Минобрнауки россии

Федеральное государственное бюджетное

образовательное учреждение

высшего профессионального образования

«Воронежский государственный университет»

(ФГБОУ ВПО «ВГУ»)

Факультет романо-германской филологии

Кафедра теоретической и прикладной лингвистики

**РАСПРЕДЕЛЕНИЕ ПРЕДЛОЖЕНИЙ ПО ДЛИНЕ**

**В ТЕКСТАХ РУССКИХ ПИСАТЕЛЕЙ XIX-XX ВЕКОВ**

ДИПЛОМНАЯРАБОТА

по специальности 031301 – Теоретическая и прикладная лингвистика

Допущено к защите в ГАК 14.05.2013

Зав. кафедрой\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *Кретов А.А, д.ф.н., проф.*

Студент\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*Коновалова О.А.*

Руководитель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*Кретов А.А, д.ф.н., проф.*

Рецензент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*\_\_Кашкин В.Б. д.ф.н., проф.*

Минобрнауки россии

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

«Воронежский государственный университет»

(ФГБОУ ВПО «ВГУ»)

Факультет Романо-германской филологии

Кафедра Теоретической и прикладной лингвистики

УТВЕРЖДАЮ

Зав. кафедрой\_\_\_\_\_\_\_\_\_\_\_\_А.А. Кретов

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2012 г.

**ЗАДАНИЕ**

на выполнение выпускной квалификационной работы

студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(фамилия, имя, отчество)

1. Тема ВКР\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Специальность 031301 – Теоретическая и прикладная лингвистика
3. Срок сдачи студентом законченной работы «\_\_» \_\_\_\_\_\_\_\_\_\_\_\_2013 г.
4. Календарный план

|  |  |  |
| --- | --- | --- |
| Наименование раздела ВКР | Срок выполнения этапов работы | Примечание |
| Глава 1 Теоретические основы распределения предложений по длине | до 15.03.2013 |  |
| Глава 2 Распределение предложений по длине в произведениях русских писателей XIX-XX вв. | до 1.05.2013 |  |
| Введение, Заключение, оформление ВКР | до 13.05.2013 |  |

Руководитель\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись) Ф.И.О.

Студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись) Ф.И.О.

Оглавление

[ВВЕДЕНИЕ 4](#_Toc358373851)

[ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ РАСПРЕДЕЛЕНИЙ ПРЕДЛОЖЕНИЙ ПО ДЛИНЕ 7](#_Toc358373852)

[1.1. Опыт изучения авторских стилей методами количественного анализа 7](#_Toc358373853)

[1.2. Исследования, посвященные изучению размера предложения 13](#_Toc358373854)

[1.3. Описание лингвистических величин с помощью теоретических распределений 20](#_Toc358373855)

[ГЛАВА 2. РАСПРЕДЕЛЕНИЕ ПРЕДЛОЖЕНИЙ ПО ДЛИНЕ В ПРОИЗВЕДЕНИЯХ РУССКИХ ПИСАТЕЛЕЙ XIX-XX ВЕКОВ 29](#_Toc358373856)

[2.1. Методика исследования 29](#_Toc358373857)

[2.1. Полученные результаты 34](#_Toc358373858)

[2.2. Анализ факторов, влияющих на распределение 38](#_Toc358373859)

[2.3. Аппроксимация экспериментальных данных 51](#_Toc358373860)

[2.4. Интерпретация полученных результатов 57](#_Toc358373861)

[СПИСОК ЛИТЕРАТУРЫ 68](#_Toc358373862)

[ПРИЛОЖЕНИЕ 1. СПИСОК ИССЛЕДОВАННЫХ ПРОИЗВЕДЕНИЙ С УКАЗАНИЕМ АВТОРА И ВРЕМЕНИ НАПИСАНИЯ ПРОИЗВЕДЕНИЯ 73](#_Toc358373863)

[ПРИЛОЖЕНИЕ 2. РАСПРЕДЕЛЕНИЕ ИССЛЕДОВАННЫХ ПРОИЗВЕДЕНИЙ ПО ГРУППАМ В ЗАВИСИМОСТИ ОТ ДЛИНЫ ТЕКСТА (В ПРЕДЛОЖЕНИЯХ). 76](#_Toc358373864)

[ПРИЛОЖЕНИЕ 3. КОЭФФИЦИЕНТЫ α, β И КАЧЕСТВО АППРОКСИМАЦИИ ДЛЯ ИССЛЕДОВАННЫХ ПРОИЗВЕДЕНИЙ 79](#_Toc358373865)

## ВВЕДЕНИЕ

Считается, что каждый известный писатель имеет свои стилистические, художественные, языковые особенности. Часто по двум-трем предложениям читатель может определить, кем написан данный текст. Данные особенности трудно формализуются. Нами выдвинута гипотеза, что помимо этих общепризнанных отличий произведения одного автора также обладают статистическими характеристиками, которые можно выявить методами количественного анализа. Одной из таких характеристик является длина предложения.

Несмотря на то, что такой аспект синтаксиса, как размер предложения давно интересует лингвистов и достаточно много исследований посвящено его изучению, остается значительное число неисследованных вопросов. Во-первых, данный параметр неравномерно изучен на материале разных языков: больше всего исследований проводилось на материале русского и немецкого языков. Русский язык оказался изучен благодаря трудам Г.А. Лесскиса, Г.Н. Акимовой, немецкий язык - благодаря работам В.Г. Адмони. Остальные языки изучены явно недостаточно. Во-вторых, размер предложения неравномерно изучен в различных аспектах. Так, этот параметр более или менее изучен для периода XVIII-XX веков в русском языке и практически не изучен для более ранних эпох. Несмотря на интерес к грамматической стороне проблемы размера предложения в трудах В.Г. Адмони и Г.Н. Акимовой, этот аспект проблемы остается малоисследованным. Отдельную проблему составляет отсутствие единой методики измерения размера предложения: использование разных единиц при подсчете длины предложения приводит к тому, что полученные разными исследователями результаты трудно сопоставлять друг с другом (Вакурова, 2011).

Предпринимаемое в данной работе исследование длины предложения основывается на идее, что распределение предложений по длине в художественном стиле во многом определяется авторскими установками, т.е. является характеристикой авторского стиля.

Главная **цель** данной дипломной работы состоит в том, чтобы проследить, как меняется распределение предложений в творчестве авторов XIX-XX вв., установить факторы, влияющие на это распределение, и описать экспериментальные данные с помощью формулы и в заключение – предложить интерпретацию полученных результатов по аналогии со сходным распределением в физике. В таком аспекте исследование длины предложения еще не проводилось.

Для достижения данной цели в ходе исследования были решены следующие **задачи**:

1. Изучение работ, посвященных исследованию размера предложения.
2. Выбор произведений для дальнейшей обработки и их группировка.
3. Получение статистических данных о длине предложения в текстах писателей XIX-XX веков и их регистрация.
4. Обработка полученных результатов с помощью методов математической статистики, построение графиков.
5. Определение факторов, влияющих на распределение и их оценка.
6. Выяснение, какое из известных в теории вероятностей распределений лучше описывает распределение предложений по длине. Оценка параметров этого распределения и их интерпретация.
7. Классификация текстов и авторов на основании параметров распределения предложений по длине.

В работе использовались статистический, описательный **метод**¸ а также метод сопоставительного анализа.

**Предметом** исследования является распределение предложений по длине (в текстовых словах) в произведениях русских писателей XIX-XX вв.

**Объектом** исследования послужилипроизведения русских писателей XIX-XX века. Всего было обработано 99 текстов 21 автора общей длиной 197.085 предложений. Все исследуемые тексты брались в сети Интернет в «Библиотеке Мошкова».

**Актуальность** проведенного исследования заключается в том, что в нём предпринята попытка объективации и формализации такого сложного явления как «стиль автора».

**Научная новизна** исследования состоит в том, что в нём проведен анализ и впервые предложена функция, которая с высокой точностью описывает экспериментальные данные распределения предложений по длине в текстах различных периодов и авторов, установлены факторы, влияющие на данное распределение, предложен (вслед за Г.Д. Селезнёвым) параметр «*синтаксической* температуры» текста, отражающий особенности распределения предложений по длине.

Работа состоит из введения, двух глав, заключения и трёх приложений, в которых в виде таблиц представлены результаты исследования.

## ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ РАСПРЕДЕЛЕНИЙ ПРЕДЛОЖЕНИЙ ПО ДЛИНЕ

* 1. . Опыт изучения авторских стилей методами количественного анализа

Основания количественного, статистического анализа языка заложены его структуре. Первым основанием применения количественных методов в языкознании является «объективная присущность языку количественных признаков, количественных характеристик» (Головин, 1971). Обращение к количественным характеристикам можно встретить и в работах представителей «традиционного языкознания». Так, В.В. Виноградов писал о том, что «по-видимому, в разных стилях книжной и разговорной речи, а также в разных стилях и жанрах художественной литературы частота употребления разных типов слов различна. Точные изыскания в этой области помогли бы установить структурно-грамматические, а отчасти и семантические различия между стилями» (Виноградов 1938, с.155-156).

Вторым основанием применения вероятностно-статистических методов в лингвистике является внутренняя зависимость, которая существует между количественными и качественными характеристиками языковой структуры (Головин, 1971). Например в языках, в которых преобладают односложные слова, всегда есть музыкальное ударение, так как при ограниченном наборе слогов возникает необходимость их различения. Если в языке в ходе развития слова становятся односложными и перестают члениться на морфемы, то в языке развивается музыкальное ударение (так объясняется ударение в тибетском и классическом китайской языках) (Иванов, 1963). Таким образом, можно предположить, что количественные различия на одном уровне дают качественные различия на другом, смежном уровне. Кроме того, изменение количественных соотношений между одними и теми же единицами языка (например, частями речи) меняет «качество» речи (Головин, 1971). Данные статистики отображают различия между языковыми и речевыми стилями. Эти различия могут быть выражены вероятностными характеристиками одних и тех же элементов языковой структуры. Все это позволяет предположить, что авторские речевые стили могут быть охарактеризованы определенными для каждого автора соотношениями частот различных элементов языка.

Третьим основанием применения количественных методов в языкознании является «подчиненность количественных характеристик языка и речи вероятностно-статистическим законам» (Головин, 1971). Поэтому одной из главных задач статистической лингвистики является описание полученных экспериментальных данных теоретическими распределениями.

Количественные методы позволяют сгруппировать и измерить эмпирически полученные лингвистические факты, а теория вероятностей помогает перейти к прогнозированию лингвистических явлений. Расширяющиеся контакты языкознания с другими науками, например, с теорией информации, могут быть осуществлены только с использованием математического языка, который универсален для различных отраслей знаний.

Статистические методы часто используются в стилеметрических исследованиях. По мнению Л.Р. Зиндера «только статистическая структура текста, как некая равнодействующая, дает возможность отнести текст к соответствующему функциональному стилю, соотнести его хронологически, определить автора и т.д. (Зиндер, 1968, с.120). В науке уже предпринимались попытки интерпретировать категорию «стиль» с использованием вероятностно-статистической терминологии. Согласно Б.Н. Головину «Функциональные стили языка – это типы его функционирования, соотнесенные с типами социальной деятельности и отличающихся друг от друга системами *существенных различий вероятностей* языковых единиц и категорий, достаточными для их качественного опознавания людьми на интуитивном уровне восприятия речи» (Головин 1979, с.233). В.Я. Слепак приводит другое определение стиля: «Стиль – это устойчивый, целостный, функционально своеобразный способ реализации коммуникативно-речевой деятельности, построенный по образцу сложных динамических систем и отличающийся от аналогичных способов, возникающих на иной экстралингвистической основе, *наборами вероятностей* и соответствующих *стохастических схем* употребления языковых средств, а также *корреляционных отношений*, количественно оценивающих взаимосвязи единиц одного и разных уровней системы языка» (Слепак, 1982 с.107). Таким образом, такое явление как стиль автора может быть описан при помощи признаков, которые имеют вероятностно-статистическую природу.

Впервые гипотезу о том, что статистические параметры текста могут отражать характеристики авторских стилей, высказал английский математик Аугустус де Морган. В 1851 году он предположил, что средние длины слов в двух тестах, написанных одним автором, даже на разные темы, должны отличаться меньше, чем средние длины слов в текстах разных авторов, даже если они относятся к одной тематике (Grzybek, 2006).

Первым отечественным исследователем попытавшимся описать особенности авторских стилей при помощи количественного анализа был Н.А. Морозов, который предложил сопоставить частоту использования служебных частиц и нейтральных слов в текстах различных авторов для установления характеристик их стилей. По его мнению, служебные слова и слова с нейтральной стилистической окраской могут отражать признаки индивидуальных стилей авторов, так как на них не влияет тематика произведения. Полученные результаты Н.А. Морозов представил в виде графиков, которые были названы «лингвистическими спектрами» (Морозов, 1915)

В 1916 году вышла статья А.А. Маркова, в которой методика, предложенная Н.А. Морозовым, была подвержена критике. По мнению математика, для того, чтобы получить значимый результат, необходимо исследовать больший по объемам материал (подсчет не пяти тысяч, а сотен тысяч слов). Но развитие вычислительной техники в начале XX века не позволяло подтвердить или опровергнуть критику ученого (Марков, 1916)

В.Фукс предпринял попытку установить авторство некоторых древних текстов, опираясь на статистические показатели грамматических структур. Он исходит из предположения о том, что частоты переходов для синтаксических классов играют в исследовании подлинности текстов чрезвычайно важную роль. Объектом его исследования стали библейские тексты. Автор провел анализ при помощи индекса различия, результаты которого показали, что фрагменты «Евангелия от Иоанна» сильно отличаются от «Апокалипсиса». Все четыре евангелических текста характеризуются внутренним сходством и имеют свои стилевые особенности (Фукс, 1975).

Норвежский математик Г. Хьетсо в 1979 году предложил методику определяющую стиль автора, также основанную на количественных методах. Суть метода сводится к вычислению семи параметров текста и их статистической оценке. К параметрам, которые предложил оценивать исследователь относятся: средняя длина слова в буквах (на основании выборки в 500 слов), общее распределение длины слова, средняя длина предложения (на основании выборки в 30 предложений), общее распределение длины предложений, лексические спектры текста на уровне словаря и текста. Для каждого параметра определялось критическое значение и проверялись две гипотезы: - текст принадлежит данному автору и альтернативная гипотеза - текст не принадлежит данному автору. Затем формулировалось заключение. При помощи такой методики Г.Хьетсо подтвердил авторство М.А. Шолохова в произведении «Тихий Дон», а также определил принадлежность Ф.М. Достоевскому некоторых анонимных статей в журналах «Время» и «Эпоха» (Хьетсо 1986, 1989)

Одним из российских лингвистов, занимающихся описанием авторских стилей при помощи количественных методов, является Г.Я.Мартыненко – основатель и руководитель Санкт-Петербургской школы стилеметрии. В 1988 году вышла его монография «Основы стилеметрии», в которой подводится итог исследований синтаксической сложности предложения (Мартыненко, 1988). Автор провел исследование на материале художественной прозы XIX – начала XX века, всего было обработано 100 авторов. Исследователем было изучено семь параметров: средний размер абзаца, средний размер предложения, средний размер словоупотребления, степень дистантности (число узлов между двумя узлам, связанными непосредственной синтаксической связью), степень перечислительности (число однородных узлов в однородной группе), степень атрибутивности словоупотреблений, степень деривационной громоздкости словоупотреблений. Г.Я. Мартыненко распределил исследуемых авторов на однородные группы по этим признакам. По его результатам некоторые изученные признаки абсолютно не чувствительны к индивидуальному стилю автора. Вариация таких признаков как степень соподчиненности (число стрелок, исходящих из данного узла) и степень гнездования (число стрелок, накрывающих данный узел, и таких, которые не имеют общих краев), например, полностью определяется размером предложения, поэтому они были исключены из классификации. Автор отмечает, что такая классификация текстов и их авторов может служить предпосылкой для «определения лингвистических аналогов поэтических школ и литературных направлений, а также для изучения эволюции литературно-художественных систем» (Мартыненко, 1988).

Количественные методы обрели широкую популярность в задачах установления авторства и описания стилей, так как учет количественных показателей обеспечивает объективность стилистических исследований.

