

кандидат филологических наук,
доцент кафедры массовых коммуникаций и медиабизнеса
факультета социальных наук и массовых коммуникаций
Финансового университета при Правительстве Российской Федерации,
профессор кафедры телевидения и радиовещания
Академии медиаиндустрии;
Россия, г. Москва.
e-mail: a.sokolov@academedia.ru

МЕТАМОДЕЛЬ ЭВОЛЮЦИИ СТИЛОМЕТРИЧЕСКИХ МЕТОДОВ НА ПРИМЕРЕ ДИАХРОНИЧЕСКОГО АНАЛИЗА ИДИОСТИЛЯ А. С. СУВОРИНА

Аннотация. В статье представлена метамодель эволюции методов компьютерной стилометрии, демонстрирующая закономерное повышение точности классификации при снижении интерпретируемости результатов. Эмпирической базой исследования послужил корпус публицистических писем А. С. Суворина 1904–1908 гг. Метамодель охватывает три парадигмальных этапа развития дисциплины: классическую стилометрию на основе метода Burrows' Delta (точность 65 %), семантическое моделирование с применением Word2Vec и графового анализа (72,3 %), гибридную архитектуру на базе трансформерной модели BERT с градиентным бустингом LightGBM (87 %). Статистическая значимость прироста точности подтверждена тестом МакНемара ($p < 0,001$) и размером эффекта Коэна ($d = 1,24$). Выявленная закономерность инверсного соотношения точности и интерпретируемости вписывается в общую логику развития количественных методов в филологии — от лингвистических спектров Н. А. Морозова до современных нейросетевых архитектур.

Ключевые слова: стилометрия, метамодель, диахронический анализ, идиостиль, А. С. Суворин, Burrows' Delta, Word2Vec, BERT, машинное обучение, цифровая гуманитаристика.

A. V. SOKOLOV,

Ph.D. of Philological Sciences,
Associate Professor
of the Department of Mass Communications and Media Business,
Financial University under the Government of the Russian Federation,
Professor at the Department of Television and Radio Broadcasting
of the Academy of Media Industry;
Moscow, Russia.
e-mail: a.sokolov@academedia.ru

METAMODEL OF STYLOMETRIC METHODS EVOLUTION: DIACHRONIC ANALYSIS OF A.S. SUVORIN'S IDIOSTYLE

Abstract. *The article presents a metamodel of computational stylometry methods evolution, demonstrating a consistent pattern of classification accuracy improvement accompanied by decreased interpretability. The empirical basis comprises the corpus of A. S. Suvorin's journalistic letters from 1904–1908. The metamodel encompasses three paradigmatic stages: classical stylometry based on Burrows' Delta (65 % accuracy), semantic modeling with Word2Vec and graph analysis (72.3 %), and a hybrid architecture combining BERT transformer with LightGBM gradient boosting (87 %). The statistical significance of accuracy gains is confirmed by McNemar's test ($p < 0.001$) and Cohen's effect size ($d = 1.24$). The identified inverse relationship between accuracy and interpretability aligns with the historical logic of quantitative methods development in philology from N. A. Morozov's linguistic spectra to contemporary neural network architectures.*

Keywords: *stylometry, metamodel, diachronic analysis, idiosyle, A. S. Suvorin, Burrows' Delta, Word2Vec, BERT, machine learning, digital humanities.*

ВВЕДЕНИЕ

Стилометрия как научная дисциплина, посвященная количественному изучению стилистических характеристик текста, насчитывает более полутора столетий истории. За это время произошла радикальная трансформация ее методологического аппарата — от простого подсчета частот служебных слов до применения глубоких нейронных сетей с сотнями миллионов параметров. Однако эта эволюция остается недостаточно отрефлексированной с точки зрения выявления системных закономерностей развития методов.

Цель настоящего исследования — построение метамодели эволюции стилометрических методов, которая позволяет не только систематизировать существующие подходы, но и выявить инвариантную закономерность: повышение точности классификации сопровождается снижением интерпретируемости результатов. Эмпирической верификацией данной метамодели служит задача диахронического анализа идиостиля А. С. Суворина (1834–1912) — выдающегося русского журналиста и издателя, чьи публицистические письма 1904–1908 гг. образуют уникальный корпус для исследования влияния социально-политических событий на индивидуальный стиль автора.

Основной исследовательский вопрос формулируется следующим образом: фиксируются ли измеримые изменения в идиостиле автора с течением времени и какой методологический инструментарий обеспечивает наиболее надежную детекцию этих изменений? Ответ на этот вопрос предполагает не только решение прикладной задачи классификации, но и осмысление более широкой проблемы соотношения точности и интерпретируемости в компьютерной филологии.

