Министерство просвещения Российской Федерации

Государственное автономное общеобразовательное учреждение

Московской области «Балашихинский лицей»

VIII Международный конкурс исследовательских работ школьников
«Research start» 2025/26

Исследовательская работа

Статистический критерий Манна-Уитни в установлении авторства романа «12 стульев»

Выполнила: Подорожняк Вероника Сергеевна

Ученица 9 класса

Руководитель: Суетин Валерий Юрьевич,

К.ф.-м.н., учитель математики

2025–2026 учебный год

Содержание

Введение

Вопросы по авторству «12 стульев» возникали давно, многих исследователей удивляло несовпадение уровней этого прекрасного романа с остальным творчеством Ильи Ильфа и Евгения Петрова. Как результат, возникли гипотезы, что роман написан кем-то другим.

Цель работы: определить, кто из современников И.Ильфа и Е.Петрова мог бы быть автором, для этого мы сравниваем числовые характеристики текстов романа «12 стульев» с текстами А.Н. Толстого, Ю.К. Олеши, С.С. Заяицкого, П. С. Романова, В.П. Катаева.

Актуальность. И математики, и лингвисты основываются на поисках авторского инварианта (числового значения некоторого параметра, характерного для того или иного автора). На филологическом факультете МГУ и на кафедре инновационных технологий языковой коммуникации Уфимского университета науки и технологий активно применяются статистические методы в исследовании языка и сравнении естественных языков. В моей работе в качестве такого инварианта берётся относительное число 55-ти служебных слов (союзов, частиц, предлогов) в текстах литературных произведений, разбитых на блоки по 16 тысяч слов. Ранее мною уже было получено существенное с точки зрения математической статистики различие наборов данных по романам «12 стульев», «Золотой телёнок» и путевых очерков «Одноэтажная Америка». Так как путевые очерки «Одноэтажная Америка» были написаны после командировки Ильи Ильфа и Евгения Петрова в США, считаем, что именно это произведение написано указанными авторами.

Задача исследования: провести статистический анализ наборов относительных частот служебных слов произведений указанных авторов, сделать вывод о возможном авторстве «12 стульев».

Объекты исследования: тексты романа «12 стульев», роман А.Н. Толстого («Пётр Первый» кн.1), произведения Ю.К. Олеши («Три толстяка», «Зависть»,

«Любовь», «Прощания», «Альдебаран», «В цирке», «Вишнёвая косточка»), С.С. Заяицкого («Красавица с острова Люлю», «Жизнеописание Степана Александровича Лососинова», «Псы господни», «Морской волчонок», «Баклажаны»), П. С. Романова («Русь», т.1), В.П. Катаева («Белеет парус одинокий», «Повелитель железа»).

Метод исследования. Для статистического анализа использовалась программа на языке Python в среде Google Colaboratory. Попарное сравнение числовых характеристик текстов проводится на основе левостороннего статистического критерия Манна-Уитни. В этом состоит основное отличие нашего метода от других работ, где оценивается дисперсия. Уровень значимости выбран равным 0,05.

Основная часть

Теоретическая часть

Современный уровень применения вычислительной техники и распределенных вычислений позволяет по-новому подойти к такой сложной лингвистической проблеме, как установление авторства литературных текстов. В наши дни стал возможен подробный математический (статистический) анализ текстов, и этим видом исследований занимаются подразделения ведущих вузов, например, лаборатория общей и компьютерной лексикологии и лексикографии филологического факультета МГУ. За последние 20 лет опубликовано много монографий и статей, посвященных поискам авторского инварианта – присущей только этому автору числовой характеристики его текста, позволяющей отличить его работы от работ других авторов. Одной из первых таких работ, выполненных ещё вручную в 1983 году, была работа [1], в которой в качестве авторского инварианта рассматривалось относительное число 50 служебных слов на выборках в 9-16 тысяч слов. Как показали многие исследования, этот параметр является устойчивым на выборках в 16 тысяч слов. Применение этого параметра к сравнению текстов романа «12 стульев» с романами М.А. Булгакова

позволило моему научному руководителю опровергнуть гипотезу И. Амлински [2] о том, что автором «12 стульев» является М.А. Булгаков [3]. Вывод сделан на основе применения критерия Манна-Уитни.

