

Завьялова И.С., Шерстинова Т.Ю.
О МОРФОЛОГИЧЕСКИХ РАЗЛИЧИЯХ В ТЕКСТАХ
РУССКОЙ МАЛОЙ ПРОЗЫ 1900–1930 гг.^{1,2}

*Национальный исследовательский университет
«Высшая школа экономики», Россия, Санкт-Петербург,
izavyalova@edu.hse.ru, tsherstinova@hse.ru*

Аннотация. Значимые социальные изменения с неизбежностью находят свое отражение в речевых практиках. Не является исключением и язык художественной литературы, оперативно адаптирующийся к новым реалиям и условиям социальной жизни. Описанное в статье исследование выполнено на материале 15 рассказов 1900–1930 гг. разных русских писателей и позволяет оценить морфологическую вариативность стиля авторов. В статье также обсуждаются некоторые общие закономерности, выявленные на морфологическом уровне (распределение частей речи и отдельных грамматических форм) для исследуемой выборки художественных текстов в диахронии.

Ключевые слова: язык русской прозы; русский рассказ; диахронические изменения; части речи; грамматическое варьирование; количественный анализ.

Поступила: 02.12.2021

Принята к печати: 24.01.2022

¹ Публикация подготовлена в результате проведения исследования по проекту № 21-04-053 «Методы искусственного интеллекта для филологических исследований» в рамках Программы «Научный фонд Национального исследовательского университета “Высшая школа экономики” (НИУ ВШЭ)» в 2021 г.

² ©Завьялова И.С., Шерстинова Т.Ю., 2022

Zavyalova I.S., Sherstinova T.Yu.
**Morphological diversity in Russian short prose
of 1900–1930s^{1,2}**

*National Research University Higher School of Economics,
Russia, St. Petersburg, izavyalova@edu.hse.ru, tsherstinova@hse.ru*

Abstract. Significant social changes are inevitably reflected in speech practices. The language of fiction, which quickly adapts to new realities and conditions of social life, is no exception. The pilot study described in the article was carried out on the basis of 15 short stories by different Russian writers written of the 1900–1930s and reflects the variability of the authors' style at the morphological level. The article also discusses some general patterns identified at the morphological level (distribution of parts of speech and individual grammatical forms) for the studied sample of literary texts in diachrony.

Keywords: language of Russian prose; Russian short stories; diachronic changes; parts of speech; grammatical variation; quantitative analysis.

Received: 02.12.2021

Accepted: 24.01.2022

Введение

Язык – явление социальное [Леонтьев, 1976], поэтому изменения, происходящие в обществе, неизбежно находят отражение в языке. Наиболее очевидно трансформации языка проявляются вследствие значимых социальных преобразований, наступающих в результате острых социальных конфликтов (революций, гражданских войн) [Селищев, 1925]. При этом наиболее подвержен изменениям лексический уровень, так как именно он отвечает за введение новых слов, отражающих новые реалии наступившей эпохи, за пересмотр семантического наполнения определенной части традиционного пласта лексики, за постепенное «списание» лексики, которая еще совсем недавно была в активном употреблении, но в условиях «новой жизни» уже не является актуальной или же ее

¹ The publication was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2021 (grant N 21-04-053 “Artificial Intelligence Methods in Literature and Language Studies”).

² ©Zavyalova I.S., Sherstinova T.Yu., 2022

использование не приветствуется новыми властями или новым социальным укладом.

Изменения, происходящие на морфологическом уровне, кажутся не столь очевидными, поскольку «ассортимент» частей речи и грамматических форм представляет собой значительно более устойчивую систему. Однако эволюция языковых стилей, которую можно наблюдать в переломные моменты истории, дает основание предполагать, что и на уровне грамматических единиц (в частности, дистрибуции частей речи или отдельных грамматических форм) можно говорить об определенных сдвигах и даже тенденциях [Frequency Word Lists ... , 2020]. Иными словами, можно выдвинуть гипотезу о том, что доля употребления тех или иных частей речи, а также грамматических форм не является некой константой, а меняется синхронно с развитием общества и его языковых предпочтений. Для проверки этой гипотезы необходимы количественные (статистические) методы анализа, выполненные на представительном объеме речевого материала.

В центре внимания нашей работы находится язык малой русской прозы (русского рассказа) одного из наиболее богатых и интересных с литературных позиций периодов, одновременно являющегося и одним из самых драматичных в истории нашей страны. Речь идет о первых трех десятилетиях XX века, которые обернулись для России чередой острых социальных потрясений и войн (русско-японская война, революция 1905–1907 гг., Первая мировая война, Февральская и Октябрьская революции 1917 г., Гражданская война, распад Российской империи, становление молодого Советского государства). Произошедшие в этот период социальные катаклизмы не только кардинально изменили социальный строй Российской империи, но и имели важные последствия для всей мировой истории. Мирные годы этого периода также были насыщены значимыми для общества событиями: в дореволюционной России это, прежде всего, столыпинские реформы и активный рост промышленного производства перед Первой мировой, а потом в советское время – НЭП, индустриализация, первые пятилетки, коллективизация.

Изменение традиционного жизненного уклада, трансформация общественных ценностей, появление новых реалий молодого советского государства нашли свое отражение в ярко выраженных

языковых изменениях, прежде всего на лексическом уровне, на которые ученые и общественные деятели обратили внимание еще в 1920-е годы [Баранников, 1919; Mazon, 1920; Jakobson, 1921; Карцевский, 1923; Поливанов, 1927; Селищев, 1925; Селищев, 1928]. Представленное в статье исследование имеет своей целью количественно оценить различия на морфологическом уровне, наблюдаемые в русских рассказах разных авторов в рассматриваемый период.

Материал

Исследование проведено на материале Корпуса русского рассказа 1900–1930 гг., специально разрабатываемого для изучения диахронических изменений языка и стиля малой русской прозы [Sherstinova, Martynenko, 2019]. Теоретической основой для создания корпуса стали стилеметрические представления, разрабатываемые в русле Санкт-Петербургской стилеметрической школы под руководством профессора Г.Я. Мартыненко [Мартыненко, 1988; Мартыненко, 2019]. Особенностью создаваемого электронного ресурса является включение в него художественных текстов большого числа русских прозаиков рассматриваемой эпохи – не только известных, но малоизвестных и даже фактически забытых [Методологические проблемы ... , 2018; О принципах создания ... , 2018]. Тем самым повышается статистическая объективность проводимых на его основе научных работ.

Корпус состоит из нескольких подкорпусов, наиболее детально аннотированным из которых является Корпус-300 общим объемом чуть более 1 млн. словоупотреблений, в который входят 310 рассказов, представляющих творчество 300 русских писателей. Принципы отбора текстов для аннотированного Корпуса-300 подробно описаны в [Sherstinova, Martynenko, 2019]. Входящие в него рассказы разделены по дате их написания на три хронологических периода, которые соотносятся со значимыми для страны историческими эпохами [О принципах создания ... , 2018]:

1) начало XX века и предреволюционные годы (1900–1913 гг.);

2) Первая мировая война и революционные годы – Февральская и Октябрьская революции и Гражданская война (1914–1922 гг.);

3) постреволюционные годы – с окончания Гражданской войны до 1930 г.

