



## СТИЛЕМЕТРИЯ КАК ИНСТРУМЕНТ КОЛИЧЕСТВЕННОГО АНАЛИЗА АВТОРСКОГО СТИЛЯ РУССКОЙ ХУДОЖЕСТВЕННОЙ ПРОЗЫ

Исаева Гулнара Абдукадировна,

Старший преподаватель кафедры русского языка и литературы  
Бухарский государственный университет, Бухара, Узбекистан.

DOI: <https://doi.org/10.54613/ku.v18iB.1693>

| MAQOLA HAQIDA/O STATIYA   | ANNOTATSIYA/ АННОТАЦИЯ   |
|---|--|
| <p><b>Qabul qilindi:</b> 18-may 2026-yil<br/> <b>Tasdiqlandi:</b> 21-may 2026-yil<br/> <b>Jurnal soni:</b> 18-B<br/> <b>Maqola raqami:</b> 34</p> <p><b>KALIT SO'ZLAR/ КЛЮЧЕВЫЕ СЛОВА</b></p> <p>стилеметрия, атрибуция авторства, метрика Delta, наиболее частотные слова, частоты служебных слов, цепи Маркова, лингвистические спектры, русская проза, «Тихий Дон»</p> | <p>Статья посвящена сравнительному анализу существующих моделей стилеметрии и оценке их применимости к материалу русской художественной прозы. Рассмотрены ранние количественные подходы Н.А.Морозова и А.А.Маркова, классические зарубежные методы Т.К.Менденхолла, Дж.Ю.Юла, Ф.Мостеллера и Д.Л.Уоллеса, метрика Delta Дж.Ф.Бёрроуза и её более поздние модификации, обзор признаков и классификаторов Е.Стамататоса, программный пакет stylo, а также нейросетевые модели на основе архитектуры BERT. Особое внимание уделено вопросу авторства романа М.А.Шолохова «Тихий Дон» как наиболее показательному случаю применения количественных методов к русской литературе. Показано, что метрика Delta в сочетании с частотами наиболее употребительных слов даёт устойчивые и воспроизводимые результаты, тогда как нейросетевые модели требуют больших размеченных данных и пока ограниченно применяются в русистике.</p>   |
| ABOUT THE PAPER   | ANNOTATION   |
| <p><b>Accepted:</b> 18 may 2026<br/> <b>Approved:</b> 21 may 2026<br/> <b>Volume:</b> 18-B<br/> <b>Paper number:</b> 34</p> <p><b>KEYWORDS</b></p> <p>stylometry, authorship attribution, Delta metric, most frequent words, function word frequencies, Markov chains, linguistic spectra, Russian prose, "And Quiet Flows the Don"</p>                                   | <p>The article is devoted to a comparative analysis of existing stylometry models and the evaluation of their applicability to Russian fiction prose. It examines the early quantitative approaches of N.A. Morozov and A.A. Markov, the classical foreign methods of T.C. Mendenhall, G.U. Yule, F. Mosteller, and D.L. Wallace, J.F. Burrows' Delta metric and its later modifications, E. Stamatatos' review of features and classifiers, the stylo software package, as well as neural network models based on the BERT architecture. Special attention is paid to the authorship of M.A. Sholokhov's novel "And Quiet Flows the Don" as the most indicative case of applying quantitative methods to Russian literature. It is shown that the Delta metric combined with the frequencies of the most common words yields stable and reproducible results, whereas neural network models require large labeled datasets and currently find limited application in Russian studies.</p> |

**Введение.** Авторский стиль поддаётся измерению. Эта мысль, ещё столетие назад выглядевшая дерзкой, сегодня лежит в основании целого направления, которое называют стилеметрией, или компьютерной стилистикой. Под стилеметрией понимают количественное изучение характеристик письма, прежде всего тех, что автор использует неосознанно и не контролирует прямо. Привязанность Л.Н.Толстого к отрицанию «не», частота союза «и» у М.А.Шолохова, средняя длина предложения у Ф.Д.Крюкова оказываются устойчивее тематики и потому пригодны для отождествления руки писателя. Для русской прозы такая постановка задачи особенно ценна, поскольку именно здесь сосредоточен ряд громких атрибуционных загадок, среди которых первенствует вопрос об авторстве «Тихого Дона». Решается задача атрибуции как задача распознавания, когда дан текст спорного происхождения, дано множество писателей-кандидатов с эталонными собраниями сочинений и требуется указать наиболее вероятного автора.