Так, М.М. Пещак исходит из предположения о том, что такой признак как наличие-отсутствие составных элементов целого может отражать авторские черты. Исследователь анализирует композиционную структуру юридических документов XIV века. В самом коротком документе насчитывалось 14 слов, а в самом длинном – 118 слов. Такое большое расхождение между объемами собранных документов поставило перед автором задачу выявить факторы, влияющие на размер документа. Исходя из текстового оформления, М.М. Пещак выделил восемь структурных компонентов в исследованных документах, каждый из которых характеризуется особым лексическим наполнением. Среди проанализированных грамот не встретилось ни одной, в которой бы реализовывались все восемь подразделов, что натолкнуло автора на мысль о том, что каждый писарь мог иметь свою излюбленную схему композиционной структуры текста. Это значит, что на основании однотипности наличия-отсутствия структурных подразделов можно создать классификацию грамот, которая может помочь в установлении авторства неатрибутированных грамот. Автор выделил пять типов композиционных структур грамот и предположил, что они могут быть связаны с различными писарскими школами, представителями которых являлись писари (Пещак, 1974).

А.Ф. Папина и Л.Н. Иванова-Маркова изучают категорию качества и категорию действенности в произведениях поэтов XIX века. Используя статистический метод, исследователи пытаются установить различия в особенностях количественного лексико-грамматического отбора, производимого поэтами. К словам категории действенности были отнесены глаголы, краткие причастия и деепричастия. К словам, обозначающим качественные признаки, относились полные и краткие прилагательные, полные причастия. По результатам статистического анализа сторонниками категории качества оказались представители пассивного романтизма (А.А. Дельвиг, К.Н. Батюшков, Е.А. Баратынский), классицизма (Г.Р. Державин). В эту же группу попал представитель активного романтизма А.А. Бестужев-Марлинский. В группу сторонников действенности вошли такие авторы как Н.М. Карамзин, В.А. Жуковский, П.А. Вяземский (Папина; Иванова-Маркова, 1974).

Исследование Н.П. Дарчук посвящено сравнению лексики двух произведений – Ю.Смолича «Мир хижинам, война дворцам» и Г. Тютюнника «Водоворот». Исследование проводилось на материале двух выборок, по двадцать тысяч словоупотреблений каждая. Были составлены частотные словари для каждой выборки. Для слов с частотой 2 вычислялись следующие статистические характеристики: средняя частота, среднее квадратичное отклонение, коэффициент вариации и коэффициент стабильности. Под «коэффициентом стабильности» исследователь понимает показатель, который характеризует равномерность распределения слова в тексте. Н.П. Дарчук исходит из предположения о том, что «своеобразие языка художественного произведения состоит в определенных способах употребления языковых элементов» (Дарчук 1974, с. 263) Автор предположил, что коэффициент равномерности может служить критерием разбиения исследуемых единиц (слов) на классы. Сравнение данных о стабильности слов позволило исследователю выявить наиболее употребительные слова в жанре художественной прозы, а также коэффициент стабильности был рассмотрен как критерий для разграничения авторских стилей (Дарчук, 1974).

Изучение текстов методами количественного анализа открывает большие возможности в исследовании авторских стилей, литературных направлений и школ. Статистический анализ языковой структуры может помочь в понимании того, как возникают и сохраняют языковую устойчивость авторские речевые стили. Многие соотношения языковых частот оказываются устойчивыми для одного автора и вместе с тем сильно отличаются от соотношения тех же языковых частот у другого автора. Это дает основание для анализа авторских стилей при помощи соотнесения частотных характеристик языковых элементов.

1.2. Исследования, посвященные изучению размера предложения

Одним из способов описания стилевых характеристик текста с помощью количественных методов, является изучение размера предложения. Этот аспект синтаксиса давно интересует исследователей, однако активное изучение этого параметра текста началось во второй половине XX в. На размер предложения влияет ряд факторов, которые можно условно подразделить на грамматические и стилистические (Вакурова, 2011).

В.Г. Адмони говорит о том, что на размер предложения в первую очередь влияет объем и характер человеческой памяти. При этом объем и характер памяти оказываются под влиянием таких факторов как форма речи (устная или письменная, монологическая или диалогическая), тематика сообщения (научная, бытовая и т.д.), степень эмоциональной насыщенности речи, особенности стилистической установки говорящего. Более сложным и интересным является процесс взаимодействия и развития этих факторов в историческом плане. «Возникновение новых сфер речевой коммуникации с новым тематическим наполнением может привести выработке таких пластов языка, которые будут, как правило, тяготеть к тем или иным размерам предложения, отличающимся от прежде преобладавшего в данном языке среднего размера предложения» (Адмони 1966, с.114).

На размер предложения также влияют грамматические факторы. По мнению С.А. Рылова на длину предложения влияет такое свойство слова, как валентность (Рылов, 1990). Если говорить о размере сложного предложения, то действует также фактор «возможности объединения двух или более предложений в целое» (Веселовская, 1987). Надо отметить, что грамматические факторы, такие как способность слова вступать в синтаксические связи с другими элементами, расширение синтагм безусловно оказывают влияние на размер предложения, однако выбор синтаксических конструкций и степень их расширения зависит от задач, которые стоят перед автором.

Г.Ф. Калашникова акцентирует внимание на экстралингвистических факторах, которые влияют на длину предложения, а именно на «психофизиологическом механизме, управляющем речевой деятельностью человека и ограничивающем размер речевых сообщений» (Калашникова 1981,с. 37). Под «психофизиологическим механизмом» Г.Ф. Калашникова понимает объем оперативной памяти человека. Исследовательница приходит к выводу, что более длинными являются предложения, в которых присутствует взаимодействие различных типов связи. Таким образом, длина предложения зависит не от количества составляющих его предикативных единиц, а от способа их сочетания. Автор статьи говорит о трех видах связи: подчинение, сочинение и соподчинение, которые в свою очередь связаны с тремя различными способами выражения мысли: развертывание логических комплексов вглубь, вширь и выделение различных свойств в одном объекте познания соответственно. В сложном предложении может быть один, два или три типа связи: среднее число предикативных единиц при наличии дополнительной связи увеличивается. Логично предположить, что с увеличением числа компонентов восприятие предложения должно затрудняться. Но «в языке участвуют механизмы, ограничивающие рост информационной емкости конструкции в процессе ее предикативного усложнения». Синтаксические модели, в которых присутствует несколько типов связи, облегчают восприятие сложных предложений, так как «в их составе происходит группировка компонентов (предикативных единиц) в более крупные блоки, которые выступают в качестве единиц оперативной памяти» (Калашникова 1981, с. 43).

Длина предложения является также важной стилевой и авторской характеристикой текста. Средний размер предложения в тексте определяется выбором тех синтаксических конструкций, которые существуют в определенном языке в каждый конкретный момент времени. Выбор синтаксических моделей организации предложения зависит, во-первых, от задач коммуникации, а во-вторых, от личных вкусовых предпочтений автора текста. Если язык обладает богатой системой синонимических средств, позволяющей один и тот же смысл передать по-разному, то такой фактор как автор текста проявляется наиболее явно. Индивидуальная свобода выбора структурных вариантов на синтаксическом уровне значительно выше, чем на каком бы то ни было другом уровне, за исключением, возможно, тех из них, которые лежат выше предложения. Таким образом, в синтаксическом строе авторские манера написания может быть проявлена лучше, чем на других уровнях. Разумно предположить, что синонимические синтаксические структуры имеют неравные размеры, что, безусловно, влияет на длину предложения.

По данным Т.М. Веселовской (Веселовская, 1987) средний размер предложения в художественной прозе составляет 21,1 слово, однако эта величина подвержена значительному разбросу в творчестве различных писателей. Применяя показатель средней длины предложения исследовательница делит писателей на «длинно пишущих» и «пишущих коротко». Такое разделение Т.М. Веселовская связывает с тем, какой тип прозы – синтагматический или расчлененный – преобладает в произведениях конкретного автора. Синтагматический тип прозы отличается развернутыми синтаксическими конструкциями, большим количеством сложных предложений. По данным Т.М. Веселовской в прозе А.А. Ананьева доля сложных предложений составляет 84,5%, в прозе Ю.В. Бондарева – 80,7 % (при среднем показателе 62,4%). Средний размер предложения в прозе А.А.Ананьева составил 34,4 слова, а в прозе Ю.В. Бондарева – 38,8 слова. В расчлененной прозе, напротив, преобладают короткие предложения, сложносочиненные предложения часто разрываются на несколько частей. Доля сложных предложений сокращается. Т.М. Веселовская приводит следующие данные: в прозе А.Б. Чаковского доля простых предложений составила 58,9%, в прозе В.И. Белова – 49,5%, в прозе А.С. Иванова – 48,4% (при среднем показателе 37,6 %). Т.М. Веселовская приходит к выводу о том¸ что «самой общей чертой распределения размера предложения в художественной прозе является его внутристилевое разнообразие» (Веселовская 1987, с. 9).

По мнению Г.А. Лесскиса «сама природа исследуемого явления не только допускает, но и предполагает некоторый произвол автора в вопросе, где поставить точку и начать новое предложение» (Лесскис 1963, с.92) Можно предположить, что складывающиеся таким образом случайные статистические характеристики отражают один из элементов авторского стиля. Г.А. Лесскис исследовал и сопоставил распределение предложений по длине в заведомо неоднородных текстах научной, публицистической, эпистолярной и художественной прозы. Автор приходит к выводу о том, что основные параметры, характеризующие распределение предложений (такие как самое длинное предложение в тексте, наиболее вероятный размер предложения, средний размер предложения, среднее квадратичное отклонение, медиана и др.) в сплошном авторском повествовании в художественной прозе лежат в статистически определенных интервалах, значительно ниже тех, которые характерны для научной прозы и несколько выше тех, которые характеризуют эпистолярную прозу. Проанализировав полученные данные, исследователь делает следующие выводы: во-первых, каждый жанр письменной речи характеризуется своим распределением размеров предложения и своими параметрами; во-вторых, типы распределения предложений различны для трех больших исторических периодов русской литературы – древней, XVIII в. и новой; в-третьих, распределение предложений по длине оказывается индивидуальным в каждом произведении и характеризует скорее не стиль автора в целом, а манеру автора в данном произведении. Последнее утверждение автор объясняет тем, что распределение предложений по длине различаются аналогично тому, как различается средний размер предложения между разными произведениями одного автора и между разными жанрами письменной речи. Г.А. Лесскис также установил, что размер предложения как случайная величина не может быть определена исключительно «авторским вкусом», она также зависит от других элементов стиля, композиционной структуры произведения и общего замысла автора.

Работы, посвященные исследованию длины предложения на материале других языков также ведутся в стилистическом ключе.

Размеры предложения в английской и средневековой прозе исследовал Дж.Юл. Он использует параметр длины предложения для атрибуции спорных текстов. Исследователь математически сформулировал гипотезу, что длина предложения и ее дисперсия вокруг средней величины является характеристикой авторского стиля. Каждый из обследованных текстов был разделен на две примерно равные части (A и B), а затем сравнивалось распределение предложений. Оказалось, что это распределение по всем параметрам приблизительно совпадает для каждого произведения в группах A и B, но всегда различаются для произведений разных авторов. Сравнивая, распределение предложений в спорных текстах и распределение предложений, принадлежащим писателям A и B, Дж. Юл смог установить авторство «Писем» Юниуса (Yule, 1939).

Б. Хорачек изучает особенности изменения размера предложения в речи разных персонажей. Объектом изучения Б. Хорачек стали поэтические произведения немецких авторов (Ноrаčеk, 1964).

П. Вашак исходил из предположения о том, что длина предложения и длина слова являются инвариантами в разных текстах одного автора и становятся характеристиками авторского стиля. Объектом исследования послужили статьи Яна Неруды в сборнике «Арабески». Вашак определяет предложение как расстояние между двумя точками и дифференцировано подходит к изучению речи автора, прямой речи и авторских предложений, входящих в прямую речь. Автор установил, что средний размер предложения в речи автора находится в интервалах 14-17 слов; средний размер вводных предложений в целом меньше среднего размера предложения в речи автора и составляет 8-9 слов. Исследователь приходит к выводам о том, что такие показатели как длина слова и длина предложения могут отражать характеристики авторских стилей. (Вашак, 1974).

С.А. Рылов вслед за Дж.Юлом и П. Вашаком, опираясь на статистические параметры текстов, ставит перед собой задачу установить авторство «Слова о полку Игореве». Исследователь сопоставляет размер предложения и некоторые другие статистические характеристики «Слова о полку Игореве» со статистическим параметрами текстов Н.М. Карамзина, А.Н. Радищева и произведений «Повесть временных лет», «Воинская повесть о походе князя Игоря на половцев в 1185 г.», «Житие протопопа Аввакума» и «Задонщина». Исследованные показатели в «Слове о полку Игореве» резко отличаются от показателей текстов XVII-XVIII вв., что позволяет А.С. Рылову сделать вывод о том, что «Слово о полку Игореве» не является подделкой XVII-XVIII вв. (Рылов 1978, 1980).

Исследования В.Г. Адмони, посвященные изучению размера предложения выполнены преимущественно в грамматическом аспекте. Ученый обращает внимание на взаимодействие двух факторов – стилистического и грамматического. Изучая размер предложения на материале немецкого языка В.Г. Адмони отмечает, что группа существительного, например, имеет способность к неограниченному развертыванию, так как существительное само по себе обладает высокой сочетаемостью, кроме того определения к существительным также могут обрастать зависимыми компонентами. (Адмони, 1963, 1966) Так, еще А. Мейе, описывая строй древних индоевропейских языков, говорил о том, что возможности увеличения размера предложения поистине безграничны (Мейе, 1938). Но, как отмечает В.Г. Адмони, это справедливо только с абстрактно смысловой точки зрения: «на каждом этапе языкового развития существуют определенные тенденции, лимитирующие использование этих возможностей, причем эти тенденции варьируются в зависимости от ряда факторов: прежде всего от стиля и тематического характера речевого отрезка, а также от грамматической функции существительного» (Адмони 1966, с 117).

Вслед за В.Г. Адмони связь между грамматическим и стилистическим аспектом попытались обнаружить и другие отечественные лингвисты, преимущественно на материале немецкого языка. Л.Н. Иноземцев на материале немецкого языка установил, что размер группы существительного зависит как от грамматических, так и от стилистических факторов (Иноземцев, 1965). К.Г. Павлова, изучая процесс субстантивации в немецком языке, приходит к выводу о том, что в языке научной публицистической прозы существительное может поглощать собой целые предложения, что ведет к увеличению емкости предложения (Павлова, 1971). С.А. Шубик сопоставляет размер предложения в художественной и научной прозе второй половины XX века. Средний размер предложения в научной прозе ниже, чем в художественной, при этом доля простых предложений в художественном стиле значительно выше. Исследователь отмечает, что на размер предложения как в научной, так и в художественной прозе влияет прежде всего размер элементарного предложения, так как длины сложных предложений в двух стилях примерно одинаковы (Шубик, 1973).

Таким образом, можно сделать вывод о том, что размер предложения в значительной степени показателен для выявления художественных стилей и стилей писателей

1.3. Описание лингвистических величин с помощью теоретических распределений

Одним из важнейших вопросов статистической лингвистики является вопрос о распределениях лингвистических единиц и анализе эмпирических распределений. При этом важно не только получить статистические данные, но и соотнести полученные данные и теоретические распределения. Такой подход позволит с одной стороны лучше понять природу распределения лингвистических единиц, а с другой – будет способствовать развитию теории распределений в статистике. По своей природе лингвистические распределения не являются чисто формальными моделями, их надо рассматривать как интерпретируемые знаковые системы. Выявление свойств и отношений, а также внутренних закономерностей, которые определяют характер лингвистического распределения в целом, должно сопровождаться качественным анализом результатов (Тулдава, 1987)

Законом распределения случайной величины называется правило, связывающее значения случайной величины и соответствующие им вероятности (Пиотровский, Бектаев, 1977). Анализ распределений экспериментальных данных является одной из основных задач в лингвистической статистике. Распределение случайной лингвистического величины можно считать квантитативной моделью лингвистического объекта – разумеется, прежде всего с количественной точки зрения. Хотя любая обнаруженная количественная закономерность в конечном счете подлежит качественному анализу. Поэтому рассматривая лингвистические распределения, сравнивая их между собой, выявляя их общие и особенные черты, можно получить информацию о лингвистическом объекте.

