical model of the fairy tale « Tereshechka»

Различные варианты одного сюжета могут включать в себя различное число персонажей и отношений между ними, все это отражено в соответствующей теоретико-графовой модели. При

этом одинаковые сюжеты могут существенно отличаться по размеру (например, текст под номером 176 включает 2760 слов, а текст под номером 177 примерно в три раза меньше: 836 слов).

Таблица 2. Описание коллекции текстов

Table 2. Description of the text collection

Номер	Название текста	Число слов	Источник
93	Ведьма и Солнцева сестра	858	Народные русские сказки А. Н. Афанасьева: в 3 т. Т. 1. М.: Гослитиздат, 1957. С. 136–138
98	Дочь и падчерица	476	- // - С. 146–147
102	Баба-яга	441	- // - С. 155–156
103	Баба-яга	725	- // - С. 156–158
104	Василиса Прекрасная	2334	- // - С. 159–165
105	Баба-Яга и заморышек	855	- // - С. 166–168
106	Баба-яга и жихарь	570	- // - С. 169–170
107	Баба-яга и жихарь	629	- // - С. 170–172
112	Терешечка	617	- // - С. 183–184
113	Гуси-лебеди	402	- // - С. 327–328
140	Зорька, вечерка и полуночка	1345	- // - С. 299–302
148	Никита Кожемяка	486	- // - С. 327–328
171	Сказка о молодце-удальце, молодильных яблоках и живой воде	1978	- // - С. 431–435
172	Сказка о молодце-удальце, молодильных яблоках и живой воде	1020	- // - С. 436–438
173	Сказка о молодце-удальце, молодильных яблоках и живой воде	1315	- // - С. 438–442
174	Сказка о молодце-удальце, молодильных яблоках и живой воде	949	- // - С. 442–444
175	Сказка о молодце-удальце, молодильных яблоках и живой воде	1504	- // - С. 444–448
176	Сказка о молодце-удальце, молодильных яблоках и живой воде	2760	- // - С. 448–455
177	Сказка о молодце-удальце, молодильных яблоках и живой воде	836	- // - С. 455–457
178	Сказка о молодце-удальце, молодильных яблоках и живой воде	972	- // - С. 457–460

В табл. 3 представлено распределение объектов сказок по группам, перечисленным в п. 1. Поскольку ни в одной сказке не содержится персонаж «Антидаритель» (W), то этот признак исключим из дальнейшего исследования (остальные признаки обозначим  $v_i$ , где индекс  $i$  принимает значения от 1 до 10). Еще четыре параметра определим по теоретико-графовой

Таблица 3. Персонажи в сказках

Table 3. Characters in fairytales

Номер текста	Число вершин (персонажей)	H ( $v_1$ )	A( $v_2$ )	P( $v_3$ )	Д( $v_4$ )	Π ( $v_5$ )	V ( $v_6$ )	G ( $v_7$ )	W	T ( $v_8$ )	N ( $v_9$ )	R ( $v_{10}$ )
93	16	1	1	1	1	4	0	0	0	1	4	3
98	8	1	2	1	0	1	0	1	0	1	1	0
102	8	1	1	1	0	1	1	1	0	2	0	0
103	15	1	1	0	0	2	1	0	0	2	2	6
104	10	1	1	0	0	1	2	0	0	3	2	0
105	14	1	2	0	1	3	1	0	0	3	2	1
106	7	1	4	0	0	0	0	0	0	2	0	0
107	7	1	2	0	0	1	0	0	0	3	0	0
112	9	1	2	0	0	4	0	0	0	2	0	0
113	10	1	2	0	0	4	0	0	0	3	0	0
140	13	3	3	0	0	0	1	0	0	5	1	0
148	6	1	1	0	0	1	0	0	0	3	0	0
171	20	1	2	0	3	3	2	3	0	1	4	1
172	15	1	1	0	3	1	0	2	0	2	4	1
173	15	1	1	0	1	0	2	2	0	1	3	4
174	18	1	1	0	3	1	1	2	0	2	6	1
175	21	1	2	1	2	0	1	2	0	2	6	4
176	39	1	5	0	3	7	1	2	0	6	12	2
177	11	1	1	0	0	4	0	1	0	1	2	1
178	10	1	1	0	0	3	0	2	0	2	0	1

Двенадцатым признаком ( $v_{12}$ ) определим параметр семантической связности текста. Он вычисляется как частное от деления числа пар вершин, связанных между собой, на величину  $n(n-1)/2$ , где  $n$  – число вершин в графе. Например, для сказки «Терешечка»  $v_{12}$  приближенно равен 0,3056.