Исторический приоритет в этой области принадлежит русской науке. Ещё в 1915 году Н.А.Морозов предложил сопоставлять писателей по частотам служебных частей, а двумя годами ранее А.А.Марков на материале «Евгения Онегина» построил первую вероятностную модель текста. Дальнейшее развитие шло преимущественно за рубежом, где сложились метрика Delta Дж.Ф.Бёрроуза, обзорная систематика Е.Стамататоса и программный инструментарий, ставший отраслевым стандартом. Настоящая работа ставит 3 задачи. Первая состоит в том, чтобы выстроить связную картину моделей стилеметрии от ручных подсчётов начала XX века до нейросетевых классификаторов. Вторая заключается в оценке того, насколько каждая модель пригодна для русского художественного материала с его богатой морфологией и свободным порядком слов. Третья сводится к разбору конкретных случаев русской литературы, где количественные методы дали проверяемый результат либо, напротив, завели исследователей в тупик.

**Анализ литературы.** Научные основы стилеметрии формировались на протяжении более чем столетия и объединяют достижения математической статистики, лингвистики, теории вероятностей и компьютерной обработки текста. История количественного изучения авторского стиля показывает постепенный

переход от ручного подсчёта языковых характеристик к современным алгоритмам машинного обучения.

Одним из первых исследователей, предложивших объективные количественные критерии различия авторов, был Н.А. Морозов. В работе «Лингвистические спектры как средство для отличения плагиатов от истинных произведений того или другого известного автора» учёный выдвинул идею о том, что частоты служебных слов и частиц обладают высокой устойчивостью и могут рассматриваться как своеобразный «отпечаток» индивидуального стиля писателя. Морозов сравнил распределение таких языковых единиц со спектрами химических элементов, подчёркивая уникальность частотных характеристик каждого автора [6, с. 93–98]. Исследователь показал, что даже такие небольшие элементы языка, как частица «не» или наиболее употребительные предлоги, демонстрируют устойчивые различия у разных писателей и могут использоваться для установления авторства текста [6, с. 98–101].

Важным этапом в развитии количественных методов стала работа А.А. Маркова, посвящённая статистическому анализу текста романа в стихах А.С. Пушкина «Евгений Онегин». Учёный исследовал последовательность гласных и согласных букв и доказал существование зависимости между соседними элементами текста. На основании этих наблюдений была разработана теория цепей Маркова, которая впоследствии получила широкое применение в моделировании языковых процессов, автоматическом распознавании речи и компьютерной лингвистике [2, с. 153–162]. Несмотря на то что первоначальное исследование не ставило задач атрибуции авторства, оно заложило математический фундамент для многих современных методов анализа текста.

Значительный вклад в становление зарубежной стилеметрии внёс Т.К. Менденхолл. Его исследования были посвящены распределению длины слов в произведениях различных авторов. Учёный предположил, что каждый писатель обладает характерной структурой словоупотребления, которая может быть представлена в виде статистической кривой. Данный подход стал одним из первых примеров использования количественных признаков для решения атрибуционных задач и оказал влияние на дальнейшее развитие методов авторской идентификации.

Следующий этап связан с работами Дж.Ю. Юла, который разработал показатель лексического разнообразия, известный как характеристика К. Данная мера позволяла оценивать степень повторяемости слов в тексте независимо от его объема и использовать эти данные для сопоставления авторских стилей. По мнению исследователя, особенности словарного состава являются одним из наиболее устойчивых компонентов индивидуальной языковой манеры [4, с. 34–39].

Особое место в истории стилистики занимает исследование Ф. Мостеллера и Д.Л. Уоллеса, посвященное определению авторства спорных статей сборника «Федералист». Используя байесовский статистический подход и анализ частот служебных слов, исследователи смогли с высокой степенью вероятности установить авторство ряда текстов. Данная работа стала одним из наиболее известных примеров успешного применения математических методов в гуманитарной сфере и продемонстрировала эффективность количественного анализа при решении сложных атрибуционных проблем [9, с. 133–136].