В литературе нам встретилось две возможные классификации лингвистических распределений (Мартыненко, 1988; Тулдава 1987). В первой представлен путь «от математики к лингвистике», во второй - «от лингвистике к математике». В классификации Г.Я. Мартыненко называются 12 аспектов, в рамках которых выделяются конкретные распределения. Каждый из аспектов представлен бинарным противопоставлением, например, гауссовость-негауссовость, строение-поведение, однопредметность-многопредметность, устойчивость-неустойчивость и т.д. Надо отметить, что список предложенных противопоставлений и соответствующих им типов распределений едва ли можно назвать полным. Не учитываются такие аспекты, как «дискретность-непрерывность», «частота-распространенность», «редкость-частость» и некоторые другие. Кроме того, предложенный перечень не устраняет частичной «синонимии» типов распределений, которая обусловлена причинно-следственными отношениями между бинарными противопоставлениями. Так ассиметрия распределения и его многовершинность могут быть внешним проявлением качественной неоднородности исследуемого признака, а степень устойчивости –количественное проявление неоднородности.

Ю.А. Тулдава считает лингвистическое распределение «моделью – описанием тех языковых объектов, которые можно представить как вероятностные системы» (Тулдава, 1987). Главное противопоставление в классификации Ю.А. Тулдава – теоретические - эмпирические распределения. Теоретические распределения описывают языковые явления, а эмпирические – речевые. Также в данную классификацию входит противопоставление пар динамических и статических распределений. Это четыре основных типа распределений, а пары дифференциальных и интегральных, дискретных и непрерывных, однообъектных и многообъектных, ранговые, комплексные и многомерные распределения исследователь относит к разновидностям основных типов распределений (Тулдава, 1987).

Далее рассмотрим основные виды распределений и статистических законов, которые встречаются в статистической лингвистике. Одним из таких законов является биноминальное распределение вида P(x)=. Оно может быть использовано для описания использования фонем, графем и их классов, а также грамматических категорий, при условии, что величины N и x не очень велики.

Надо отметить, что в реальности это условие редко соблюдается, что приводит к использованию других законов для описания лингвистических единиц. Распределение Пуассона часто применяют для описания редких лингвистических единиц. Примером может служить появление слова *море*, или словоформы *море*, или словосочетания *у самого синего моря* в тексте. Распределение Пуассона можно представить формулой: P(x, λ) = . Единственным параметром этого распределения является величина λ, которая отражает среднее число употреблений интересующего нас слова в тексте (Пиотровский, Бектаев, 1977).

Применяя для аппроксимации распределение Пуассона и биноминальное распределение, исследователь исходит из того, что речь представляет собой поток простейших лингвистических событий. Однако формирование слов, словосочетаний и предложений представляет собой взаимодействие как случайных, так и детерминированных событий. Детерминированными событиями, с точки зрения лингвистики, являются правила и нормы, заложенные в системе языка. Случайными событиями можно считать описываемые в тексте ситуации объективной действительности. Это значит, что для описания лингвистических объектов следует применять такие распределения, которые учитывали бы взаимодействие случайных и детерминированных процессов.

Одним из таких распределений является распределение Чебанова-Фукса (Пиотровский, Бектаев, 1977). Это распределение основывается на предположении о том, что в языке помимо величин, которые могут принимать любое положительное значение, а также значение нуль, существуют единицы, которые никогда не могут принимать значение нуль (в связи с законами функционирования системы языка). К таким величинам можно отнести слоги, морфемы, слова, предложения и т.д. Каждый слог состоит по крайней мере из одной фонемы, каждое слово состоит по крайней мере из одного слога или морфемы и т.д. Р.Г. Пиотровский называет первые величины бескомпонентными, а вторые, соответственно, небескомпонентными. Если небескомпонентная случайная величина встречается в тексте достаточно часто, то её можно описать с помощь распределения Пуассона, при условии, что заданный системой языка обязательный элемент будет исключен из рассмотрения. т.е. вместо аргумента x, следует взять разность x–1, вместо средней λ – разность λ – 1. Таким образом, формула Пуассона преобразуется формулу Чебанова-Фукса: P(x, λ – 1)=.

Распределение Чебанова-Фукса, аналогично распределению Пуассона, определяется единственным параметром – параметром λ. С помощью распределения Чебанова-Фукса описывались процессы образования слов в английском, немецком, румынском, арабском, древнегреческом, русском и грузинском языках. Но были обнаружены значительные расхождения опытных данных и теоретической модели для румынского, арабского и английского языков. Следовательно, можно сделать вывод, что распределение Чебанова-Фукса не является универсальным законом, описывающим процесс образования лингвистических единиц.

При описании лингвистических единиц, необходимо принимать во внимание детерминированные правила, которые задаются системой языка. Например, в румынском языке спектр вероятностей для длины слова в поэтической речи отличается от спектра в деловом стиле. Это говорит о том, что процесс образования лингвистических единиц характеризуется более тонкой статистической структурой, чем та, которая представлена распределением Чебанова-Фукса. Эту структуру более точно описывает распределение Фукса-Гачечиладзе (Пиотровский, Бектаев, 1977). Основная идея этого распределения заключается в том, что лингвистические единицы могут включать в себя ν обязательных элементов. Тогда, мы можем исключить эти обязательные элементы из функции распределения и получить распределение вероятностей появления случайной величины различной длины. Распределение Фукса-Гачечиладзе имеет следующий вид:

P(x, λ-.

Другим важным законом, описывающем случайные величины, является нормальное распределение или закон Гаусса. Нормальное распределение является предельным, к которому стремятся остальные теоретические распределения. Форма графика нормального закона идеально симметрична относительно центра. Реальные распределения лингвистических величины плохо аппроксимируются этим законом, т.к. часто характеризуются правосторонней ассиметрией. В связи с этим предпринимаются попытки описывать экспериментальные распределения с помощью распределений Кэпптейна, Шарлье, выравнивающих кривых Пирсона и Бородачева (Романовский, 1961). Такое моделирование основывается не на внешнем сходстве экспериментальных данных и теоретического распределения, а на лингвистической сущности случайного явления. С этой точки зрения особый интерес представляет логнормальное распределение. Основная идея этого распределения в лингвистическом смысле состоит в том, что значение случайной величины складывается не из независимых внутриязыковых и лингвистических величин, как в нормальном распределении. Значение случайной величины является результатом действия ряда причин, производящих последовательные «импульсы». Величина этих импульсов зависит во-первых от самих импульсов, во-вторых - от величины x, созданной действием этих импульсов. В этом случае нормально распределена не сама случайная величина, а ее логарифм: f(x)=.

Так выбор длины слова, например, зависит не только от семантики предыдущего контекста, но и от длины предшествующего слова. Г. Хердан использует логнормальное распределение для математической экспликации вероятностного построения словаря языка и его реализации в тексте. По его мнению это распределение отражает присущий языку принцип оптимального кодирования информации (Herdan, 1964). Дж.Б. Кэрролл предложил на основе логнормального распределения модель, описывающую частоту слов и связь между объемом словаря и текста (J.B. Carrol, 1970). Интересным представляется тот факт, что данная модель позволяет получить текстовую частоту из словарной. Однако теоретического обоснования использования именно этого распределения при описании частот слов в работе предложено не было.

Одним из исследователей, попытавшимся объяснить природу статистической закономерности в тексте, был Дж. Ципф (Zipf, 1949). Согласно закону Ципфа, произведение частот слов в текстах и их ранга (в порядке убывания частоты) для всех рангов в каком-либо тексте постоянно. Закон Ципфа можно представить формулой: r , где r- ранг слова, f –частота, с – приблизительно постоянное число, зависящее от длины текста. Исследователь объясняет эту связь существованием фундаментального закона человеческого поведения. Этот закон он назвал «принципом минимального усилия» по аналогии с принципом максимального действия в физике. Однако такая интерпретация соотношения частоты и ранга не позволяла строить математические модели, которые соответствовали бы эмпирическим данным (Биллингслей, 1969). Позднее закон Ципфа был уточнен Б.Мандельбротом, определившим понятие затрат через длину слова, что позволило ему получить математическое выражение экономии затрат. Эта формула, позволяющая лучше, чем закон Ципфа, описать взаимосвязь между частотой и рангом, известна под названием закона Ципфа-Мандельброта. Результаты исследования Мандельброта представлены в работе о «макролингвистике» (Mandelbrot, 1961). Макролингвистика, по его мнению, должна изучать «крупномасштабные» языковые явления. Роль макролингвистики по отношению к микролингвистики (грамматике), как считает Мандельброт, должна быть аналогична роли термодинамики по отношению к механики индивидуальных молекул газа: не все закономерности могут быть описаны на микроскопическом уровне. Исходя из этой точки зрения Мандельброт внес некоторые изменения в формулу закона Ципфа, чтобы приблизить ее к реально наблюдаемым данным. Преобразованная формула имеет следующий вид: , где r – ранг слова, p – относительная частота слова рангом, ρ и B – константы данного текста. При этом ρ дает поправку для слов низкого ранга, B – соответствует скорости убывания логарифма Р в зависимости от r. (Mandelbrot, 1961). Теоретическая модель Ципфа-Мандельброта позволила вывести так называемый объем Ципфа Z. С помощью этого параметра становится возможным сопоставить тексты различной длины с точки зрения их частотной структуры. Объем текста определяется как оптимальная длина и дает прогноз информационного потока в тексте (Orlov, Boroda, Nadarejsvili, 1982). Интересно, что действие закона Ципфа-Мандельброта было также обнаружено в музыкальных текстах (Boroda, 1982b) и в художественных картинах (рассматривалось распределение цветов) (Vološin, Orlov, 1082). Зависимость между длиной слова и его частотой была обнаружена и другими лингвистами. Например, Х. Гитэ обнаружил подобную зависимость для романских языков (Guiter, 1974). Подобные распределения встречаются также в математике, географии, социологии, психологии и других науках.

Г. Альтман (Altmann, 1980) обосновал зависимость числа значений слова от его длины. Эта гипотеза нашла подтверждение на материале десяти типологически разных языков (Fickermann, Marker-Jaeger, Rothe 1984).

Еще одним законом, описывающим структуру текста является закон Менцерата, который устанавливает зависимость между длиной слова и длиной слогов. Позже Альтман вывел уравнение, которое отображает зависимость между длиной компонента и длиной конструкта. (Altmann, 1980). Эта зависимость справедлива также и для других уровней описания.

Особый интерес представляют работы Э.Келиха (Kelih, 2004). Исследователь изучает распределение длин предложений на материале словенских тестов. В работе оценивалось распределение предложений по длине как в отдельных текста, так и в корпусах. Для описания были выбраны три вида распределений: отрицательное биноминальное, гипер-пуассоновское распределение и гипер-распределение Паскаля. Наилучший результат описания достигался отрицательным биноминальным распределением.

За последнее столетие было открыто большое статистических количество закономерностей в лингвистике. Несмотря на это, многие вопросы остаются открытыми. В основном они касаются интерпретации полученных зависимостей и получения некоторого теоретического фона, который бы объединил их и установил взаимосвязи.

Замечательно, что распределения, подобные тем, что были обнаружены лингвистами, встречаются и в других дисциплинах. Интерес для науки представляют те закономерности, которые могут быть перенесены на широкий класс аналогичных явлений. А. Эйнштейн в своем очерке "Физика и реальность" говорит о том, что "Критический ум физика не может ограничиваться рассмотрением только его собственной теории. Он не может двигаться вперед без критического рассмотрения значительно более сложной проблемы: анализа природы повседневного мышления" (цит. по: Грибанов, 1979, с. 10 ). Диалог между дисциплинами позволяет взглянуть на достижения других наук и попытаться каким-то образом использовать их для объяснения языковых явлений.

Одной из попыток объяснить лингвистические процессы через физические явления является работа Ю.А. Шрейдер (Шрейдер, 1967). В ней была высказана гипотеза, что каждому элементу *х* словаря *V* имеет смысл сопоставить некоторую характеристику сложности *Е(х)* – аналог энергии в термодинамике. Исследователь приходит к выводу, что вероятность появления в тексте ансамбля слова можно описать формулой Гиббса: =p()=k,

где p() – вероятность появления в данном тексте ансамбля слова , E()– энергетическая характеристика этого слова, k и α – постоянные, характеризующие данный ансамбль текстов. Таким образом, энергия, по мнению автора, является абстрактным понятием сложности.

Также пользуясь «термодинамической аналогией» Г.Д. Селезнев предложил модель исторической динамики лексической системы (Селезнев, 2005). По данным автора статьи распределение слов по числу значений подчиняется экспоненциальному распределению (по крайней мере для романских языков, на материале которых проводилось исследование). В термодинамике также встречается распределение этого вида – распределение Больцмана. Это распределение является следствием двух законов физики: закона сохранения энергии и закона возрастания энтропии. По аналогии с законами термодинамики Г.Д. Селезнев обозначил две закономерности, действительные уже для языка: закон сохранения общего, суммарного количества значений всех слов некоторой замкнутой языковой системы и закон возрастания энтропии распределения значений, содержащихся в тексте или словаре по однозначным, двузначным, трехзначным, …, *i-*значным словам. Исходя из этих соображений была предложена «семантическая температура» и «семантическая энтропия» языка.

В статистической теории информации, например, объяснения явлений по аналогии с физикой являются вполне привычными. В лингвистике, по видимому, такая методика пока не обрела популярность. Н.В. Крушевский писал о том, что в языке действуют законы, совершенно тождественные с законами, действующими в других сферах существующего, т.е. так называемые законы природы, не допускающие никаких исключений и уклонений. (Крушевский, 1998). В данной дипломной работе предпринята попытка описать распределение предложений по длине одним из известных в статистике распределений и интерпретировать его параметры, опираясь на аналогичное распределение в физике.

## ГЛАВА 2. РАСПРЕДЕЛЕНИЕ ПРЕДЛОЖЕНИЙ ПО ДЛИНЕ В ПРОИЗВЕДЕНИЯХ РУССКИХ ПИСАТЕЛЕЙ XIX-XX ВЕКОВ

2.1. Методика исследования

Квантитативный анализ отличается тем, что в нем применяется совокупность методов, которые уже разработаны в рамках других наук (математика, теория информации и т.д.) Одной из важнейших задач, которая стоит перед исследователем является обоснованный выбор и комбинация этих методов в исследованиях, особенно в исследованиях «на стыке наук», в том числе в работах по квантитативной лингвистике.

Ведущим методом нашего исследования является статистический метод, заключающийся в использовании подсчетов и измерений при изучении языковых явлений. Главным основанием применения статистического метода в исследовании следует признать «объективную присущность языку количественных признаков, количественных характеристик и внутреннюю зависимость, существующую между ними» (Головин, 1965). При этом лингвистическим объектам, имеющим качественную природу приписываются количественные оценки, вследствие чего эти явления могут быть изучены как количественные объекты.

Основным инструментом квантитативного анализа служит моделирование с помощью распределений, что позволяет изучить внутреннее строение и поведение объекта. Распределение, как обобщающее, интегральное понятие является важнейшей структурной характеристикой вероятностных систем.

Мы использовали метод «распределения исследуемого параметра в тексте». Его суть в следующем: в исследуемом тексте выделяется параметр, в нашем случае, длина предложения. Для того чтобы выяснить, как распределен данный параметр в тексте, высчитывается длина каждого предложения текста. Выделяются все предложения, которые имеют длину равную один и подсчитывается количество таких предложений. Аналогичная процедура выполняется с предложениями длиной в два слова и т.д.

Способы измерения размера предложения были описаны в работах В.Г. Адмони и Г.А. Лесскиса. Однако надо отметить, что единицы измерения длины предложения варьируются в разных работах. Чаще всего длину предложения измеряют в словах, хотя встречаются исследования, в которых единицей измерения длины предложения служит стихотворная строка (Horaček, 1964). Такой своеобразный способ подсчета возможен только в поэзии. Подсчет длины предложения в словах возможнее как в поэтических произведениях, так и в прозе. Единственный вопрос, который встает перед исследователем: что принять за границы слова? Во встретившихся нам исследованиях, посвященных изучению размера предложения, авторы по-разному отвечают на этот вопрос. П. Вашак считает слово «всякую группу графем между двумя пробелами» (Вашак, 1974). Г.А. Лесскис полагает, что при подсчете длины предложения необходимо исключить имена собственные и сложные грамматические формы (Лесскис, 1963). С.А. Рылов предлагает исключить из подсчета служебные слова. Если производить подсчет исключительно в знаменательных словах, то становится возможным отвлечься от морфологических свойств слова ... и представить протяженность речи на более высоком уровне - уровне позиционных свойств слова» (Рылов 1990, с.60).

