

КОЛИЧЕСТВЕННЫЕ ИССЛЕДОВАНИЯ ЛЕКСИКИ И СЕМАНТИКИ

QUANTITATIVE RESEARCH ON LEXIS AND SEMANTICS

УДК: 81'23; 81'37; 81'42

DOI: 10.31249/chel/2025.02.04

Комалова Л.Р.

СЕМАНТИЧЕСКАЯ КАРТА СКАЗКИ «СНЕЖНАЯ КОРОЛЕВА»: О ЧЕМ ГОВОРЯТ ЧАСТОТЫ?¹

*Институт научной информации по общественным наукам
Российской академии наук,
Московский государственный лингвистический университет,
Россия, Москва, komalova@inion.ru*

Аннотация. В статье представлены результаты анализа текста сказки «Снежная королева» датского писателя Г.Х. Андерсена, который считается одним из основоположников жанра литературной сказки. В исследовании ставится задача проверить гипотезу о возможности представить семантическую карту произведения на основе частотного анализа представленной в нем лексики. Ручная и при помощи автоматических средств обработка текста позволила определить его уровень сложности, выявить его смысловое ядро и охарактеризовать его как сказку для девочек в возрасте 9–14 лет. В познавательных целях сказка будет полезна, так как на уровне лексики представлен широкий спектр семантических категорий, описывающих макро- и микросреду, которыми окружены персонажи, познавательно-навыковые характеристики и чувственный мир действующих лиц. Можно заключить, что формальный анализ текста сказки на уровне лексики дает родителям лишь некоторые основания для осуществления выбора, оставляя не освещенными ценностный и морально-воспитательный потенциалы сказки.

¹ Исследование выполнено в рамках государственного задания ФГБУН «Институт научной информации по общественным наукам РАН» по теме «Новейшие тенденции в развитии гуманитарной мысли России и Запада: философские течения, развитие культурологии и духовная жизнь общества в цифровую эпоху».

Ключевые слова: волшебная сказка; Г.Х. Андерсен; Снежная королева; Герда; Кай; частота; семантическая карта.

Поступила: 01.09.2024

Принята к печати: 28.12.2024

Komalova L.R.

**Semantic map of “The Snow Queen” fairy tale:
what do word frequencies show us?**¹

*Institute of Scientific Information for Social Sciences
of the Russian Academy of Sciences,
Moscow State Linguistic University,
Russia, Moscow, komalova@inion.ru*

Abstract. The paper analyzes the fairy tale “The Snow Queen” by the Danish writer G.H. Andersen, who is considered one of the founders of the literary fairy tale genre. The research aims to test the hypothesis about the possibility of presenting a semantic map of the literature work based on the frequency analysis of the vocabulary presented in it. Manual and automatic text processing made it possible to determine the level of complexity of the fairy tale, identify its semantic core and characterize it as a fairy tale for girls aged 9–14. The “Snow Queen” fairy tale will be useful for educational purposes, since at the lexical level a wide range of semantic categories are presented describing the macro- and microenvironment that surrounds the characters, their cognitive characteristics and skills, and their emotional sphere. It can be concluded that a formal analysis of the fairy tale at the lexical level gives a parent only some grounds for making a choice, leaving uncovered the value and moral-educational potential of the fairy tale.

Keywords: magic fairy tale; G.H. Andersen; Snow Queen; Gerda; Kai; frequency; semantic map.

Received: 01.09.2024

Accepted: 28.12.2024

Введение

«Снежная королева» Г.Х. Андерсена относится к жанру литературной сказки. Появившись в конце XVII – начале XVIII в., литературная сказка (чаще всего волшебная) окончательно оформ-

¹ The research is carried out within the framework of the state assignment to the Federal State Budgetary Institution of Science “Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences”, project “The latest trends in the development of humanitarian thought in Russia and the West: philosophical movements, the development of Cultural Studies and the spiritual life of society in the digital era”.

ляется как жанр в XIX столетии в творчестве братьев Гримм, Э.Т.А. Гофмана, Л. Кэрролла, О. Уайльда и, конечно же, Г.Х. Андерсена. Сформировавшийся, как и его современники, под влиянием эстетики и поэтики романтизма, датский писатель сделал решительный шаг навстречу литературной сказке, став одним из основоположников этого жанра [Водясова, Уткина, 2018, с. 11].

На основании ряда словарных статей и теоретико-литературных изысканий, посвященных литературной сказке, можно отметить, что одной из важнейших характеристик этого жанра является наличие особого ирреального (волшебного, фантастического) хронотопа, взаимодействие с которым и определяет жизнь и судьбу ее героев [Брауде, 1979, с. 74; Поэтика ..., 2008, с. 235]. В свою очередь, герои литературной сказки, оказываясь по воле автора в ситуации, требующей сложного, неоднозначного решения, ставят читателя, «болеющего» за них перед нравственным выбором, который он потом так или иначе экстраполирует на свой жизненный опыт. Л.П. Водясова и Т.В. Уткина выделяя отличительные черты литературной сказки [Водясова, Уткина, 2018, с. 23], делают акцент на том, что для литературной сказки свойственна четко выраженная авторская позиция: авторское отношение, оценки, благодаря которым читатель сразу понимает, кому из героев автор симпатизирует, кому сопереживает, а к кому относится отрицательно. Таким образом, в силу выраженной художественными средствами авторской позиции литературная сказка способна непосредственно оказывать влияние на реципиента, формируя картину мира и понимание роли выбора и поступка, то есть выполняя воспитательную функцию без ее чрезмерного педалирования.