Сравнение числовых характеристик текстов романов «12 стульев», «Золотого телёнка» и путевых очерков «Одноэтажная Америка», основанное на применении программы из работы [4] в среде GoogleColab, я представила как доклад на Фестивале профильного образования «ПрофГоризонт» 12.12.2025 (Технопарк «Исток- РТУ МИРЭА», г.Фрязино). В работе получено статистически значимое различие наборов относительных частот появления служебных слов исследуемых текстов. Результаты приведены в Приложении А.

Мы решили проверить, кто из авторов 30-х годов 20-го века мог бы быть автором с точки зрения близких значений авторского инварианта. Этому исследованию и посвящена настоящая работа.

В моей работе, кроме машинной обработки параметров текста, применяется математическая статистика, точнее, один из критериев, позволяющих сделать вывод о принадлежности двух выборок одной генеральной совокупности – критерий Манна-Уитни. Такое применение математики для не-математиков – очень интересный приём, показывающий ценность математики для широкого круга специалистов, далёких от математики.

Обзор современных теоретических подходов

На нынешнем этапе развития проблематики авторства текстов исследователи сходятся на том, что используемые методы должны быть а) бессознательными, б) универсальными, в) верифицируемыми.

Бессознательность подразумевает отсутствие контроля автора за соблюдением устойчивости\неустойчивости параметра. Поэтому, например, такой параметр, как средняя длина предложения или частотность применения частицы «не» (явно легко контролируемый параметр) не годится.

Универсальность означает, что инструмент должен быть пригоден для любого автора. Понятно, что исследуемые нами методы не применимы к переводным текстам, так как они будут «работать» на переводчика, а не на автора.

Верифицируемость позволяет другим исследователям воспроизвести эксперимент, чтобы убедиться в его корректности.

Наиболее часто используемые сейчас параметры статистического описания текстов это *повторяемость слов (активный словарный запас), ранг местоимения я, корреляция средней длины слова и средней длины предложения, сравнение частотностей букв.* Эти параметры обладают всеми тремя указанными атрибутами подходящих инструментов. Кроме того, энтузиасты разрабатывают программы *Лингвистический анализатор* [5], *Морфологический анализатор* [6] в которых исследуются десятки других параметров.

На филологическом факультете МГУ активно работают над статистическим анализом текстов, создавая частотные грамматико-семантические словари русских писателей [7]. На кафедре инновационных технологий языковой коммуникации Уфимского университета науки и технологий также применяются статистические методы в исследовании языка и сравнении естественных языков [8].

Основная идея использования статистических методов в определении авторства следующая: если надо выбрать, Автор1 или Автор 2 является автором текста X, мы сравниваем текст X и тексты других произведений этих авторов. Далее с использованием функций распределения параметров мы делаем вывод об авторстве. Если есть новый текст с неизвестным автором, мы подбираем, к каким наборам характеристик он ближе всего и не является ли один из уже известных авторов автором этого текста.

Идея использования 50 служебных слов (союзов, частиц, предлогов) была выдвинута в работе [1], где опытным путём показана устойчивость

параметра на произведениях Л.Н.Толстого, А.П.Чехова, А.И. Куприна и пр. В работе [3] показано, что на генеральной выборке (на всех крупных произведениях) Л.Н. Толстого имеется, тем не менее, некая изменчивость этого параметра. Ещё большая разница наблюдается в ранних и поздних произведениях М.А. Шолохова. Это не удивительно: после авиакатастрофы 1942 года, когда М.А. Шолохов получил сильнейшую травму головы, стиль его письма резко изменился. Вообще, именно авторство «Тихого Дона» породило массу новых приёмов и методов изучения статистических характеристик текстов.

Материалы и методы исследования

Работу по статистическому анализу литературных текстов я начала со сравнения числовых характеристик текстов романов «12 стульев», «Золотой телёнок» и путевых очерков «Одноэтажная Америка» И. Ильфа и Е. Петрова. В исследовательском проекте [9] я получила статистически значимые отличия в наборах относительных частот служебных слов в этих произведениях. Эти результаты (таблицы и расчёты) приведены в Приложении.