Для всех текстов аннотированной выборки были построены частотные словари, отражающие разнообразие используемой лексики и выявляющие наиболее характерные слова языка эпохи [Гребенников, Скребцова, 2019; Frequency Word Lists ... , 2020; Using TXM Platform, 2021; Скребцова, 2021; Скребцова, Гребенников, Шерстинова, 2021], автоматический синтаксический анализ позволил определить основные тренды в динамике синтаксической структуры [Martynenko, Sherstinova, 2019; Sherstinova, Ushakova, Melnik, 2020], а с помощью инструментов автоматического тематического моделирования и ручного экспертного аннотирования были выделены основные группы рассказов по тематике [Skrebtsova, 2020; Скребцова, 2020; Zamiraylova, Mitrofanova, 2020; Topic Modelling ... , 2020].

Данное исследование также проведено на основе текстовой выборки из Корпуса-300. Все рассказы аннотированного корпуса были размечены с помощью двух парсеров – MyStem и UDPipe, осуществляющих лемматизацию и морфологическую разметку, а последний в том числе и синтаксическую. Но поскольку автоматическая грамматическая разметка для русского языка дает большое количество омонимичных форм, качественный морфологический анализ требует предварительного ручного снятия омонимии. Поэтому для проведения этого пилотного анализа, основанного на выборке со снятой грамматической омонимией, в случайном порядке были отобраны 15 текстов разных авторов, в равной доле относящихся к каждому из трех исторических периодов. Выбор авторов внутри каждого из периодов осуществлялся случайно, что обеспечивает «непредвзятость» выборки. Для каждого из периодов в Корпусе-300 писатели представлены одним единственным рассказом (подробнее о выборке рассказов аннотированного корпуса см. [Martynenko, Sherstinova, 2019]), поэтому задача выбора конкретного текста для каждого из писателей не стояла.

Таблица 1

Тексты исследуемой подвыборки рассказов
(«с/у» – словоупотребления)

№	Период	Автор	Название	Год написания	Нарратив (лицо)	Размер (с/у)
1	I (1900–1913)	Брюсов Валерий Яковлевич	Мраморная головка (Рассказ бродяги)	1903	1	1195
2		Свенцицкий Валентин Павлович	Солдат задумался...	1906	3	599
3		Лейкин Николай Александрович	На хрен да на редьку, на кислую капусту	1906	3	1086
4		Ауслендер Сергей Абрамович	Занятые люди	1912	3	1159
5		Лазарев-Грузинский Александр Семенович	Незабудки	1913	3	1251
6	II (1914–1922)	Булгаковский Дмитрий Гаврилович	На крестинах	1914	3	843
7		Сандомирский Михаил Борисович	Верочка	1915	1	613
8		Оршер Иосиф Львович	Ничего не случилось	1917	3	633
9		Новиков-Прибой Алексей Силыч	Попался	1917	3	962
10		Романов Пантелеймон Сергеевич	Хороший комитет	1918	3	1071
11	III (1923–1930)	Добychин Леонид Иванович	Нинон	1924	3	574
12		Зорич А. (Локоть Василий Тимофеевич)	Болото	1926	3	886
13		Олеша Юрий Карлович	Лиомпа	1927	3	1257
14		Козин Владимир Романович	Самарская степь, девятнадцатый год, чалый Вор и Вильям Шекспир	1928	1	1939
15		Андреев Василий Михайлович	Голубенькие письма	1929	3	1129

Тексты, на материале которых проведено исследование, представлены в табл. 1. Обращает на себя внимание относительно меньший объем рассказов второго, военно-революционного, периода, приводящий к меньшему суммарному объему выборки для 1914–1922 гг. Однако эта особенность отражает общую картину, свойственную литературной продукции этой эпохи в целом: наши предыдущие исследования показали, что в этот период средняя длина всех текстов малой прозы уменьшается в сравнении с первым и третьим периодами [Frequency Word Lists ... , 2020], что может объясняться тем фактом, что период острых социальных потрясений (войны, революции) не способствует написанию развернутых объемных текстов, и само время диктует необходимость обращения к более лаконичным формам.

Поскольку на распределение личных форм глагола может оказывать влияние вид повествования (от какого лица оно ведется – от первого или от третьего), тексты выборки сбалансированы по этому признаку (для каждого периода один текст написан от первого лица, а остальные четыре – от третьего).

Методика

Все тексты исследовательской выборки были размечены с помощью парсера UDPipe [UDPipe] на базе СинТагРус [СинТагРус]. Выбор данного парсера обусловлен тем фактом, что он показывает высокое качество автоматической морфологической разметки, а также тем, что синтаксический анализ текстов Корпуса-300 базировался на UDPipe-разметке.

В UDPipe выделяется 15 лингвистических категорий, большинство из которых является частеречными: предлог ADP, частица PART, сочинительный союз CONJ, подчинительный союз SCONJ, междометие INTJ, имя существительное NOUN, имя собственное PROP, числительное NUM, местоимение PRON, глагол VERB, вспомогательный глагол AUX, имя прилагательное ADJ, наречие ADV, детерминатор DET и знак пунктуации PUNCT.

Из всех перечисленных категорий было принято решение отредактировать только одну категорию детерминаторов (DET), которую, как отмечает Я.Г. Тестелец, заполняют притяжательные и указательные местоимения, а также артикли [Тестелец, 2001,

с. 519]. В размеченных текстах позицию детерминатора заняли и некоторые другие разряды местоимений. Например, местоимения-прилагательные «любой» и «никакой». Поскольку в русском языке артикли не используются, было решено заменить категорию детерминатора на категорию местоимения.

Кроме того, для анализа стилистики литературного текста показалось целесообразным выделить две дополнительные категории – вводные слова (INTRO) и предикативы (STAT). По умолчанию вводные слова размечаются в UDPipe морфологически соответствующими частями речи или наречиями, нами же было принято решение выделять их, маркируя как INTRO, как, например, в следующих примерах нашей выборки:

Через две минуты Петр Петрович поднял бокал и начал:

– Первый тост мой будет, **конечно** (ADV-> INTRO)¹, за нашу матушку Русь...

И.Л. Оршер «Ничего не случилось»²

– Я, **можно** ((ADV) **сказать**³ (VERB) -> INTRO), всю семью одел.

П.С. Романов «Хороший комитет»

Что касается предикативов, они тегируются программой как краткие прилагательные за исключением тех слов, у которых нет краткой формы, например, «надо» и «можно». В этом случае им приписывается категория наречия. Разделение нами наречий и предикативов при экспертной разметке осуществлялось на основе правил, описанных в работе М.В. Дегтяревой [Дегтярева, 2007, с. 37], предлагающей выделять предикатив как особую гибридную часть речи. Морфологически предикатив не изменяется, поэтому наравне со служебными частями речи и наречием, он не рассматривается ниже при подсчете дистрибуции грамматических форм.

¹ В скобках приводится исходный код разметки программы, а после символов -> код исправленной категории.

² Здесь и далее цит. по [Корпус русского рассказа, 1900–1930].

³ При экспертном аннотировании все вводные конструкции и междометия, состоящие из двух и более слов, объединялись в один элемент, соответствующий отдельной записи базы данных.

В табл. 2 представлен список грамматических признаков, выделяемых UDPipe для знаменательных частей речи. Отметим некоторые особенности разметки UDPipe, а также правила, которых мы придерживались при редактировании автоматической разметки грамматических признаков.

Таблица 2.