Обращение к формальному анализу текста литературной сказки как к объекту исследования связано с важным для родителей запросом современной культуры воспитания на оснащение диагностическим инструментарием определения семантики сказочного текста в автоматизированном формате, что, в свою очередь, требует разработки методики построения аналитического алгоритма. Работа с частотным словарем произведения, по нашему мнению, является базовой процедурой для автоматизированных систем. В связи с этим на данном этапе следует минимизировать ошибки методологии.

Методика исследования

Будем основываться на предположении, что родитель сознательно выбирает сказку. Предположительно, этот выбор основан на таких параметрах, как биологический пол ребенка (условно можно разделить тексты на сказки для девочек, сказки для мальчиков и сказки-униセックス), возраст ребенка (наличие в сказке множества основных персонажей и сюжетных линий или небольшое количество протагонистов, следующих одному сюжету), морально-воспитательный и ценностный потенциалы сказки (чему учит сказка, какие качества и поведение персонажей поощряются, а какие порицаются), объяснительный потенциал сказки (ситуации поиска, развития, фрустрации, конфликта и их разрешение в сказке) и др. В ходе исследования ставится задача проверки возможности осуществления подобного выбора с опорой на формальные характеристики текста сказки на уровне лексики.

В качестве материала для исследования была взята сказка Ганса Христиана Андерсена «Снежная королева»¹ на русском языке. Сказка получила 4032 положительных отзыва и 1605 отрицательных у пользователей портала «Ну-ка дети». На основе балльной оценки был построен рейтинг сказки (рис. 1).



Рис. 1. Рейтинг сказки «Снежная королева»
на основе оценок пользователей портала «Ну-ка дети»

В ходе исследования применялись методы корпусной лингвистики и дистрибутивной семантики. Обработка текста проводилась вручную с обращением к открытым онлайн-сервисам автоматической обработки естественного языка. Исследовательская процедура состояла из следующих шагов:

¹ Текст сказки размещен на портале «Ну-ка дети»: https://nukadeti.ru/skazki/andersen_snezhnaya_koroleva

- подсчет формальных характеристик текста;
- анализ сложности текста;
- вычисление индекса лексического разнообразия текста и индекса удобочитаемости текста;
- построение частотного словаря на основе словоформ;
- соотнесение словоформ с частями речи с обращением к контекстам словоупотребления;
- лемматизация словоформ;
- сопоставление лемм словаря сказки с леммами частотного словаря русской лексики;
- объединение лемм сказки в семантические категории и ранжирование семантических категорий;
- кластеризация семантических категорий и ранжирование семантических кластеров;
- реконструкция семантической карты текста сказки.

Проведение аналитических процедур на лексическом уровне обусловлено тем, что репрезентация темы как «смыслового ядра текста, конденсированного и обобщенного содержания текста» [Москальская, 1981, с. 17], реализуется на уровне слова, так как «репрезентация собственно “текстовых” микротем может осуществляться дискретно по всему тексту без четкой и однозначной локализации» [Белоусов, Ичкинеева, 2011, с. 29].

Исследуемый массив составляет 48 060 символов с пробелами, что составляет 7878 словоформ, в том числе 5595 знаменательных слов (имен существительных, глаголов, имен прилагательных, наречий, числительных, местоимений, а также имен собственных) и 2283 служебных слова (союзы, предлоги, послелоги, частицы, междометия и др.). Словарь сказки состоит из 1603 уникальных лемм.

Для определения уровня сложности текста сказки мы обратились к онлайн-сервису «Текстометр»¹, в котором «уровень сложности текста по шкале CEFR (от A1 до C2) определяется автоматически на основании результатов работы математической модели, обученной на коллекции из 800 текстов из пособий по РКИ, информация об уровне сложности которых <...> известна, и более чем 100 лингвистических признаков текста» [Лапошина, Лебедева, 2021, с. 334]. Индекс сложности учитывает структурную и лексическую сложность текста, его динамичность и описательность. Для

¹ <https://textometr.ru/>

текста сказки «Снежная королева» он составил 37, что свидетельствует о том, что текст подходит для читателей в возрасте 9–14 лет.

Индекс лексического разнообразия текста сказки высчитывался по формуле: $TRR = V / N$, в которой V обозначает количество уникальных слов, N обозначает общее количество словоформ, и составил 0,2037. Индекс удобочитаемости Флеша¹, адаптированный для русского языка, для текста сказки составил 68, что свидетельствует о том, что текст легко читать.

Далее аналитические операции проводились только со знаменательными словами (5595 единиц принято за 100%). Анализировались частоты и частотность употребления слов в тексте сказки, что позволяет выявить коммуникативно значимые языковые единицы внутри одного корпуса [Борискина, Шилихина, 2017]. Построение частотного словаря² продемонстрировало разброс рангов слов от 1 до 133 (рис. 2). Распределение частей речи в анализируемой сказке выявило значительное количество местоимений, что не характерно для русского языка (рис. 3).

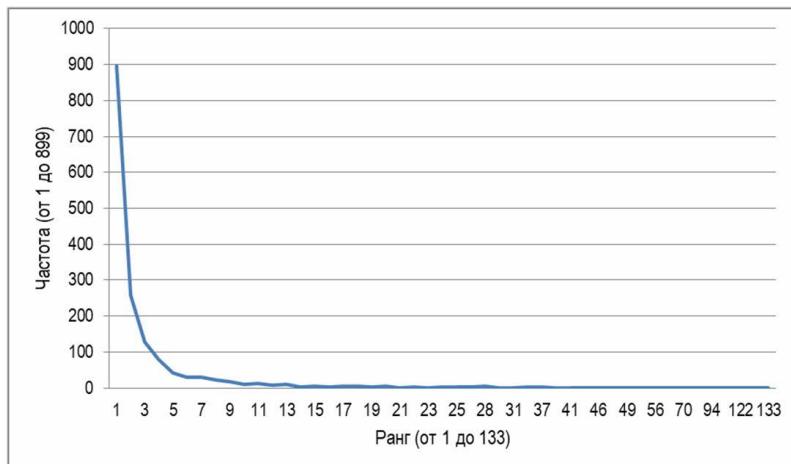


Рис 2. Частотное распределение лемм в тексте сказки «Снежная королева»