Параметры  $v_{13}$  и  $v_{14}$  представляют собой коэффициенты гиперболической регрессии, которой можно аппроксимировать распределение степеней вершин сказки (предварительно выполнив сортировку и нормирование исходного массива). Например, на рис. 2 изображено подобное распределение для сказки «Терешечка». Здесь зависимость приближена гиперболической кривой  $y=0,443056/x - 0,02815$  (подробнее об этих признаках в [14]).

модели (похожим образом это делалось в [14]). Одиннадцатым признаком ( $v_{11}$ ) выберем максимальную степень вершины теоретико-графовой модели, т. е. количество связей у наиболее значимого персонажа. Например, в сказке «Терешечка»  $v_{11}=27$  (на рис. 1 вершина, соответствующая главному герою Терешечке (H1), инцидентна 27 ребрам).

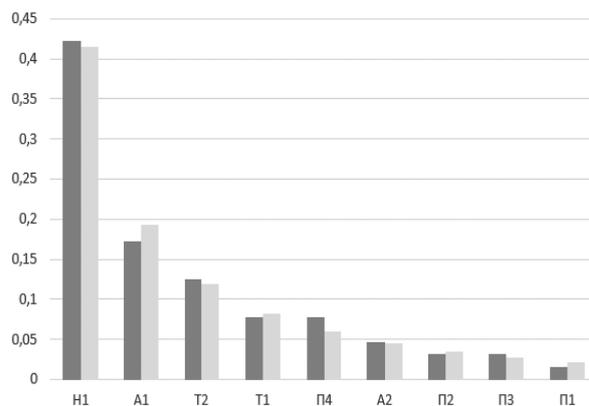


Рис. 2. Распределение нормированных степеней вершин сказки «Терешечка»

Fig. 2. The distribution of the normalized degrees of the vertices of the fairy tale «Tereshchka»

На основе вышеперечисленных признаков построим классификатор, который мог бы рас-

познать принадлежность сказки к первой или второй группе (т. е. по формальным параметрам идентифицировать сюжет сказки о молодце-удальце, молодильных яблоках и живой воде). Для этого с помощью информационной системы «Фольклор» [15] были построены теоретико-графовые модели текстов и рассчитаны необходимые числовые характеристики. Далее в пакете Statistica 6 были рассчитаны дискриминантные функции  $f_1$  и  $f_2$ , которые имеют вид (коэффициенты дискриминантных функций представлены в табл. 4):

$$f_i = a_{i1}v_1 + a_{i2}v_2 + a_{i3}v_3 + \dots + a_{i14}v_{14} + \text{const.}$$

Таблица 4. Коэффициенты дискриминантных функций

Table 4. Coefficients of discriminant functions

Признаки	$a_{1j}$ для $f_1$	$a_{2j}$ для $f_2$
$v_1$	110,146	131,702
$v_2$	-12,007	-22,441
$v_3$	29,919	-15,443
$v_4$	41,525	21,439
$v_5$	48,435	48,956
$v_6$	68,048	55,892
$v_7$	70,357	138,538
$v_8$	27,878	29,055
$v_9$	48,641	71,446
$v_{10}$	29,027	29,736
$v_{11}$	-9,37	-11,2
$v_{12}$	1042,68	1132,959
$v_{13}$	1840,43	2063,56
$v_{14}$	1906,89	1810,551
const	-587,56	-737,523

Подставляя вместо  $v_1, v_2, \dots, v_{14}$  признаки для конкретных сказок, можно сравнить значения  $f_1$  и  $f_2$ . Если  $f_1 > f_2$ , то классификатор относит текст в первой группе, т. е. этот сюжет существенно отличается от сказки о молодце-удальце, молодильных яблоках и живой воде. В против-

ном случае – наоборот (относит ко второй группе).

Эксперименты показали, что модель верно классифицировала все тексты, т. е. без ошибок первые 12 текстов отнесла к первой группе, а оставшиеся 8 текстов ко второй группе. Важным показателем качества модели является значение Лямбды Уилкса (Wilks' Lambda), которое равно 0,0275781. Это число меньше 0,05, что свидетельствует о хорошем качестве построенного классификатора.

#### Заключение

Таким образом, можно сделать вывод о том, что теоретико-графовые модели могут быть использованы для решения задачи классификации сказочных сюжетов. На их основе при помощи математических методов можно не только исследовать вариативность сказок, но и выделить наиболее значимые и типичные для определенной географической местности или временного периода составляющие.

Кроме того, теоретико-графовая модель обеспечивает большую визуальную доступность, что важно для интерпретации текстов в филологических исследованиях. В дальнейшем, используя современные методы машинного обучения и другие перспективные технологии (например, GNN – GraphNeuralNetwork), на большей выборке текстов можно построить более совершенные классификаторы для систематизации и классификации фольклорных текстов разных типов.

