e-ISSN 3060-5059
DOI: 10.67227

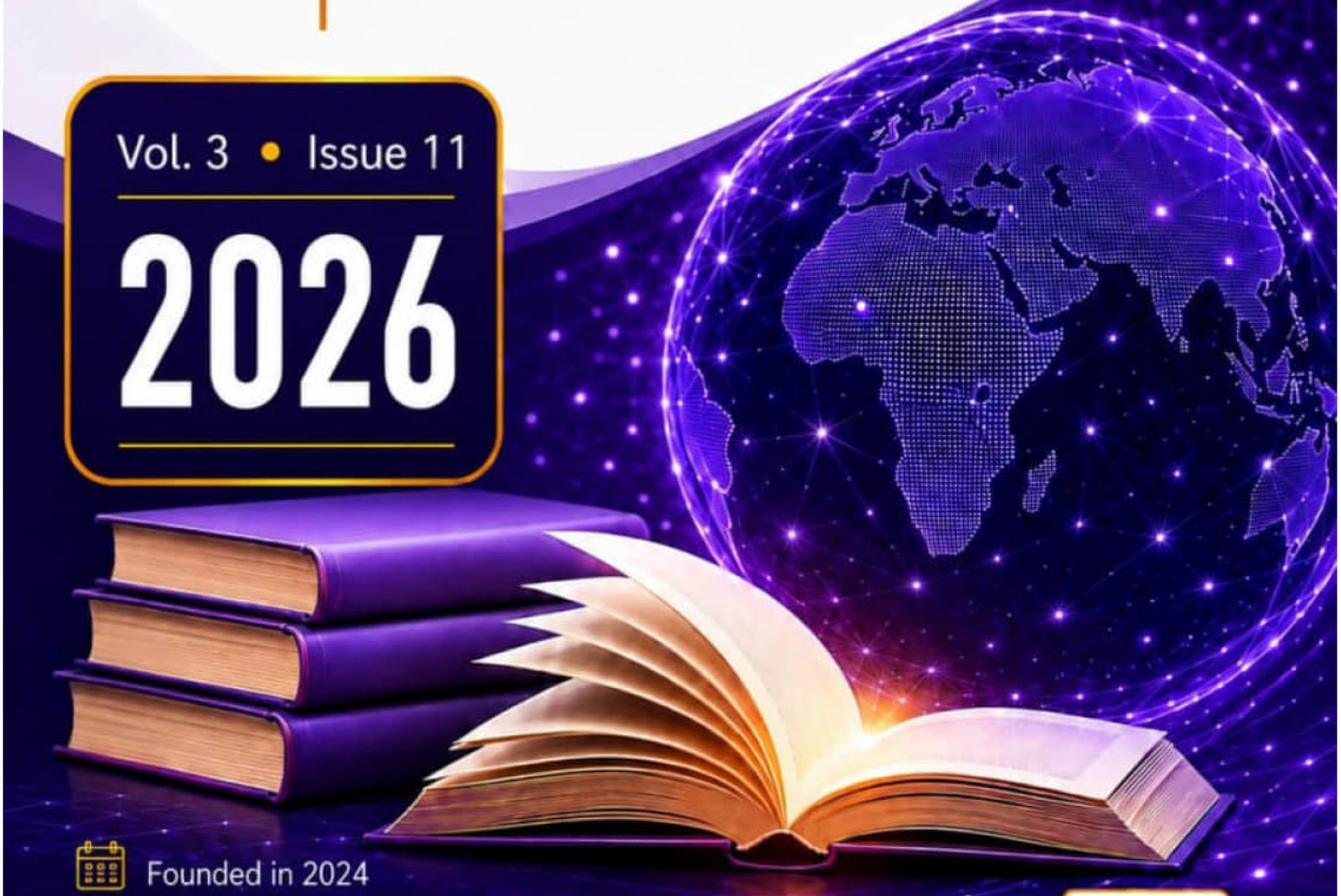
IJTIMOIY-GUMANITAR SOHADA ILMIY-INNOVATSION TADQIQOTLAR

ILMIY METODIK JURNAL

Scientific and Innovative Research in Social and Humanitarian Studies

Vol. 3 • Issue 11

2026



Founded in 2024

publishscience.uz | publish_science_journal@mail.ru

PEER REVIEWED • OPEN ACCESS

ISSN 3060-5059 (Online)

Article / Original paper https://doi.org/10.67227/hgm3a128

M. A. SHOLOXOVNING «TINCH DON» ROMANINI YAGONA MATNLAR KORPUSI ASOSIDA STILOMETRIK ATRIBUTSIYA QILISH

Avezov Suxrob Sobirovich
Nasulloyeva Shaxinabonu Sheraliyevna
Buxoro davlat universiteti, dotsent, (PhD).
Buxoro davlat universiteti, magistrant.
e-mail: 1990senigama@gmail.com **ORCID**: https://orcid.org/0000-0003-1137-3403
e-mail: sahinabonu1@gmail.com **ORCID**: https://orcid.org/0009-0001-6528-6287

Annotatsiya.
Maqolada M. A. Sholoxovning «Tinch Don» romani muallifligi besh nafar yozuvchining 29 ta matnidan iborat, qariyb ikki million so‘zni o‘z ichiga olgan yagona korpus asosida tekshiriladi. Tadqiqotda avvalgi izlanishlarda qo‘llanilgan bir parametrli mezonlarning beqaror natijalar berishi aniqlanadi. V. A. Bubnov tomonidan keltirilgan predloglar chastotasiga oid natijalar esa so‘zlar o‘rniga bo‘shliqlar sanalgandagina takrorlanishi ko‘rsatiladi. Berrouzning Delta usuli asosida o‘tkazilgan stilometrik tahlil natijalari romanning barcha kitoblarini M. A. Sholoxovga nisbat beradi.