¹ <https://intexty.com/chitabelnost-teksta/>

² <https://www.i2text.com/ru/word-frequency-counter>

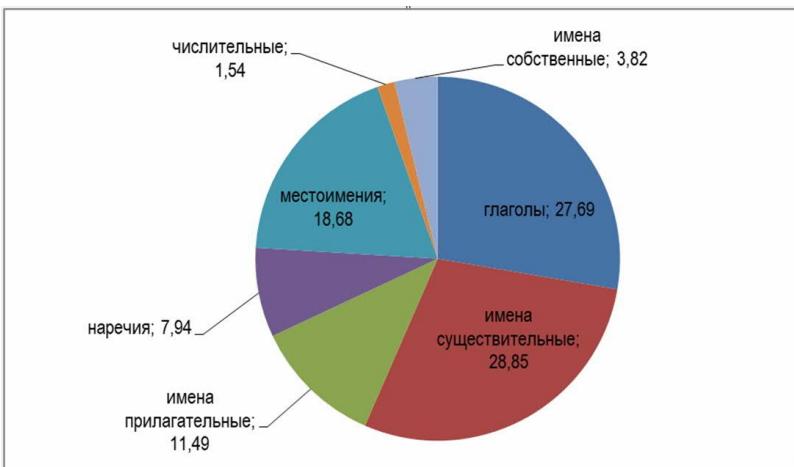


Рис. 3. Соотношение частей речи в тексте сказки «Снежная королева»
(доля в %)

Результаты исследования

Для интерпретации частоты встречаемости лемм в тексте мы опирались на следующее суждение: при нормативном размещении 1800 типографических символов на одной странице текста в тексте анализируемой сказки будет 26,7 страниц. Следовательно, леммы с частотой 26 и более встречаются в среднем на каждой странице сказки, а леммы с частотой от 13 до 25 встречаются в среднем на каждом развороте. Оговоримся, что в данном случае рассуждение строится на предположении, что текст напечатан без иллюстраций стандартным кеглем и что читатель работает с ним самостоятельно, то есть читает про себя (например, выполняя домашнее задание), а не воспринимает текст на слух. Здесь также не учтен формат чтения с экрана, который, как правило, не подразумевает разбиения на страницы. При таком чтении интерпретация частоты встречаемости лемм потребует расчета в зависимости от времени и скорости чтения.

Ограничение по пределу частоты позволило выделить семантическую доминанту текста сказки «Снежная королева» (табл. 1). Далее проводилось сопоставление лемм из сказки с теми же леммами из подкорпуса художественной литературы частотного

словаря русского языка [Ляшевская, Шаров, 2009] на основе метрики $i\text{pm}$ (*instances per million words*), которая характеризует число употреблений на миллион слов корпуса. Метрика рассчитывается по формуле $i\text{pm} = f * 10^6 / N$, в которой f обозначает абсолютную частоту леммы в корпусе, N обозначает объем корпуса лемм. Список наиболее типичных лемм для каждого функционального типа текстов (значимая лексика) составителями словаря был выделен на основе сравнения частоты лемм в данном подкорпусе текстов и в остальном корпусе. Словари значимой лексики включают по 500 лемм. Имена собственные Герда и Кай в частотном словаре русского языка отсутствуют.

Таблица 1

Сопоставление наиболее частотных лемм сказки
 «Снежная королева» и тех же лемм
 в частотном словаре русского языка

Часть речи	Лемма	Частота	$i\text{pm}$ по сказке	$i\text{pm}$ по словарю	разница между значениями $i\text{pm}$
имя собств.	Герда	117	20 911,5	—	20 911,5
имя собств.	Кай	71	12 689,9	—	12 689,9
имя собств.	(Снежная) королева	26	4647,0	21,3	4625,7
имя собств. + сущ.	(северный) олень	24	4289,5	15,6*	4273,9
имя собств.	(маленькая / молодая) разбойница	20	3753,4	0,6*	3574
имя сущ.	ворон	27	4825,7	5,3*	4825,7
имя сущ.	ворона	12	2144,8	12,6	2122,2
имя сущ.	глаз	22	3932,1	1580,3	2351,8
имя сущ.	голова	15	2681,0	1326,7	1354,3
имя сущ.	дверь	13	2323,5	892,3	1431,2
имя сущ.	девочка	46	8221,6	297,6	7841,8
имя сущ.	день	13	2323,5	<i>1281,4</i>	1042,1
имя сущ.	дом	17	3038,4	<i>1174,6</i>	1863,8
имя сущ.	история	13	2323,5	<i>261,1</i>	2062,4
имя сущ.	окно	18	3217,2	517,1	2651,3
имя сущ.	осколок	17	3038,4	<i>32,4</i>	3006,0
имя сущ.	принцесса	23	4110,8	<i>21,9</i>	4088,9