#### Библиографические ссылки

1. Праведников С. П. Лингвофольклористика в научном пространстве гуманитаристики и вузовском учебном плане // Профессорский журнал. Серия: Русский язык и литература. 2021. №4(8). С. 42–45. DOI: 10.18572/2687-0339-2021-4-42-45.
2. Куран М. С. Специфика анализа сказки в семиотическом аспекте // Мир науки, культуры, образования. 2010. № 5. С. 238–240.
3. Андреев Н. П. Указатель сказочных сюжетов по системе Ааарне. Л. : Изд-во Государственного русского географического общества, 1929. 120 с.
4. Сравнительный указатель сюжетов. Восточнославянская сказка. Л. : Наука, 1979. 437 с.
5. Бобунова М. А., Хроленко А. Т. Русская народно-поэтическая речь на фоне иноэтнических традиций: кросскультурный аспект // Вестник славянских культур. 2021. № 61. С. 226–237. DOI: 10.37816/2073-9567-2021-61-226-237.
6. Пропт В. Морфология сказки. Л. : Academia, 1928. 152 с.
7. Гаазе-Рапопорт М. Г. Поиск вариантов в сочинении сказок // Зарипов Р. Х. Машинный поиск

вариантов при моделировании творческого процесса. М. : Наука, 1983. С. 213–223.

8. Нечай Ю. П., Поверенная А. А. Языковые средства реализации сюжета русской волшебной сказки // Гуманитарные и социальные науки. 2022. Т. 90, № 1. С. 83–88. DOI 10.18522/2070-1403-2022-90-1-83-88.

9. Афанасьев А. М. Народные русские сказки А. Н. Афанасьева: в 3 т. М. : Государственное издательство художественной литературы (Гослитиздат), 1957.

10. Bozhenyuk A., Knyazeva M., Bozheniuk V., Kacprzyk J. Fuzzy clique set determination method as an example of fuzzy temporal graph invariant // Advances in Intelligent Systems and Computing. 2021. Vol. 1306. Pp. 1–9. DOI: 10.1007/978-3-030-04164-9\_8.

11. Debrouvier A., Parodi E., Perazzo M., Soliani V., Vaisman A. A modeland query language for temporal graph databases // The VLDB Journal. 2021. Vol. 30. No. 5. Pp. 825–858. DOI: 10.1007/s00778-021-00675-4.

12. Ma Yu., Yuan Ye., Liu M., Wang G., Wang Y. Graph simulation on large scale temporal graphs // GeoInformatica. 2020. Vol. 24. No. 1. Pp. 199–220. DOI: 10.1007/s10707-019-00381-y.

13. Zhang T., Gao Yu., Qiu L., Chen L., Linghu Q., Pu Sh. Distributed time-respecting flow graph pattern matching on temporal graphs // World Wide Web. 2020. Vol. 23. No. 1. Pp. 609–630.

14. Щеголева Л. В., Лебедев А. А., Москин Н. Д. Методы анализа данных в задаче разграничения фольклорных и авторских текстов // Вопросы языкознания. 2020. № 2. С. 61–74. DOI: 10.31857/S0373658X0008823-4.

15. Rogov A. A., Kulakov K. A., Moskin N. D. Программная поддержка в решении задачи атрибуции текстов // Программная инженерия. 2019. Т. 10. № 5. С. 234–240. DOI: 10.17587/prin.10.234-240.

### References

1. Pravednikov S.P. [Linguistic and folkloristics in the scientific space of humanities and the university curriculum]. *The professorial journal. Series: Russian language and Literature*, 2021, no. 4, pp. 42-45 (in Russ.). DOI: 10.18572/2687-0339-2021-4-42-45.

2. Kuran M.S. [Specificity of fairy tale analysis in the semiotic aspect]. *The world of science, culture, and education*, 2010, no. 5, pp. 238-240 (In Russ.).

3. Andreev N.P. *Ukazatel' skazochnyhsyuzhetov-posistemeAarne* [Index of fairy tales according to the Aarne system]. Leningrad, Publishing house of the State Russian Geographical Society, 1929, 120 p. (in Russ.).

4. *Sravnitel'nyj ukazatel' syuzhetov. Vostochnoslavianskaya skazka* [Comparative index of plots. East

Slavic fairy tale]. Leningrad, Nauka Publ., 1979, 437 p. (in Russ.).

5. Bobunova M.A., Khrolenko A.T. [Russian popular poetic speech in the aspect of crosscultural linguistics]. *Bulletin of Slavic Cultures*, 2021, no. 61, pp. 226-237 (In Russ.). DOI: 10.37816/2073-9567-2021-61-226-237.