Признаки знаменательных частей речи,
выделяемых UDPipe

Часть речи	Выделяемые признаки	Часть речи	Выделяемые признаки
<i>Существительное</i>	Род Одушевленность Число Падеж	<i>Глагол</i>	Время Число Наклонение Род Вид Залог ¹ Форма ² Лицо
<i>Прилагательное</i>	Род Число Падеж Одушевленность Степень ³ Форма	<i>Наречие</i>	Степень
<i>Числительное</i>	Род Падеж Одушевленность	<i>Местоимение</i>	Род Падеж Число Лицо

Говоря о признаках имен существительных, в первую очередь следует обратить внимание на то, что UDPipe использует для русского языка несколько расширенный, по сравнению с традици-

¹ Введен дополнительный медиальный (англ. *middle*) залог.

² Выделяются финитная форма, инфинитив, деепричастие, причастие.

³ Определяется для всех прилагательных.

онным списком из 6 падежей, перечень. В табл. 3 представлено соответствие обозначений всех падежных форм в UDPipe.

Таблица 3.

Русские падежи и их соответствия
в разметке UDPipe

Падежи в русском языке	UDPipe
Именительный	nominative Nom
Родительный	generative Gen
Второй родительный* (насыпать сахару)	partitive Par
Дательный	dative Dat
Винительный	accusative Acc
Творительный	instrumental Ins
Предложный	locative Loc
Второй предложный* (на пляжу)	
Вокатив / звательный*	–

Отдельного внимания заслуживает вопрос выделения таких особых форм, как второй родительный и второй предложный падежи. Ссылаясь на А.А. Зализняка, В.А. Плунгян [Плунгян, 2002, с. 231] отмечает слабую дифференцированность второго предложного падежа в связи с совпадением форм первого и второго предложного у подавляющего большинства лексем в русском языке. Помимо этого, парадигма множественного числа не характеризуется противопоставлением двух предложных падежей [там же]. Поэтому при обработке выборки первый и второй предложный падежи специально не помечались. Так как в разметке UDPipe есть обозначение для специальных форм количественно-отделительного родительного, все ошибочно размеченные примеры были вручную исправлены и учитывались при итоговом подсчете ошибок.

Относящийся в русском языке к квазиграммемам звательный падеж не маркировался при экспертной коррекции разметки, если его форма полностью совпадала с формой именительного падежа. Все формы «старого» вокатива были размечены как междометия (*Господи, Боже*), так как они были употреблены вне контекста об-

ращения с целью выражения эмоциональной реакции. Остальные падежные формы не вызвали затруднений при разборе.

Что касается рода и числа существительных, следует отметить формы *singularia tantum* («дичь», «молодежь») и *pluralia tantum* («щи», «детвора»). Отдельные обозначения для соответствующих примеров не вводились, поэтому в итоговой разметке у всех подобных существительных указано число, но для форм множественного числа отсутствует род.

Таблица 4.

Разметка глаголов в UDPipe

Признак	Описание признака	Обозначение
Род Gender	Мужской Женский Средний	Masc, masculine Fem, feminine Neut, neutral
Число Number	единственное множественное	Sing, Singular Plur, Plural
Лицо Person	первое второе третье	1 2 3
Вид Aspect	совершенный несовершенный	Perf, Perfect Imp, Imperfect
Наклонение Mood	изъявительное повелительное	Ind, Indicative Imp, Imperative
Время Tense	прошедшее настоящее будущее	Past Pres, Presence Fut, Future
Форма глагола VerbForm	личная инфинитив причастие деепричастие	Fin, Finite Inf, Infinitive Part, Participle Conv
Залог Voice	активный пассивный медиальный	Act, active Pass, passive Mid, middle

Существительные в примерах с олицетворением / персонификацией размечались как неодушевленные, если по умолчанию и вне контекста они не являются одушевленными, как в следующих примерах:

Кран тихо сморкался. (Anim -> Inan)

Одеяло сидело рядом, ложилось рядом, уходило, сообщало новости.
(Anim -> Inan)

Ю.К. Олеша «Лиомпа»

В табл. 4 представлены все характеристики глагола, используемые в разметке UDPipe.

Говоря о категориях глагола, следует отметить, что UDPipe не распознает в предложениях сослагательное наклонение. Все примеры глаголов в контексте с частицей «бы» программой размечены как глаголы в изъявительном наклонении. В ходе экспертной проверки все эти случаи были исправлены на Subj (*subjunctive*).

Примеры с омографами не исправлялись при условии, что в приведенном контексте возможно два и более произнесений:

Зато западных сразу **узнаешь**.

Узнаешь (узнаёшь), Tense=Pres/Fut

П.С. Романов «Хороший комитет»

Как видно из табл. 4, UDPipe дополнительно выделяет медиальный залог. Этот вид залога отсутствует в *общепринятой*, по словам В.И. Гавриловой, бинарной системе современной русистики «действительный залог – страдательный залог» [Гаврилова, 2015, с. 43]. Это противопоставление сделано на основе семантики представленных форм: так, «при действительном залоге семантический субъект всегда агенс, <...>, а при страдательном залоге – объект, пациенс» [Кокорина, 2012, с. 20]. Медиальный же залог объединяет в себе признаки агенса и пациенса. С.И. Кокорина приводит в пример клаузу «Автобус остановился», в которой подразумевается наличие субъекта, побудившего действие, в данном контексте – остановку автобуса. Однако в этом случае в синтаксической структуре не прослеживается явного «воздействия», как в пассивной конструкции «Автобус остановлен» [там же, с. 21]. Поэтому при экспертной редакции рассматривалась собственно-возвратность глаголов, у которых обозначен медиальный залог.

Далее оказалось, что автоматическая разметка не выявляет вежливой формы местоимений – все примеры с «вы» носят тег множественного числа. Было принято решение оставить эти случаи без исправления, так как, во-первых, в рассказах не встречается написание данной формы местоимения с заглавной буквы, во-

вторых, это нарушило бы сочетаемость с формой множественного числа у глаголов:

– Виктор Петрович, что **вы** (Plur) **сидите** там в духоте?

А.С. Лазарев-Грузинский «Незабудки»

Следует обратить внимание на омонимичные части речи: например, порядковое числительное «первый» чаще всего употребляется в контексте «лучший / главный» или «наступивший раньше», из-за чего получает тег прилагательного. Все временные и пространственные существительные, проявляющие признаки наречия («гулять *днем*», «гулять *лесом*»), в экспертной разметке, как правило, исправлялись на наречия. Однако при наличии зависимого прилагательного тег существительного оставался без изменений, поскольку в таких примерах «трудно говорить об аморфности» [Руделева, Руделев, 2010, с. 18]:

Прошлой ночью (NOUN) Верочка пришла ко мне.

М.Б. Сандомирский «Верочка»

Автоматическая морфологическая разметка всей выборки текстов была вручную исправлена с опорой на описанные в этом разделе правила.

Дистрибуция частей речи

В табл. 5 приведены данные об абсолютной и относительной частоте встречаемости частеречных категорий в рассматриваемой выборке.

Прежде чем перейти к анализу распределения грамматических форм в динамике, необходимо убедиться, что отобранные тексты действительно отражают «морфологические свойства» изучаемой эпохи. Объективное решение этой задачи возможно через оценку соответствия дистрибуции частей речи в текстах нашей выборки со статистическими данными, полученными на текстовом материале значительно большего объема.

В качестве таких реферативных данных можно использовать статистику о дистрибуции частей речи, посчитанную для всего аннотированного Корпуса-300 из Корпуса русского рассказа 1900–

1930 гг. [Frequency Word Lists ... , 2020]. Данная статистика была получена на материале разметки, выполненной MyStem¹.