*Семантическая карта сказки «Снежная королева»:
о чём говорят частоты?*

Часть речи	Лемма	Частота	ірм по сказке	ірм по словарю	разница между значениями ірм
имя сущ.	роза	19	3395,9	49,9	3346,0
имя сущ.	рука	18	3217,2	2137,9	1079,3
имя сущ.	сад	13	2323,5	<i>113,8</i>	2209,7
имя сущ.	сани	25	4468,3	<i>20,6</i>	4447,7
имя сущ.	свет	18	3217,2	<i>515,2</i>	2702,0
имя сущ.	сердце	17	3038,4	<i>318,1</i>	2720,3
имя сущ.	слово	17	3038,4	<i>992,2</i>	2046,2
имя сущ.	старуха	20	3574,6	139,6	3435,0
имя сущ.	цветок	17	3038,4	<i>108,9</i>	2929,5
глагол	быть	133	23 771,2	<i>12 887,1</i>	10 884,1
глагол	видеть	16	2859,7	1179,9	1679,8
глагол	говорить	20	3574,6	<i>2178,4</i>	1396,2
глагол	знать	22	3932,1	2537,5	1394,6
глагол	мочь	24	4289,5	<i>2862,4</i>	1427,1
глагол	отвечать / ответить	13	2323,5	628,4	1695,1
глагол	рассказывать / рассказать	25	4468,3	<i>523,4</i>	3944,9
глагол	сидеть	28	5004,5	922,6	4081,9
глагол	сказать	61	10902,6	4115,7	6786,9
глагол	спрашивать / спросить	20	3574,6	1172,5	2402,1
глагол	стоять	15	2681,0	756,4	1924,6
имя прилаг.	белый	14	2502,2	<i>495,1</i>	-2997,3
имя прилаг.	больший	13	2323,5	<i>78,1</i>	2245,4
имя прилаг.	большой	27	4825,7	<i>811,1</i>	4014,6
имя прилаг.	другой	17	3038,4	<i>1529,3</i>	1509,1
имя прилаг.	красный	13	2323,5	<i>313,7</i>	2009,8
наречие	где	19	3395,9	<i>1317,9</i>	2078,0
наречие	опять	16	2859,7	576,7	2283,0
наречие	совсем	18	3217,2	668,9	2548,3
наречие	там	13	2323,5	<i>1456</i>	867,5
наречие	теперь	13	2323,5	1267,7	1055,8
наречие	тут	20	3574,6	1409,2	2165,4

Часть речи	Лемма	Частота	іртм по сказке	іртм по словарю	разница между значениями іртм
местоим.	все	70	12 511,2	4028,9	8482,3
местоим.	им / ним / ними / них	41	7328,0	—	7328,0
местоим.	их	31	5540,7	722,9	4817,8
местоим.	каждый	15	2681,0	758,4	1922,6
местоим.	какой	19	3395,9	1401,7	1994,2
местоим.	который	25	4468,3	3102,6	1365,7
местоим.	мой / меня / мне	50	8936,6	2136,7	6799,9
местоим.	мы	14	2502,2	4311,8	-1809,6
местоим.	он / его	139	24 843,6	20 371,8	4471,8
местоим.	она / её	199	35 567,5	10 502,7	25064,8
местоим.	они	56	10 008,9	5502,4	4506,5
местоим.	сам / сама / само / сами	28	5004,5	2337,9	2666,6
местоим.	свой / своя / свое / свои	48	8579,1	3906,8	4672,3
местоим.	такой / такая / такое / такие	29	5183,2	2635,8	2547,4
местоим.	твой / твоя / твоё / твои	33	5898,1	645,1	5253,0
местоим.	тот / та / то / те	49	8757,8	2687,6	6070,2
местоим.	ты	37	6613,0	6153,6	459,4
местоим.	эта / этот / эти	28	5004,5	10 814,7	-5810,2
местоим.	я	39	6970,5	17 800,2	-10 829,7
числิต.	один / одна / одни	28	5004,5	588,6	4415,9
числิต.	раз	15	2681,0	1325,6	1355,4

Легенда: полужирным шрифтом выделены леммы, относящиеся к словарю значимой лексики художественной литературы; курсивом – леммы, относящиеся к словарю лексики художественной литературы; знаком «*» отмечены леммы, которые отсутствуют в подкорпусе художественной литературы, но присутствуют в алфавитном списке словаря; знаком «↔» отмечены леммы, которые отсутствуют в словаре; значения іртм полужирным шрифтом и курсивом приведены по подкорпусу художественной литературы.

Частота коллокации «Снежная королева» составляет 26; лемма «королева» вне коллокации встречается два раза и не относится к персонажу «Снежная королева», лемма «снежный» вне данной коллокации встречается в тексте сказки пять раз. Частота коллокации «северный олень» равна девяти, лемма «олень» встречается в тексте 15 раз, лемма «северный» вне привязки к коллокации встречается в тексте пять раз. Частота коллокации «маленькая разбойница» составляет 18,

«молодая разбойница» – два, лемма «маленький» вне привязки к коллокации встречается в тексте девять раз.

Состав персонажей (главных действующих лиц, к которым относятся герой и протагонисты) представлен людьми (*Герда, Кай, бабушка, старуха, принцесса, принц, разбойница, лапландка и финка*), антропоморфными животными (*ворон, ворона, олень, голуби* умеют говорить, ведут себя как люди, сочувствуют, мыслят, сохраняя при этом физический облик животных), природными явлениями (*говорящая река*) и сказочными персонажами (*Снежная королева и тролль*). В основном действующие лица – это персонажи женского пола, частотнее представленные в тексте сказки (346 упоминаний женских персонажей по сравнению с 170 упоминаниями мужских персонажей).

В тексте сказки почти отсутствуют абстрактные имена существительные, которые описывают характеристики героев и дают оценку их действиям на ценностном уровне. В силу того, что сказка подходит для чтения детям в возрасте 9–14 лет, а абстрактное мышление начинает активно развиваться с 8–9 лет, отсутствие абстрактных понятий следует отдельно разобрать. По нашему мнению, они представлены имплицитно, например абстрактное понятие «любовь» в тексте сказки вербализуется частотным употреблением им. сущ. *сердце*, категория «прекрасное / красота» передается им. сущ. *роза и цветок*, категория «рациональное познание» эксплицируется им. сущ. *глаз* и глаголами *знать* и *видеть*.