В нашей работе в качестве единицы подсчета будет использоваться слово, как его определяет П. Вашак – «всякая группа графем между двумя пробелами». Это так называемое «графическое слово». Оно может материально совпадать со словом, определенным в соответствии с грамматическими критериями и интуицией носителей языка, но такого совпадения может и не быть. Это особые единицы, лишь косвенно соотносящиеся с «настоящим» словом, т. е. словом с грамматической точки зрения. Мы приняли данное определение слова исходя из практических задач: для упрощения автоматической обработки текста. Предложением считалась группа слов между двумя точками, или знаками, которые выступают в функции точки (вопросительный и восклицательный знак, многоточие), что также обусловлено удобством автоматической обработки текстов.

Описанный метод можно применить для получения различных характеристик одного текста. Так мы можем получить распределение предложений в тексте по количеству используемых слов, распределение предложений в тексте по количеству используемых в них знаках препинания и т.д.

Важно отметить, для получения значимого результата, выбираемый параметр должен «демонстрировать повторяемость, периодичность своих свойств, иметь закономерное распределение своих параметров и т.д.» (Садовский 1974, с.42-43). Так исследовать частоту использования предложений разной длины в словах, на наш взгляд правомерно, так как предложение каждой длины, встречаются в тексте достаточно часто. Именно повторяемость является важным условием квантификации языкового материала и его анализа при помощи математических методов. Кроме того, данный параметр должен слабо контролироваться автором на сознательном уровне. А если бы даже автор и попытался, то не смог бы долго его контролировать и в результате довольно быстро вернулся бы в прежнее устойчивое и типичное для него состояние.

Дополнительными методами исследования являются сравнительно-сопоставительный и описательный методы. Применение сравнительно-сопоставительного метода заключается в установлении сходств и различий между полученными распределениями предложений по длине. Применение описательного метода с характерными приёмами наблюдения, обобщения, обработки, интерпретации результатов наблюдения дает возможность передать особенности распределения предложений по длине в исследуемых тестах .

Мы предположили, что распределение предложений по длине в тексте может отражать специфику авторского стиля. В этом случае, во-первых, параметры должны сохранять "постоянное значение" для произведений, написанных одним автором. То есть, иметь небольшое отклонение от среднего значения (слабо колебаться). И во-вторых, данные параметры должны уверенно различаться у разных авторов. Индивидуальный стиль является результатом целого ряда сил, которые в разной мере формируют автора и направляют его к определенным повторяющимся схемам. С точки зрения длины предложения можно утверждать, что под влиянием индивидуальных стилеобразующих факторов автор создает определенную схему (“норму”), синтаксический шаблон, который реализуется в ходе постоянного нарушения и выполнения.

Но проблема определения специфики авторского стиля заданного текста осложняется тем, что стили авторов не однообразны, а зависят от хронологического, жанрового и тематического факторов. Так известно, что средний размер длины предложения примерно одинаков для текстов, написанных разными писателями в одном жанре, но существенно различен для произведений одного и того же писателя выполненных в разных жанрах.

Для определения факторов, влияющих на распределение предложений по длине, а также степени этого влияния был проведен многофакторный дисперсионный анализ.

Дисперсионным анализом называют статистический метод анализа результатов, зависящих от действия количественных или качественных факторов. Дисперсионный анализ может быть использован для выявления совместного влияния как лингвистических, так и экстралингвистических факторов, не поддающихся количественному измерению, на изучаемый показатель. Суть метода в том, что общая вариация результирующего показателя расчленяется на части, соответствующие раздельному и совместному влиянию различных качественных факторов, и остаточную вариацию, аккумулирующую влияние всех неучтенных факторов. Статистическое изучение этих частей позволяет делать выводы о том, действительно ли оказывает влияние на результирующий показатель тот или иной фактор, а также оценивать силу его влияния.

Следующим этапом исследования было определение теоретического распределения, которое наилучшим образом описывало бы экспериментальные данные. Полученные распределения вероятностей предложений по длине p(n), выраженной в количестве слов в предложении n, аппроксимировались различными стандартными распределениями.

Термин аппроксимация происходит от лат. "approximate" и означает приближение. Функция f является аппроксимацией функции g, если она в том или ином смысле близка к g. Аппроксимация позволяет выразить исследуемый объект через более удобные или известные в науке функции. (Носенко, 1981). В научных исследованиях аппроксимация применяется для описания, анализа, обобщения и дальнейшего использования эмпирических результатов.

Между величинами может существовать точная (функциональная) связь, когда одному значению аргумента соответствует одно определенное значение, и менее точная (корреляционная) связь, когда одному конкретному значению аргумента соответствует приближенное значение или некоторое множество значений функции, в той или иной степени близких друг к другу. При ведении научных исследований, обработке результатов наблюдения или эксперимента обычно приходиться сталкиваться со вторым вариантом. При изучении количественных зависимостей различных показателей, значения которых определяются эмпирически, как правило, имеется некоторая их вариабельность. Частично она задается неоднородностью самих изучаемых объектов неживой и, особенно, живой природы, частично обуславливается погрешностью наблюдения и количественной обработке материалов. Последнюю составляющую не всегда удается исключить полностью, можно лишь минимизировать ее тщательным выбором адекватного метода исследования и аккуратностью работы. Поэтому при выполнении любой научно-исследовательской работы возникает проблема выявления подлинного характера зависимости изучаемых показателей, этой или иной степени замаскированных неучтенностью вариабельности значений. Для этого и применяется аппроксимация - приближенное описание корреляционной зависимости переменных подходящим уравнением функциональной зависимости, передающим основную тенденцию зависимости (или ее "тренд")

Далее выбиралось то теоретическое распределение, качество аппроксимации которого было ближе к 1 (или к 100%).

2.1. Полученные результаты

На рисунках 1 и 2 представлены графики распределения предложений по длине в произведениях выбранных писателей XIX и века соответственно. Надо отметить, что пики большинства графиков различаются. Все распределения имеют колоколообразную форму, тем не менее у каждого из авторов форма графика индивидуальна. Это говорит о том, что на такой параметр как длина предложения оказывает влияние автор текста.

Рис.1. Распределение предложений по длине (в словах) в произведениях писателей XIX века.

Рис.2. Распределение предложений по длине (в словах) в произведениях писателей XX века.

Выбор синтаксических моделей организации предложения определяется, в том числе, и личными вкусовыми предпочтениями автора текста. Этот фактор особенно ярко проявляется в том случае, если язык обладает развитой системой синонимических средств, позволяющей один и тот же смысл передать по-разному. Разумно предположить, что синонимические синтаксические модели имеют разную длину, и, таким образом, употребление автором одних синтаксических моделей при игнорировании других совершенно определенным образом влияет на распределение предложений по длине. А.М. Пешковский указывает для примера 32 синонимических типа разных синтаксических конструкций. Главной сокровищницей синтаксической синонимики русского языка являются вариации словорасположения, свободный порядок слов русского языка (Пешковский, 1959).

Рис. 3. График с накоплением распределения предложений различной длины

в романах И.С. Тургенева

На рисунке 3 представлен график с накоплением для следующих романов И.С. Тургенева: *Новь, Дым, Рудин, Дворянское гнездо, Накануне, Отцы и дети.* При использовании такого способа отображения пики каждого графика складываются, что позволяет видеть наиболее значимые суммарные параметры исследуемого параметра (длина предложения).

На приведенном графике мы видим, что в исследуемых текста наиболее часто встречаются предложения длиной в 5 слов. Также наблюдаются общие пики для предложений длиной в 15 и 33 слова.

В приведенном выше графике предложение длиной в 5 слов занимает первое место по частоте употребления, но может оказаться, что предложение с данной длиной очень часто используются в некоторых текстах И.С. Тургенева, что и дает ему первое место, но в других текстах того же автора частота его использования относительно невелика. Для того, чтобы увидеть значимые пики, характерные для большинства текстов нами было подсчитано среднеквадратическое отклонение или стандартное отклонение (иногда его называют просто стандартом) исследуемого параметра в группе текстов.

Рис.4. Среднеквадратическое отклонение в романах И.С. Тургенева

Стандартное отклонение равно корню квадратному из суммы квадратов разностей между отдельными значениями случайной величины и ее средней (т.е. квадратов отклонений), поделенной на число этих разностей. Выборочное стандартное отклонение является оценкой генерального стандартного отклонения, причем точность этой оценки зависит, как и в случае средней, от числа значений случайной величины. Другими словами, это способ, с помощью которого возможно измерить степень изменчивости (разброса или рассеяния) чисел в наборе данных.

Среднеквадратическое отклонение высчитывалось по формуле:

 , (1)

где n – количество предложений в текстах;

– длина предложения;

- математическое ожидание случайной величины (длины предложения).

Среднеквадратическое отклонение показывает, каким может быть среднее отклонение выборочных частот от значения средней. График на рисунке 4 показывает, что наибольшему разбросу значений в исследованных текстах подвержены предложения длиной в 4 слова. Интересно, что этот пик совпадает с одной из вершин по длине. Также наблюдаются пики для предложения в 7, 9, 12 и 22 слова.

2.2. Анализ факторов, влияющих на распределение

Нами высказано предположение, что на особенности распределения предложений по длине помимо автора текста, могут влиять такие факторы, как длина текста и время написания произведения.

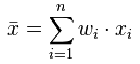
Для того чтобы подтвердить или опровергнуть гипотезу о влиянии длины текста, мы условно разделили все исследованные произведения И.С. Тургенева на 5 групп.

Для исследования были намеренно выбраны произведения одного писателя, чтобы на результат не повлияли авторские различия в текстах. Основанием для деления на группы послужил такой признак, как длина текста, которая измерялась в предложениях. Распределение произведений по группам, а также длина тексты в предложениях представлены в таблице 1.

|  |  |  |
| --- | --- | --- |
| Номер группы | Название произведения | Длина текста в предложениях |
| 1 | Поездка в полесье | 408 |
| Жид | 481 |
| Муму | 519 |
| 2 | Дневник лишнего человека | 828 |
| Яков Пасынков | 981 |
| Ася | 1034 |
| Часы | 1065 |
| Фауст | 916 |
| 3 | Первая любовь | 1327 |
| Затишье | 1659 |
| Степной король Лир | 1727 |
| Несчастная | 1868 |
| 4 | Рудин | 3108 |
| Накануне | 3658 |
| Дворянское гнездо | 2784 |
| Дым | 3269 |
| 5 | Отцы и дети | 4466 |
| Новь | 6292 |
| Записки охотника | 7598 |

Табл. 1. Распределение произведений И.С. Тургенева по группам

Далее было подсчитано среднее значение для каждой группы (Рисунки 4, 5). Среднее значение является самым распространенным статистическим показателем, который используется для измерения центра или середины в наборе числовых данных. Среднее рассчитывалось как среднее арифметическое взвешенное по формуле

, (2)

где – возможные значения длин предложений

– количество предложений длиной в относительных величинах

На рисунке 5 представлено среднее значение для группы 1 (короткие тексты до 600 предложений)

Рис.5. Среднее значение длины предложения для группы 1

В первой группе представлены тексты малой длины (около 500 предложений). На графике можно отметить две вершины 4 и 6. Нами высказана гипотеза о том, что здесь накладываются два распределения, два разных множества. Эта тенденция прослеживается и далее по графику на промежутке 7-9, 10-12 можно наблюдать “плечи”, и далее также наблюдается скачкообразное поведение графика. Все это лишь подтверждает нашу гипотезу о двух накладывающихся распределениях. Мы предположили, что эти два распределения – распределение предложений, содержащих слова автора и распределение предложений, содержащих прямую речь, которые при обработке текстов не исключались, так как именно в прямой речи отражается своеобразие речи героев, а значит, и авторов, которые являются первичными генераторами этих текстов. Но поскольку прямая речь и слова автора имеют разную структуру и, что для нас наиболее интересно, различную длину (предложения, содержащие прямую речь в среднем короче, чем предложения, содержащие слова автора), их можно отнести к разным множествам.

Рис. 5. Среднее значение длины предложения для групп 2,3,4,5

На данном графике представлены средние для остальных групп. Здесь также можно увидеть признаки наложения двух распределений. Однако здесь они видны менее отчетливо по сравнению с первой группой. На графике второй группы отчетливо видны так называемые “зубчики”. Это также говорит о наложении двух распределений. На графиках третьей, четвертой и пятой групп эти “зубчики” сглаживаются. Вероятно, это связано с тем, что в этих группах находятся тексты большей длины. Количество предложений, содержащих авторскую речь в таких текстах значительно превышает количество предложений с прямой речью. Этим и объясняется “сглаживание” так называемых “зубчиков”.

Таким образом, наша гипотеза о том, что *размер текста* влияет на характер распределение предложений по длине в текстах одного автора подтвердилась.

Мы предположили, что на длину предложения может также влиять такой фактор, как *время написания* произведения. Здесь можно рассмотреть два вида влияния: во-первых, влияние времени написания произведения в рамках творчества одного автора. Автор, как и любой человек, меняется в течение жизни, меняется стиль написания произведений, а значит, могут изменятся и статистические особенности его произведений. Во-вторых, влияние времени написания произведения на более крупном промежутке времени. Эти изменения отражают изменения в синтаксических конструкциях, которые преобладают на разных промежутках времени.

Для того, чтобы подтвердить или опровергнуть гипотезу об изменении средней длины предложения в рамках творчества одного автора, мы упорядочили произведения И.С. Тургенева, Ф.М. Достоевского, Л.Н. Толстого в хронологическом порядке. Всего было проанализировано 35 произведений, куда вошли и рассказы, и повести, и романы. Затем была подсчитана средняя длина предложения в словах для каждого произведения и построены диаграммы.

Диаграмма на рисунке 7 показывает, что средняя длина большинства текстов И.С.Тургенева лежит в интервале от 10,5 до 13. Это говорит о том, что средняя длина является определенным статистическим выражением стремления автора к использованию определенной синтаксической нормы.

Нужно отметить, что в длинных текстах выравнивается различие в частотах употребления предложений определенной длины. В более кратких текстах это различие выражается наиболее явно: выбиваются произведения со средней длиной предложения от 7 до 8,5 слов. Эта длина характерна для маленьких произведений И.С. Тургенева, написанных в начале его творчества (“Жид”, “Бретер”, “Петушков”). Произведения со средней длиной предложения больше 14 слов являются достаточно крупными произведениями И.С. Тургенева. Возможно, этим и объясняется такая большая средняя длина.