6. Propp V. *Morfologiya skazki* [Morphology of a fairy tale]. Leningrad, Academia, 1928, 152 p. (in Russ.).

7. Gaaze-Rapoport M.G. *Poiskvariantov v sochneniiskazok* [Search for variants of the writing of fairy tales]. In: Zaripov R.H. *Mashinnyj poisk variantov pri modelirovanii tvorcheskogo processa* [Machine search for variants of modeling the creative process]. Moscow, Nauka Publ., 1983, pp. 213-223 (in Russ.).

8. Nechai Yu.P., Poverennaya A.A. [Language means of implementing the plot of Russian fairy tale]. *Humanities and Social Sciences*, 2022, vol. 90, no. 1, pp. 83-88 (in Russ.). DOI 10.18522/2070-1403-2022-90-1-83-88.

9. Afanasyev A.M. *Narodnye russkie skazki A. N. Afanas'eva* [Russian folk tales by A. N. Afanasyev]: in 3 volumes. Moscow, State Publishing House of Fiction (Goslitizdat), 1957 (in Russ.).

10. Bozhenyuk A., Knyazeva M., Bozheniuk V., Kacprzyk J. [Fuzzy clique set determination method as an example of fuzzy temporal graph invariant]. In *Advances in Intelligent Systems and Computing*, 2021, vol. 1306, pp. 1-9. DOI: 10.1007/978-3-030-04164-9\_8.

11. Debrouvier A., Parodi E., Perazzo M., Soliani V., Vaisman A. A modeland query language for temporal graph databases. In *The VLDB Journal*, 2021, vol. 30, no. 5, pp. 825-858. DOI: 10.1007/s00778-021-00675-4.

12. Ma Yu., Yuan Ye., Liu M., Wang G., Wang Y. Graph simulation on large scale temporal graphs. In *GeoInformatica*, 2020, vol. 24, no. 1, pp. 199-220. DOI: 10.1007/s10707-019-00381-y.

13. Zhang T., Gao Yu., Qiu L., Chen L., Linghu Q., Pu Sh. Distributed time-respecting flow graph pattern matching on temporal graphs. In *World Wide Web*, 2020, vol. 23, no. 1, pp. 609-630.

14. Shchegoleva L.V., Lebedev A.A., Moskin N.D. [Methods of data analysis in the task of distinguishing folklore and author's texts]. *Questions of linguistics*. 2020, no. 2, pp. 61-74 (in Russ.). DOI: 10.31857/S0373658X0008823-4.

15. Rogov A.A., Kulakov K.A., Moskin N.D. [Software support in solving the problem of attribution of texts]. *Software Engineering*. 2019, vol. 10, no. 5, pp. 234-240 (in Russ.). DOI: 10.17587/prin.10.234-240.

## Graph-theoretical Approach to the Analysis of the Variability of Russian Fairy Tales

N. D. Moskin, DScin Engineering, Associate Professor, Petrozavodsk State University, Petrozavodsk, Russia

A. A. Lebedev, PhD in Philology, Petrozavodsk State University, Petrozavodsk, Russia

*The paper considers the graph-theoretical approach used to analyze Russian folk fairy tales. Such a formal structure-redescription of the analyzed folklore texts is based on the patterns outlined by V. Ya. Propp in a number of his works and in his famous book "Morphology of a Fairy Tale", in particular. This allows systematization and analysis of the key events occurring in texts. The vertices of the graph-theoretical model represent the fairy tale characters being classified into certain fixed groups, while the edges indicate meetings between the characters, accompanied by the actions significant for the fairy tale plot. Using the example of the text «Tereshchka» from the collection by A. M. Afanasyev, it is shown how the model is built and the numerical graph characteristics are calculated. For this purpose, the tools of the «Folklore» information system are also used. By means of this program, you can not only store text collections and their graph-theoretical models, but also create formal structures, perform their visualization, aggregation and comparison automatically. Discriminant analysis showed that according to formal parameters it is possible to identify the plot of a fairy tale about a brave fellow, rejuvenating apples and living water. The Wilks' Lambda coefficient value is 0.0275781 (less than 0.05), which indicates the good quality of the classifier. Based on this, it should be concluded that graph-based mathematical models can be used to systematize and compare the plots of Russian folk tales. This opens up prospects for studying both the fairy tale variability and identifying the most significant and typical plots for a certain geographical area or time period.*

**Keywords:** Russian fairy tale, graph-theoretic model, plot variability, discriminant analysis, information system «Folklore».

Получено: 10.06.24

### Образец цитирования

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.

### For Citation

Moskin N.D., Lebedev A.A. [A Graph-Theoretical Approach to the Analysis of Variability in Russian Fairy Tales]. *Intellectual'nye sistemy v proizvodstve*. 2024, vol. 22, no. 4, pp. 90-97. DOI: 10.22213/2410-9304-2024-4-90-97.