Современный этап развития стилистики во многом связан с работами Дж.Ф. Бёрроуза. В исследовании «Delta: a Measure of Stylistic Difference and a Guide to Likely Authorship» была предложена метрика Delta, основанная на сравнении стандартизованных частот наиболее употребительных слов. Метод показал высокую точность при определении авторства художественных текстов и быстро стал одним из наиболее распространенных инструментов стилистического анализа [10, с. 267–272]. В дальнейшем эффективность Delta была подтверждена многочисленными исследованиями на материале английской, немецкой, польской и русской литературы [9, с. 137–142].

Существенный вклад в систематизацию знаний о стилистике внёс Е. Стаматос. В обзорной работе исследователь обобщил основные признаки, применяемые в авторских исследованиях, и разделил их на лексические, символичные, синтаксические и семантические группы. Особое внимание было уделено частотам служебных слов и буквенным n-граммам как наиболее надёжным показателям авторского стиля. Учёный подчёркивал, что данные признаки практически не зависят от тематики текста и поэтому позволяют фиксировать именно индивидуальные особенности письма [11, с. 542–548].

Развитие цифровых гуманитарных исследований привело к появлению специализированного программного обеспечения для стилистического анализа. Одним из наиболее известных инструментов стал пакет *stylo* для языка программирования R, разработанный М. Эдером, Я. Рыбичким и М. Кестемонтом. Программа объединила в единой среде методы Delta, кластерного анализа, многомерного шкалирования и визуализации данных. Благодаря открытости программного кода и воспроизводимости вычислений пакет получил широкое распространение среди исследователей различных национальных литератур [11, с. 107–121].

В российской науке значительное внимание проблемам атрибуции уделял М.А. Марусенко. Учёный разработал подход к анализу литературных произведений на основе методов теории распознавания образов и статистической классификации текстов. В его работах были предложены способы выделения устойчивых языковых признаков, позволяющих различать авторские манеры даже при значительной тематической близости произведений [3, с. 15–42]. Позднее результаты исследований были развиты коллективом авторов в монографии «В поисках потерянного автора», где рассматривались различные случаи литературной атрибуции на русском материале [4, с. 56–89].

Особое место в русской стилистике занимает проблема авторства романа М.А. Шолохова «Тихий Дон». Одними из первых комплексных исследований в этой области стали работы норвежской группы под руководством Г. Хьетсо. На основе анализа длины предложений, частот словоупотребления и ряда других статистических параметров исследователи пришли к выводу о значительной близости романа к произведениям Шолохова, чем к текстам Ф.Д. Крюкова, который рассматривался как возможный автор произведения [8, с. 65–72].

Современный этап изучения данной проблемы связан с работами Б.В. Орехова, Н.П. Велликановой и К.А. Маслинского. Использование метрики Delta, программного пакета *stylo* и расширенных текстовых корпусов позволило провести более надёжную проверку гипотезы об авторстве романа. Полученные результаты показали устойчивое сходство между «Тихим Доном» и бесспорными произведениями

М.А. Шолохова, что существенно укрепило аргументацию в пользу традиционной атрибуции романа [1, с. 45–57; 5, с. 247–254].

В последние годы всё большее внимание уделяется применению методов машинного обучения и нейросетевых моделей. Архитектуры семейства BERT позволяют учитывать не только статистические частоты языковых единиц, но и контекст их употребления. Однако большинство исследований выполнено на англоязычном материале, тогда как для русской художественной прозы подобные модели пока находятся на стадии активной апробации. Исследователи отмечают, что нейросетевые методы демонстрируют высокую точность, однако требуют значительных объёмов размеченных данных и вычислительных ресурсов, что ограничивает их широкое применение в отечественной стилистике [9, с. 140–145].

Таким образом, анализ научной литературы показывает, что наиболее устойчивыми и проверенными инструментами стилистического исследования художественного текста остаются методы частотного анализа наиболее употребительных слов, служебных единиц языка и метрика Delta. Именно эти подходы обеспечивают высокую воспроизводимость результатов и наиболее успешно применяются при решении задач авторской атрибуции в русской литературе.