Рис. 7. Распределение средней длины предложения в произведениях

И.С. Тургенева

На рисунках 7 и 8 представлены распределения текстов Ф.М. Достоевского и Л.Н. Толстого. Произведения также расположены в хронологическом порядке.

|  |  |  |
| --- | --- | --- |
| Рис.8. Распределение средней длины предложения  в текстах Ф.М. Достоевского |  |  |
|  |  |  |
|  |  |  |
| Рис.9. Распределение средней длины предложения в текстах Л.Н. Толстого |  |  |
|  |  |  |
| Надо сказать, что единой тенденции не было обнаружено. Это говорит о том, что средняя длина предложения является статистическим выражением стремления автора к использованию определенной синтаксической нормы. Кроме того, на данный параметр влияет множество других трудно учитываемых факторов, что, безусловно, усложняет установление закономерности на протяжении короткого промежутка времени в рамках творчества одного писателя.  Для того чтобы проверить гипотезу о влиянии времени написания на длину предложения в рамках XIX-XX вв. все произведения были распределены на 6 групп, в зависимости от времени создания и подсчитана средняя длина предложения для каждой из групп. В таблице 2 представлены исследованные хронологические срезы. Основанием для деления на группы послужило время написания произведения. Всего было проанализировано 96 произведений 20 авторов. Общее количество исследованных предложений составляет 1246074 шт.   |  |  |  |  | | --- | --- | --- | --- | | № группы | Век | Треть века | Годы | | 1 | XIX | 1 | 1800-1833 | | 2 | XIX | 2 | 1834-1866 | | 3 | XIX | 3 | 1867-1899 | | 4 | XX | 1 | 1900-1933 | | 5 | XX | 2 | 1934-1966 | | 6 | XX | 3 | 1967-1999 |   Табл.2. Хронологические срезы  Исследованные произведения и распределение их по группам представлено в Приложении 1. Затем была подсчитана средняя длина и доверительный интервал для каждой группы. Результаты представлены в таблице 3.   |  |  |  |  | | --- | --- | --- | --- | | Срез по времени | Среднее | 95% доверительный интервал | | | Нижняя граница | Верхняя граница | | первая треть XIХ в. | 13,339 | 12,930 | 13,748 | | вторая треть XIХ в. | 13,399 | 13,301 | 13,498 | | третья треть XIХ в. | 12,626 | 12,560 | 12,692 | | первая треть XX в. | 10,748 | 10,669 | 10,828 | | вторая треть XX в. | 13,318 | 13,228 | 13,409 | | третья треть XX в. | 12,555 | 12,364 | 12,745 |   Табл. 3. Среднее значение длины предложения и доверительный интервал для исследованных групп  Для наглядности проиллюстрируем полученные данные (Рисунок 10)  Рис.10. Изменение средней длины предложения на протяжении XIX-XX вв.  На диаграмме видно, что на протяжении XIX - начала XX века происходит снижение исследуемого параметра. По данным Веселовской Т.М. на этом этапе развития литературы происходит не только уменьшение размера простого предложения, но также и уменьшение доли сложного предложения при одновременном уменьшении среднего размера сложного предложения (Веселовская, 1987). В первой трети XX века получает распространение актуализирующая проза, с чем, по-видимому, можно связать снижение размера предложения. С 1940-х годов синтагматическая проза существует на ряду с расчлененной, что может служить объяснение некоторого увеличения среднего размера предложения во второй половине XX века.  Актуализирующая проза противопоставляется синтагматической. Это противопоставление строится на основании разных типов повествования. Если в синтагматической прозе преобладают подчинительные связи, т.е. «организация предложения представлена как иерархическая система, где элементы сообщения находятся в определенной взаимосвязи» (Акимова, 1990), а границы предложения и высказывания, как правило, совпадают, то актуализирующая проза характеризуется синтаксической расчлененностью высказывания. Размер предложения в этом типе прозе значительно сокращается, что приводит к несовпадению границ предложения и высказывания: «синтагматическая цепочка нарушается в различных синтаксических звеньях» (Акимова, 1973). Уменьшается количество сложноподчиненных предложений, так как актуализирующая проза основана не на логике, а на риторике. Преобладают сложносочиненные предложения и бессоюзные. Этот новый тип прозы Г.Н.Акимова связывает с тенденцией к аналитизму в синтаксисе.  Итак, можно сделать вывод, что средняя длина предложения изменяется с течением времени и отражает изменения в синтаксисе.  Для подтверждения гипотезы о том, что на распределение предложений по длине в прозе влияют такие факторы, как автор теста, время написания произведения и длина текста (в предложениях), их взаимодействие, а также для измерения силы влияния каждого из факторов нами был проведен многофакторный дисперсионный анализ. Дисперсионный анализ проводился с помощью пакета статистических программ SPSS 13 на материале 96 произведений 20 авторов. Группировка текстов по времени написания проводилась аналогично, по срезам (таблица 4). По длине тексты распределялись по следующему принципу:   |  |  | | --- | --- | | Длина текста  (в предложениях) | № группы | | до 1000 | 1 | | 1001-3000 | 2 | | 3001-5000 | 3 | | 5001-7000 | 4 | | 7001-9000 | 5 | | 9001-более | 6 |   Табл. 4. Распределение по группам в зависимости от длины текста  Распределение произведений по группам с указанной длиной в представлено в Приложении 2.  Схема многофакторного эксперимента представлена на рисунке 11. Анализ влияния факторов на распределение предложений по длине был проведен в два этапа. На первом этапе важно было выяснить, оказывают ли обозначенные нами факторы существенное влияние на распределение предложений по длине, т.е. являются ли эти факторы статистически значимыми. Исследовалось как наличие или отсутствие влияния на результирующий признак каждого качественного фактора, так и наличие или отсутствие взаимодействия пар факторов. |  |  |

Рис.11. Схема многофакторного анализа

Прежде чем проводить дисперсионный анализ, необходимо установить однородность дисперсий между группами. Для этого был проведен тест Ливиня, т.к. его результаты не зависит от предположения о нормальности распределения. Его результаты представлены в таблице 5.

|  |  |  |  |
| --- | --- | --- | --- |
| F | ст.св.1 | ст.св.2 | Р |
| 556,186 | 51 | 492389 | <0,00001 |

Табл. 5. Критерий Ливиня проверки равенства дисперсий

Тест Ливиня на равенство дисперсий показывает значимый результат со значением вероятности ошибки р<0,00001≤0,05. Это означает, что выполняется условие однородности дисперсий между группами, которое является основной предпосылкой для возможности проведения дисперсионного анализа.

Результаты дисперсионного анализа приведены в таблице 6.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Иcточник | Сумма квадратов типа III | ст.св. | Средний квадрат | F | P |
| *Главные эффекты* | | | | | |
| Время написания произведения | 116517,465 | 4 | 29129,366 | 258,918 | <0,00001 |
| Длина произведения | 22932,005 | 5 | 4586,401 | 40,766 | <0,00001 |
| Автор | 701507,465 | 14 | 50107,676 | 445,385 | <0,00001 |
| *Двухфакторные взаимодействия* | | | | | |
| Время написания произведения \* Длина произведения | 229013,497 | 2 | 114506,749 | 1017,800 | <0,00001 |
| Время написания произведения \* Автор | 87360,011 | 3 | 29120,004 | 258,835 | <0,00001 |
| Длина произведения \* Автор | 132459,418 | 5 | 26491,884 | 235,475 | <0,00001 |

Табл. 6. Результаты дисперсионного анализа

Показатели таблицы позволяют сделать вывод, что все три фактора: автор, время написания произведения и длина текста оказывают статистически значимое влияние на длину предложения, также как и взаимодействие пар факторов (p≤0,05). Когда мы оцениваем взаимодействие двух факторов, то фактически нас интересует, будет ли одинаковым влияние на зависимую переменную одного из факторов на всех уровнях другого фактора. Очевидно, что если это влияние неодинаково, то второй фактор каким-то образом опосредует влияние первого фактора, и, следовательно, можно говорить о существовании взаимодействия между ними.

Следующим этапом анализа является определение степени влияния каждого из факторов, а также их взаимодействия на изучаемую величину. На рисунках 12 и 13 представлены факторы и их взаимодействия по степени их влияния на исследуемый признак: длину предложения.

Рис. 12. Распределение основных факторов по силе влияния на длину предложения

Рис. 13. Распределение взаимодействий факторов по силе влияния.

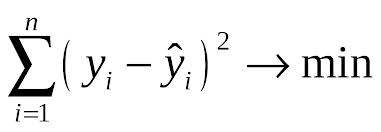
Из результатов дисперсионного анализа следует, что наибольшее влияние на распределение предложений по длине оказывает такой фактор, как автор текста. Кроме того было обнаружено значительное влияние взаимодействия факторов. Межфакторное взаимодействие представляет особый интерес: такой фактор как длина текста в отдельности может оказывать на исследуемую характеристику не такое значительное влияние как в сочетании с другими факторами.

2.3. Аппроксимация экспериментальных данных

Для нас было важно понять, не может ли данное распределение быть описано некоторой аналитической формулой для того, чтобы определить природу процесса Моделирование данных с помощью какого-либо теоретического распределения позволяет перенести исследованные данные на более высокую ступень обобщения. Известна важная роль теоретических распределений в отображении закономерностей материального мира, и можно предположить, что в отношении лингвистических объектов такое моделирование открывает широкие возможности не только для решения прикладных задач, но и для выявления глубоких квантитативных закономерностей структуры и функционирования языка. (Тулдава, 1987).

Задача аппроксимации (сглаживания) экспериментально полученных зависимостей заключается в подборе аппроксимирующей функции и определении её параметров. При этом необходимо, чтобы аппроксимирующая кривая наилучшим образом отражала экспериментальные данные, т. е. общую тенденцию экспериментальной зависимости, и вместе с тем, сглаживала случайные отклонения.

Аппроксимация была проведена средствами программного пакета Excel. Использовался метод наименьших квадратов. Суть метода заключается в подборе таких параметров теоретического распределения, при которых сумма квадратов отклонений наблюдаемых значений от теоретических является минимальной:



, (3)

где - экспериментальное значение вероятности;

– теоретическое значение вероятности.

Исходя из колоколообразной ассиметричной формы графиков распределений предложений по длине p(n) для их аппроксимации были выбраны следующие характерные распределения: гамма-распределение, распределение Вейбулла, распределение Пуассона. Далее высчитывался коэффициент достоверности для каждого распределения.

Коэффициент достоверности аппроксимации  показывает степень соответствия теоретического распределения исходным данным. Его значение может лежать в диапазоне от 0 до 1. Чем ближе  к 1, тем точнее теоретическое распределение описывает имеющиеся данные.

Качество аппроксимации высчитывалась по формуле:

; (4)

,

,

где  - экспериментальное значение вероятности;

 - теоретическое значение вероятности;

n – количество предложений.

Аппроксимация проводилась на материале текстов М.А. Булгакова, И.А. Гончарова, М.Горького, И.С. Тургенева, К.А. Федина, М.А. Шолохова, М.Е. Салтыкова-Щедрина, А.И.Эртеля, И.С. Шмелева. Полученные данные представлены в таблице 7. Наилучшее качество аппроксимации достигается для гамма-распределения.

|  |  |  |  |
| --- | --- | --- | --- |
| Автор | Качество аппроксимации | | |
| γ-распределение | распределение Вейбулла | распределение Пуассона |
| Булгаков | **99,62%** | 99,53% | 97,50% |
| Гончаров | **99,03%** | 98,97% | 96,78% |
| Горький | **99,31%** | 99,25% | 96,65% |
| Тургенев | **96,94%** | 96,84% | 95,92% |
| Федин | **99,18%** | 99,15% | 98,34% |
| Шолохов | **99,23%** | 99,10% | 98,57% |
| Щедрин | **99,78%** | 99,67% | 98,54% |
| Эртель | **97,31%** | 97,25% | 95,37% |
| Шмелев | **98,95%** | 98,90% | 96,87% |

Табл.7. Качество аппроксимации экспериментального распределения предложений по длине различными теоретическими распределениями

Вид гамма-распределения можно представить следующей формулой:

 (5)

Запись означает, что речь идет о семействе функций, каждая из которых определяется двумя параметрами — α и β. В этом смысле гамма-распределение является двухпараметрическим семейством распределений. Параметр α>0 задает масштаб кривой, а β>0 — форму. В отличие от графиков, формулы позволяют проведение различных операций, которые могут служить основанием для раскрытия новых параметров исследуемого явления. Наибольшая достоверность аппроксимации R2 достигается при значениях подбираемых параметров α и β, представленных в Приложении 3.

Пример аппроксимации распределения предложений по длине для текстов М.Е. Салтыкова-Щедрина Г-распределением приведен на рисунке 14. Модельные данные изображены розовой линией, экспериментальные – синим пунктиром. Поскольку гамма-функция аппроксимирует статистические данные с хорошей точностью, возникает вопрос об интерпретации как функционального вида, так и параметров распределения.

α=1,83; β=8,12; =99,78%

Рис.14. Распределение предложений по длине (по данным текстов М.Е. Салтыкова-Щедрина).

Можно подумать, что рядом с автором стоит толпа математиков, и они разрешают составлять ему предложения из строго определенного количества слов, чтобы в итоге получилось такое распределение. Но это, конечно же, не так. Возможное объяснение тому, что исследованные данные подчиняются именно Г- распределению кроется в особенностях человеческой психологии. Согласно В.Г. Суходольскому, «…большинство случаев временных интервалов, так или иначе характеризующих психическую активность, подчиняются именно этому распределению. Ему следуют также общие (суммарные) затраты времени на последовательность рабочих действий, слабо зависящих (или независящих друг от друга), причем величина этих затрат для разных действий может варьироваться». В рассматриваемом нами случае упоминаемые временные интервалы отражают время, затрачиваемое на написание, произнесение или чтение предложений того или иного автора (Суходольский, 1972). Однако, не стоит забывать, что мы имеем дело с вероятностным подходом к изучению длины предложения. При таком подходе мы должны исходить из вероятностной концепции причинности, согласно которой «причинность есть нечто, могущее присутствовать в большей или меньшей степени, а не только быть или не быть» (Винер 1964, с.309). Таким образом, можно предположить, что существует определенная связь между особенностями такой квантитативной характеристики текста как распределение предложений по длине и некоторыми закономерностями деятельности мозга.

Положительная асимметрия распределения случайной величины говорит о том, что на случайную величину влияет множество факторов, при этом воздействие одного или нескольких из них значительно превосходит по силе воздействия все остальные. В нашем случае основным фактором, который по силе воздействия превосходит остальные, является автор текста.

Гамма-распределение применялось также для описания времени жизни, продолжительности производственного цикла, величины персональных доходов, поведения населения в близком к устойчивому равновесном состоянии, интервала между наступлениями событий и времени обслуживания. Интересно отметить, что по данным Г.Д. Селезнева гамма-распределению подчиняются также распределения длин слов в романских языках (Селезнев, 2011)

В результате, мы можем теперь сопоставить каждому произведению два числа alpha и beta, отражающие характер поведения функции длины предложения. Пользуясь параметрами α и β можно расположить каждого автора на плоскости (Рисунок 15). Вертикальная и горизонтальная оси проведены в соответствии со значениями параметров *alpha* и *beta*. Эти параметры отражают характер поведения графика распределения предложений по длине. Назовем *alpha* - параметром масштаба графика распределения, а *beta* - параметром формы графика распределения предложений по длине. Таким образом, для нас наиболее важен параметр *beta*. Именно он, как показали статистические эксперименты, лучше «чувствует» характер распределения графика. Параметр *alpha* отвечает скорее за масштаб графика. Но, тем не менее, необходимо отметить, что между параметрами α и β существует определенная связь. На графике прослеживается тенденция возрастания параметра α при уменьшении параметра β.

Рис.15. Расположение авторов в соответствии с параметрами α и β

На рисунке видно, что из общей картины несколько выделяются произведения И.С. Шмелева, и В.М. Шукшина. У этих авторов нетипично большой коэффициент α и очень маленький β. Это говорит о том, что для них нетипично употребление очень длинных предложений, и в то же время они очень часто использует короткие предложения и предложения средней длины. Интересно заметить, что их произведения отличались от всех распределений и на рисунках 1 и 2, где были изображены графики распределения длин предложений.

Так же выделяются произведения В.В. Набокова. У него нетипично маленький коэффициент α и сравнительно большой коэффициент β. Интересно, что сразу за ним (по величине параметра β) идет Л.Н. Толстой, который традиционно относится к авторам, предпочитающим длинные предложения. Наиболее близко друг к другу расположены М.А. Булгаков, А. Веселый, К.А. Федин и А.И. Эртель. Это означает, что распределения предложений по длине в их текстах похожи и форма графиков практически одинакова. Возможно расстояние, на котором находятся друг от друга точки, и выражает «степень синтаксической индивидуальности» автора.

2.4. Интерпретация полученных результатов

Интересно отметить, что распределения подобные Г-распределению встречается и в физике. Вслед за Г.Д. Селезневым, который провёл аналогию между распределением скоростей частиц в идеальном газе, описываемое формулой распределения Максвелла и распределением слов по длине (Селезнев, 2011) , нами была проведена аналогия между распределением длин предложений и распределением плотности излучения нагретого физического тела.

Закон Вина, который описывает плотность излучения нагретого физического тела можно представить следующей формулой (Роджерс, 1972):

, где (6)

Uν — плотность энергии излучения

h — постоянная Планка,

k — постоянная Больцмана,

c — скорость света в вакууме

ν — частота излучения,

T — температура излучающего тела

Преобразуем формулу Г-распределения в формулу закона Вина. Произведем замену переменных:

;

;

n = ;

Г =*;*

p(n)=.