В настоящей работе в качестве объектов исследования мы взяли тексты романа «12 стульев», роман А.Н. Толстого («Пётр Первый» кн.1), произведения Ю.К. Олеши («Три толстяка», «Зависть», «Любовь», «Прощания», «Альдебаран», «В цирке», «Вишнёвая косточка»), С.С. Заяицкого («Красавица с острова Люлю», «Жизнеописание Степана Александровича Лососинова», «Псы господни», «Морской волчонок», «Баклажаны»), П. С. Романова («Русь», т.1), В.П. Катаева («Белеет парус одинокий», «Повелитель железа», рассказы, фельетоны). Поясню выбор авторов. О сходстве некоторых сцен романов «12 стульев» и «Мастер и Маргарита» говорили многие, подчёркивая даже одинаковую ритмику фраз. Так как уже считаю доказанным, что М.А. Булгаков не является автором «12 стульев», я искала, кто ещё из современников И. Ильфа и Е. Петрова мог быть автором. Алексей Николаевич Толстой был ведущим писателем той эпохи. Валентин Петрович Катаев – один из успешных писателей

страны в то время – родной брат Евгения Петрова, вполне мог, написав сатирический роман, отдать его брату. А мог найти талантливого автора, остро нуждающегося в деньгах, и уговорить его отдать своё творение. Таким вполне мог быть его друг - Юрий Карлович Олеша, «безмерно талантливый, вечно голодный и пьяный и без гроша в кармане», по свидетельству его знакомых. Ю.К. Олеша был знаком и с И. Ильфом по литературной деятельности в Одессе. Пантелеймон Романов привлёк моё внимание тем, что такого яркого и талантливого писателя совсем не изучают на уроках литературы. Чувство юмора Сергея Заяицкого, его опыт в написании фельетонов, также привели к выбору этого интересного автора.

Основной программой, используемой в работе, была программа, созданная А.В. Астафуровым и В.Ю. Суетиным [4] на языке Python в среде Google Colaboratory специально для изучения частотных характеристик текста (см. Приложение). Программа удаляет из текста знаки препинания, разбивает текст на блоки по n тысяч слов, подсчитывает в них служебные слова в соответствии с [3]:

предлоги — в, на, с, за, к, по, из, у, от, для, во, без, до, о, через, со, при, про, об, ко, над, из-за, из-под, под;

союзы — и, что, но, а, да, хотя, когда, чтобы, если, тоже, или, то есть, зато, будто;

частицы — не, как, же, даже, бы, ли, только, вот, то, ни, лишь, ведь, вон, то есть, нибудь, уже, либо.

Подсчитывается относительная частота всех служебных слов в каждой выборке из n(=16) тысяч слов, результат выводится в Excel-таблицу. Выбор Google Colaboratory обусловлен удобством использования: нет необходимости ставить Python на локальный компьютер, можно проводить исследования прямо в браузере.

Сравнение проводится на основе левостороннего статистического критерия Манна-Уитни, как и в работе [3]. В этом состоит основное отличие нашего метода от работы [1], где оценивается дисперсия (рассеяние относительно среднего значения). Уровень значимости выбран равным 0,05, такой стандартный уровень помогает принимать обоснованные решения на основе данных.

Структура статистического исследования практически одинакова в любых исследованиях и состоит из нескольких последовательных этапов:

1. *Формулировка гипотез*. Нужно сформулировать обе статистические гипотезы: нулевую и альтернативную. Нулевая гипотеза утверждает, что различия между сравниваемыми наборами данных случайны, альтернативная — что различия не случайны, существенны. В нашем случае нулевая гипотеза означает, что исследуемые тексты написаны одним автором, альтернативная гипотеза – авторы сравниваемых произведений различны.

2. *Выбор уровня значимости.* Традиционно порогом значимости считается 0,05 (5%), но в некоторых областях (например, в медицине, генетике) применяют более строгие критерии (0,01 или 0,001). Итак, вероятность отвергнуть верную нулевую гипотезу в нашем случае равна 0,05.

3. *Выбор статистического теста, оценка его параметров и вычисление эмпирических статистик*. Перед запуском любого теста важно оценить его ключевые параметры, поскольку они напрямую влияют на статистическую мощность. Наш выбор – непараметрический критерий Манна-Уитни. Критерий левосторонний, то есть если эмпирическое значение получается меньше критического, отвергаем нулевую гипотезу о том, что выборки взяты из одной генеральной совокупности (для нас – исследуемые тексты, скорее всего, не являются текстами одного автора). Если же эмпирическое значение больше критического, то нет причин отвергать нулевую гипотезу (для нас – эти тексты принадлежат одному автору). Выбор этого критерия обусловлен тем, что данные малых объёмов распределены не по

нормальному закону, а, практически, равномерно (поэтому мы не используем критерий Стьюдента). Понятно, что статистическими методами нельзя доказать нулевую гипотезу, но можно её опровергнуть или не опровергнуть.