ИСТОРИКО–ТЕОРЕТИЧЕСКИЕ ОСНОВАНИЯ СТИЛОМЕТРИИ

Истоки количественного изучения стиля восходят к середине XIX века. Как отмечает Г. Я. Мартыненко, стилометрия формировалась в широком междисциплинарном контексте наряду с антропометрией, биометрией, психометрикой и эконометрией¹. Основоположителем дисциплины принято считать немецкого филолога Вильгельма Диттенбергера (Wilhelm Dittenberger), опубликовавшего в 1881 г. статью «Sprachliche Kriterien für die Chronologie der Platonischen Dialoge»². Диттенбергер исследовал частотность употребления греческих частиц $\mu\eta\nu$ и ее производных в диалогах Платона, установив, что определенные комбинации частиц систематически отсутствуют в ранних диалогах и появляются только в поздних. Это позволило разделить корпус Платона на хронологические группы исключительно на основании лингвостатистических критериев.

Независимо от Диттенбергера шотландский филолог Льюис Кэмпбелл (Lewis Campbell) применял аналогичный подход к датировке диалогов «Софист» и «Политик» (1867), анализируя ритмические каденции и редкую лексику³. Термин «stylometry» был впервые введен в научный оборот польским ученым Винцентом Лютославским (Wincenty Lutoslawski) в 1897 г. в статье «Principes de stylométrie appliqués à la chronologie des œuvres de Platon»⁴. Лютославский также сформулировал «стилометрический закон», согласно которому стилистическая близость текстов свидетельствует об их хронологической близости.

Параллельно в англоязычной традиции развивались статистические методы анализа авторского стиля. Американский физик Томас Корвин Менденхолл (Thomas Corwin Mendenhall) в 1887 г. предложил метод «характеристических кривых композиции» (characteristic curves of composition), основанный на распределении длин слов⁵. Менденхолл рассматривал такие распределения как своеобразные «спектры», уникальные для каждого автора — метафора, которую впоследствии подхватит Н. А. Морозов.

СТИЛОМЕТРИЯ В РОССИИ

Н. А. Морозов и его последователи. В России пионером стилометрии стал Николай Александрович Морозов (1854–1946) — революционер-народник и разносторонний ученый. В 1915 г. он опубликовал работу «Лингвистические спектры: Средство для отличия плагиатов от истинных произведений

¹ Мартыненко Г. Я. Стилометрия: возникновение и становление в контексте междисциплинарного взаимодействия. Ч. 1. Первые шаги: XIX век // Структурная и прикладная лингвистика. Вып. 10. СПб. : Изд-во СПбГУ, 2014. С. 3–22.

² Dittenberger W. Sprachliche Kriterien für die Chronologie der platonischen Dialoge // Hermes. 1881. Bd. 16. S. 321–345.

³ Campbell L. The sophistes and politicus of plato. Oxford : Clarendon Press, 1867.

⁴ Lutoslawski W. Principes de stylométrie appliqués à la chronologie des œuvres de Platon // Revue des études grecques. 1897. Vol. 10. P. 61–81.

⁵ Mendenhall T. C. The characteristic curves of composition // Science. 1887. Vol. 9, №. 216. P. 237–249.

того или другого известного автора»⁶. Морозов первым в отечественной науке поставил задачу атрибуции текста на основе статистического метода. Вслед за Диттенбергером он полагал, что служебные части речи («распорядительные частицы») не зависят от тематики текста, а определяются исключительно индивидуальным складом речи автора.

Важнейшим вкладом Морозова стало введение понятия лингвистических спектров — графических представлений колебаний частотных характеристик в разных произведениях, которые можно рассматривать как прообразы современных статистических распределений. Морозов атрибутировал не только древние, но и современные тексты — произведения Гоголя, Пушкина, Толстого, Тургенева, существенно расширив область применения метода.

Эксперименты Морозова привлекли внимание выдающегося статистика А. А. Чупрова, который на текстовом материале (произведения Гете, гекзаметры Овидия и Вергилия) пересмотрел теорию устойчивости статистических рядов В. Лексиса, показав существование патологической устойчивости с поднормальной дисперсией⁷. Переписка Чупрова с академиком А. А. Марковым привела к эпохальному открытию — теории цепей Маркова, возникшей при исследовании чередования гласных и согласных в романе «Евгений Онегин»⁸. Таким образом, филологический материал послужил источником фундаментальных математических достижений.

Русская формальная школа и математическое стиховедение. Особую роль в развитии квантитативных подходов к анализу текста сыграла русская формальная школа «Общество изучения поэтического языка (ОПОЯЗ), сложившаяся в 1910–1920-х годах. Андрей Белый (Б. Н. Бугаев, 1880–1934) стал основоположником математического стиховедения, организовав масштабные статистические исследования русской поэзии⁹. Его книга «Символизм» (1910) получила высокую оценку А. А. Чупрова как пример «искусно выполненных статистических исследований» ритмики стиха¹⁰.