Подставим эти значения в формулу Г-распределения:

;

;

Таким образом, мы подтвердили тот факт, что между распределением предложений по длине, которое подчиняется законам гамма-распределения, и распределением плотности излучения нагретого физического тела прослеживается глубокая аналогия. При этом величина аналогична вероятности встречаемости предложения определенной длины в тексте. А параметр β Г-распределения предложений в формуле (5) обратно пропорционален температуре *kT*. Это дает нам возможность сопоставить два явления: длину волны и длину предложения (рисунок 16, 17). На рисунке 9 видно, что длина волны обратно пропорциональна температуре *kT*, т.е. чем больше длина волны, тем меньше величина *kT*. Исходя из того, что , мы можем предположить, что длина предложения и параметр β находятся в прямо пропорциональной зависимости.

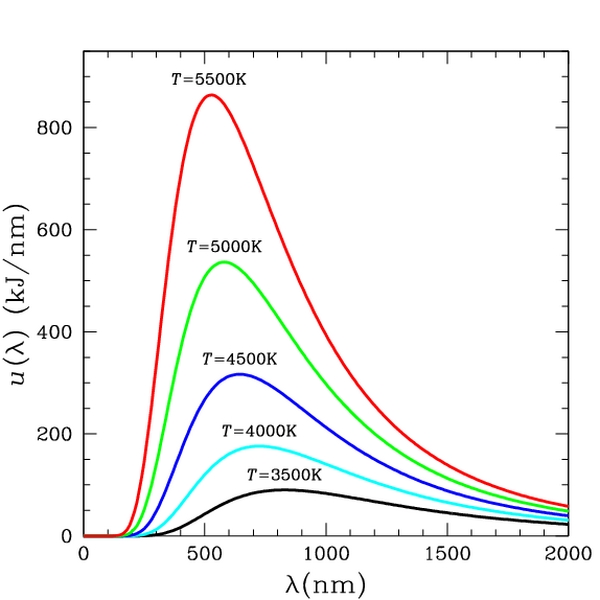


Рис.16. Зависимость между плотностью излучения () и длиной волны λ.

Рис.17. Зависимость между вероятностью встречаемости предложения и их длиной в словах.

Для того, чтобы подтвердить нашу гипотезу, построим график изменения длины предложения и коэффициента β (Рисунок 18). По оси абсцисс отложим значения коэффициента β, а по оси ординат – значение средней длины предложения.

Рис. 18. Зависимость между средней длиной предложения и коэффициентом β

Среднее арифметическое значение не совсем объективно отражает статистические данные, потому что на него очень просто повлиять с помощью выбросов (очень большого или очень маленького значения в наборе данных, которое не является типичным). Поэтому средняя длина предложения рассчитывалась как математическое ожидание случайной величины. Математическое ожидание является мерой концентрации (центрирования) случайной величины. Смысл математического ожидания заключается в следующем: это точка на числовой оси, являющаяся координатой центра тяжести распределения единичной массы на этой числовой оси:=1.

Математическое ожидание высчитывалось по формуле:

мат ожидание.png

, где (7)

– возможные значения случайной величины (в нашем случае – длины предложений);

– вероятности событий, что случайная величина X примет значение (т.е. вероятность появления в тексте предложения длиной ).

На графике видно, что коэффициент β находится в прямой зависимости от средней длины предложения, т.е. наша гипотеза подтвердилась. Принимая во внимание тот факт, что мы обозначили , можно ввести параметр «синтаксической температуры» текста, который будет равен .

Подобный параметр был получен в работах Г.Д. Селезнева (Селезнев, 2005; Селезнев, 2007; Селезнев, 2011): была выведена «семантическая» и «функциональная температура» для романских языков по аналогии с распределением Больцмана и Максвелла соответственно.

Анализ «синтаксической температуры» можно провести не только на материалах художественной прозы, но, например научной и деловой. Мы получим такого же типа распределения и тогда, можно будет оценить как изменяется «синтаксическая температура» в зависимости от стилей и жанров. Можно предположить, что «синтаксическая температура» текстов в научном и официально-деловом стиле будет ниже, чем в художественной прозе, кроме того будет меньше варьироваться в зависимости от автора текста. Это связано с тем, что во-первых, в подобных текстах необходимо исключить возможность неверного толкования, что ведет к большому количеству уточняющих конструкций, и, как следствие, увеличению средней длины предложения. Во-вторых, в научном и деловом стиле выбор синтаксических конструкций ограничен существованием так называемых «клише», что ведет в снижению проявления авторских особенностей в тестах.

Закон Вина, с помощью которого был выведен параметр «синтаксической температуры» описывает зависимость между длиной волны и потоком излучаемой энергии для абсолютно черных тел. Абсолютно чёрное тело — физическая идеализация, применяемая в термодинамике, тело, поглощающее всё падающее на него электромагнитное излучение во всех диапазонах и ничего не отражающее. Несмотря на название, абсолютно чёрное тело само может испускать электромагнитное излучение любой частоты и визуально иметь цвет. Спектр излучения абсолютно чёрного тела определяется только его температурой. Сплошной спектр излучения звезды близок к излучению абсолютно чёрного тела с температурой, равной температуре её фотосферы, которую можно оценить по закону Вина.

В 1890-1924 годах в Гарвардской обсерватории была разработана Современная (гарвардская) спектральная классификация звёзд (Таблица 8).



Табл. 8. Основная (Гарвардская) спектральная классификация звезд.

Данная классификация является температурной классификацией, основанной на виде и относительной интенсивности линий поглощения и испускания спектров звёзд. Спектральный класс звезды указывает на температуру фотосферы, которая коррелирует с абсолютной массой. Спектральные классы О, В, А часто называют горячими или ранними, классы F и G – солнечными, а классы К и М – холодными или поздними спектральными классами (Бакулин, 1983).

Принимая во внимание тот факт, что при соотнесении закона Вина и гамма-распределения параметру температуре *kT* соответствовал параметр , мы можем условно разделить исследованных писателей по спектральным классам в зависимости от величины коэффициента или «синтаксической температуры» (Рисунок 19).

Мы не учитывали голубой и бело-голубой спектральные классы, так как их температура значительно превышает минимальную (в десять и в двадцать раз соответственно). Белый, желто-белый, желтый, оранжевый и красный спектр наносились на график пропорционально распределению температур в Гарвардской спектральной классификации звезд.

|  |  |
| --- | --- |
| **1** | Шукшин В.М |
| **2** | Шмелев И.С. |
| 3 | Пикуль В.С. |
| 4 | Пушкин А.С. |
| 5 | Шолохов М.А. |
| 6 | Эртель А.И. |
| **7** | А. Веселый |
| **8** | Булгаков М.А. |
| **9** | Федин К.А |
| **10** | Лажечников И.И. |
| **11** | Гончаров И.А. |
| **12** | Тургенев И.С. |
| **13** | Горький М. |
| **14** | Достоевский Ф.М. |
| **15** | Нарежный В. Т. |
| **16** | Салтыков-Щедрин М.Е. |
| **17** | Азольский А. А. |
| **18** | Пелевин В. О. |
| **19** | Толстой Л.Н. |
| **20** | Набоков В.В. |

Рис. 19. Распределение авторов по «спектральным классам».

На графике видно, что большинство авторов сконцентрировано в «желто-оранжевом спектре». Это говорит о том, что распределения предложений по длине данных авторов являются своеобразной нормой для языка. В то время как распределения В.М. Шукшина и И.С. Шмелева являются скорее исключениями и достаточно сильно отличаются по форме графика от остальных авторов. Для их произведений характерна высокая «синтаксическая температура». Средняя длина предложения в их текстах короче, чем в произведениях остальных писателей и вероятность появления короткого предложения больше. Возможно, это связано с тяготением В.М. Шукшина и И.С. Шмелева к сказовой (разговорной) форме повествования. Интересно, что самыми «холодными» оказались произведения наиболее рациональных авторов нашей выборки – В.В. Набокова и Л.Н. Толстого, у которых самые высокие показатели средней длины предложения. Для этих писателей важно передать все богатство, все оттенки мысли и переживания, показать мысль в ее динамике и становлении. Такие авторы, как А.С. Пушкин и В.С. Пикуль, по видимому, тяготеют к выражению мыли более компактным способом, собранно и сконцентрировано.

Оппозицию «высокая ~ низкая» синтаксическая температура можно соотнести с широко использовавшейся ещё Ш. Балли оппозицией «эмоциональность ~ рациональность». Эмоциональная речь кратка и дискретна, рациональная – пространна и континуальна. В традиционных для русской литературы терминах это оппозиция «сердца» и «ума». Поскольку одно не встречается без другого, а лишь в той или иной степени преобладает, наша классификация позволяет разместить авторов на шкале «*сердцеумия*».

В таблице 9 представлены значения средней длины предложения, коэффициента β и синтаксической температуры (1/β) для исследованных авторов.

|  |  |  |  |
| --- | --- | --- | --- |
| Автор | Средняя длина предложения | β | «Синтаксическая температура» |
| Шукшин В.М. | 5,55 | 2,84 | 0,35 |
| Шмелев И.С. | 8,64 | 3,43 | 0,29 |
| Пикуль В.С. | 10,50 | 5,25 | 0,19 |
| Пушкин А.С. | 12,09 | 5,51 | 0,18 |
| Шолохов М.А. | 11,63 | 6,44 | 0,16 |
| Эртель А.И. | 11,66 | 6,59 | 0,15 |
| А. Веселый | 10,40 | 6,88 | 0,15 |
| Булгаков М.А. | 10,25 | 6,92 | 0,14 |
| Федин К.А. | 10,70 | 6,98 | 0,14 |
| Лажечников И.И. | 15,09 | 7,34 | 0,14 |
| Гончаров И.А. | 13,08 | 7,36 | 0,14 |
| Тургенев И.С. | 12,60 | 7,49 | 0,13 |
| Горький М. | 10,92 | 7,74 | 0,13 |
| Достоевский Ф.М. | 13,56 | 7,79 | 0,13 |
| Нарежный В.Т. | 16,23 | 7,94 | 0,13 |
| Салтыков-Щедрин М.Е. | 14,85 | 8,12 | 0,12 |
| Азольский А.А. | 12,92 | 8,18 | 0,12 |
| Пелевин В.О. | 12,26 | 8,31 | 0,12 |
| Толстой Л.Н. | 15,21 | 9,60 | 0,10 |
| Набоков В.В. | 17,44 | 12,41 | 0,08 |

Табл. 9. Параметры средней длины предложения, коэффициента β и «синтаксической температуры» для исследованных авторов

**ЗАКЛЮЧЕНИЕ**

Проведенное исследование позволило выявить закономерности распределения предложений по длине в исследуемых текстах. Нами были выявлены факторы, влияющие на данное распределение: длина текста, автор и время написания произведения. Произведения исследованных авторов отличаются вершинами и формой графика, что говорит о том, что на длину предложения влияет такой фактор, как автор текста.

Для распределения предложений по длине в более коротких текстах характерны так называемые “зубчики” и в некоторых случаях наблюдается две вершины, что говорит о наложении двух распределений. Мы предположили, что эти два распределения – распределение предложений, содержащих слова автора, и распределение предложений, содержащих прямую речь. В коротких произведениях противопоставление прямой речи и слов автора выражено наиболее отчетливо. Распределение предложений для длинных текстов выглядит более однородным, так как количество предложений, содержащих слова автора, значительно превышает количество предложений, содержащих прямую речь. Во всех длинных текстах присутствует одна вершина.

Влияние фактора времени было рассмотрено в двух аспектах: влияние времени написания произведения в рамках творчества одного автора и влияние времени на длину предложения на протяжении двух веков. В первом случае определенной тенденции не было обнаружено. Средняя длина предложения изменялась на протяжении XIX-XX вв. Можно отметить, что на протяжении XIX - начала XX века происходит постепенное уменьшение среднего размера предложения. Это связано с тем, что на смену синтагматической прозе пришла актуализирующая. В дальнейшем этот процесс усложнился, так как наряду с расчлененной прозой вторую жизнь получает синтагматическая проза. Именно этим, судя по всему, можно объяснить некоторое увеличение размера сложного предложения во второй половине XX века.

По результатам многофакторного дисперсионного анализа все перечисленные выше факторы, также как и их взаимодействие оказывают статистически значимое влияние на длину предложения. По степени влияния факторы распределены следующим образом в порядке убывания силы влияния: автор произведения, время написания и длина текста.

Также нам удалось доказать, что распределение предложений по длине подчиняется законам гамма-распределения, что вероятно связано с особенностями человеческой психологии. Качество аппроксимации во всех случаях было больше 96%, коэффициенты α и β при этом являются уникальным для всех авторов.

Распределения, подобные гамма-распределению, встречаются также и в физике. Нами была проведена аналогия между распределением предложений по длине и законом Вина. На основании этого сопоставления была выведена «синтаксическая температура» текста, которая вычисляется как 1/β. Этот параметр отражает характер распределения предложений по длине и обратно пропорционален средней длине предложения в тексте.

Рассмотренное исследование наглядно иллюстрирует возможности методов математической статистики и теории вероятностей в задачах лингвистики. Там, где одной только интуиции читателя недостаточно, так как она всегда субъективна и недостаточно достоверна, применяется математический подход – строгий, объективный, основывающийся на математической модели стиля определенного вида.