Таблица 5.

Дистрибуция частей речи анализируемой выборки
(по автоматической разметке UDPipe с последующей ручной редактурой)

№	Категория	Обозначение	Абс. частота	%
1	существительное	NOUN	3544	23,35
2	глагол	VERB	2875	18,94
3	местоимение	PRON	1803	11,88
4	предлог	ADP	1602	10,55
5	прилагательное	ADJ	1232	8,12
6	наречие	ADV	1104	7,27
7	сочинительный союз	CCONJ	817	5,38
8	частица	PART	787	5,18
9	имя собственное	PROPN	630	4,15
10	подчинительный союз	SCONJ	292	1,92
11	вспомогательный глагол	AUX	194	1,28
12	числительное	NUM	140	0,92
13	предикатив	STAT	87	0,57
14	междометие	INTJ	42	0,28
15	вводное слово	INTRO	30	0,20

Поскольку частеречные категории автоматической разметки MyStem и UDPipe полностью не совпадают, в первую очередь было необходимо привести их «к общему знаменателю». Сопоставление частеречной дистрибуции тестовой выборки с реферативными данными Корпуса-300 для 10 традиционно выделяемых частей речи представлено в табл. 6.

Из табл. 6 видно, что между исследуемой выборкой и данными Корпуса-300 наблюдается хорошая согласованность. Для всех категорий разница между значениями, полученными разными

¹ Егошина И.С. Русская проза первой трети XX века : дистрибуция частей речи в синхронии и диахронии : выпускная квалификационная работа бакалавра : по направлению подготовки № 45.03.01: «Филология», образовательная программа «Филология». – Санкт-Петербург : НИУ ВШЭ, 2020. – 57 с.

морфоанализаторами, не превосходит по модулю 1%. Это, с одной стороны, свидетельствует о «типичности» отобранных текстов для исследуемой эпохи, а с другой – еще раз демонстрирует статистически быструю сходимость частеречной дистрибуции к некоторым общим для языка в заданный период времени значениям [Копотев, 2014].

Таблица 6.

Сопоставление частеречной дистрибуции тестовой выборки с данными, полученными на материале Корпуса-300

№	Часть речи	Исследуемая выборка (UDPipe, %)	Корпус-300 (MyStem, %)
1	существительное	27,50	26,75
2	глагол	20,22	19,95
3	местоимение	11,89	12,25
4	предлог	10,55	10,35
5	прилагательное	8,12	7,86
6	наречие	8,04	8,33
7	союз	7,31	8,77
8	частица	5,19	4,7
9	числительное	0,92	0,57
10	междометие	0,28	0,24

Следующая табл. 7 показывает распределение частей речи (в категориях разметки UDPipe) исследуемой выборки рассказов по рассматриваемым периодам.

Согласно табл. 7 в раннесоветском периоде доля существительных тестовой выборки значительно увеличивается, наблюдается незначительный рост доли глаголов, одновременно происходит снижение использования личных местоимений, союзов, наречий. В целом названные тенденции совпадают с трендами, описанными на материале всего Корпуса-300 [Frequency Word Lists ... , 2020], однако отмеченное для реферативного корпуса постепенное уменьшение к третьему периоду прилагательных и предлогов на нашей выборке не наблюдается, а доля глаголов в разных периодах реферативного корпуса фактически не меняется.

Таблица 7.

**Распределение частей речи в рассказах
тестовой выборки в динамике, %**

№	Категория	I период (1900–1913)		II период (1914–1922)	III период (1923–1930)
1	существительное	NOUN	21,57	21,36	26,39
2	глагол	VERB	17,49	19,54	19,84
3	местоимение	PRON	13,45	12,28	10,15
4	предлог	ADP	10,26	10,89	10,58
5	прилагательное	ADJ	8,08	6,97	8,97
6	наречие	ADV	7,38	8,04	6,63
7	сочинительный союз	CCONJ	5,98	4,73	5,30
8	частица	PART	5,92	6,99	3,22
9	имя собственное	PROPN	4,72	3,23	4,28
10	подчинительный союз	SCONJ	2,00	1,95	1,84
11	вспомогательный глагол	AUX	1,21	1,34	1,30
12	числительное	NUM	0,74	1,32	0,81
13	предикатив	STAT	0,53	0,66	0,55
14	междометие	INTJ	0,45	0,32	0,09
15	вводное слово	INTRO	0,21	0,39	0,05

Таким образом, можно сделать вывод о том, что с позиций частей речи литературный стиль малой русской прозы за исследуемые 30 лет становится более субстантивным, несколько менее описательным и менее «личным».

**Динамика грамматических форм
знаменательных частей речи**

Дистрибуция грамматических форм была получена в целом по выборке для каждого из трех рассматриваемых периодов, а также для всех рассказов по отдельности. В табл. 8 приведены количественные данные о дистрибуции основных грамматических признаков в целом по выборке и в диахронии. Эти данные посчитаны для всех частей речи, для которых та или иная категория является релевантной.

Таблица 8.

**Дистрибуция грамматических признаков в целом
по выборке и в диахронии, %**

№	Грамматические признаки		Всего	I период	II период	III период
1	падеж	Nom (именительный)	37,16	36,97	38,74	36,32
		Acc (винительный)	22,72	24,11	20,65	22,84
		Gen (родительный)	17,26	16,39	19,21	16,77
		Ins (творительный)	8,78	8,11	8,13	9,77
		Loc (предложный)	7,77	7,64	6,58	8,64
		Dat (дательный)	6,09	6,70	5,99	5,62
		Par (паритив)	0,22	0,08	0,70	0,03
2	одушевленность	Anim (одушвл.)	29,22	31,89	31,40	25,74
		Inan (неодушвл.)	70,78	68,11	68,60	74,26
3	род	Fem (женский)	32,14	30,94	24,93	37,30
		Masc (мужской)	51,91	53,18	58,09	47,32
		Neut (средний)	15,95	15,88	16,98	15,39
4	число	Plur (множественное)	21,61	21,71	21,45	21,63
		Sing (единственное)	78,39	78,29	78,55	78,37
5	степень	Cmp (сравнительная)	2,76	2,83	3,01	2,56
		Pos (положительная)	96,59	96,93	96,84	97,23
		Sup (превосходная)	0,20	0,24	0,15	0,21
6	лицо	1	31,74	29,81	34,34	32,36
		2	19,32	20,26	21,46	16,37
		3	48,94	49,93	44,19	51,27
7	вид	Imp (несовершенный)	54,78	57,79	49,47	56,05
		Perf (совершенный)	45,22	42,21	50,53	43,95
8	наклонение	Imp (повелительное)	4,43	5,59	4,89	3,20
		Ind (изъявительное)	94,60	92,98	93,69	96,49
		Subj (сослагательное)	0,97	1,43	1,42	0,31
9	время	Fut (будущее)	5,20	7,27	5,44	3,47
		Past (прошедшее)	73,81	60,97	75,24	82,54
		Pres (настоящее)	20,99	31,76	19,32	13,99
10	форма глагола	Conv (деепричастие)	6,33	4,30	8,78	6,26
		Fin (личная)	78,12	78,71	75,21	79,67
		Inf (инфинитив)	9,13	11,16	9,13	7,49
		Part (причастие)	6,43	5,83	6,88	6,58
11	залог	Act (действительный)	78,82	79,34	80,19	77,46
		Mid (медиальный)	17,63	17,98	16,47	18,17
		Pass (страдательный)	3,55	2,69	3,34	4,38

В распределении падежных форм можно наблюдать относительно устойчивую картину: во всех периодах сохраняется один и тот же ранговый порядок частот падежных форм, и даже близкие в

процентном отношении падежи не меняют своего ранга по частоте от первого периода к третьему. Как и следовало ожидать, самым распространенным падежом является именительный (что характерно и для каждого текста выборки в отдельности).