Kalit so‘zlar: Sholoxov, «Tinch Don», atributsiya, stilometriya, Berrouz Deltasi, yordamchi so‘zlar, predloglar spektri, Kryukov, korpus lingvistikasi.

СТИЛОМЕТРИЧЕСКАЯ АТРИБУЦИЯ РОМАНА М. А. ШОЛОХОВА «ТИХИЙ ДОН» НА ОСНОВЕ ЕДИНОГО ТЕКСТОВОГО КОРПУСА

Авезов Сухроб Собирович
Насуллоева Шахинабону Шералиевна
Бухарский государственный университет, доцент, (PhD).
Бухарский государственный университет, магистрант.

Аннотация.
В статье исследуется проблема авторства романа М. А. Шолохова «Тихий Дон» на основе единого корпуса, включающего 29 текстов пяти писателей общим объёмом около двух миллионов слов. Установлено, что однопараметрические критерии, использованные в предыдущих исследованиях, демонстрируют нестабильные результаты. Показано, что приведённые В. А. Бубновым показатели частотности предлогов воспроизводятся лишь в том случае, когда вместо слов подсчитываются пробелы. Результаты стилометрического анализа с применением метода Дельты Берроуза позволяют атрибутировать все книги романа М. А. Шолохову.

Ключевые слова: Шолохов, «Тихий Дон», атрибуция, стилометрия, Дельта Берроуза, служебные слова, спектр предлогов, Крюков, корпусная лингвистика.

STYLOMETRIC AUTHORSHIP ATTRIBUTION OF M. A. SHOLOKHOV’S NOVEL AND QUIET FLOWS THE DON BASED ON A UNIFIED TEXT CORPUS

Avezov Suxrob Sobirovich
Nasulloyeva Shaxinabonu Sheraliyevna
Bukhara State University, Associate Professor, (PhD).
Bukhara State University, Master's Student.

Abstract.
This article investigates the authorship of M. A. Sholokhov’s novel And Quiet Flows the Don using a unified corpus comprising 29 texts by five writers, totaling approximately two million words. The study finds that single-parameter criteria employed in previous research produce unstable results. It also demonstrates that the preposition frequency figures reported by V. A. Bubnov can be reproduced only when spaces are counted instead of words. The results of stylometric analysis using Burrows’s Delta method attribute all books of the novel to M. A. Sholokhov.

Keywords: Sholokhov, And Quiet Flows the Don, authorship attribution, stylometry, Burrows’s Delta, function words, preposition frequency spectrum, Kryukov, corpus linguistics.

Введение. Спор об авторстве романа «Тихий Дон» начался в 1928 году. Слухи о плагиате появились тогда же, когда журнал «Октябрь» опубликовал первые две книги [2, с. 60]. В 1999 году Институт мировой литературы им. А. М. Горького РАН обнаружил рукописи первой и второй книг, считавшиеся утраченными. Сторонники

ISSN 3060-5059 (Online)

авторства М. А. Шолохова сочли свою позицию доказанной, однако их оппоненты от прежних взглядов не отказались [2, с. 61]. Для количественной филологии роман стал испытательным полигоном, подобным текстам Шекспира в англоязычной науке [5, с. 72].

Точные методы долгое время казались выходом из этого спора. Ожидалось, что количественный анализ позволит освободить аргументацию от политических и эстетических пристрастий исследователя. Однако Г. Хьетсо с соавторами и А. Т. Фоменко, опираясь на количественные данные, пришли к противоположным выводам, после чего доверие к количественным аргументам заметно снизилось [5, с. 72–73]. Н. П. Великанова и Б. В. Орехов описывают типичный ход подобных исследований формулой «Мы подсчитали параметр X в некоторых текстах, и это должно значить следующее...» [5, с. 73]. При этом признак объявлялся авторским без проверки на текстах писателей, авторство произведений которых не вызывает сомнений.

Таким образом, сформировался целый ряд однопараметрических критериев. Г. Хьетсо сравнивал среднюю длину предложения, А. Т. Фоменко — долю служебных слов, В. А. Бубнов и А. В. Сурвило опирались на частоты предлогов «в», «на», «с», а А. В. Зенков — на распределение первых значащих цифр в числительных. Каждый признак выдвигался как самостоятельное доказательство. Однако ни один из них до сих пор не проверялся на общем корпусе вместе с остальными.

Актуальность исследования обусловлена именно этим пробелом. Метод Delta Дж. Берроуза уже применялся к роману [5, с. 77–80], однако прежние критерии на тех же текстах не воспроизводились. Поэтому остаётся неясным, где в них заканчивается авторский сигнал и начинается погрешность подготовки материала. Цель статьи заключается в сопоставлении однопараметрических критериев и метода Delta на едином корпусе. В него вошли роман, бесспорные произведения М. А. Шолохова, тексты Ф. Д. Крюкова как главного претендента на авторство [5, с. 77] и контрольные тексты трёх писателей той же эпохи.