## СПИСОК ЛИТЕРАТУРЫ

1. Адмони В.Г. Количественный и качественный анализ грамматических явлений // Вопросы языкознания. 1963. № 4. С. 57-63.
2. Адмони В.Г. Размер предложения и словосочетания как явление синтаксического строя // Вопросы языкознания. 1966. № 4. С. 111-118.
3. Акимова Г.Н. Размер предложения как фактор стилистики и грамматики //Вопросы языкознания. 1973. № 2. С. 67-79.
4. Акимова Г.Н. Новое в синтаксисе современного русского языка.- М., 1990. – 168 с.
5. Биллингслей П. Эргодическая теория и информация. М.: Мир, 1969. – с.159-160
6. Бакулин П. А., Кононович Э. В., Мороз В. И.. Курс общей астрономии. М.: Наука, 1983. – 512 с.
7. Вакурова Т.И. Размер предложения в поэзии и прозе М.Ю. Лермонтова: автореф. дис. канд. фил. наук, Санкт-Петербургский Государственный университет, Санкт-Петербург, 2011.
8. Вашак П. Длина слова и длина предложения в текстах одного автора // Вопросы статистической стилистики / Под ред. Б.Н. Головина. Киев, 1974. – С. 314-329.
9. Веселовская Т.М. Размер предложения как фактор грамматики и стилистики. автореф. дис. канд. фил. наук. Л., 1987.
10. Виноградов В.В. Современный русский язык. Пособие для лит.-лингвист. фак-тов высш. учебн. заведений. Вып. 1-2. М., Учпедгиз, 1938.
11. Головин Б.Н. Опыт вероятностно-статистического изучения некоторых явлений истории русского литературного языка XIX-XX вв. // Вопросы языкознания, № 3, 1965, С.137-146.
12. Головин Б.Н. Язык и статистика. И., “Просвещение”, 1971. 190с.
13. Головин Б.Н., М.Ф. Березин Общее языкознание. М., 1979.
14. Грибанов Д.П. Философское мировоззрение Эйнштейна // Эйнштейн и философские проблемы физики XX века. М., 1979. С.44-45.
15. Дарчук Н.П. Статистические параметры лексики в произведениях Ю.Смолича и Г.Тютюнника // Вопросы статистической стилистики / Под ред. Б.Н. Головина. Киев, 1974. – С. 262-275.
16. Зиндер Л.Р., Строева Т.В. К вопросу о применении статистики в языкознании // Вопросы языкознания. 1968. №6 С.120-123
17. Иванов В.В. Некоторые проблемы современной лингвистики. – Народы Азии и Африки, 1963, №4, С.176-177
18. Иноземцев Л.Н. Емкость группы существительного как функция от сферы употребления речи (На материале современного немецкого языка) // Ученые записки Горьковского государственного педагогического института им. М. Горького. 1965. Вып. 65. С. 3-15
19. Калашникова Г.Ф. Факторы, определяющие протяженность сложного предложения // Научные доклады высшей школы. Филологические науки. 1981. № 5. С. 37-46.
20. Крушевский Н. В. Избранные труды по языкознанию. М.: Наследие, 1998.
21. Лесскис Г. А. О зависимости между размером предложения и его структурой в разных видах текста. // Вопросы языкознания», 1963, №3. – С. 92-112.
22. Марков А. А. Об одном применении статистического метода /А. А. Марков // Известия Императорской Академии наук. Сер. 6. – 1916. – Т. 10, № 4. – С. 239–242.
23. Мартыненко Г. Я. Основы стилеметрии. — Л., 1988.
24. Мейе А. Введение в сравнительное изучение индоевропейских языков, М. — Л., 1938. – 359с.
25. Морозов Н. А. Лингвистические спектры: Средство для отличения плагиатов от истинных произведений того или другого известного автора : Стилеметрический этюд / Н. А. Морозов // Известия Отдела русского языка и словесности Императорской Академии наук. – 1915. – Т. 20, кн. 4. – С. 93–127.
26. Носенко И.А. Начала статистики для лингвистов: Учеб. Пособие. – М. – Высш.школа, 1981. – 157с.
27. Павлова К.Г. Проблема объема предложения (словосочетания) в лингвистической литературе // Иностранные языки в школе. 1971. № 2. С. 12-18.
28. Пешковский А. М.. Избранные труды. М., 1959 – 800с.
29. Пещак М.М. Наличие-отсутствие составных элементов целого – одна из стилистических черт произведения // Вопросы статистической стилистики / Под ред. Б.Н. Головина. Киев, 1974. – С. 217-228.
30. Пиотровский Р. Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. М., 1977. – 383 с.
31. Роджерс Э. Физика для любознательных/ Э.Роджерс – М.,1972
32. Романовский В.И. Математическая статистика. Кн.1. Основы теории вероятностей и математической статистики. Ташкент, из-во АН Узб. ССР, 1963.
33. Рылов С.А. О применении параметров средней длины высказывания при определении авторства «Слова о полку Игореве» // Эволюция и предыстория русского языкового строя. Горький, 1978.
34. Рылов С.А. «Слово о полку Игореве» и памятники русской литературы XVII-XVIII вв. Сопоставительный анализ грамматического развертывания высказывания // Эволюция и предыстория русского языкового строя. Горький, 1980. С. 93-104.
35. Рылов С.А. Синтаксическая организация древнерусской речи. Н. Новгород, 1990.
36. Садовский В.Н. Основания общей теории систем. – М.: Наука, 1974. – 129 с.
37. Селезнев Г.Д. Природа экспоненциального распределения слов по числу значений. Проблемы компьютерной лингвистики: Сб-к. научн. трудов / Под ред. А.А.Кретова. – Вып.2. – Воронеж: РИЦ ЕФ ВГУ, 2005. с. 169-173.
38. Селезнев Г.Д. Природа экспоненциального распределения слов по числу значений. Вестник ВГУ. Сер. Лингвистика и межкультурная коммуникация № 2 (Ч. 1), 2007, – С. 42-45.
39. Селезнев Г.Д. [Природа распределения длин слов в словарях романских языков](file:///C:\Users\Оля\AppData\Local\Opera\Opera\temporary_downloads\Parametric%20studies%20(2).doc#_Toc308192887) // Проблемы компьютерной лингвистики: Сборник научных трудов/ под ред. А.А. Кретова. – Вып. 5. – Воронеж, 2011. 329-335.
40. Слепак Б.Я. Некоторые теоретико-методологические предпосылки качественно-количественной концепции стиля // Linguistica. Тарту, 1982. С. 107-117
41. Суходольский Г.В. Основы математической статистики для психологов / Г.В. Суходольский .— Л. : Изд-во Ленингр. ун-та, 1972 . – 464 c.
42. Тулдава Ю. А. Проблемы и методы квантитативно-системного исследования лексики / Ю. А. Тулдава. – Тарту : Тартуский гос. ун-т, 1987. – 203 с.
43. Фукс В. По всем правилам искусства (точные методы в исследованиях литературы, музыки и изобразительного искусства // Искусство и ЭВМ. – Гл. VI. – М., 1975.
44. Хетсо Г. Принадлежность Достоевскому: к вопросу об атрибуции Ф.М.Достоевскому анонимных статей в журналах "Время" и "Эпоха". - Oslo, 1986
45. Хетсо Г. Кто написал «Тихий Дон»? : (проблема авторства « Тихого Дона») / Г. Хетсо, С. Густавссон, Б. Бекман. — М. : Книга, 1989.
46. Шpeйдep Ю.А. О возможности теоретического вывода статистических закономерностей текста.– В кн.: Проблемы передачи информации, М., 1967 – 57-63.
47. Шубик С.А. О размерах предложения в современном немецком языке// Статистика речи и автоматический анализ текста, 1972. – Л., 1973. – с.86-98.
48. Altmann, G. Prolegomena to Menzerath's Law. In: Grotjahn, R. [Hrsg.]: Glottometrika 2. Bochum, 1980 – 1-10.
49. Boroda, M.G. Die melodische Elementareinheit. In : Orlov, Ju.K.; Boroda, M. G.; Nadarejsvili, I.S.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 1982 – 205-221
50. Boroda, M.G. Haufigkeitsstrukturen musikalischer Texte. In : Orlov, Ju.K.; Boroda, M. G.; Nadarejsvili, I.S.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 1982 – 231-262
51. Fickermann, I.; Markner-Jager, B.; Rothe, U. Wortlange und Bedeutungskomplexitat. In: Boy, J.; Kohler, R. [Hrsg.]: Glottometrika 6. Bochum, 1984 – 115-126
52. Grzybek, Peter: Contributions to the Science of Text and Language. Word Length Studies and Related Issues. Dordrecht, NL: Springer, 2006 – 15-90.
53. Guiter, H. Les relationes frequence - longueur - sens des mots (langues romanes et anglais). XIV Congresso Internazionale di linguistica e filologia romanza. Napoli 1974, 15-20.
54. Kelih E., Grzybek P.. Häufigkeiten von Satzlängen: Zum Faktor der Intervallgröße als Einflussvariable // Glottometrics 8, 2004. – P. 23-41.
55. Herdan G. Quantitative Linguistics, London, Butterworths, 1964.
56. Horacek B. Kunstprinzipien der Satzgestaltung. Studien zu einer inhaltbezogenen Syntax der Deutschen Dichtersprache. Österreichische Akademie der Wissenschaften. Wien, 1964.
57. Mandelbrot B. Word frequencies and markovian models of discourse / Proceedings of the twelfth symposium in applied mathematics. Vol.XII, 1961, p.190 – 220.
58. Meier H. Deutsche Sprachstatistik. Hildesheim, 1964. Bd. 1
59. Orlov, Ju.K.; Boroda, M.G.; Nadarejsvili, I.S. Sprache, Text, Kunst. Quantitative Analysen. Bochum, 1982.
60. Yule G.U.On sentence-length as a characteristics of stile in prose: with application to two cases of disputed authorship // Biometrica. 1939. XXX. 3­4.
61. Zipf G.R., Human Behavior and the Principle of Least Effort, 1949

## ПРИЛОЖЕНИЕ 1. СПИСОК ИССЛЕДОВАННЫХ ПРОИЗВЕДЕНИЙ С УКАЗАНИЕМ АВТОРА И ВРЕМЕНИ НАПИСАНИЯ ПРОИЗВЕДЕНИЯ

|  |  |  |  |
| --- | --- | --- | --- |
| Автор | Произведение | Век | Треть века |
| А.Веселый | Гуляй Волга | XХ | 1 |
| А.Веселый | Россия, кровью умытая | XХ | 1 |
| Азольский А.А. | Степан Сергеич | XX | 3 |
| Азольский А.А. | Лопушок | XX | 3 |
| Азольский А.А. | Затяжной выстрел | XX | 3 |
| Булгаков М.А. | Белая Гвардия | XХ | 1 |
| Булгаков М.А. | Записки Юного Врача | XХ | 1 |
| Булгаков М.А. | Собачье сердце | XХ | 1 |
| Булгаков М.А. | Мастер и Маргарита | XХ | 2 |
| Гончаров И.А. | Обломов | XIХ | 2 |
| Гончаров И.А. | Обыкновенная история | XIХ | 2 |
| Гончаров И.А. | Фрегат Паллада | XIХ | 2 |
| Гончаров И.А. | Обрыв | XIХ | 3 |
| Достоевский Ф.М. | Бедные люди | XIХ | 2 |
| Достоевский Ф.М. | Бесы | XIХ | 3 |
| Достоевский Ф.М. | Игрок | XIХ | 3 |
| Достоевский Ф.М. | Идиот | XIХ | 3 |
| Достоевский Ф.М. | Подросток | XIХ | 3 |
| Достоевский Ф.М. | Преступление и наказание | XIХ | 3 |
| Достоевский Ф.М. | Униженные и оскорбленные | XIХ | 2 |
| Лажечников И. И. | Последний Новик | XIХ | 1 |
| Лажечников И. И. | Ледяной дом | XIХ | 2 |
| Лажечников И. И. | Басурман | XIХ | 2 |
| Лажечников И. И. | Внучка панцирного боярина | XIХ | 3 |
| М. Горький | Фома Гордеев | XIХ | 3 |
| М. Горький | Трое | XХ | 1 |
| М. Горький | Мать | XХ | 1 |
| М. Горький | Дело Артамоновых | XХ | 1 |
| Набоков В.В. | Дар | XХ | 2 |
| Набоков В.В. | Король, дама, валет | XХ | 1 |
| Набоков В.В. | Лолита | XX | 3 |
|  |  |  |  |
| *Продолжение таблицы 1.* | | | |
| Набоков В.В. | Защита Лужина | XХ | 1 |
| Набоков В.В. | Приглашение на казнь | XХ | 2 |
| Нарежный В. Т. | Бурсак | XIХ | 1 |
| Нарежный В. Т. | Российский Жилблаз | XIХ | 1 |
| Нарежный В. Т. | Гаркуша, малороссийский разбойник | XIХ | 1 |
| Нарежный В. Т. | Два Ивана, или Страсть к тяжбам | XIХ | 1 |
| Пелевин В.О. | Омон Ра | XX | 3 |
| Пелевин В.О. | Жизнь насекомых | XX | 3 |
| Пелевин В.О. | Чапаев и Пустота | XX | 3 |
| Пелевин В.О. | Generation «П» | XX | 3 |
| Пикуль В.С. | Океанский патруль | XХ | 2 |
| Пикуль В.С. | Баязет | XХ | 2 |
| Пикуль В.С. | Слово и дело | XX | 3 |
| Пушкин А.С. | Арап Петра Великого | XIХ | 1 |
| Пушкин А.С. | Дубровский | XIХ | 1 |
| Пушкин А.С. | Капитанская дочка | XIХ | 2 |
| Толстой Л.Н. | Анна Каренина | XIХ | 3 |
| Толстой Л.Н. | Воскресенье | XIХ | 3 |
| Толстой Л.Н. | Казаки | XIХ | 3 |
| Толстой Л.Н. | Крейцерова соната | XIХ | 3 |
| Толстой Л.Н. | Отрочество | XIХ | 2 |
| Толстой Л.Н. | Поликушка | XIХ | 3 |
| Толстой Л.Н. | Смерть Ивана Ильича | XIХ | 3 |
| Толстой Л.Н. | Хаджи-Мурат | XIХ | 3 |
| Толстой Л.Н. | Юность | XIХ | 2 |
| Толстой Л.Н. | Семейное счастье | XIХ | 2 |
| Толстой Л.Н. | Детство | XIХ | 2 |
| Тургенев И.С. | Рудин | XIХ | 2 |
| Тургенев И.С. | Дворянское гнездо | XIХ | 2 |
| Тургенев И.С. | Накануне | XIХ | 2 |
| Тургенев И.С. | Отцы и дети | XIХ | 2 |
| Тургенев И.С. | Дым | XIХ | 3 |
| Тургенев И.С. | Новь | XIХ | 3 |
| Тургенев И.С. | Ася | XIХ | 2 |
| Тургенев И.С. | Затишье | XIХ | 2 |
| Тургенев И.С. | Жид | XIХ | 2 |
| Тургенев И.С. | Муму | XIХ | 2 |
| Тургенев И.С. | Первая любовь | XIХ | 2 |
| *Продолжение таблицы 1.* | | | |
| Тургенев И.С. | Несчастная | XIХ | 3 |
| Тургенев И.С. | Поездка в полесье | XIХ | 2 |
| Тургенев И.С. | Степной король Лир | XIХ | 3 |
| Тургенев И.С. | Фауст | XIХ | 2 |
| Тургенев И.С. | Часы | XIХ | 3 |
| Тургенев И.С. | Яков Пасынков | XIХ | 2 |
| Федин К.А | Города и годы | XХ | 1 |
| Федин К.А | Первые радости | XХ | 2 |
| Федин К.А | Необыкновенное лето | XХ | 2 |
| Шмелев И.С. | Человек из ресторана | XХ | 1 |
| Шмелев И.С. | Солнце мертвых | XХ | 1 |
| Шмелев И.С. | Няня из Москвы | XХ | 2 |
| Шмелев И.С. | Лето Господне | XХ | 2 |
| Шолохов М.А. | Поднятая целина | XХ | 1 |
| Шолохов М.А. | Тихий Дон | XХ | 2 |
| Шолохов М.А. | Они сражались за Родину | XХ | 2 |
| Шукшин В.М | Печки-лавочки | XX | 3 |
| Шукшин В.М | Там вдали… | XX | 3 |
| Шукшин В.М | До третьих петухов | XX | 3 |
| Щедрин М.Е. | Господа ташкентцы | XIХ | 3 |
| Щедрин М.Е. | Дневник провинциала в Петербурге | XIХ | 3 |
| Щедрин М.Е. | Помпадуры и помпадурши | XIХ | 3 |
| Щедрин М.Е. | Господа Головлевы | XIХ | 3 |
| Щедрин М.Е. | Современная идиллия | XIХ | 3 |
| Эртель А.И. | Волхонская барышня | XIХ | 3 |
| Эртель А.И. | Гарденины, их дворня, приверженцы и враги | XIХ | 3 |
| Эртель А.И. | Карьера Струкова | XIХ | 3 |