По табл. 1 видно, что самый частотный глагол *быть* относится к категории «состояние», большинство частотных глаголов раскрывают семантику речепорождения (*сказать, рассказать, спросить, отвечать*), что логично, так как в процессе развития сюжета персонажи много общаются друг с другом. При том что по сюжету сказки Герда предпринимает путешествие, то есть пребывает в постоянном движении, глаголы движения не вошли в перечень лемм с наибольшими частотами.

Наречия *тут, где, там* и им. сущ. *дом, сад* обозначают пространственные координаты сказки. Далее будет показано, что в зависимости от истории эти координаты будут меняться. Частотные им. сущ. *дом, сад* ассоциативно связаны с ценностными ориентациями «семья / развитие / забота / безопасность» (см., например: [Двуязычный ..., 2020]).

По данным табл. 1 видно, что большинство частотных лемм входит в подкорпус лексики художественной литературы, при этом 22 из 69 лемм относятся к числу значимой лексики в этом корпусе. Этот факт подтверждает статус литературной (авторской) сказки как художественного произведения.

Обратимся к данным табл. 1 для выявления ключевых слов для сказки через сопоставление двух текстовых массивов. «Анализ ключевых слов позволяет обозначить содержательную направленность текста или корпуса, а также определить основные идеи и стилистические предпочтения автора» [Измаилян, 2022, с. 3]. По определению М. Скотта ключевые это такие слова, которые обладают необычно или неожиданно высокой частотностью в сравнении с некой нормой. О.Г. Горина подчеркивает, что «ключевые слова не есть самые частотные слова в тексте – они скорее самые “неожиданно частотные”» [Горина, 2011, с. 31].

Обозначим частотный словарь русского языка в качестве нормативного корпуса. В качестве исследуемого корпуса выступает текст сказки «Снежная королева». Как правило, ключевыми словами являются имена существительные, поэтому будем анализировать леммы этой части речи. Исключим из анализа леммы, которые соотносятся примерно в равной мере (разница меньше 1000), а также леммы, для которых показатели *iрт* по словарю больше показателя *iрт* по сказке (отрицательные разницы), так как они не отражают специфики семантики текста сказки.

Если выстроить список лемм (исключая им. собственные и коллокации) на основе разниц в значениях метрики *iрт* в порядке убывания от большего значения к меньшему, то разброс значений будет варьироваться от 7000 до 1000. Уникальными для сказки «Снежная королева» являются им. собственные *Герда*, *Кай* и коллокации, ставшие в сказке почти им. собственными *Снежная королева*, *маленькая разбойница*, *северный олень*. Необычно частотными можно также назвать леммы *ворон* и *осколок*.

В связи с тем, что текст сказки организован по принципу наименования сюжетных локаций и состоит из семи композиционно завершенных историй, построим частотные списки для каждой истории по-отдельности (табл. 2). Для этого воспользуемся семантическим онлайн-анализатором текста¹.

¹ <https://advego.com/text/seo>

*Семантическая карта сказки «Снежная королева»:
о чём говорят частоты?*

Таблица 2

Распределение частотных лемм по историям

№ истории	Персонажи (частота)	Действия персонажей (частота)	Среда и обстановка (частота)
1	тролль (4)	делаться (2), попадать (2)	осколок (9), зеркало (7), очки (2), (сторона) свет (2), сердце (2)
2	Кай (19), мальчик (11), Герда (9), дети (6), бабушка (5), девочка (5), (Снежная) королева (5)	знать (4), (по)целовать (4), передразнивать (3), попасть (3)	сан(к)и (17), окно (10), роза (7), ящик (7), снег (6), площадь (5), сердце (5), стекло (5), шуба (5), желоб (4), лед (4), зима (3), картинка (3), куст (3), лето (3), осколок (3), садик (3), снежинка (3)
3	Герда (31), старушка (17), Кай (16), девочка (14), бабушка (3), братец (3)	знать (6), умереть (5), вернуться (4), искать (4), отвечать (4), бросить (3), верить (3), горевать (3), колдовать (3), уметь (3)	цветок (11), роза (9), башмачок (7), лодка (7), река (7), сад (7), берег (7), земля (6), дом (5), солнце (5), дверь (4), золото (4), цветник (4), двор (3), клюка (3), куст (3), одуванчик (3)
4	Герда (22), принцесса (21), (лесной) ворон (20), Кай (14), девочка (12), принц (8), ручная ворона (8), невеста (6), слуга (5), служанка (5), лакей (4), жених (3)	говорить (7), рассказать (7), знать (6), (по)думать (6), (у)видеть (6), войти (5), мочь (4), сделать (4), прийти (4), явиться (4), жениться (3), скрипеть (3), дать (3)	дворец (8), карета (6), сон (5), дверь (4), золото (4), ворота (5), газета (3), гвардия (3), дама (3), двор (3), должность (3), лампа (3), лестница (3), лошадь (3), сапог (3), (белый) свет (3), серебро (3), спальня (3)
5	Герда (21), маленькая разбойница (17), (северный) олень (11), девочка (9), голуби (8), Кай (5), вороны (3), матя (5), разбойник (4), Снежная королева (4)	убить (6), знать (4), рассказать (4), спать (4), закричать (3), отвечать (3), привязать (3)	нож (7), карета (5), Лапландия (4), лес (4), огонь (3), стена (3)
6	Герда (16), (северный) олень (11), финка (10), Снежная королева (8), девочка (6), лапландка (5)	написать (3), приняться (3), рассказать (3), сделать (3), читать (3)	сила (5), история (3), треска (4), небо (3), руканица (3), сердце (3), хлопья (3), Лапландия (2), Финляндия (2)
7	Герда (18), Кай (17), Снежная королева (8), молодая / маленькая разбойница (3)	узнать (6), сложить (7), поцеловать (3)	чертог (5), ветер (4), лыдина (4), (сторона) свет (4), фигура (4), зеркало (3), край (3), осколок (3), сердце (3), слеза (3)

Первая история посредством персонажа *тролль* вводит *зеркало* как механизм преобразования человеческого сознания и создает условия для начала сюжета сказки (леммы *осколок*, *очки*, *попадать*, *делаться*, проявляясь в завершающей истории, складываются в циклическую конструкцию). В последующих пяти историях перед читателем предстают различные персонажи, исполняющие различные

действия в привязке к определенной локации. Завершающая история кратко описывает обратный путь Герды и Кая.