АНАЛИЗ ЛИТЕРАТУРЫ И МЕТОДОЛОГИЯ

Монография Г. Хьетсо и его соавторов основывалась на анализе средней длины предложения. У Ф. Д. Крюкова она составила 13,9 слова, у М. А. Шолохова — 12,9, а в «Тихом Доне» — 12,4 слова [5, с. 73]. А. Т. Фоменко принял за авторский инвариант долю служебных слов при объёме выборки не менее 16 000 слов. По его расчётам, первая и вторая книги романа вместе с началом третьей не принадлежат автору «Донских рассказов» [5, с. 74]. Б. В. Ковалёв указывает, что длина слова и предложения не служит надёжным показателем индивидуального стиля. Даже безупречный подсчёт теряет смысл, если исследователь выбрал лингвистически нерелевантный параметр [7, с. 71]. Исследование Л. А. Вдовиной на материале произведений 50 авторов наглядно подтверждает этот вывод. Средняя длина предложения обеспечила лишь 5 % верных атрибуций, частоты 265 служебных слов — 35,3 %, а символьные четырёхграммы с весами TF-IDF — 82,5 % [4, с. 97]. Д. В. Хмелёв, построивший метод атрибуции на цепях А. А. Маркова, отмечал, что метод авторского инварианта принципиально не позволяет различать более десяти писателей [8].

Работы В. А. Бубнова и А. В. Сурвило восходят к «лингвистическим спектрам» Н. А. Морозова. Каждый том романа делился на фрагменты объёмом по 10 000 слов, после чего вычислялись средние частоты предлогов «в», «на», «с» [2, с. 62]. Интегральный спектр «Тихого Дона» составил 0,0171, 0,0144 и 0,0092. Для Ф. Д. Крюкова были получены значения 0,0220, 0,0165 и 0,0136, а для А. С. Серафимовича и И. А. Родионова — значения того же порядка [3, с. 79]. Эти различия авторы назвали «формальным доказательством» того, что на авторство романа может претендовать лишь М. А. Шолохов [3, с. 79]. Дополнительным аргументом служила форма спектра: у «Тихого Дона» она была выпуклой вверх, а у остальных писателей — вниз [3, с. 80]. Тем же способом В. А. Бубнов вычислил частоты союзов и частицы «не» по четырём томам [1, с. 57]. При этом тексты романа были взяты из сети Интернет [2, с. 63].

Противоположный вывод сделал А. В. Зенков. На основании распределения первых значащих цифр числительных он установил близость текстов Ф. Д. Крюкова к начальным частям романа. Принадлежность «Донских рассказов» и трёх крупных романов одному автору показалась ему крайне сомнительной [9, p. 15]. Тексты Ф. Д. Крюкова из-за малого объёма пришлось объединить в один файл [9, p. 16].

Метод Delta был предложен Дж. Берроузом в 2002 году. Для каждого текста определяются относительные частоты N наиболее частотных слов корпуса. Частоты преобразуются в z-оценки, а расстояние между двумя текстами рассчитывается как среднее абсолютное значение разности этих оценок [7, с. 73–74]. В основе метода лежит эмпирическое наблюдение, согласно которому тексты одного автора располагаются ближе друг к другу, чем тексты разных авторов [7, с. 72]. На корпусе русской межвоенной прозы Ф. Н. Двинятин и Б. В. Ковалёв получили верный результат атрибуции в 16 случаях из 17 [6, с. 224]. Н. П. Великанова и Б. В. Орехов, применив метод Delta к роману, установили расстояние 0,59 между «Донскими рассказами» и первым томом «Тихого Дона». Для сравнения: расстояние между романами «Мастер и Маргарита» и «Белая гвардия» составляет 0,73 [5, с. 80]. Ограничения метода хорошо известны. Delta не позволяет обнаружить универсальный «отпечаток пальца» автора вне конкретного корпуса, поскольку отношения между текстами изменяются в зависимости от состава выборки [7, с. 77]. Для получения надёжного результата необходима жанровая однородность материала при объёме текста не менее 10 000 слов [5, с. 77].

Материал исследования составили 29 текстов пяти писателей общим объёмом 2 067 528 словоупотреблений. Произведения М. А. Шолохова представлены восемью единицами. В их число вошли четыре книги «Тихого Дона», 25 ранних рассказов и фельетонов под общим условным названием «Донские рассказы»,

ISSN 3060-5059 (Online)

обе книги «Поднятой целины» и главы романа «Они сражались за Родину». Ф. Д. Крюков представлен произведениями «Зыбь», «Шквал» и «На Тихом Дону» (в общей сложности 74 183 словоупотребления). Контрольную группу составили пять текстов М. А. Булгакова, пять текстов М. Горького и восемь русскоязычных романов В. В. Набокова, написанных в период с 1926 по 1938 год. Тексты взяты из открытого набора RussianNovels, подготовленного для стилеметрических исследований (https://github.com/JoannaBy/RussianNovels). Рассказы и поздняя проза М. А. Шолохова взяты из открытого репозитория, созданного для проверки авторства романа (https://github.com/donRumata03/Sholokhov). Электронная версия романа в обоих источниках одна и та же. Сверка с научным изданием ИМЛИ РАН, на котором основано исследование Н. П. Великановой и Б. В. Орехова [5, с. 77], остаётся задачей следующего этапа.