## ПРИЛОЖЕНИЕ 2. РАСПРЕДЕЛЕНИЕ ИССЛЕДОВАННЫХ ПРОИЗВЕДЕНИЙ ПО ГРУППАМ В ЗАВИСИМОСТИ ОТ ДЛИНЫ ТЕКСТА (В ПРЕДЛОЖЕНИЯХ).

|  |  |  |  |
| --- | --- | --- | --- |
| Автор | Произведение | Длина текста в предложениях | № группы в зависимости от длины текста |
| А.Веселый | Гуляй Волга | 4548 | 3 |
| А.Веселый | Россия, кровью умытая | 12920 | 6 |
| Азольский А.А. | Степан Сергеич | 7441 | 5 |
| Азольский А.А. | Лопушок | 3249 | 3 |
| Азольский А.А. | Затяжной выстрел | 5924 | 4 |
| Булгаков М.А. | Белая Гвардия | 7035 | 5 |
| Булгаков М.А. | Записки Юного Врача | 3586 | 3 |
| Булгаков М.А. | Собачье сердце | 2845 | 2 |
| Булгаков М.А. | Мастер и Маргарита | 8547 | 5 |
| Гончаров И.А. | Обломов | 11870 | 6 |
| Гончаров И.А. | Обыкновенная история | 8611 | 5 |
| Гончаров И.А. | Фрегат Паллада | 13118 | 6 |
| Гончаров И.А. | Обрыв | 18242 | 6 |
| Достоевский Ф.М. | Бедные люди | 2958 | 2 |
| Достоевский Ф.М. | Бесы | 14338 | 6 |
| Достоевский Ф.М. | Игрок | 3425 | 3 |
| Достоевский Ф.М. | Идиот | 13741 | 6 |
| Достоевский Ф.М. | Подросток | 12273 | 6 |
| Достоевский Ф.М. | Преступление и наказание | 13474 | 6 |
| Достоевский Ф.М. | Униженные и оскорбленные | 9396 | 6 |
| Лажечников И. И. | Последний Новик | 9639 | 6 |
| Лажечников И. И. | Ледяной дом | 819 | 1 |
| Лажечников И. И. | Басурман | 7978 | 5 |
| Лажечников И. И. | Внучка панцирного боярина | 6123 | 4 |
| М. Горький | Фома Гордеев | 7385 | 5 |
| М. Горький | Трое | 7973 | 5 |
| М. Горький | Мать | 8173 | 5 |
| М. Горький | Дело Артамоновых | 5519 | 4 |
| Набоков В.В. | Дар | 4278 | 3 |
| Набоков В.В. | Король, дама, валет | 4220 | 3 |
| Набоков В.В. | Лолита | 5205 | 4 |
| Набоков В.В. | Защита Лужина | 2962 | 2 |
| *Продолжение таблицы 2* | | | |
| Набоков В.В. | Приглашение на казнь | 2663 | 2 |
| Нарежный В. Т. | Бурсак | 5199 | 4 |
| Нарежный В. Т. | Российский Жилблаз | 9311 | 6 |
| Нарежный В. Т. | Гаркуша, малороссийский разбойник | 2663 | 2 |
| Нарежный В. Т. | Два Ивана, или Страсть к тяжбам | 2394 | 2 |
| Пелевин В.О. | Омон Ра | 2240 | 2 |
| Пелевин В.О. | Жизнь насекомых | 3380 | 3 |
| Пелевин В.О. | Чапаев и Пустота | 7900 | 5 |
| Пелевин В.О. | Generation «П» | 5786 | 4 |
| Пикуль В.С. | Океанский патруль | 23996 | 6 |
| Пикуль В.С. | Баязет | 14279 | 6 |
| Пикуль В.С. | Слово и дело | 17138 | 6 |
| Пушкин А.С. | Арап Петра Великого | 631 | 1 |
| Пушкин А.С. | Дубровский | 2009 | 2 |
| Пушкин А.С. | Капитанская дочка | 2675 | 2 |
| Толстой Л.Н. | Анна Каренина | 19376 | 6 |
| Толстой Л.Н. | Воскресенье | 8978 | 5 |
| Толстой Л.Н. | Казаки | 4341 | 3 |
| Толстой Л.Н. | Крейцерова соната | 1807 | 2 |
| Толстой Л.Н. | Отрочество | 1175 | 2 |
| Толстой Л.Н. | Поликушка | 1288 | 2 |
| Толстой Л.Н. | Смерть Ивана Ильича | 1094 | 2 |
| Толстой Л.Н. | Хаджи-Мурат | 2487 | 2 |
| Толстой Л.Н. | Юность | 2059 | 2 |
| Толстой Л.Н. | Семейное счастье | 1609 | 2 |
| Толстой Л.Н. | Детство | 1435 | 2 |
| Тургенев И.С. | Рудин | 3107 | 3 |
| Тургенев И.С. | Дворянское гнездо | 2772 | 2 |
| Тургенев И.С. | Накануне | 3658 | 3 |
| Тургенев И.С. | Отцы и дети | 4461 | 3 |
| Тургенев И.С. | Дым | 3256 | 3 |
| Тургенев И.С. | Новь | 6285 | 4 |
| Тургенев И.С. | Ася | 1039 | 2 |
| Тургенев И.С. | Затишье | 1659 | 2 |
| Тургенев И.С. | Жид | 481 | 1 |
| Тургенев И.С. | Муму | 440 | 1 |
| Тургенев И.С. | Несчастная | 1739 | 2 |
| Тургенев И.С. | Первая любовь | 1327 | 2 |
| *Продолжение таблицы 2* | | | |
| Тургенев И.С. | Поездка в полесье | 408 | 1 |
| Тургенев И.С. | Степной король Лир | 1727 | 2 |
| Тургенев И.С. | Фауст | 916 | 1 |
| Тургенев И.С. | Часы | 1066 | 2 |
| Тургенев И.С. | Яков Пасынков | 1065 | 2 |
| Федин К.А | Города и годы | 9279 | 6 |
| Федин К.А | Первые радости | 8715 | 5 |
| Федин К.А | Необыкновенное лето | 17489 | 6 |
| Шмелев И.С. | Человек из ресторана | 5256 | 4 |
| Шмелев И.С. | Солнце мертвых | 7116 | 5 |
| Шмелев И.С. | Няня из Москвы | 5618 | 4 |
| Шмелев И.С. | Лето Господне | 11126 | 6 |
| Шолохов М.А. | Поднятая целина | 18446 | 6 |
| Шолохов М.А. | Тихий Дон | 42785 | 6 |
| Шолохов М.А. | Они сражались за Родину | 3398 | 3 |
| Шукшин В.М | Печки-лавочки | 2984 | 2 |
| Шукшин В.М | Там вдали… | 2484 | 2 |
| Шукшин В.М | До третьих петухов | 2369 | 2 |
| Щедрин М.Е. | Господа ташкентцы | 6139 | 4 |
| Щедрин М.Е. | Дневник провинциала в Петербурге | 6185 | 4 |
| Щедрин М.Е. | Помпадуры и помпадурши | 5166 | 4 |
| Щедрин М.Е. | Господа Головлевы | 6312 | 4 |
| Щедрин М.Е. | Современная идиллия | 6674 | 4 |
| Эртель А.И. | Волхонская барышня | 4132 | 3 |
| Эртель А.И. | Гарденины, их дворня, приверженцы и враги | 16382 | 6 |
| Эртель А.И. | Карьера Струкова | 3848 | 3 |

## ПРИЛОЖЕНИЕ 3. КОЭФФИЦИЕНТЫ α, β И КАЧЕСТВО АППРОКСИМАЦИИ ДЛЯ ИССЛЕДОВАННЫХ ПРОИЗВЕДЕНИЙ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Автор | Произведение | α | β | Качество аппроксимации |
| А.Веселый | Гуляй Волга | 1,57 | 7,36 | 98,46% |
| А.Веселый | Россия, кровью умытая | 1,44 | 6,16 | 99,89% |
| А.Веселый | В ЦЕЛОМ ПО АВТОРУ | 1,47 | 6,88 | 99,67% |
| Азольский А.А. | Степан Сергеич | 1,74 | 5,42 | 99,55% |
| Азольский А.А. | Лопушок | 1,58 | 9,33 | 98,77% |
| Азольский А.А. | Затяжной выстрел | 1,40 | 8,99 | 98,76% |
| Азольский А.А. | В ЦЕЛОМ ПО АВТОРУ | 1,50 | 8,18 | 99,25% |
| Булгаков М.А. | Белая Гвардия | 1,27 | 7,60 | 99,46% |
| Булгаков М.А. | Записки Юного Врача | 1,61 | 5,47 | 99,32% |
| Булгаков М.А. | Собачье сердце | 1,45 | 5,75 | 99,99% |
| Булгаков М.А. | Мастер и Маргарита | 1,82 | 7,24 | 100,00% |
| Булгаков М.А. | В ЦЕЛОМ ПО АВТОРУ | 1,46 | 6,92 | 99,62% |
| Гончаров И.А. | Обломов | 1,48 | 7,17 | 96,80% |
| Гончаров И.А. | Обыкновенная история | 1,78 | 6,28 | 98,83% |
| Гончаров И.А. | Фрегат Паллада | 2,25 | 6,60 | 99,67% |
| Гончаров И.А. | Обрыв | 1,75 | 6,65 | 98,45% |
| Гончаров И.А. | В ЦЕЛОМ ПО АВТОРУ | 1,68 | 7,36 | 99,03% |
| Достоевский Ф.М. | Бедные люди | 1,98 | 6,60 | 98,09% |
| Достоевский Ф.М. | Бесы | 1,71 | 7,62 | 99,27% |
| Достоевский Ф.М. | Игрок | 1,78 | 6,72 | 98,27% |
| Достоевский Ф.М. | Идиот | 1,59 | 8,68 | 99,37% |
| Достоевский Ф.М. | Подросток | 1,79 | 7,71 | 99,26% |
| Достоевский Ф.М. | Преступление и наказание | 1,43 | 8,36 | 98,85% |
| Достоевский Ф.М. | Униженные и оскорбленные | 1,74 | 6,88 | 98,41% |
| Достоевский Ф.М. | В ЦЕЛОМ ПО АВТОРУ | 1,64 | 7,79 | 99,31% |
| Лажечников И. И. | Последний Новик | 1,80 | 8,82 | 98,99% |
| Лажечников И. И. | Ледяной дом | 1,86 | 7,94 | 99,93% |
| Лажечников И. И. | Басурман | 2,19 | 6,47 | 99,17% |
| Лажечников И. И. | Внучка панцирного боярина | 2,33 | 6,24 | 99,42% |
| Лажечников И. И. | В ЦЕЛОМ ПО АВТОРУ | 1,96 | 7,34 | 99,27% |
| М. Горький | Фома Гордеев | 1,34 | 7,49 | 99,38% |
| М. Горький | Трое | 1,60 | 6,22 | 99,73% |
| М. Горький | Мать | 1,76 | 5,77 | 99,36% |
| М. Горький | Дело Артамоновых | 1,02 | 12,55 | 96,33% |
| *Продолжение таблицы* | | | | |
| М. Горький | В ЦЕЛОМ ПО АВТОРУ | 1,35 | 7,74 | 99,31% |
| Набоков В.В. | Дар | 1,31 | 16,25 | 97,98% |
| Набоков В.В. | Король, дама, валет | 1,72 | 6,53 | 97,39% |
| Набоков В.В. | Лолита | 1,19 | 14,73 | 97,23% |
| Набоков В.В. | Защита Лужина | 1,47 | 10,84 | 97,04% |
| Набоков В.В. | Приглашение на казнь | 1,29 | 9,20 | 96,96% |
| Набоков В.В. | В ЦЕЛОМ ПО АВТОРУ | 1,23 | 12,41 | 99,92% |
| Нарежный В. Т. | Бурсак | 2,55 | 6,47 | 99,08% |
| Нарежный В. Т. | Российский Жилблаз | 1,94 | 7,26 | 99,51% |
| Нарежный В. Т. | Гаркуша, малороссийский разбойник | 1,83 | 8,63 | 96,88% |
| Нарежный В. Т. | Два Ивана, или Страсть к тяжбам | 2,11 | 8,58 | 96,72% |
| Нарежный В. Т. | В ЦЕЛОМ ПО АВТОРУ | 2,05 | 7,94 | 99,26% |
| Пелевин В. О. | Омон Ра | 1,13 | 11,06 | 97,71% |
| Пелевин В. О. | Жизнь насекомых | 1,48 | 8,17 | 97,53% |
| Пелевин В. О. | Чапаев и Пустота | 1,62 | 6,72 | 98,78% |
| Пелевин В. О. | Generation «П» | 1,73 | 6,43 | 98,77% |
| Пелевин В. О. | В ЦЕЛОМ ПО АВТОРУ | 1,42 | 8,31 | 98,67% |
| Пикуль В.С. | Океанский патруль | 1,76 | 6,30 | 99,57% |
| Пикуль В.С. | Баязет | 1,74 | 5,49 | 97,99% |
| Пикуль В.С. | Слово и дело | 2,38 | 4,45 | 99,41% |
| Пикуль В.С. | В ЦЕЛОМ ПО АВТОРУ | 1,99 | 5,25 | 99,65% |
| Пушкин А.С. | Арап Петра Великого | 2,34 | 5,52 | 91,84% |
| Пушкин А.С. | Дубровский | 2,17 | 5,50 | 98,58% |
| Пушкин А.С. | Капитанская дочка | 2,16 | 4,47 | 96,81% |
| Пушкин А.С. | В ЦЕЛОМ ПО АВТОРУ | 2,06 | 5,51 | 97,91% |
| Толстой Л.Н. | Анна Каренина | 1,57 | 8,23 | 97,64% |
| Толстой Л.Н. | Воскресенье | 1,45 | 9,31 | 97,58% |
| Толстой Л.Н. | Казаки | 1,59 | 5,97 | 97,85% |
| Толстой Л.Н. | Крейцерова соната | 1,77 | 7,01 | 97,42% |
| Толстой Л.Н. | Отрочество | 1,77 | 10,09 | 95,75% |
| Толстой Л.Н. | Поликушка | 1,57 | 6,87 | 95,63% |
| Толстой Л.Н. | Смерть Ивана Ильича | 1,30 | 10,12 | 96,73% |
| Толстой Л.Н. | Хаджи-Мурат | 1,96 | 6,18 | 98,16% |
| Толстой Л.Н. | Юность | 1,82 | 12,05 | 97,18% |
| Толстой Л.Н. | Семейное счастье | 2,28 | 7,34 | 97,40% |
| Толстой Л.Н. | Детство | 2,21 | 8,27 | 96,69% |
| Толстой Л.Н. | В ЦЕЛОМ ПО АВТОРУ | 1,52 | 9,60 | 98,69% |
| *Продолжение таблицы* | | | | |
| Тургенев И.С. | Рудин | 1,90 | 5,23 | 98,53% |
| Тургенев И.С. | Дворянское гнездо | 1,75 | 7,82 | 97,45% |
| Тургенев И.С. | Накануне | 1,72 | 6,03 | 97,50% |
| Тургенев И.С. | Отцы и дети | 1,72 | 6,40 | 96,33% |
| Тургенев И.С. | Дым | 1,34 | 9,68 | 95,25% |
| Тургенев И.С. | Новь | 1,29 | 7,70 | 93,92% |
| Тургенев И.С. | В ЦЕЛОМ ПО АВТОРУ | 1,50 | 7,49 | 96,94% |
| Федин К.А | Города и годы | 1,57 | 6,40 | 99,62% |
| Федин К.А | Первые радости | 1,65 | 6,33 | 98,61% |
| Федин К.А | Необыкновенное лето | 1,51 | 6,88 | 98,48% |
| Федин К.А | В ЦЕЛОМ ПО АВТОРУ | 1,49 | 6,98 | 99,18% |
| Шмелев И.С. | Человек из ресторана | 2,40 | 3,16 | 99,12% |
| Шмелев И.С. | Солнце мертвых | 2,36 | 2,85 | 99,31% |
| Шмелев И.С. | Няня из Москвы | 2,85 | 3,50 | 99,03% |
| Шмелев И.С. | Лето Господне | 2,23 | 4,28 | 98,38% |
| Шмелев И.С. | В ЦЕЛОМ ПО АВТОРУ | 2,27 | 3,43 | 99,12% |
| Шолохов М.А. | Поднятая целина | 1,79 | 5,91 | 99,46% |
| Шолохов М.А. | Тихий Дон | 1,53 | 6,09 | 99,33% |
| Шолохов М.А. | Они сражались за Родину | 1,82 | 6,93 | 97,81% |
| Шолохов М.А. | В ЦЕЛОМ ПО АВТОРУ | 1,67 | 6,44 | 99,23% |
| Шукшин В.М | Печки-лавочки | 1,80 | 2,82 | 99,38% |
| Шукшин В.М | Там вдали… | 1,84 | 2,89 | 98,85% |
| Шукшин В.М | До третьих петухов | 2,03 | 2,46 | 98,38% |
| Шукшин В.М | В ЦЕЛОМ ПО АВТОРУ | 1,84 | 2,84 | 99,37% |
| Щедрин М.Е. | Господа ташкентцы | 2,00 | 7,86 | 98,70% |
| Щедрин М.Е. | Дневник провинциала в Петербурге | 1,69 | 9,02 | 98,86% |
| Щедрин М.Е. | Помпадуры и помпадурши | 1,78 | 8,93 | 99,16% |
| Щедрин М.Е. | Господа Головлевы | 2,08 | 6,47 | 98,99% |
| Щедрин М.Е. | Современная идиллия | 1,71 | 8,08 | 99,37% |
| Щедрин М.Е. | В ЦЕЛОМ ПО АВТОРУ | 1,83 | 8,12 | 99,78% |
| Эртель А.И. | Волхонская барышня | 2,03 | 4,85 | 97,89% |
| Эртель А.И. | Гарденины, их дворня, приверженцы и враги | 1,51 | 5,93 | 98,29% |
| Эртель А.И. | Карьера Струкова | 1,17 | 10,80 | 94,23% |
| Эртель А.И. | В ЦЕЛОМ ПО АВТОРУ | 1,52 | 6,59 | 97,31% |