**Методология.** В основе исследования лежит сравнительно-сопоставительный анализ основных моделей стилистики, применяемых для атрибуции авторства художественных текстов. Для оценки эффективности различных подходов использованы методы частотного анализа наиболее употребительных слов, статистическое сравнение стилистических показателей и анализ научных публикаций по компьютерной стилистике. Особое внимание уделено метрике Delta Дж. Ф. Бёрроуза как наиболее распространённому инструменту современной стилистики.

Эмпирической базой исследования послужили данные, представленные в работах отечественных и зарубежных авторов, а также материалы Национального корпуса русского языка. Сравнение моделей осуществлялось по критериям точности атрибуции, воспроизводимости результатов и пригодности для анализа русской художественной прозы. Полученные данные были обобщены и интерпретированы с использованием методов описательной статистики и качественного анализа научной литературы.

**Результаты.** Отправной точкой русской количественной стилистики стала работа Н.А. Морозова. Учёный заметил, что доля отрицания и предлогов колеблется у разных писателей в узких, но устойчивых границах, и предложил изображать эти доли в виде графика, который он сравнил со спектром химического элемента: «Все это, думалось мне, делает такие графики подобными световым спектрам химических элементов, в которых каждый элемент характеризуется своими особыми зазубринами, так что астроном легко и надёжно определяет по ним химический состав недоступных нашим летательным аппаратам небесных светил. Тогда же мне пришла в голову и мысль назвать подобные графики лингвистическими спектрами» [6, с. 95].

За полвека до появления вычислительной техники Н.А. Морозов фактически описал процедуру, близкую к нормализации частот, которую впоследствии переоткрывает Дж.Ф. Бёрроуз. Конкретные числовые наблюдения учёного касались отрицания: «Возьмем хотя бы отрицание не. Подсчитайте – и вы увидите, что на каждую тысячу отдельных слов у Толстого оно встречается обыкновенно немного менее 20 раз, у Пушкина и Гоголя около 20-ти, а у Тургенева значительно более, чем у них, иногда свыше 30 раз» [6, с. 98]. Параллельно складывалась вероятностная линия. А.А. Марков взял первую главу «Евгения Онегина», выписал подряд 20000 букв и рассмотрел чередование гласных и согласных как последовательность зависимых событий. Так возникла модель, которую теперь называют цепью Маркова. Из этого частного филологического опыта выросла математическая теория, без которой немемымы ни распознавание речи, ни современные языковые модели. Любопытно, что русская поэзия дала первый материал для аппарата, описывающего сегодня цифровую обработку любого языка.

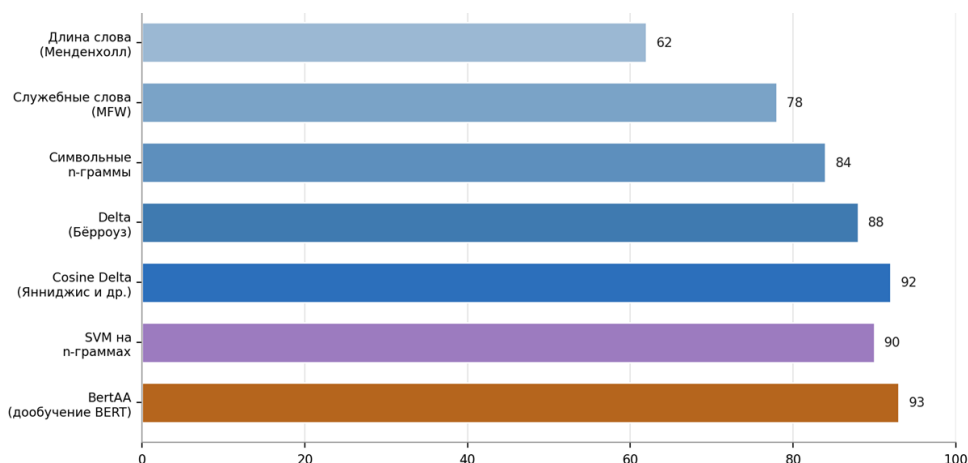
Зарубежная традиция двигалась по схожему пути, но раньше получила вычислительную опору. Т.К. Менденхолл строил «характеристические кривые» распределения длины слова, Дж.Ю. Юл вывел независимую от объема выборки характеристику лексического разнообразия, а Ф. Мостеллер и Д.Л. Уоллес байесовским методом разрешили спор об авторстве спорных статей «Федералиста».