Героем сказки, бесспорно, является Герда. Это подтверждается не только частотой употребления леммы *Герда*, но и ее частотным присутствием в каждой из шести историй. Герду частотно называют *девочкой*. Вторая история строится вокруг Кая (в совокупности леммы *Кай* и *мальчик* составляют 30 по сравнению с 9 леммами *Герда* в этой истории). В завершающей истории соотношение частоты лемм *Герда* и *Кай* составляет 18 к 17. Таким образом автор возвращает Кая в амплуа протагониста, ставя его в один ряд с героиней.

Действие перемещается с городской *площади* и из импровизированного домашнего *садика* в заколдованный *сад*, в котором располагаются *дом* и *цветник*, затем – во *дворец*, после – в *лес*, в *Финляндию*, *Лапландию*, *чертоги Снежной королевы* и, наконец, обратно в *родной город*. В каждой из локаций Герду встречают персонажи-проводники: в родном городе – это *бабушка*, в заколдованным саду – *старушка*, во дворце – лесной *ворон* и ручная *ворона*, *принцесса* и *принц*, северный *олень*, в лесу – *маленькая разбойница* и северный *олень*, в других странах – *финка*, *лапландка* и *северный олень*.

Интересен и способ передвижения Герды из локации в локацию: из родного города она пешком в *башмачках* направляется к реке, затем на *лодке* – к заколдованныму *саду*, снова пешком во *дворец*, откуда на *карете*, запряженной *лошадьми*, она въезжает в темный *лес*, далее верхом на северном *олене* она направляется в *Финляндию*, *Лапландию* и в *чертоги Снежной королевы*. Эти способы передвижения могут свидетельствовать о географических особенностях локаций, их удаленности друг от друга, о посильности перемещения между ними. Кроме того, они указывают на социальный статус этих мест: межличностная коммуникация со старушкой возможна своими силами, социальная иерархия дворца требует соответствующей атрибутики, лес угрожает своей необузданной природной силой в совокупности с обитающими в нем асоциальными личностями, погружение в пустынное и неизведанное пространство далеких холодных стран требует сопровождения со стороны необычных проводников.

Труднее всего интерпретации поддаются выделенные частотные глаголы. Лемма *знать* является частотной для пяти историй, вероятно, это указывает на то, что автор делает акцент на рацио-

нальное познание мира. Эмоциональная сторона познания представлена леммами (*по*)целовать, горевать и переодевать. Наличие частотной леммы убить в истории про лес еще раз свидетельствует об опасности, потенциально ожидаемой от этой локации.

Вернемся к массиву знаменательных слов (5595 словоформ). Для построения семантической карты сказки исключим из него местоимения. Итоговый массив составит: $5595 - 1035 = 4560$ лемм (примем его за 100%). Из него удалось структурировать в семантические категории 3842 леммы (что составило 57,88% от всех словоформ в тексте, или 68,67% от общего числа всех знаменательных слов в тексте, или 84,25% от всех сгруппированных лемм). Совокупность семантических категорий позволяет представить семантическую карту текста сказки «Снежная королева». Моделирование по предложенной методике успешно реализовывалось нами ранее на разноплановом материале (см., например: [Потапова, Комалова, 2015, 2017; Комалова, 2023, 2024]). Обобщение категорий по семантической близости позволило выделить 13 кластеров (табл. 3).

Преобладание кластеров «макросреда» (например, *солнце*, *земля*, *окно*, *зал*, *тогда*, *вечно*, *здесь*, *вперед*, *ледяной*, *замерший*, *сверкающий*, *ярко*, *стекло* и др. – всего 18,97%) и «персонажи и их роли» (друг, *дети*, *мать*, *гости* и др. – 10,28%) подтверждает характеристики литературной сказки, выделенные в [Водясова, Уткина, 2018]. Эти кластеры представляют статические элементы истории, они дополняются описанием динамики ее развития (*выходить*, *бежать*, *кивать*, *вставать*, *останавливаться*, *пуститься* и др. – 8,98%) и познавательно-навыковыми характеристиками персонажей (*закричать*, *воскликнуть*, *думать*, *забыть*, *смотреть*, *увидеть*, *слушать*, *треск*, *книжка*, *сказка*, *буква*, *написанный* и др. – 8,79%). Отметим, что в тексте сказки сопоставимо представлена семантика природы (6,63%) и семантика рукотворной среды (5,17%), времени (2,65%) и пространства (2,32%).

Выделение автономной категории «холод» (1,04%) обусловлено не только частотностью лемм, но и сюжетообразующей функцией этой категории: примечательно, что не было выявлено оснований для выделения категории «тепло», которая могла бы составить диахотическую пару «холоду», символически обозначая противоборство зла и добра. Вероятно, в сказке добро олицетворяется через категорию «свет» (0,71%).