U-критерий Манна-Уитни был разработан в 1945 году Ф.Уилкоксоном и в 1947 существенно переработан Х.Б.Манном и Д.Р.Уитни. Условия применения:

- *Независимая переменная* должна состоять из двух категориальных независимых групп.
- *Наблюдения* должны быть независимыми (не должно быть никаких отношений между двумя группами или внутри каждой группы).
- *Не требует наличия нормального распределения*.
- *Ограничения*: в каждой из выборок должно быть не менее 3 значений признака, допускается, чтобы в одной выборке было два значения, но во второй — не менее пяти. В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало (до 10)

Суть критерия Манна-Уитни: ранжируются два набора (в нашем случае - наборы относительных частот служебных слов в текстах, разбитых на блоки). Из этих наборов в n_1 и n_2 данных составляется и ранжируется новый набор данных, вычисляются суммы индексов первого и второго наборов в объединенном столбце: R_1 и R_2, выбирается наибольшая из этих сумм $R_{max} = T$ и соответствующий n_{max}. Статистика Манна-Уитни имеет вид

$$U = n_1 \cdot n_2 + \frac{n_{max} \cdot (n_{max} + 1)}{2} - T.$$

Чем меньше значение U, тем вероятнее, что различия между значениями параметра в выборках достоверны. Таблица критических значений для уровня значимости 0,05 приведена ниже:

Таблица1. Критические значения критерия Манна-Уитни (а=0,05)

n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
n_2	ρ=0,05																		
3	-	0																	
4	-	0	1																
5	0	1	2	4															
6	0	2	3	5	7														
7	0	2	4	6	8	11													
8	1	3	5	8	10	13	15												
9	1	4	6	9	12	15	18	21											
10	1	4	7	11	14	17	20	24	27										
11	1	5	8	12	16	19	23	27	31	34									
12	2	5	9	13	17	21	26	30	34	38	42								

Практическая часть

Описание проводимого исследования

Мною проведено попарное сравнение наборов данных по «12 стульям» и указанным выше текстам (несколько произведений Ю.К. Олеши «склеены» в один файл, так же поступили и с текстами С.С. Заяицкого и В.П. Катаева, остальные источники достаточно объёмные). Выбор объёма разбиения текста на блоки в 16 тысяч слов обусловлен доказанной в работе [1] устойчивостью авторского инварианта на таких объёмах выделяемых из текстов блоков. Я проводила сравнения и на объёмах в 10 тыс., результаты получились те же. В работе [1] авторы делали пропуски в текстах, что было вызвано ручной обработкой огромного числа данных. Мы отказались от пропусков, что увеличило число сравниваемых блоков и позволило применять критерий Манна-Уитни.

Тексты скачивались из бесплатных библиотек, из текстов удалялись сноски, предисловия, послесловия, комментарии издателей, выходные данные, рисунки и т.д. Затем тексты переводились в .txt -файлы в кодировке UTF-8, подгружались в упомянутую выше программу в Google Colab, на выходе получался Excel-файл с таблицей, содержащей столбец относительных частот употребления служебных слов в каждом блоке разбитого на 16 тысяч слов

текста. Пользуясь функциями Excel, я сводила эти столбцы в один, сортировала обобщенный набор по возрастанию величин и ранжировала данные, то есть приписывала ранг каждому значению. В случае равных значений ранг получается дробным. Для удобства подсчёта суммы рангов данных, относящихся к разным текстам, я выделяла данные разных текстов разными цветами. Затем подсчитывала сумму рангов каждого из сравниваемых наборов данных, вычисляла статистику Манна-Уитни и сравнивала с критическими значениями. Учитывая левосторонний характер критерия, делала вывод о возможном авторстве романа «12 стульев».

Полученные результаты представлены в Таблице 2: желтым цветом выделены данные по «12 стульям», зеленым – тексты В.П.Катаева, голубым – текст П.С.Романова, оранжевым – текст А.Н.Толстого, коричневым – С.С.Заяицкого, красным – текст Ю.К.Олеши.