Перед проведением подсчётов буква «ё» заменялась на «е», а редакторские примечания в квадратных скобках удалялись. Имена собственные исключались автоматически. Словоформа считалась именем собственным, если не менее чем в 90 % случаев употребления она писалась с прописной буквы. Таких словоформ было выявлено 2 210. Без этой процедуры имена героев попадали в верхнюю часть частотного списка. К. А. Маслинский показал, что исключение имён «Григорий» и «Давыдов» заметно изменяет результат атрибуции романа [7, с. 89]. Лемматизация не проводилась, что соответствует практике применения метода Delta [7, с. 90].

Исследование включало четыре серии опытов. В первой серии для каждого текста вычислялись средняя длина предложения, частоты предлогов «в», «на», «с» и доля служебных слов по закрытому списку из 125 предлогов, союзов и частиц. Вторая серия была направлена на проверку воспроизводимости частот, опубликованных В. А. Бубновым и А. В. Сурвило. Третья серия основывалась на классическом варианте метода Delta с использованием словарей объёмом от 100 до 1000 наиболее частотных слов с шагом 100 и кластеризацией текстов методом Уорда. В рамках этой серии все тексты также разделялись на последовательные фрагменты объёмом по 10 000 словоупотреблений. Всего было получено 183 фрагмента. Фрагмент романа относился к автору ближайшего фрагмента из других текстов. Параллельно применялось правило наименьшего среднего расстояния до фрагментов каждого автора. Точность процедуры проверялась на 145 фрагментах бесспорных текстов, причём каждый фрагмент сравнивался только с фрагментами других произведений. В четвёртой серии восемь частей романа по отдельности сопоставлялись с остальным корпусом. Расчёты выполнены на языке Python с использованием библиотек NumPy, SciPy и Matplotlib.

РЕЗУЛЬТАТЫ

Первая серия показала, что однопараметрические признаки разводят тексты по времени создания легче, чем по авторам (табл. 1).

Таблица 1. Однопараметрические признаки текстов корпуса

Текст	Словоупотреблений	Средняя длина предложения, слов	«в», %	«на», %	«с», %	Служебные слова, %
М.А.Шолохов						
«Тихий Дон», кн. 1	92 737	8,38	2,76	2,34	1,36	21,26
«Тихий Дон», кн. 2	94 128	9,26	2,58	2,06	1,33	21,55
«Тихий Дон», кн. 3	107 781	9,39	2,55	2,20	1,41	23,16
«Тихий Дон», кн. 4	126 875	9,67	2,21	2,09	1,40	24,58
«Донские рассказы»	82 782	9,07	2,64	2,25	1,29	23,47
«Поднятая целина», кн. 1	93 428	9,17	2,37	1,95	1,29	24,69
«Поднятая целина», кн. 2	111 311	11,93	2,26	1,96	1,28	27,15
«Они сражались за Родину»	47 251	13,09	1,96	1,97	1,20	26,47
Ф.Д.Крюков						
«Зыбь»	21 630	8,84	2,40	1,61	1,24	24,62
«Шквал»	22 489	9,21	2,49	1,49	1,01	25,01
«На Тихом Дону»	30 064	12,61	2,79	1,98	1,47	25,41
Контрольная группа (наименьшее и наибольшее значение)						
М.А.Булгаков, 5 текстов	277 312	8,00 / 11,94	3,13 / 3,56	1,57 / 2,04	0,93 / 1,15	24,15 / 25,98
М.Горький, 5 текстов	548 173	8,40 / 12,35	2,20 / 2,47	1,45 / 1,69	0,86 / 1,13	21,80 / 24,48

В.В.Набоков, 8 текстов	411 567	11,63 / 20,83	2,34 / 3,03	1,48 / 1,88	1,12 / 1,48	23,59 / 24,87

Средняя длина предложения в книгах романа находится в пределах от 8,38 до 9,67 слова, а в текстах Ф. Д. Крюкова — от 8,84 до 12,61 слова. Полученные нами абсолютные значения ниже, чем у Г. Хьетсо, поскольку каждая реплика диалога считалась отдельным предложением. Правило сегментации одинаково для всех текстов, поэтому сравнение внутри корпуса корректно. Разброс значений между тремя текстами Ф. Д. Крюкова (3,8 слова) превышает разницу в 1,0 и 1,5 слова, на которой Г. Хьетсо основывал свой вывод. По средней длине предложения «Тихий Дон» не отделяется ни от произведений Ф. Д. Крюкова, ни от текстов М. А. Булгакова и М. Горького.