Таблица 3

**Семантические категории и кластеры,
сконструированные на основе текста
сказки «Снежная королева»**

Семантические кластеры (доля в %)	Семантические категории	Абсолютное число лемм	Доля (в %)
макросреда (18,97%)	природа	371	6,63
	строения, интерьер	289	5,17
	время	148	2,65
	пространство	130	2,32
	холод	58	1,04
	свет	40	0,71
	фрагмент стекла	25	0,45
персонажи и их роли (10,28%)	персонажи и их роли	575	10,28
динамика (8,98%)	(средства) передвижени(е/я)	259	4,63
	положение, движение тела	157	2,81
	начало / завершение действия	86	1,54
познавательно- навыковые характеристики персонажей (8,79%)	коммуникация	206	3,68
	когниции	141	2,52
	зрение	67	1,2
	слух, звуки	50	0,89
	сообщение	20	0,36
	письменность	7	0,14
микросреда (6,4%)	части тела	159	2,84
	индивидуально-личностные качества персонажей	76	1,36
	одежда	63	1,13
	питание, еда	43	0,77
	возраст	17	0,3
состояния (4,08%)	состояния	228	4,08
размер, величина, мера, интенсивность (3,36%)	размер, величина, мера, интен- сивность	188	3,36
эмоции, чувства, настроение, переживания (2,3%)	эмоции, чувства, настроение, переживания	131	2,3
числа (1,86%)	числа	104	1,86
модальность (1,3%)	модальность	73	1,3
цвет (1,05%)	цвет	59	1,05
качества объекта (1,02%)	материалы, форма, текстура	33	0,59
	богатство	24	0,43
часть и целое (0,28%)	часть и целое	15	0,28

В кластере «познавательно-навыковые характеристики персонажей» преобладает речевая коммуникация (3,68%) и когниции (2,52%).

Кластер «микросреда» в основном описывает физическое тело (например, *спина, лицо, шея* и др. – 2,84%) и индивидуально-личностные качества (*милый, умный, добрый* и др. – 1,36%) персонажей. По сравнению с познавательно-навыковыми характеристиками персонажей их чувственный мир описан мало (например, *плакать, обрадоваться, поцеловать* и др. – 2,3%). Половина кластера «состояния» (4,08%) представлена глаголом *быть*, частотны также леммы *стать, умереть, спать, жить* и др. В кластере «качества объекта» (1,02%) особо выделяется подмножество, объединяющее леммы *золото* и *золотой*, которые ассоциируются с богатством. Преобладающей модальностью (1,3%) в тексте выступает возможность, например, *мочь, можно*. В тексте представлены также абстрактные понятия, которые формируют кластеры «размер, величина, мера, интенсивность» (например, *огромный, много, точно, бедный, крепко, высоко* и др. – 3,36%), «числа» (*два, первый* и др. – 1,86%), «цвет» (*чёрный, зелёный, голубой* и др. – 1,05%) и «часть и целое» (*кусок, краешек* и др. – 0,28%).

Заключение

На основе полученных в ходе исследования результатов можно сформулировать следующие выводы.

Формальные показатели сложности, лексического разнообразия и удобочитаемости текста свидетельствуют о том, что сказка ориентирована на детей 9–14 лет. В силу преобладания женских персонажей в позициях героя (Герда) и protagonистов (Снежная королева, бабушка, старуха, принцесса, разбойница), это сказка для девочек. На основе анализа лексики затруднительно составить представление о том, чему учит эта сказка, однако в познавательных целях она будет полезна, так как на уровне лексики представлен широкий спектр семантических категорий, описывающих макро- и микросреду, которыми окружен человек, познавательно-навыковые характеристики и чувственный мир действующих лиц. В этом смысле формальный анализ текста сказки на уровне лексики дает родителям некоторые основания для осуществления выбора сказки.

Для дальнейшей работы над поиском параметров для автоматизированного анализа текста сказок представляется необходимым по примеру формирования корпуса в зависимости от функционального стиля формировать корпус частотной лексики и по жанровому признаку, тогда появится возможность сопоставлять

тексты на одном основании, что приведет к пересмотру значимости частотных лемм. Предполагаем, что, например, глаголы с семантикой речепорождения станут жанровыми маркерами и уйдут из частотного списка как ключевые слова.

Список литературы

- Белоусов К.И., Ичкинеева Д.А. Графосемантическое моделирование структурного синтеза тематического пространства текста // Филология и человек. – 2011. – № 1. – С. 26–37.
- Борискина О.О., Шилихина К.М. Корпусные исследования политического дискурса в лингвистике // Политическая наука. – 2017. – № 2. – С. 30–53.
- Брауде Л.Ю. Скандинавская литературная сказка. – Москва : Наука, 1979. – 206 с.
- Водясова Л.П., Уткина Т.В. Литературная сказка Мордовии: инвариантность жанров, речевые параметры, педагогический потенциал. – Саранск : Мордовский государственный педагогический институт, 2018. – 105 с.
- Горина О.Г. Применение методов корпусной лингвистики для определения контекстно-специфических слов и коллокаций // Вестник ЛГУ им. А.С. Пушкина. – 2011. – № 3. – С. 27–36.
- Двуязычный ассоциативный словарь базовых ценностей / Пиццальникова В.А., Адамова З.Г., Кошелева Ю.П., Панарина Н.С., Хлопова А.И., Пэй Ц., Яо Ч. ; под общ. ред. В.А. Пиццальниковой. – Москва : Спутник+, 2020. – 261 с.
- Комалова Л.Р. Возможна ли экстернализация киноопыта посредством комментирующего дискурса (на основе отзывов и квазирецензий пользователей сайтов «Афиша» и «Кинопоиск»)? // Человек: образ и сущность. Гуманитарные аспекты. – 2024. – № 2(58). – С. 72–93. DOI: 10.31249/chel/2024.02.04
- Комалова Л.Р., Зуева В.А. Метод ассоциативного эксперимента применительно к становящимся практикам (на материале понятия «Интернет поведения») // Человек: образ и сущность. Гуманитарные аспекты. – 2023. – № 4(56). – С. 118–143. DOI: 10.31249/chel/2023.03.07
- Краюшкина Т.В. Физиология и психология персонажей русских народных волшебных сказок. – Владивосток : Дальнаука, 2011. – 258 с.
- Лапошина А.Н., Лебедева М.Ю. Текстометр: онлайн-инструмент определения уровня сложности текста по русскому языку как иностранному // Русистика. – 2021. – Т. 19, № 3. – С. 331–345. DOI: <http://dx.doi.org/10.22363/2618-8163-2021-19-3-331-345>
- Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). – Москва : Азбуковник, 2009. – URL: <http://dict.ruslang.ru/freq.php?>
- Москальская О.И. Грамматика текста. – Москва : Высшая школа, 1981. – 183 с.
- Потапова Р.К., Комалова Л.Р. Вербальная структура коммуникативного акта агрессии : тематический толковый словарь. Вып. 1. – Москва : ИНИОН РАН, 2015. – 146 с.