Если говорить о динамике, то можно отметить следующую вариацию частоты падежных форм: в военно-революционный период увеличивается доля именительного («страна») и родительного падежей («стране»), а также «просторечного» второго родительного / партитива («стакан чаю»). В советский период распределение падежей близко к началу века, отмечается лишь некоторое увеличение доли творительного падежа («страной»). Динамика употребления дательного падежа («стране») на протяжении трех периодов незначительно снижается.

Следует обратить внимание на то, что представленные в табл. 8 усредненные показатели сглаживают вариативность этих характеристик, наблюдаемую у разных авторов. Например, у И.Л. Оршера (II период) и Ю.К. Олеши (III период) отмечается высокий процент вхождений родительного падежа, который почти равен доле винительного (22,8% Acc vs 21,24% Gen и 16,88% Acc vs 16,40% Gen соответственно). А в рассказе А. Зорича (III период) количество родительного падежа в процентном соотношении даже превышает количество винительного. Более того, в этом рассказе меньше всего используется именительный падеж (25,58%) по сравнению с остальной выборкой, а на творительные и предложные формы в сумме приходится около 24%.

Доля «самого редкого» из традиционно выделяемых падежей – дательного – в рассказах А.С. Лазарева-Грузинского и Н.А. Лейкина (I период), а также у Д.Г. Булгаковского возрастает до 8%. Дистрибуция творительного и предложного падежей также варьируется от автора к автору.

Такой признак, как одушевленность / неодушевленность, показал по выборке существенное преобладание неодушевленных существительных над одушевленными (70,78% vs 29,22%). Тем не менее показатели для отдельных рассказов в зависимости от тематики и стиля могут существенно варьироваться. Так, в трех рассказах наблюдается несколько более высокий процент одушевленности (около 41%). Это рассказы С.А. Ауслендера, В.А. Свенцицкого (I период) и И.Л. Оршера (II период). Самым «неодушевленным»

оказался рассказ Ю.К. Олеши «Лиомпа» с долей в 81,53%. Интересно, что изначально именно для этого рассказа многие слова были размечены UDPipe как одушевленные «по контексту» (см. раздел «Методика»).

Относительно категории рода превалирующим по всей выборке является употребление мужского рода (51,91%). Если рассматривать отдельные периоды, то можно отметить преобладание форм мужского рода над женскими во второй, военно-революционный, период. Возможно, наблюдаемая диспропорция объясняется преобладанием героев-мужчин в рассказах этого времени. Стабильной является доля слов среднего рода, которая составляет от 15% до 17% для каждого периода. Наоборот, наиболее изменчивой оказывается процентная доля женского рода с минимумом 24,93% во второй период и максимумом 37,30% в третий период. На такой высокий процент грамматических форм женского рода в третьем периоде, по-видимому, влияет рассказ Л.И. Добычина «Нинон», доля форм женского рода в котором достигает 68,75% и является аномальной для выборки. По-видимому, полученный результат обусловлен тем фактом, что в произведении представлены сразу несколько главных женских персонажей (матушка Олимпиада, Нинон, Мари, Барб Собакина). Результаты для рассказов первого периода полностью отражают полученную дистрибуцию для всей выборки в целом.

По признаку числа, как и предполагалось, преобладает единственное (в соотношении 78,39% к 21,61%). Можно утверждать, что доля форм единственного / множественного числа в текстах малой русской прозы стремится к инвариантному значению, сохраняется от периода к периоду с разницей меньше 0,5%. От общей выборки несколько отличаются результаты для рассказов В.П. Свенцицкого (I период) и П.С. Романова (II период), где доля множественного числа вырастает до 32%. Однако самым выдающимся по признаку числа оказался рассказ А.С. Новикова-Прибоя про офицера и матроса, в котором отмечается самый высокий процент единственного числа, а именно 86,44%.

Сравнительная степень как категория оказалась не очень частотной и составила всего 2,76%. Самыми высокими показателями характеризуются рассказы «Нинон» Л.И. Добычина (4,60%) и «Самарская степь, девятнадцатый год, чалый Вор и Вильям

Шекспир» П.С. Романова (4,44%). По сравнению со сравнительной доля превосходной степени по полученным данным еще меньше – 0,2%. Что касается признака лица, характеризующего глаголы и местоимения, формы третьего лица встречаются значительно чаще, чем формы первого или второго лица. Как уже было отмечено выше, на частоту употребления тех или иных личных форм сильное влияние оказывает вид повествования. Однако тексты выборки по этому признаку сбалансированы – для большинства рассказов читатель занимает позицию наблюдающего со стороны, что подтверждается дистрибуцией форм лица. У В.Р. Козина, М.Б. Сандомирского и В.Я. Брюсова доля употреблений первого лица значительно выше (57,81%; 63,64% и 65,64%, соответственно), из чего следует, что повествование в этих рассказах ведется от первого лица. Самая низкая доля «присутствия» автора в тексте, т.е. форм третьего лица, наблюдается у А.С. Лазарева-Грузинского (5,77%), а наименьшим количеством обращений в виде второго лица характеризуются рассказы Ю.К. Олеши и М.Б. Сандомирского (5,34% и 3,31% соответственно).

В распределении глаголов по виду в среднем наблюдается незначительное преобладание форм несовершенного вида (54,48% vs 45,22%). Однако во второй период мы наблюдаем обратную картину – глаголы совершенного вида показали большую частоту использования (50,53%), в то время доля глаголов несовершенного вида уменьшилась до 49,47%. Как было обнаружено при анализе статистики по конкретным рассказам, высокий процент употребления глаголов совершенного вида в этот период встретился только в одном тексте – А.С. Новикова-Прибоя (61,78%), именно это привело к увеличению средней доли глаголов совершенного вида. Что касается несовершенного вида, самая высокая доля отмечается в рассказе Ю.К. Олеши «Лиомпа». Для большинства остальных рассказов соотношение совершенного и несовершенного вида оценивается как постоянное независимо от времени написания текста.

Самым употребительным наклонением ожидаемо является изъявительное. В нашей выборке доля глаголов изъявительного наклонения составляет 94,6%, причем наблюдается постепенный рост этой доли от первого периода к третьему. Соответственно в советский период процент глаголов повелительного и особенно

сослагательного наклонения падает. Выборка первого периода показывает наибольшее содержание повелительного (5,59%) и условного (1,43%) наклонений. Среди авторов особенно выделяется В.П. Свенцицкий, в рассказе которого все глаголы изъявительного наклонения. Императивов больше всего в рассказе Н.А. Лейкина «На хрен да на редьку, на кислую капусту» (23,24%), что объясняется большим количеством приказов и советов, которые по сюжету даются повару. Наиболее «условным» оказывается рассказ Д.Г. Булгаковского «На крестинах», но тем не менее и здесь доля глаголов сослагательного наклонения остается не очень высокой (5,66%).