Таблица 2. Относительные частоты употребления служебных слов

12 стульев	Катаев	Романов	Толстой	Заяицкий	Олеша
0,2001	0,1815	0,2407	0,1966	0,1911	0,1888
0,2011	0,1908	0,2444	0,2029	0,1959	0,1895
0,2012	0,1933	0,2456	0,203	0,2028	0,1976
0,2031	0,2092	0,2523	0,2077	0,2046	0,1996
0,2066	0,2188	0,2548	0,2081	0,2072	0,2005
	0,222	0,2556	0,2084	0,2086	
		0,2559	0,2128	0,2134	
		0,2573	0,2141	0,2145	
		0,2587	0,2156	0,2159	
			0,2158	0,2199	
			0,2167		
			0,2169		
			0,2198		

Анализ полученных результатов

Объединённые ранжированные наборы данных имеют вид

Таблица 3. Ранжированные обобщенные ряды

12\Заяицкий	Ранги	12\Толстой	Ранги	12\Романов	Ранги	12\Катаев	Ранги	12\Олеша	Ранги
0,1911	1	0,1966	1	0,2001	1	0,1815	1	0,1888	1
0,1959	2	0,2001	2	0,2011	2	0,1908	2	0,1895	2
0,2001	3	0,2011	3	0,2012	3	0,1933	3	0,1976	3
0,2011	4	0,2012	4	0,2031	4	0,2001	4	0,1996	4
0,2012	5	0,2029	5	0,2066	5	0,2011	5	0,2001	5
0,2028	6	0,203	6	0,2407	6	0,2012	6	0,2005	6
0,2031	7	0,2031	7	0,2444	7	0,2031	7	0,2011	7
0,2046	8	0,2066	8	0,2456	8	0,2066	8	0,2012	8
0,2066	9	0,2077	9	0,2523	9	0,2092	9	0,2031	9
0,2072	10	0,2081	10	0,2548	10	0,2188	10	0,2066	10
0,2086	11	0,2084	11	0,2556	11	0,222	11		
0,2134	12	0,2128	12	0,2559	12				
0,2145	13	0,2141	13	0,2573	13				
0,2159	14	0,2156	14	0,2587	14				
0,2199	15	0,2158	15						
		0,2167	16						
		0,2169	17						
		0,2198	18						

Вычислим значения параметров критерия Манна-Уитни для каждого такого попарного сравнения.

1. «12 стульев» (1) и С.С. Заяицкий (2):

$n_1 = 5, n_2 = 10, R_1 = 28,\ R_2 = 92$, т.е.

$R_{max} = T = 92, n_{max} = n_2 = 10.$

Статистика Манна-Уитни равна

$$U = n_1 \cdot n_2 + \frac{n_{max} \cdot (n_{max} + 1)}{2} - T = 14.$$

Критическое значение с уровнем значимости 0,05 для наших n_1, n_2 выбираем по таблице 1, оно равно 11. Так как критерий левосторонний, то эмпирическое значение статистики не попало в критическую область, и у нас нет повода отклонить нулевую гипотезу о том, что оба набора исследуемых данных

взяты из одной генеральной совокупности. Другими словами, Сергей Сергеевич Заяицкий, вероятно, мог бы быть автором «12 стульев».

2. «12 стульев» и А.Н. Толстой.

$n_1 = 5, n_2 = 13, R_1 = 24,\ R_2 = 147$, т.е.

$R_{max} = T = 147, n_{max} = n_2 = 13.$

Статистика Манна-Уитни равна

$$U = n_1 \cdot n_2 + \frac{n_{max} \cdot (n_{max} + 1)}{2} - T = 9.$$

Критическое значение с уровнем значимости 0,05 для наших n_1, n_2 равно 15. Так как критерий левосторонний, то эмпирическое значение статистики попало в критическую область, и у нас есть основания отклонить нулевую гипотезу о том, что оба набора исследуемых данных взяты из одной генеральной совокупности. U-тест показал статистически значимое различие в частотах употребления служебных слов в текстах А.Н. Толстого и в романе «12 стульев». Другими словами, Алексей Николаевич Толстой, скорее всего, не может быть автором «12 стульев».

3. «12 стульев» и Ю.К. Олеша.