Доля служебных слов в первой книге романа составляет 21,26 %, во второй — 21,55 %, а в «Донских рассказах» — 23,47 %. Расхождение между двумя первыми книгами и ранними рассказами подтверждает наблюдение А. Т. Фоменко. Однако дальнейшая динамика этого признака позволяет объяснить данное расхождение иначе. Начиная с первой книги романа, доля служебных слов увеличивается почти монотонно в соответствии с хронологией создания произведений. В третьей книге она достигает 23,16 %, в четвёртой — 24,58 %, а во второй книге «Поднятой целины» — 27,15 %. Размах значений внутри бесспорного творчества одного писателя составил 5,9 процентного пункта. Все 18 текстов трёх контрольных писателей укладываются в интервал шириной 4,2 процентного пункта (от 21,80 до 25,98 %). Признак, который изменяется внутри творчества одного автора сильнее, чем между произведениями разных авторов, не может считаться авторским инвариантом.

Из трёх предлогов различительную силу сохраняет только «на». Во всех восьми текстах М. А. Шолохова его частота составляет от 1,95 до 2,34 %, а в произведениях Ф. Д. Крюкова — от 1,49 до 1,98 %. Частоты предлогов «в» и «с» у двух писателей перекрываются. Высокую частоту предлога «на» у М. А. Шолохова независимо отметили Ф. Н. Двинятин и Б. В. Ковалёв на основании данных Национального корпуса русского языка (2,1 % при среднем уровне от 1,6 до 1,7 %) [6, с. 217].

Вторая серия выявила расхождение с опубликованными данными (табл. 2). Подсчёт по всему тексту романа показал частоты предлогов «в», «на» и «с», равные 0,0250, 0,0216 и 0,0137 соответственно, что примерно в полтора раза превышает значения интегрального спектра, приведённые В. А. Бубновым и А. В. Сурвило. Для повести «Зыбь», единственного текста Ф. Д. Крюкова, представленного в обоих исследованиях, полученные нами значения практически совпали с опубликованными.

Таблица 2. Частоты служебных слов в «Тихом Доне» и повести «Зыбь» при разных способах подсчёта

Источник значений	**«в»**	**«на»**	**«с»**	**«и»**	**«не»**
«Тихий Дон», данные В.А.Бубнова и А.В.Сурвило [2; с. 63], [1; с. 57]	0,0171	0,0144	0,0092	0,0199	0,0105
«Тихий Дон», наш подсчёт относительно числа слов	0,0250	0,0216	0,0137	0,0318	0,0166
«Тихий Дон», наш подсчёт относительно числа отрезков между пробелами	0,0170	0,0147	0,0093	0,0203	0,0112
«Зыбь», данные В.А.Бубнова и А.В.Сурвило [2; с. 64]	0,0232	0,0165	0,0124		
«Зыбь», наш подсчёт относительно числа слов	0,0240	0,0161	0,0124		

Причину расхождения удалось установить. В распространённой электронной версии романа слова отделены друг от друга двойными пробелами. Если считать единицей каждый отрезок между одиночными пробелами, число «слов» завышается в 1,465 раза. При таком подсчёте частоты предлогов принимают значения 0,0170, 0,0147 и 0,0093, почти совпадающие с интегральным спектром романа [2, с. 63]. Тот же множитель позволяет с высокой точностью воспроизвести частоты союза «и» и частицы «не» из статьи 2015 года [1, с. 57]. Совпадение сразу по пяти служебным словам позволяет с высокой вероятностью предположить, что знаменатель при расчёте частот для романа был завышен. Различие абсолютных уровней между «Тихим Доном» и текстами Ф. Д. Крюкова оказывается следствием технической ошибки, а не стилистических особенностей.

Форма спектра при исправленном подсчёте сохраняется. Вторая конечная разность, которую В. А. Бубнов и А. В. Сурвило вычисляли как разность первых разностей частот [3, с. 80], отрицательна во всех восьми текстах М. А. Шолохова (от −0,78 до −0,21 процентного пункта). Во всех трёх текстах Ф. Д. Крюкова она положительна (от 0,31 до 0,53 процентного пункта). В контрольной группе положительное значение наблюдается в 17 текстах из 18. Единственное исключение — «Дело Артамоновых» М. Горького — имеет значение, близкое к нулю (−0,07). Выпуклость спектра вверх обусловлена высокой частотой предлога «на», то есть тем же признаком, который выделил М. А. Шолохова в первой серии.

Третья серия выполнена с использованием метода Delta (рис. 1). При словарях объёмом от 400 до 1000

ISSN 3060-5059 (Online)

слов разбиение дерева на пять кластеров в точности соответствует авторской принадлежности текстов. Все восемь текстов М. А. Шолохова образуют отдельную ветвь в девяти опытах из десяти. Исключением стал словарь из 300 слов. В этом случае тексты Ф. Д. Крюкова присоединяются к группе поздних произведений М. А. Шолохова, но не к первым книгам романа, авторство которых оппоненты приписывают Ф. Д. Крюкову.