За исследуемые 30 лет наиболее значимые отличия проявились в употреблении временных форм глагола – произошло существенное увеличение доли глаголов прошедшего времени (с 61% до 82%) за счет снижения форм настоящего и будущего. Если в довоенный период настоящее время характеризует треть от всех глагольных форм, то в советский период доля глаголов настоящего времени сократилась до одной пятой, а доля форм будущего времени уменьшилась в два раза по сравнению с началом века. Таким образом, повествование советского периода «сместилось в прошлое», и можно предположить, что «наступившее будущее» заставило писателей чаще писать «о прошедшем».

Отмеченные пропорции временных форм глагола свойственны и всем рассказам выборки по отдельности за исключением текста А.С. Лазарева-Грузинского, где преобладают глаголы настоящего времени. Также есть рассказ, в котором ни разу не встретились глаголы в форме будущего времени, – это «Солдат задумался...» В.П. Свенцицкого.

Среди выделяемых UDPipe четырех форм глагола (личная, инфинитив, причастие и деепричастие) самой частотной является личная форма (78,12% по всей выборке). Распределение причастий и деепричастий примерно одинаковое (6,33% деепричастий и 6,43% причастий). Средняя доля инфинитива постепенно снижается от первого периода к третьему. Других тенденций по этому признаку выявить не удалось. По количеству причастий лидирует рассказ А. Зорича «Болото» с 15,71%, а наибольшее количество деепричастий встретилось у А.С. Новикова-Прибоя (19,56%). Лидером по использованию инфинитивов стал рассказ Д.Г. Булгаков-

ского (22,38%), в котором наблюдается более чем двукратное их превышение по сравнению со средним значением.

Залоги распределяются по периодам в приблизительно равной пропорции: самый частотный – действительный (78,8%), на втором месте находится медиальный (17,6%), а самым редким является страдательный (3,5%). У каждого автора присутствуют в тексте глаголы всех трех залогов, при этом на долю действительного залога приходится не менее 70% глаголов. По медиальному залогу главенствующее положение занимает А.С. Лазарев-Грузинский (26,51%), а по глаголам страдательного залога – А. Зорич (10,79%). Никаких тенденций изменения частоты форм залога во времени выявить не удалось.

Заключение

В результате исследования, проведенного на тестовой выборке из 15 рассказов 1900–1930 гг., принадлежащих перу разных русских писателей и входящих в аннотированную часть (Корпус-300) Корпуса русского рассказа указанного периода, получены статистические данные о морфологической вариативности грамматических форм рассмотренных текстов. Эти количественные данные можно считать предварительной оценкой дистрибуции частеречных и грамматических форм русской малой прозы изучаемого периода. Полученные результаты отражают особенности литературного стиля, свойственные отдельным произведениям (а, возможно, и шире – авторским стилям), а также позволяют предположить наличие некоторых общих закономерностей, связанных с диахронической вариативностью. Поскольку относительные частоты грамматических классов слов наряду с синтаксическими особенностями текста являются важным признаком авторского стиля в стилиметрическом анализе [Мартыненко, 1988], полученные данные расширяют эмпирический задел стилиметрических исследований, проводимых на материале русских литературных текстов, и могут быть в дальнейшем использованы для изучения общего и частного в морфологической вариативности малой прозы.

Исследование показало, что в отношении распределения частей речи наблюдается стабильность в ранговых порядках между рассматриваемыми временными периодами: для всех трех эпох

самая частотная часть речи – это существительное, на втором месте всегда идет глагол, на третьем – предлог, потом союз, личные местоимения и т.д. В то же время относительные частоты для большинства частей речи в большей или меньшей степени менялись от периода к периоду. Самая заметная разница наблюдается для существительных: в раннесоветский период процент их употребления увеличился почти на 5%, и произошло это в основном за счет уменьшения доли местоимений, союзов и наречий. С позиций грамматических форм наиболее значимые различия проявились в употреблении временных форм глагола – за исследуемые три десятилетия произошло существенное увеличение доли глаголов прошедшего времени за счет снижения форм настоящего и будущего. Гипотезу о том, что русская малая проза раннесоветского периода характеризуется преобладанием в ней глаголов прошедшего времени, стоит проверить на большем объеме литературных текстов. Если она подтвердится, будут основания утверждать, что не только лексический, но отчасти и морфологический уровень литературного стиля способен адаптироваться к внешним событиям, происходящим в обществе.

В то же время исследование показало, что есть грамматические категории и формы, частота употребления которых фактически не меняется независимо от тематики литературных текстов и даты их написания. Такие категории можно считать неким «грамматическим инвариантом», подобно практически инвариантной дистрибуции, наблюдаемой для ритмических структур языка малой художественной прозы [Kazartsev, Davydova, Sherstinova, 2020]. На морфологическом уровне к таким инвариантам можно отнести относительную частоту использования предлогов, числительных, междометий и, возможно, глаголов. Что касается грамматических признаков, наибольшую стабильность на нашем материале показало распределение форм единственного и множественного числа, среднего рода, положительной степени прилагательных.

Отметим, что небольшой объем исследовательской выборки позволяет говорить о сделанных выводах только как о предварительных. Планируется расширение текстового материала для получения статистически более достоверных результатов. Для упрощения процесса корректуры автоматической разметки – самой трудоемкой части работы – рассматривается возможность обраче-

ния к списку всех исправлений, которые потребовалось внести в аннотированные с помощью UDPipe данные¹. Этот перечень отражает наиболее типичные ошибки программы и может быть использован для частичной автоматизации процедуры проверки при дальнейшей работе.

Список литературы

- Баранников А.П.* Из наблюдений над развитием русского языка в последние годы // Учёные записки Самарского университета. – 1919. – № 2. – С. 64–80.
- Гаврилова В.И.* Залоговая характеристика возвратных глаголов с характеризующе-качественным значением // Вестник московского университета. Серия 9: Филология. – 2015. – № 5. – С. 42–58.
- Гребенников А.О., Скребцова Т.Г.* Языковая картина мира в русском рассказе начала XX века // Философия и гуманитарные науки в информационном обществе. – 2019. – № 3(25). – С. 82–92.
- Дегтярева М.В.* Частеречный статус предикатива : монография. – Москва : МГОУ, 2007. – 161 с.
- Карцевский С.И.* Язык, война и революция. – Берлин : Русское универсальное издательство, 1923. – 72 с.
- Кокорина С.И.* Ещё немного о залогах русского глагола // Вестник центра междunarодного образования московского государственного университета. Филология. Культурология. Педагогика. Методика. – 2012. – № 2. – С. 20–25.
- Копотев М.В.* Введение в корпусную лингвистику. Электронное учебное пособие для студентов филологических и лингвистических специальностей университетов. – Praha : Animedia, 2014. – 230 с.
- Корпус русского рассказа 1900–1930 гг. – URL: <https://russian-short-stories.ru>
- Леонтьев А.А.* Язык как социальное явление (к определению объекта языкознания) // Известия АН СССР. Отделение литературы и языка. – 1976. – Т. 35, вып. 4. – С. 299–307.
- Мартыненко Г.Я.* Методы математической лингвистики в стилистических исследованиях. – Москва : Нестор-История, 2019. – 296 с.
- Мартыненко Г.Я.* Основы стилеметрии. – Санкт-Петербург : Изд-во СПбГУ, 1988. – 176 с.
- Методологические проблемы создания Компьютерной антологии русского рассказа как языкового ресурса для исследования языка и стиля русской художественной прозы в эпоху революционных перемен (первой трети XX века) /

¹ *Завьялова И.С.* Дистрибуция грамматических форм в русской прозе (на материале рассказа 1900–1930 гг.) : курсовая работа : по направлению подготовки № 45.03.01: «Филология», образовательная программа «Филология». – Санкт-Петербург : НИУ ВШЭ, 2021. – 40 с.