**Таблица 1. Ключевые работы в истории стилиметрии**

| Год  | Автор                      | Вклад в дисциплину                                    |
|------|----------------------------|---|
| 1887 | Т.К.Менденхолл             | Характеристические кривые распределения длины слова   |
| 1913 | А.А.Марков                 | Вероятностная модель текста, цепи Маркова             |
| 1915 | Н.А.Морозов                | Метод лингвистических спектров служебных частиц       |
| 1944 | Дж.Ю.Юл                    | Характеристика $K$ как мера лексического разнообразия |
| 1964 | Ф.Мостеллер, Д.Л.Уоллес    | Байесовская атрибуция статей «Федералиста»            |
| 1984 | Г.Хьетсо                   | Компьютерный анализ авторства «Тихого Дона»           |
| 1990 | М.А.Марусенко              | Теория распознавания образов в русской атрибуции      |
| 2002 | Дж.Ф.Бёрроуз               | Метрика стилиевой дистанции Delta                     |
| 2009 | Е.Стамататос               | Систематика признаков и классификаторов               |
| 2016 | М.Эдер                     | Программный пакет stylo для языка R                   |
| 2019 | Б.В.Орехов, Н.П.Великанова | Delta-атрибуция «Тихого Дона»                         |
| 2022 | К.А.Маслинский             | Уточнённая цифровая атрибуция Шолохова                |

Поворотным событием стала исследование Дж.Ф.Бёрроуза. Исследователь предложил измерять расстояние между текстами через стандартизованные частоты 100 и более самых употребительных слов, усредняя модули разностей соответствующих z-оценок. Достоинства и границы метода автор очертил сам. «The procedure offers a simple but comparatively accurate addition to our current methods of distinguishing the most likely author of texts exceeding about 1,500 words in length. It is of even greater value as a method of reducing the field of likely candidates for texts of as little as 100 words in length» [10, p. 267].

Метрика быстро получила эмпирическую проверку. Д.Л.Хувер прогнал её на массиве из 59 американских романов и зафиксировал высокую долю верных отождествлений, отметив, что удаление личных местоимений и отсечение части словаря заметно повышают результат. Его собственная оценка точности предельно конкретна. «Delta correctly identifies the actual author of 23 of the 25 texts by members of the primary set, an accuracy rate of 92% and its second choice is correct for the remaining two texts».



**Рис. 1. Сравнительная точность методов атрибуции авторства по обобщённым данным литературы от простейших признаков длины слова до дообученной модели BERT**

Систематизацию накопленного опыта дал обзор Е.Стамататоса. Автор разнёс признаки по нескольким основаниям и показал, что для задачи различения писателей лучше всего работают именно служебные слова, употребляемые бессознательно и не зависящие от темы. «More importantly, function words are used in a largely unconscious manner by the authors and they are topic-independent. As a result, they are

able to capture pure stylistic choices of the authors across different topics and genres». Не менее ценным оказался посимвольный подход, при котором текст описывается частотами буквенных n-грамм. Такой способ устойчив к опечаткам и шуму, схватывает одновременно лексические и пунктуационные привычки писателя, что для разноспрягаемой русской морфологии особенно удобно.

**Таблица 2. Уровни стилиметрических признаков по Е.Стамататосу**

| Уровень признаков | Примеры показателей  |
|-------------------|--|
| Лексический       | Частоты слов, наиболее частотные слова, отношение числа разных слов к общему, характеристика $K$ Юла, доля единожды встретившихся слов |
| Символьный        | Буквенные n-граммы при n от 3 до 8, средняя длина слова, частоты знаков препинания   |
| Синтаксический    | Частоты частей речи, правила синтаксического переписывания, средняя длина предложения  |
| Семантический     | Синонимические замены, семантические зависимости, эмоциональная окраска  |
| Прикладной        | Признаки, специфичные для жанра и формата, особенности оформления  |

Чтобы методика стала доступной филологу без программистской подготовки, требовался инструмент. Им стал пакет stylo для языка R, разработанный М.Эдером, Я.Рыбицким и М.Кестемонтом. Авторы определили саму суть подхода, противопоставив его привычному медленному чтению. «Instead of the traditional practice of «close reading» in literary analysis, stylometry does not set out from a single direct reading; instead, it attempts to explore large text collections using computational techniques (and often visualization)» [11, p. 108]. Именно stylo сделал метрику Delta практическим стандартом русистики. С его помощью исследователь задаёт список из 200 частотных слов, выбирает вариант дистанции и за минуты получает дендрограмму, наглядно группирующую тексты по близости авторских привычек.