$n_1 = 5, n_2 = 5,\ R_1 = 39,\ R_2 = 16$, т.е. $R_{max} = T = 39\ , n_{max} = n_1 = 5.$

Статистика Манна-Уитни равна

$$U = n_1 \cdot n_2 + \frac{n_{max} \cdot (n_{max} + 1)}{2} - T = 12.$$

Критическое значение с уровнем значимости 0,05 для наших n_1, n_2 выбираем по таблице 1, оно равно 4. Так как критерий левосторонний, то эмпирическое значение статистики не попало в критическую область, и у нас нет повода отклонить нулевую гипотезу о том, что оба набора исследуемых данных взяты из одной генеральной совокупности. Другими словами, Юрий Карлович Олеша вполне мог бы быть автором «12 стульев».

4. «12 стульев» и П.С. Романов.

Здесь хорошо видно, что наибольшее значение из набора данных по «12 стульям» меньше наименьшего значения для текста П.С. Романова, а тогда эмпирическое значение статистики Манна-Уитни равно нулю, то есть Пантелеймон Сергеевич Романов, скорее всего, не мог быть автором «12 стульев».

5. «12 стульев» и В. П. Катаев

$n_1 = 5, n_2 = 6,\ R_1 = 30,\ R_2 = 36$, т.е. $R_{max} = T = 36, n_{max} = n_2 = 6.$

Статистика Манна-Уитни равна

$$U = n_1 \cdot n_2 + \frac{n_{max} \cdot (n_{max} + 1)}{2} - T = 21.$$

Критическое значение с уровнем значимости 0,05 для наших n_1, n_2 выбираем по таблице 1, оно равно 5. Так как критерий левосторонний, то эмпирическое значение статистики не попало в критическую область, и у нас нет повода отклонить нулевую гипотезу о том, что оба набора исследуемых данных взяты из одной генеральной совокупности. Другими словами, В.П. Катаев также вполне мог быть автором «12 стульев».

Заключение

Как показало исследование, применённый метод может успешно применяться при доказательстве того, что кто-то не является автором того или иного русского литературного текста. Однако указать, кто именно является автором, этот метод не может: мы можем лишь оценить величину отклонения от критического значения критерия Манна-Уитни.

Из исследуемых текстов, эмпирическая статистика которых не попала в критическую область, наибольшее расхождение с критическим значением статистики Манна-Уитни среди исследуемых текстов у текстов Валентина

Катаева (родного брата Евгения Петрова). То есть мы можем сделать вывод, что с б**о**льшей вероятностью среди писателей, с текстами которых мы проводили сравнение характеристик одного из лучших романов 20-го века – «12 стульев» - автором является Валентин Катаев. Понятно, что, расширив круг писателей того времени, мы можем получить другие выводы.

Интересно также заметить, что данные по «12 стульям» очень мало отличаются от своего среднего значения (стандартное отклонение $2{,}6 \cdot 10^{-3}$). Если бы роман писали два автора, были бы части текста с различными частотными характеристиками. Складывается ощущение, что этот прекрасный роман написан одним автором, а вот кем именно – неизвестно.

Во время исследования я получила истинное удовольствие сразу от многих моментов:

1. Мне удалось разобраться в применении критерия Манна-Уитни.
2. Очень интересно было применить математические методы к нематематической задаче.
3. Я познакомилась с творчеством писателей начала 20-го века, о которых раньше не слышала, и оказалось, что эти авторы создали много красивых и интересных произведений.
4. Я впервые получила свой результат, которого ещё не было у других.
5. Я научилась работать с научной литературой, искать специфическую информацию, увидела, как много специалистов и энтузиастов работают вместе.

Надеюсь, приобретённый опыт позволит мне заняться исследовательской деятельностью в дальнейшем. Интересно было бы проанализировать таким методом тексты, составленные нейросетями, выявить для каждой нейросети её авторский инвариант. Это могло бы иметь приложение

в улучшении сервиса антиплагиата. Я планирую продолжить статистическое сравнение русских литературных текстов с использованием других параметров – активного словарного запаса, средней длины предложения и пр.