- Мартыненко Г.Я., Шерстинова Т.Ю., Мельник А.Г., Попова Т.И. // Компьютерная лингвистика и вычислительные онтологии. (Труды XXI международной объединенной конференции «Интернет и современное общество», IMS-2018, Санкт-Петербург, 30 мая – 2 июня 2018 г. : сборник научных статей). – Санкт-Петербург : Университет ИТМО, 2018. – Вып. 2. – С. 97–102.
- О принципах создания корпуса русского рассказа первой трети XX века / Мартыненко Г.Я., Шерстинова Т.Ю., Попова Т.И., Мельник А.Г., Замирайлова Е.В. // Труды XV международной конференции по компьютерной и когнитивной лингвистике «TEL 2018». – Казань, 2018. – С. 180–197.
- Плунгян В.А. К семантике русского локатива «второго предложного падежа» // Семиотика и информатика. – 2002. – № 37. – С. 229–254.
- Полыванов Е.Д. Русский язык сегодняшнего дня // Литература и марксизм. – 1927. – № 4. – С. 167–180.
- Руделева О.А., Руделев В.Г. Существительное и наречие (на материале русского языка) // Вестник Воронежского государственного университета. – 2010. – № 2. – С. 17–23.
- Селищев А.М. Революция и язык // Селищев А.М. На путях к педагогическому самообразованию. – Москва : Мир, 1925. – С. 141–146.
- Селищев А.М. Язык революционной эпохи : из наблюдений над русским языком последних лет (1917–1926). – Москва : Работник просвещения, 1928. – 248 с.
- СинТагРус = Синтаксически размеченный корпус русского языка : информация для пользователей // Ruscorpora. – URL: <https://ruscorpora.ru/new/instruction-syntax.html>
- Скребцова Т.Г. Динамика тем русских рассказов начала XX века // Философия и гуманитарные науки в информационном обществе. – 2020. – № 3. – С. 45–60.
- Скребцова Т.Г. Новые реалии общественно-политической жизни 1920-х гг. и их отражение в русской литературе и лексикографии (на примере сложносокращенных слов) // Политическая лингвистика. – 2021. – № 2. – С. 146–154. – URL: https://doi.org/10.26170/1999-2629_2021_02_14
- Скребцова Т.Г., Гребенников А.О., Шерстинова Т.Ю. Динамика лексического состава русской художественной прозы (на материале частотных словарей корпуса русских рассказов 1900–1930) // Компьютерная лингвистика и интеллектуальные технологии : по материалам ежегодной международной конференции «Диалог». – 2021. – № 20. – С. 646–660.
- Тестелец Я.Г. Введение в общий синтаксис. – Москва : РГТУ, 2001. – 805 с.
- Frequency Word Lists and Their Variability (the Case of Russian Fiction in 1900–1930) / Sherstinova T., Grebennikov A., Skrebtsova T., Guseva A., Gukasian M., Egoshina I., Turygina M. // 27th Conference of Open Innovations Association FRUCT. – Trento : University of Trento, 2020. – P. 366–373. – URL: <https://fruct.org/publications/acm27/files/She.pdf>
- Jakobson R. Vliv revoluce na ruský jazyk = Влияние революции на русский язык. – Praha : 1921. – 34 p.
- Kazartsev E., Davydova A., Sherstinova T. Rhythmic Structures of Russian Prose and Occasional Iambs (a Diachronic Case Study) // Speech and Computer. 22nd Interna-

- tional Conference, SPECOM 2020. – Deutschland GmbH : Springer Science and Business Media, 2020. – Vol. 12335, LNAI. – P. 194–203.
- Martynenko G., Sherstinova T. Analytical Distribution Model for Syntactic Variables Average Values in Russian literary Texts // Proc. of the Int. Conf. DTGS-2019. Digital Transformation and Global Society (4th International Conference, DTGS 2019, St. Petersburg, Russia, June 19–21, 2019), Revised Selected Papers. – 2019. – Vol. 1038 : Communications in Computer and Information Science. Springer International Publishing. – P. 719–731.
- Mazon A. Lexique de la guerre et de la révolution en Russie (1914–1918). – Paris : E. Champion, 1920. – 63 p.
- Sherstinova T., Martynenko G. Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century // R. Piotrowski's Readings in Language Engineering and Applied Linguistics (Proc. of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019), Saint Petersburg, Russia, November 27, 2019), CEUR Workshop Proceedings. – 2020. – Vol. 2552. – P. 105–120.
- Sherstinova T., Ushakova E., Melnik A. Measures of Syntactic Complexity and their Change over Time (the Case of Russian) // 27th Conference of Open Innovations Association FRUCT. – Trento : University of Trento, 2020. – P. 221–229. – URL: <https://fruct.org/publications/fruct27/files/She.pdf>
- Skrebtsova T. Thematic Tagging of Literary Fiction: The Case of Early 20th Century Russian Short Stories // Proceedings of the International Conference “Internet and Modern Society” IMS-2020. CEUR Workshop Proceedings. – 2020. – Vol. 2813. – P. 265–276.
- Topic Modelling with NMF vs. Expert Topic Annotation: the Case Study of Russian Fiction / Sherstinova T., Mitrofanova O., Skrebtsova T., Zamiraylova E., Kirina M. // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Advances in Computational Intelligence: 19th Mexican International Conference on Artificial Intelligence, MICAI 2020 (Mexico City; Mexico, 12–17 October 2020). – Deutschland GmbH : Springer Science and Business Media. – 2020. – Vol. 12469, LNAI. – P. 134–151.
- UDPipe = морфологический парсер // Lindat. – URL: <http://lindat.mff.cuni.cz/services/udpipe/run.php>
- Using TXM Platform for Research on Language Changes over Time: the Dynamics of Vocabulary and Punctuation in Russian Literary Texts / Lavrentiev A.M., Sherstinova T.Yu., Chepovskiy A.M., Pincemin B. // Vestnik TSU. – 2021. – № 70. – P. 69–89. – URL: <http://journals.tsu.ru/engine/download.php?id=212443&area=files>
- Zamiraylova E., Mitrofanova O. Dynamic Topic Modeling of Russian Prose of the First Third of the XXth Century by Means of Non-Negative Matrix Factorization // R. Piotrowski's Readings in Language Engineering and Applied Linguistics (Proc. of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019), Saint Petersburg, Russia, November 27, 2019; CEUR Workshop Proceedings). – 2020. – Vol. 2552. – P. 321–339.