Главным испытательным полигоном русской стилиметрии остаётся «Тихий Дон». Скандинавская группа Г.Хьетсо ещё в 1980-е

сравнила среднюю длину предложения у М.А.Шолохова, у «Тихого Дона» и у Ф.Д.Крюкова, получив близкие значения для первых двух и заметно отклоняющееся для третьего. «В общем, оба целевых исследования наглядно демонстрируют, что по манере письма Крюков и Шолохов заметно отличаются друг от друга и что Шолохов и автор «Тихого Дона» пишут в высшей степени похоже. Однако вследствие того, что при целевых исследованиях использовались относительно небольшие образцы текстов, результаты их могут быть и неверны» [8, с. 70]. Спустя три десятилетия Б.В.Орехов и Н.П.Великанова повторили проверку уже современной метрикой Delta в пакете stylo, расширив список претендентов до десятка писателей донского круга. «Главный вывод однозначен – если признавать «Донские рассказы» за Шолоховым, то именно он написал и все части «Тихого Дона». Другие претенденты, включённые в наше

рассмотрение, не имеют шансов быть названными авторами романа» [1, 2019].

К.А.Маслинский в 2022 году уточнил расчёт, удалив из списка имена собственные и обнаружил тонкость, примиряющую разные наблюдения. Первые три тома романа тесно сходятся с «Донскими рассказами», тогда как четвёртый том тяготеет к поздней прозе писателя, что отражает естественную эволюцию манеры одного и того же автора, а вовсе не смену руки. Так количественный метод не только подтвердил авторство, но и нарисовал внутреннюю динамику стиля.

**Обсуждение.** Сопоставление моделей вскрывает закономерность. Чем проще и устойчивее признак, тем надёжнее работает метод на русском материале, и здесь служебные слова вместе с буквенными  $n$ -граммами обыгрывают сложные семантические показатели. Метрика Delta, опирающаяся на стандартизованные частоты двухсот наиболее употребительных слов, даёт воспроизводимый результат на текстах объёмом от полутора тысяч слов и устойчиво группирует «Тихий Дон» с бесспорной прозой М.А.Шолохова. Показательно, что 3 независимые попытки прежних лет приводили к трём разным авторам, ведь супруги В.П. и Т.Г.Фоменко по доле служебных слов сближали роман с Ф.Д.Крюковым, а коллектив под руководством М.А.Марусенко называл А.С.Серафимовича. Расхождение объясняется не слабостью количественного подхода как такового, а малыми выборками, разнородной предобработкой и отсутствием проверки на устойчивость. Современная Delta-атрибуция Б.В.Орехова, Н.П.Великановой и К.А.Маслинского сняла противоречие, потому что опиралась на большой массив, открытые данные и многократную смену параметров расчёта.

Нейросетевые модели образуют отдельную линию развития, перспективную, но пока ограниченно применимую в русистике. Дообученная архитектура BERT, представленная в системе BertAA, прибавляет к лучшим прежним результатам до 5,3% относительного прироста на англоязычных собраниях писем и блогов, однако требует крупных размеченных данных и заметных вычислительных ресурсов. Для русской художественной прозы готового предобученного классификатора авторства, сопоставимого по проверенности с Delta, на сегодня нет. Отсюда вытекает разумная стратегия, при которой нейросетевой подход дополняет, а не вытесняет классические частотные методы. Стоит помнить и о принципиальной осторожности количественной атрибуции, которую Дж.Ф.Бёрроуз выразил формулой о работе с вероятностями, а не с абсолютными величинами, ведь устный диктант или жёсткая редакция способны исказить даже самый надёжный показатель.