Потапова Р.К., Комалова Л.Р. Вербальная структура коммуникативного акта агрессии : тематический толковый словарь. Вып. 2. – Москва : ИНИОН РАН, 2017. – 124 с.

Поэтика: словарь актуальных терминов и понятий / гл. науч. ред. Н.Д. Тамарченко. – Москва : Интрагда, 2008. – 358 с.

Романова Л.Г. Сравнительный анализ степени адекватности перевода посредством моделирования тематического пространства текста // Система ценностей современного общества. – 2010. – № 10-1. – С. 215–219.

References

- Belousov, C.I., Ichkineeva, D.A. (2011). Graphosemantic modelling of structural synthesis of thematic space of the text. *Filologija i chelovek*, 1, 26–37.
- Boriskina, O.O., Shilikhina, K.M. (2017). Corpus methods in the political discussion in linguistics. *Political science*, 2, 30–53.
- Braude, L.Ju. (1979). *Skandinavskaja literaturnaja skazka*. Moscow: Nauka.
- Vodjasova, L.P., Utkina, T.V. (2018). *Literaturnaja skazka Mordovii: invariantnost' zhanrov, rechevye parametry, pedagogicheskij potentsial*. Saransk: Mordovskij gosudarstvennyj pedagogicheskij institut.
- Gorina, O.G. (2011). Implementation of corpus linguistic techniques to identify a text or genre-specific key words and collocations. *Vestnik LGU imeni A.S. Pushkina*, 3, 27–36.
- Pishchalnikova, V.A., Adamova, Z.G., Kosheleva, Ju.P., Panarina, N.S., Khlopova, A.I., Pei, C., Yao, Ch. (2020). *Dvujazychnyj assotsiativnyj slovar' bazovyh tsennostej*. Moscow: Sputnik+.
- Komalova, L.R. (2024). Is it possible to externalize filmic experience through commenting discourse (based on “Afisha” and “Kinopoisk” users’ reviews)? *Human Being: Image and Essence. Humanitarian Aspects*, 2(58), 72–93. DOI: 10.31249/chel/2024.02.04
- Komalova, L.R., Zueva, A.V. (2023). Method of associative experiment applied to emerging practices (research on Internet of Behavior concept). *Human Being: Image and Essence. Humanitarian Aspects*, 4(56), 118–143. DOI: 10.31249/chel/2023.03.07
- Krayushkina, T.V. (2011). *Physiology and psychology of characters in magic Russian folk tales*. Vladivostok: Dal'nauka.
- Laposhina, A.N., Lebedeva, M.Yu. (2021). Textometr: an online tool for automated complexity level assessment of texts for Russian language learners. *Russian Language Studies*, 19(3), 331–345. DOI: <http://dx.doi.org/10.22363/2618-8163-2021-19-3-331-345>
- Ljashevskaja, O.N., Sharov, S.A. (2009). *Chastotnyj slovar' sovremenennogo russkogo jazyka (na materialah Natsional'nogo korpusa russkogo jazyka)*. Moscow: Azbukovnik. Retrieved from: <http://dict.ruslang.ru/freq.php>
- Moskal'skaja, O.I. (1981). *Grammatika teksta*. Moscow: Vysshaja shkola.
- Potapova, R.K., Komalova, L.R. (2015). *Verbal structure of communicative act of aggression: thematic explanatory dictionary. Issue 1*. Moscow: INION RAN.
- Potapova, R.K., Komalova, L.R. (2017). *Verbal structure of communicative act of aggression: thematic explanatory dictionary. Issue 2*. Moscow: INION RAN.

Tamarchenko, N.D. et al. (2008). *Poetics: a dictionary of current terms and concepts*. Moscow: Intrada.

Romanova, L.G. (2010). Sravnitelnyj analiz stepeni adekvatnosti perevoda posredstvom modelirovaniya tematicheskogo prostranstva teksta. *Sistema tsennostej sovremennoogo obshhestva*, 10–1, 215–219.

Об авторе

Комалова Лилия Ряшиитовна – доктор филологических наук, доцент, руководитель и ведущий научный сотрудник Центра эмерджентных практик, Институт научной информации по общественным наукам Российской академии наук, профессор кафедры прикладной и экспериментальной лингвистики, Московский государственный лингвистический университет, Россия, Москва, komalova@inion.ru

About the author

Komalova Liliya Ryashitovna – Doctor Habil. (Linguistics), Associate Professor, Head and Leading Research Fellow of the Centre for Emerging Practices, Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences, Professor at the Department of Applied and Experimental Linguistics, Moscow State Linguistic University, Russia, Moscow, komalova@inion.ru