Список литературы

1. *Фоменко В.П., Фоменко Т.Г.* Авторский инвариант русских литературных текстов. Методы количественного анализа текстов нарративных источников. М.: АН СССР, Ин-т Истории СССР, 1983. – С. 86–109.
2. *Амлински И.* 12 стульев от Михаила Булгакова. – Берлин – Kirschner Verlag, 2013. – 328 с.
3. *Суетин В.Ю.* Применение частотных характеристик для определения авторства литературных текстов. // Вестник ТвГУ, Серия: Прикладная математика, 2022. – №2. – С. 84–89.
4. *Суетин В.*Ю., *Астафуров А.В.* Относительная частота служебных слов в текстах Л.Н. Толстого и М.А. Шолохова // Теория и практика языковой коммуникации. Уфа. 19-20 июня 2023 г. – С. 204-212.
5. *Львов А.* Лингвистический анализатор [Электронный ресурс] https://fantlab.ru/article374 *дата обращения 26.11.2025*
6. *Сокирко А.В.* Морфологический анализатор [Электронный ресурс] http://www.aot.ru/docs/sokirko/ Dialog2004.htm *дата обращения 26.11.2025*
7. Частотный грамматико-семантический словарь языка художественных произведений А. П. Чехова [Электронный ресурс] https://www.philol.msu.ru/~lex/chehov.html *дата обращения 20.10.2025*
8. *Рогожникова, Т.М., Астафуров А.В.* Сверхмедленная электрическая активность мозга как инструмент анализа суггестивного потенциала вербальной модели. *Теория языка и межкультурная коммуникация: электронный научный журнал Курского государственного университета, (4), 294–309.*

9. *Подорожняк В.С.* Статистический анализ текстов романов «12 стульев», «Золотой телёнок» и путевого очерка «Одноэтажная Америка». Фестиваль профильного образования «ПрофГоризонт» 12.12.2025 (Технопарк «Исток-РТУ МИРЭА», г. Фрязино).

Приложение

Сравнительный анализ текстов «12 стульев», «Золотого телёнка» и «Одноэтажной Америки»

Результаты сравнения относительных частот употребления служебных слов в текстах романов «12 стульев», «Золотой телёнок» и путевых очерков «Одноэтажная Америка» И. Ильфа и Е. Петрова на выборках по 10 тысяч слов (доклад на Фестивале профильного образования «ПрофГоризонт» 12.12.2025 (Технопарк «Исток- РТУ МИРЭА», г.Фрязино).

Относительное число служебных слов в выборках по 10 тыс слов

12 стул	ЗолТел	ОднАмер
0,1963	0,2066	0,1871
0,1977	0,2068	0,1898
0,1981	0,2083	0,1902
0,1995	0,2093	0,1914
0,2011	0,2098	0,193
0,2021	0,2099	0,1965
0,2043	0,2155	0,1971
0,2065	0,216	0,1972
0,2098		0,1975
		0,1992

Ранжированные обобщенные наборы данных попарного сравнения

12/ЗолТел	Ранг	12/ОдАм	Ранг	ОдАм/Зол	
0,1963	1	0,1871	1	0,1871	1
0,1977	2	0,1898	2	0,1898	2
0,1981	3	0,1902	3	0,1902	3
0,1995	4	0,1914	4	0,1914	4
0,2011	5	0,193	5	0,193	5
0,2021	6	0,1963	6	0,1965	6
0,2043	7	0,1965	7	0,1971	7
0,2065	8	0,1971	8	0,1972	8
0,2066	9	0,1972	9	0,1975	9
0,2068	10	0,1975	10	0,1992	10
0,2083	11	0,1977	11	0,2066	11
0,2093	12	0,1981	12	0,2068	12
0,2098	13,5	0,1992	13	0,2083	13
0,2098	13,5	0,1995	14	0,2093	14
0,2099	15	0,2011	15	0,2098	15
0,2155	16	0,2021	16	0,2099	16
0,216	17	0,2043	17	0,2155	17
		0,2065	18	0,216	18
		0,2098	19		

«12 стульев» и «Золотой телёнок». Статистика Манна-Уитни U равна 4,5, критическое значение 15.

«12 стульев» и «Одноэтажная Америка». U=17, критическое значение 20.

«Золотой телёнок» и «Одноэтажная Америка». ». U=0, критическое значение 23.

Вывод: ***все наборы различаются существенно: скорее всего, авторы этих произведений различны***.

Программа Астафурова-Суетина расположена в Google Colaboratory по адресу https://colab.research.google.com/drive/1C5d3D2-jV-NFZnxktKtspV_GsEX_N0F2#scrollTo=dXt6sdQ74cuJ