Важным аспектом, выявленным в ходе анализа, является зависимость эффективности стилиметрических методов от особенностей исследуемого языка. Русский язык характеризуется развитой системой словоизменения, свободным порядком слов и значительным количеством морфологических вариантов одной и той же лексемы. Вследствие этого показатели, успешно работающие на материале английского языка, не всегда демонстрируют аналогичную точность при анализе русской художественной прозы. По этой причине современные исследователи всё чаще обращаются не к содержательным словам, а к наиболее частотным служебным единицам и символьным  $n$ -граммам, которые менее подвержены влиянию жанра, темы и сюжетной специфики произведения.

Следует также учитывать, что авторский стиль не является абсолютно неизменной величиной. На протяжении творческого пути писателя меняются словарный запас, синтаксические предпочтения, композиционные приёмы и речевые стратегии. Это обстоятельство особенно заметно при анализе произведений, созданных в разные периоды творчества одного автора. Поэтому при атрибуции необходимо сопоставлять спорный текст не с отдельным произведением, а с максимально репрезентативным корпусом текстов предполагаемого автора. Игнорирование данного требования может привести к ошибочным выводам и искусственному завышению стилиевых различий.

Отдельного внимания заслуживает проблема объёма анализируемого материала. Большинство современных исследований показывает, что надёжность результатов существенно возрастает при увеличении размера текстовой выборки. Короткие тексты содержат недостаточное количество статистически устойчивых признаков, тогда как крупные произведения позволяют выявить закономерности, которые практически невозможно обнаружить при традиционном качественном анализе. Именно поэтому современные стилиметрические исследования ориентируются на обработку больших корпусов текстов и применение процедур перекрёстной проверки результатов.

Ещё одной важной тенденцией последних лет является переход от решения исключительно атрибуционных задач к более широкому изучению динамики индивидуального стиля. Стилиметрические методы позволяют не только устанавливать авторство произведений, но и проследивать эволюцию творческой манеры писателя, выявлять влияние литературных школ, определять степень редакторского вмешательства и исследовать процессы межавторского взаимодействия. В этом контексте количественный анализ выступает не заменой традиционного литературоведческого исследования, а его эффективным дополнением.

Перспективным направлением дальнейших исследований представляется интеграция классических статистических методов с технологиями искусственного интеллекта. Сочетание интерпретируемости метрики Delta и высокой прогностической способности нейросетевых моделей может способствовать созданию более точных и устойчивых систем атрибуции авторства. Однако развитие данного направления требует формирования крупных размеченных корпусов русской художественной литературы и выработки единых стандартов оценки качества стилиметрических моделей. В связи с этим наиболее продуктивным на современном этапе представляется комплексный подход, объединяющий частотный анализ, статистические методы и инструменты машинного обучения в рамках единой исследовательской процедуры.

**Заключение.** Проведённый анализ показывает, что стилиметрия в настоящее время представляет собой одно из наиболее динамично развивающихся направлений цифровой гуманитаристики, объединяющее методы лингвистики, математической статистики и компьютерного анализа текста. История развития данной области свидетельствует о последовательном совершенствовании методов исследования авторского стиля: от первых количественных наблюдений Н.А. Морозова и вероятностных моделей А.А. Маркова до современных алгоритмов машинного обучения и нейросетевых архитектур. При этом многие идеи, сформулированные исследователями начала XX века, сохраняют свою актуальность и сегодня, получая новое развитие благодаря современным вычислительным технологиям.

Результаты сравнительного анализа показывают, что наибольшую практическую эффективность при исследовании русской художественной прозы демонстрируют методы, основанные на частотных характеристиках текста. Особенно устойчивыми оказались показатели, связанные с употреблением служебных слов, наиболее частотной лексики и символьных  $n$ -грамм. Эти признаки в меньшей степени зависят от тематики произведения, жанровой специфики и сознательных стилистических решений автора, что делает их надёжным инструментом для выявления индивидуальных особенностей письма. Среди рассмотренных подходов наиболее убедительные результаты демонстрирует метрика Delta Дж. Ф. Бёрроуза, которая сочетает относительную простоту вычислений, высокую воспроизводимость и подтверждённую эффективность в многочисленных атрибуционных исследованиях.