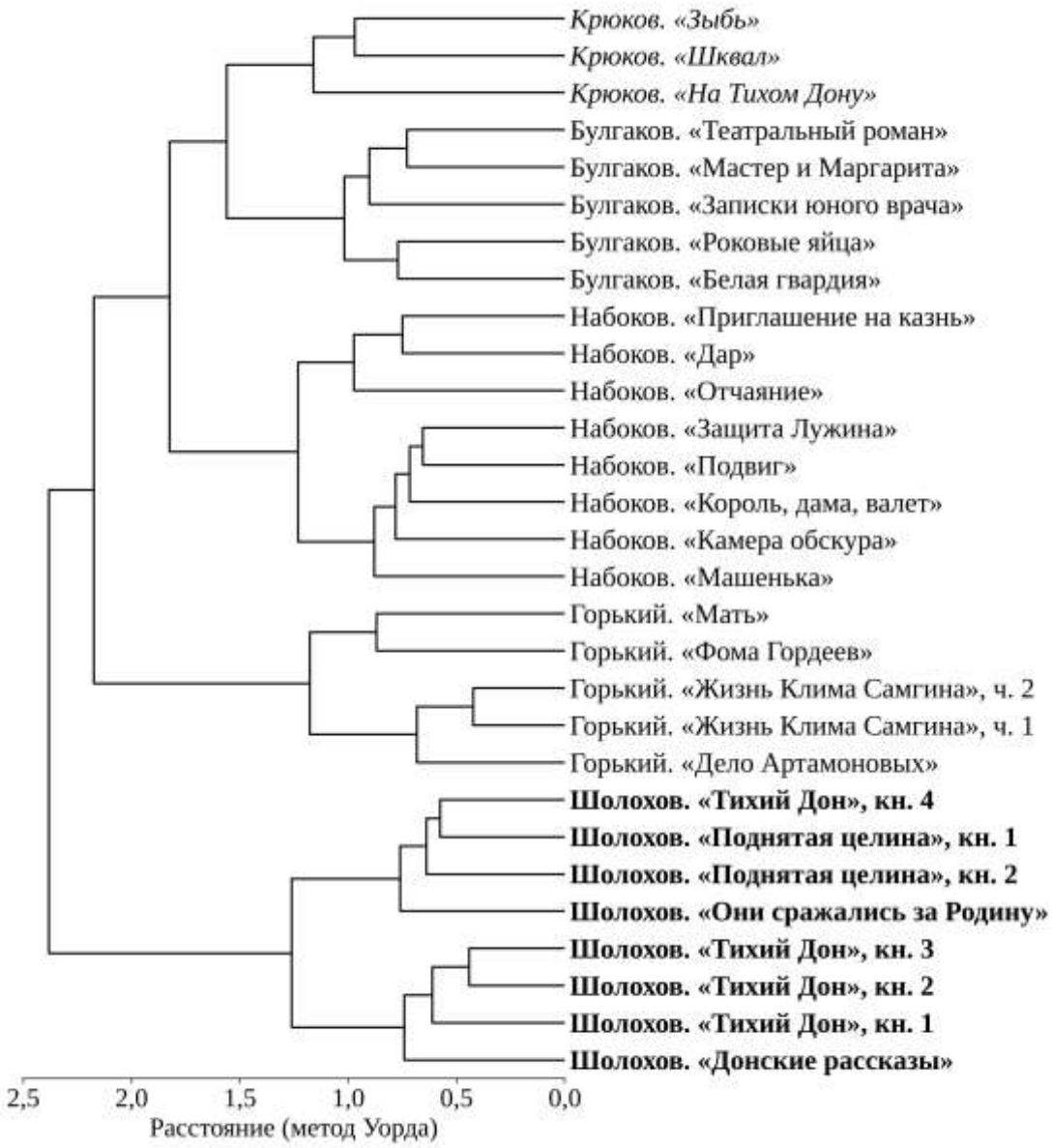


Рис. 1. Дендрограмма текстов корпуса (Delta, 500 наиболее частотных слов, метод Уорда)

Внутри шолоховской ветви во всех десяти опытах выделяются две группы. Первая объединяет «Донские рассказы» с первыми тремя книгами романа. Во вторую вошли четвёртая книга, обе книги «Поднятой целины» и «Они сражались за Родину». Деление совпадает с хронологией. Четвёртая книга завершена в 1940 году, через восемь лет после выхода первой книги «Поднятой целины».

Средние расстояния подтверждают картину дендрограммы (табл. 3). Каждая книга романа ближе к бесспорным текстам М.А.Шолохова (от 0,708 до 0,859), чем к текстам любого другого автора (от 1,017 до 1,234).

Таблица 3. Средние расстояния Delta от книг «Тихого Дона» до текстов корпуса и атрибуция фрагментов

Кн.	**Ближайший текст вне романа**	**Шолохов[1]**	**Крюков**	**Булгаков**	**Горький**	**Набоков**	**Фрагменты Шолохова[2]**
1	«Донские рассказы» (0,625)	0,859	1,203	1,145	1,234	1,230	40 из 40
2	«Поднятая целина», кн. 1	0,814	1,055	1,057	1,152	1,144	39 из 40

ISSN 3060-5059 (Online)

	(0,698)						
3	«Поднятая целина», кн. 1 (0,556)	0,714	1,017	1,029	1,160	1,145	50 из 50
4	«Поднятая целина», кн. 1 (0,576)	0,708	1,045	1,043	1,121	1,117	60 из 60

Примечание. Словарь из 500 наиболее частотных слов. [1] Среднее расстояние до «Донских рассказов», двух книг «Поднятой целины» и произведения «Они сражались за Родину». [2] Число фрагментов объёмом по 10 000 словоупотреблений, отнесённых к М. А. Шолохову по правилу ближайшего соседа, суммарно при словарях из 200, 400, 600, 800 и 1000 слов.