References

- Barannikov, A.P. (1919). Iz nabljudenij nad razvitiem russkogo jazyka v poslednie gody. *Uchjonye zapiski Samarskogo universiteta*, 2, 64–80.
- Gavrilova, V.I. (2015). Zalogovaja harakteristika vozvratnyh glagolov s harakterizuju-shhe-kachestvennym znacheniem. *Vestnik moskovskogo universiteta*, 5, 42–58.
- Grebennikov, A.O., Skrebcova, T.G. (2019). Jazykovaja kartina mira v russkom rasskaze nachala XX veka. *Filosofija i gumanitarnye nauki v informacionnom obshhestve*, 25, 82–92.
- Degtjareva, M.V. (2007). *Chasterechnyj status predikativa*. Moscow: Moskovskiy gosudarstvennyy oblastnoy universitet.
- Karcevskij, S.I. (1923). *Jazyk, vojna i revoljucija*. Berlin: Russkoe universal'noe izdatel'stvo.
- Kokorina, S.I. (2012). Eshhjo nemnogo o zalogah russkogo glagola. *Vestnik centra mezhdunarodnogo obrazovanija moskovskogo gosudarstvennogo universiteta. Filologija. Kul'turologija. Pedagogika. Metodika*, 2, 20–25.
- Kopotev, M.V. (2014). *Vvedenie v korpusnuju lingvistiku*. Praha: Animedia.
- Korpus russkogo rasskaza 1900–1930*. Retrieved from: <https://russian-short-stories.ru>
- Leont'jev, A.A. (1976). Jazyk kak social'noe javlenie (k opredeleniju objekta jazykoznaniya). *Izvestija AN SSSR. Otdelenie literatury i jazyka*, 35, 299–307.
- Martynenko, G.Ja (2019). *Metody matematicheskoj lingvistiki v stilisticheskikh issledovanijah*. Moscow: Nestor-Istorija.
- Martynenko, G.Ja (1988). *Osnovy stilemetrii*. Leningrad : Leningrad State University.
- Martynenko, G.Ja, Sherstinova, T.Ju., Mel'nik, A.G., Popova, T.I. (2018). Metodologicheskie problemy sozdaniya Komp'juternoj antologii russkogo rasskaza kak jazykovogo resursa dlja issledovanija jazyka i stilja russkoj hudozhestvennoj prozy v jepohu revoljucionnyh peremen (pervoj treti XX veka). *Komp'juternaja lingvistika i vychislitel'nye ontologii*, 2, 97–102.
- Martynenko, G.Ja, Sherstinova, T.Ju., Mel'nik, A.G., Popova, T.I., Zamirajlova, E.V. (2018). O principah sozdaniya korpusa russkogo rasskaza pervoj treti XX veka. *Trudy XV Mezhdunarodnoj konferencii po komp'juternoj i kognitivnoj lingvistike* (pp. 180–197). Kazan: Akademiya nauk respubliky Tatarstan.
- Plungjan, V.A. (2002). K semantike russkogo lokativa “vtorogo predlozhnogo padezha”. *Semiotika i informatika*, 37, 229–254.
- Polivanov, E.D. (1927). Russkij jazyk segodnjashnego dnja. *Literatura i marksizm*, 4, 167–180.
- Rudeleva, O.A., Rudelev, V.G. (2010). Sushhestvitel'noe i narechie (na materiale russkogo jazyka). *Vestnik Voronezhskogo gosudarstvennogo universiteta*, 2, 17–23.
- Selishhev, A.M. (1925). Revoljucija i jazyk. *Na putjah k pedagogicheskomu samoobrazovaniju* (pp. 141–146). Moscow: Mir.
- Selishhev, A.M. (1928). Jazyk revoljucionnoj jepohi: iz nabljudenij nad russkim jazykom poslednih let (1917–1926). Moscow: Rabotnik prosveshhenija.

- SinTagRus* = Sintaksicheski razmechennyj korpus russkogo jazyka: informacija dlja pol'zovatelej. Ruscorpora. Retrieved from: <https://ruscorpora.ru/new/instruction-syntax.html>
- Skrebcova, T.G. (2020). Dinamika tem russkikh rasskazov nachala XX veka. *Filosofija i gumanitarnye nauki v informacionnom obshhestve*, 3, 45–60.
- Skrebcova, T.G. (2021). Novye realii obshhestvenno-politicheskoy zhizni 1920-h gg. i ih otrazhenie v russkoj literature i leksikografii (na primere slozhnosokrashennykh slov). *Politicheskaja lingvistika*, 2, 146–154. Retrieved from: https://doi.org/10.26170/1999-2629_2021_02_14
- Skrebcova, T.G., Grebennikov, A.O., Sherstinova, T.Ju. (2021). Dinamika leksicheskogo sostava russkoj hudozhestvennoj prozy (na materiale chastotnykh slovarej korpusa russkikh rasskazov 1900–1930). *Komp'juternaja lingvistika i intellektual'nye tehnologii*, 20, 646–660.
- Testelec, Ja.G. (2001). *Vvedenie v obshhij sintaksis Rossijskij gosudarstvennyj gumanitarnyj universitet*. Moscow: RGGU.
- Sherstinova, T., Grebennikov, A., Skrebtsova, T., Guseva, A., Gukasian, M., Egoshina, I., Turygina, M. (2020). Frequency Word Lists and Their Variability (the Case of Russian Fiction in 1900–1930). In: *27th Conference of Open Innovations Association FRUCT* (pp. 366–373). Trento: University of Trento.
- Jakobson, R. (1921). *Vliv revoluce na ruský jazyk* (pp. 28–60). Praha.
- Kazartsev, E., Davydova, A., Sherstinova, T. (2020). Rhythmic Structures of Russian Prose and Occasional Iambs (a Diachronic Case Study). In: *Speech and Computer* (pp. 194–203). Deutschland GmbH: Springer Science and Business Media.
- Martynenko, G., Sherstinova, T. (2019). Analytical Distribution Model for Syntactic Variables Average Values in Russian literary Texts. In: *Proc. of the Int. Conf. DTGS-2019. Digital Transformation and Global Society* (pp. 719–731). Cham: Springer.
- Mazon, A. (1920). *Lexique de la guerre et de la révolution en Russie (1914–1918)*. Paris: E. Champion.
- Sherstinova, T., Martynenko, G. (2020). Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century. In: *R. Piotrowski's Readings in Language Engineering and Applied Linguistics*, 2552, 105–120.
- Sherstinova, T., Ushakova, E., Mel'nik, A. (2020). Measures of Syntactic Complexity and their Change over Time (the Case of Russian). In: *27th Conference of Open Innovations Association FRUCT* (pp. 221–229). Trento: University of Trento.
- Skrebtsova, T. (2020). Thematic Tagging of Literary Fiction: The Case of Early 20th Century Russian Short Stories. *Proceedings of the International Conference "Internet and Modern Society"*, 2813, 265–276.
- Sherstinova, T., Mitrofanova, O., Skrebtsova, T., Zamiraylova, E., Kirina, M. (2020). Topic Modelling with NMF vs. Expert Topic Annotation: the Case Study of Russian Fiction. *Advances in Computational Intelligence* (pp. 134–151). Deutschland GmbH: Springer Science and Business Media.
- UDPipe = morfologicheskij parser. Retrieved from: <http://lindat.mff.cuni.cz/services/udpipe/run.php>

- Lavrentiev, A.M., Sherstinova, T.Yu., Chepovski, A.M., Pincemin, B. (2021). Using TXM Platform for Research on Language Changes over Time: the Dynamics of Vocabulary and Punctuation in Russian Literary Texts. *Vestnik TSU*, 70, 69–89.
- Zamiraylova, E., Mitrofanova, O. (2019). Dynamic Topic Modeling of Russian Prose of the First Third of the XXth Century by Means of Non-Negative Matrix Factorization. In: *R. Piotrowski's Readings in Language Engineering and Applied Linguistics*, 2552, 321–339.