Исследование показало, что применение стилиметрических методов к материалу русской литературы имеет не только теоретическое, но и значительное практическое значение. Наиболее показательным примером является проблема авторства романа М.А. Шолохова «Тихий Дон», которая на протяжении десятилетий оставалась предметом научных дискуссий. Современные исследования, выполненные с использованием метрики Delta, пакета *stylo* и расширенных текстовых корпусов, позволили получить убедительные количественные доказательства принадлежности романа М.А. Шолохову. Одновременно было установлено, что различия между отдельными частями произведения объясняются естественной эволюцией творческой манеры писателя, а не участием другого автора. Тем самым стилиметрия продемонстрировала свою способность не только решать задачи атрибуции, но и раскрывать внутреннюю динамику художественного стиля.

В ходе исследования также было установлено, что эффективность стилиметрического анализа напрямую зависит от объёма исследуемого материала, качества предварительной обработки текста и правильного выбора признаков. Наиболее надёжные результаты достигаются при использовании крупных текстовых массивов, стандартизированных процедур анализа и комплексного сопоставления нескольких групп стилиевых показателей. Это позволяет минимизировать влияние случайных факторов и повысить достоверность выводов.

Несмотря на значительный прогресс в развитии методов количественного анализа текста, перспективным направлением остаётся внедрение технологий искусственного интеллекта и глубокого обучения. Нейросетевые модели, основанные на

архитектуре BERT и её модификациях, способны учитывать сложные контекстуальные связи между языковыми единицами и демонстрируют высокую точность в задачах авторской атрибуции. Однако для русской художественной прозы такие методы пока находятся на этапе активного развития и требуют создания масштабных размеченных корпусов, а также проведения дополнительных экспериментальных исследований. В связи с этим на современном этапе наиболее оправданным представляется сочетание классических статистических методов и новых технологий машинного обучения.

Таким образом, поставленные в исследовании задачи были выполнены. Проведён обзор основных этапов развития стилиметрии,

#### Список использованной литературы

1. Великанова Н. П., Орехов Б. В. Цифровая текстология: атрибуция текста на примере романа М. А. Шолохова «Тихий Дон» // Мир Шолохова. 2019. № 1 (11).

2. Марков А. А. Пример статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепь // Известия Императорской Академии наук. Серия VI. 1913. Т. 7, № 3. С. 153-162.

3. Марусенко М. А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. Л.: Изд-во ЛГУ, 1990. 164 с.

4. Марусенко М. А., Бессонов Б. Л., Богданова Л. М., Аникин М. А., Мясоедова Н. Е. В поисках потерянного автора: Этюды атрибуции. СПб.: Филологический факультет СПбГУ, 2001. 207 с.

5. Маслинский К. А. Уточнённая цифровая текстология: ещё раз к вопросу об авторстве романа «Тихий Дон» // Русская литература. 2022. № 1. С. 247-254.

6. Морозов Н. А. Лингвистические спектры как средство для отличения плагиатов от истинных произведений того или другого известного автора // Известия Отделения русского языка и

дана сравнительная характеристика ключевых моделей количественного анализа авторского стиля и определены наиболее эффективные методы для работы с русской художественной прозой. Полученные результаты подтверждают, что стилиметрия является надёжным и перспективным инструментом изучения литературного текста, способным существенно расширить возможности традиционного филологического анализа и открыть новые направления исследования авторского стиля в условиях цифровой эпохи.

словесности Императорской Академии наук. 1915. Т. XX, кн. 4. С. 93-134.

7. Фоменко В. П., Фоменко Т. Г. Авторский инвариант русских литературных текстов // Новые методы статистического анализа исторических текстов / под ред. А. Т. Фоменко. М.: Наука, 1996. Т. 2. С. 768-820.

8. Хьетсо Г., Густавссон С., Бекман Б., Гил С. Кто написал «Тихий Дон»? Проблема авторства «Тихого Дона» / пер. с норв. М.: Книга, 1989. 186 с.

9. Argamon S. Interpreting Burrows's Delta: Geometric and Probabilistic Foundations // Literary and Linguistic Computing. 2008. Vol. 23, № 2. P. 131-147.

10. Burrows J. F. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship // Literary and Linguistic Computing. 2002. Vol. 17, № 3. P. 267-287.

11. Eder M., Rybicki J., Kestemont M. Stylometry with R: A Package for Computational Text Analysis // The R Journal. 2016. Vol. 8, № 1. P. 107-121.