Тексты Ф. Д. Крюкова не ближе к роману, чем тексты М. А. Булгакова, а для первой книги они даже дальше (1,203 против 1,145). Материал Ф. Д. Крюкова при этом неоднороден. Среднее расстояние между тремя его текстами составляет 1,068, тогда как между бесспорными текстами М. А. Шолохова — 0,796. Поэтому решающим доводом служит не удалённость романа от текстов Ф. Д. Крюкова, а его близость к произведениям М. А. Шолохова.

Атрибуция фрагментов дала однозначный результат. Из 38 фрагментов романа при пяти размерах словаря (190 решений) правило ближайшего соседа отнесло к М. А. Шолохову 189. Единственный отклонившийся фрагмент второй книги при словаре из 400 слов был отнесён к М. А. Булгакову. По правилу наименьшего среднего расстояния все 190 решений получены в пользу М. А. Шолохова. Контрольная проверка на 145 фрагментах бесспорных текстов показала точность от 97,9 до 99,3 % по первому правилу и от 97,2 до 100 % по второму. Хуже других распознаются фрагменты Ф. Д. Крюкова (четыре или пять из шести), что объясняется малым объёмом его текстов.

Четвёртая серия касалась восьми частей романа (рис. 2). Три первые части, составляющие первую книгу, ближе всего к «Донским рассказам» (от 0,719 до 0,751 при словаре из 500 слов). Начиная с четвёртой части ближайшим текстом становится первая книга «Поднятой целины». Расстояние до «Донских рассказов» в седьмой и восьмой частях увеличивается до 0,895 и 0,943 соответственно. Тексты Ф. Д. Крюкова во всех частях оказываются дальше обоих шолоховских ориентиров. Наибольшее расстояние до них наблюдается в первой и второй частях (1,353 и 1,295) — именно тех, которые противники шолоховского авторства чаще всего связывают с Ф. Д. Крюковым. Аналогичные результаты получены при словарях из 300 и 700 слов.

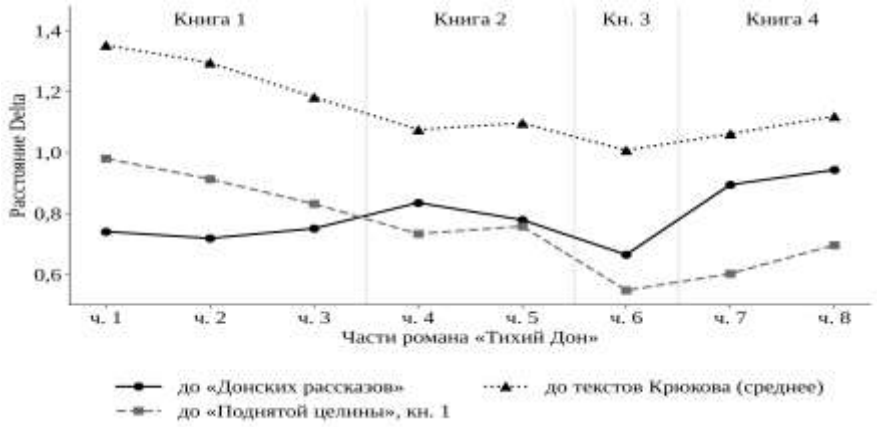


Рис. 2. Расстояние Delta от восьми частей «Тихого Дона» до текстов сравнения (500 наиболее частотных слов)

ОБСУЖДЕНИЕ

Результаты первой серии конкретизируют тезис Н. П. Великановой и Б. В. Орехова о мифологизации количественных подходов [5, с. 73]. Однопараметрические признаки не являются ложными сами по себе. Они фиксируют реальные свойства текста, однако эти свойства зависят от времени создания, жанра и редакторской правки в большей степени, чем от личности автора. Доля служебных слов действительно позволяет отделить первые книги романа от «Донских рассказов». Однако тот же признак позволил бы выявить более существенное различие между произведениями «Мать» и «Фома Гордеев» М. Горького (2,7 процентного пункта), чем между первой книгой романа и ранними рассказами (2,2 процентного пункта). Низкая точность показателя длины предложения в исследовании Л. А. Вдовиной [4, с. 97] и замечание Д. В. Хмелёва об ограничениях авторского инварианта [8] согласуются с полученными нами данными.

Вторая серия показывает, что воспроизводимость подготовки текста важнее сложности математического аппарата. Различие абсолютных частот, принятое за формальное доказательство, объясняется наличием двойных пробелов в электронной версии текста. Итоговый вывод В. А. Бубнова и А. В. Сурвило тем не менее подтверждается, хотя и по другой причине. Отрицательная вторая разность спектра характерна для всех текстов

М. А. Шолохова и отсутствует в произведениях Ф. Д. Крюкова. Её источник — повышенная частота предлога «на» — зафиксирован и по данным НКРЯ [6, с. 217]. Тематическое объяснение маловероятно, поскольку Ф. Д. Крюков пишет о том же донском быте, но предлог «на» употребляет реже. Тем не менее частотность одного предлога остаётся лишь дополнительным аргументом.

Результаты применения метода Delta согласуются с данными Н. П. Великановой и Б. В. Орехова. При словаре из 200 слов расстояние от первой книги романа до «Донских рассказов» составило в нашем исследовании 0,546, а в их работе — 0,59 [5, с. 80], хотя корпуса и редакции текстов различаются. Совпадает и вывод о том, что поздние произведения М. А. Шолохова хуже подходят для сравнения из-за эволюции стиля [5, с. 80]. Профиль частей романа наглядно демонстрирует эту эволюцию. От первой книги к четвёртой близость к ранним рассказам сменяется близостью к «Поднятой целине». Именно такой стилистический дрейф однопараметрические методы интерпретировали как смену автора, тогда как метод Delta относит к М. А. Шолохову 189 случаев из 190.

С выводами А. В. Зенкова наши данные расходятся. Начальные части романа, которые он сблизил с произведениями Ф. Д. Крюкова [9, p. 15], в нашем исследовании оказались наиболее удалёнными от текстов Ф. Д. Крюкова. Расхождение может объясняться природой используемого признака. Числительные встречаются в тексте редко, причём наибольшую информативность имеют частоты первых цифр 1, 2 и 3 [9, p. 11]. Кроме того, тексты Ф. Д. Крюкова в его работе были объединены в один файл [9, p. 16].

Ограничения исследования определяют направления дальнейшей работы. Роман анализировался по распространённой электронной версии, а не по научному изданию. Корпус текстов Ф. Д. Крюкова невелик и неоднороден по жанровому составу. В него не вошли произведения других кандидатов, рассмотренных Н. П. Великановой и Б. В. Ореховым, в частности А. С. Серафимовича и В. Севского [5, с. 77–78]. Сам по себе метод Delta не доказывает авторства. Он формирует вторичную атрибуционную гипотезу, которую необходимо проверять средствами текстологии [7, с. 77]. В случае «Тихого Дона» такую проверку обеспечивают обнаруженные рукописи первой и второй книг [2, с. 61].

ЗАКЛЮЧЕНИЕ

Проверка на едином корпусе из 29 текстов позволяет сформулировать четыре основных вывода.

Средняя длина предложения и доля служебных слов изменяются внутри творчества М. А. Шолохова сильнее, чем между произведениями разных писателей, поэтому не могут рассматриваться как устойчивые авторские инварианты.

Частоты предлогов, опубликованные для романа В. А. Бубновым и А. В. Сурвило, воспроизводятся только при подсчёте отрезков между пробелами вместо слов.

Форма спектра с повышенной частотой предлога «на» при этом действительно позволяет отличить тексты М. А. Шолохова от произведений Ф. Д. Крюкова.

Метод Delta относит к М. А. Шолохову все четыре книги и все восемь частей романа. Первая книга наиболее близка к «Донским рассказам», а четвёртая — к «Поднятой целине», что отражает эволюцию индивидуального авторского стиля.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Бубнов В. А. Анализ логической структуры романа «Тихий Дон» с помощью информационных технологий // Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2015. – № 3. – С. 50–59. – URL: https://journals.rudn.ru/informatization-education/article/view/13261.

2. Бубнов В. А., Сурвило А. В. Сравнительный компьютерный анализ текста романа «Тихий Дон» с текстами четырех рассказов Фёдора Крюкова // Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2016. – № 1. – С. 60–69. – URL: https://journals.rudn.ru/informatization-education/article/view/13279.

3. Бубнов В. А., Сурвило А. В. Сравнительный компьютерный анализ текста романа «Тихий Дон» с текстами Фёдора Крюкова, Александра Серафимовича и Ивана Родионова // Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2016. – № 2. – С. 73–82. – URL: https://journals.rudn.ru/informatization-education/article/view/13217.

4. Вдовина Л. А. Атрибуция авторства на основе оценки стилеметрических параметров текста (на материале англоязычных туристических блогов): магистерская диссертация. – Красноярск: Сибирский федеральный университет, 2021. – 121 с. – URL: https://elib.sfu-kras.ru/handle/2311/141791.

5. Великанова Н. П., Орехов Б. В. Цифровая текстология: атрибуция текста на примере романа М. А. Шолохова «Тихий Дон» // Мир Шолохова. – 2019. – № 1 (11). – С. 70–82. – URL: https://publications.hse.ru/articles/314793960.

6. Двинятин Ф. Н., Ковалев Б. В. «Сирин не уступает Леонову»: метод Delta для стилеметрического анализа русских романов межвоенного периода // Шаги / Steps. – 2024. – Т. 10, № 3. – С. 207–229. – DOI: https://doi.org/10.22394/2412-9410-2024-10-2-207-229.

7. Ковалев Б. В. Стилеметрический анализ вариативности индивидуального авторского стиля в романах Марио Варгаса Льосы: дис. … канд. филол. наук: 5.9.8. – Санкт-Петербург: Санкт-Петербургский государственный университет, 2025. – 177 с.

8. Хмелёв Д. В. Распознавание автора текста с использованием цепей А. А. Маркова // Вестник Московского

ISSN 3060-5059 (Online)

университета. Серия 9: Филология. – 2000. – № 2. – С. 115–126. – URL: http://www.philol.msu.ru/~lex/khmelev/published/vestnik/a.pdf.

9. Zenkov A. V. A New Statistical Method of Text Attribution // International Journal of Professional Science. – 2017. – № 1. – P. 6–21. – URL: https://cyberleninka.ru/article/n/new-statistical-method-of-text-attribution.