В. Б. Шкловский, Б. М. Эйхенбаум, Ю. Н. Тынянов, Р. О. Jakobson и другие представители формальной школы развивали идеи системного изучения литературы, динамики текста, ритма и синтаксиса с применением статистических методов. Тынянов выдвинул концепцию синхронических и диахронических литературно-художественных систем, ядра и периферии — идеи, созвучные теории статистической совокупности Чупрова¹¹.

⁶ Морозов Н. А. Лингвистические спектры: Средство для отличия плагиатов от истинных произведений того или другого известного автора // Известия Отд. рус. языка и словесности Имп. Академии наук. СПб., 1915. Т. XX. Кн. 4. С. 93–134.

⁷ Чупров А. А. Очерки по теории статистики. М., 1910.

⁸ Марков А. А. Об одном применении статистического метода // Известия Имп. Академии наук. СПб., 1913. Сер. 6. Т. 7. С. 239–242.

⁹ Белый А. Символизм. М.: Мусагет, 1910.

¹⁰ Чупров А. А. Указ. соч.

¹¹ Тынянов Ю. Н. Архаисты и новаторы. Л.: Прибой, 1929.

КОМПЬЮТЕРНАЯ ЭРА: ОТ МОСТЕЛЛЕРА ДО ТРАНСФОРМЕРОВ

Поворотным моментом в истории стилометрии стала работа американских статистиков Фредерика Мостеллера и Дэвида Уоллеса «Inference and Disputed Authorship: The Federalist» (1964), посвященная атрибуции спорных статей «Федералиста»¹². Применяв байесовский анализ частот служебных слов, они установили авторство Джеймса Мэдисона для всех двенадцати спорных эссе. Эта работа не только решила историческую загадку, но и продемонстрировала мощь компьютерных методов в гуманитарных исследованиях.

В 2002 г. Джон Берроуз (John Burrows) предложил метод Delta — меру стилистического расстояния, ставшую «золотым стандартом» в атрибуции авторства¹³. Формула Delta вычисляет среднее манхэттенское расстояние между z-нормализованными частотами наиболее употребимых слов: $\Delta(A, B) = \sum |z_i(A) - z_i(B)| / n$. Метод отличается простотой, воспроизводимостью и высокой интерпретируемостью.

Революционным событием стало появление архитектуры Word2Vec (Т. Миколов и др., 2013), позволившей перейти от подсчета дискретных единиц к анализу семантических связей в непрерывном векторном пространстве¹⁴. Н. Трипто и М. Али (2023) развили графовую модель Word2Vec для атрибуции авторства, где узлами графа являются слова, а ребрами — их семантическая близость¹⁵. Наконец, трансформерная архитектура BERT (Дж. Девлин и др., 2019)¹⁶ и ее русскоязычная адаптация RuBERT¹⁷ открыли возможности использования контекстуализированных представлений со 110 миллионами параметров.

МЕТАМОДЕЛЬ ЭВОЛЮЦИИ СТИЛОМЕТРИЧЕСКИХ МЕТОДОВ

Анализ истории стилометрии позволяет выделить три парадигмальных этапа, характеризующихся принципиально различными эпистемологическими установками и техническими решениями. Предлагаемая метамодель фиксирует инвариантную закономерность: каждый последующий этап обеспечивает прирост точности классификации при одновременном снижении интерпретируемости результатов (см. табл.).

¹² Mosteller F. Inference and Disputed Authorship : The Federalist. Readingio / F. Mosteller, D. L. Wallace. MA : Addison-Wesley, 1964.

¹³ 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. DOI: 10.1093/lilc/17.3.267.

¹⁴ Mikolov T. Efficient estimation of word representations in vector space / T. Mikolov, K. Chen, G. Corrado, J. Dean // Proceedings of ICLR. 2013. arXiv:1301.3781.

¹⁵ Tripto N. I. The Word2vec graph model for author attribution and genre detection in literary analysis / N. I. Tripto, M. E. Ali // arXiv preprint. 2023. arXiv:2310.16972.

¹⁶ Devlin J. BERT : Pre-training of deep bidirectional transformers for language understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // Proceedings of NAACL-HLT. 2019. P. 4171–4186.

¹⁷ Kuratov Y. Adaptation of deep bidirectional multilingual transformers for Russian language / Y. Kuratov, M. Arkhipov // Computational Linguistics and Intellectual Technologies: Dialogue 2019. Vol. 18. P. 333–340.

Метамодель эволюции стилометрических методов

Параметр	Парадигма 1: Частотная	Парадигма 2: Семантическая	Парадигма 3: Гибридная
Период	1880–1960-е	2010-е	с 2018
Философия	Стиль = частоты слов	Стиль = семант. связи	Стиль = смысл + форма
Метод (в данном иссл.)	Burrows' Delta	Word2Vec + Graph	BERT + LightGBM
Точность, %	65	72,3	87
Интерпретируемость	Высокая	Средняя	Низкая
Сложность модели	~10 признаков	~100 признаков	110M параметров

Первая парадигма (конец XIX — середина XX в.) — частотная стилометрия, основанная на «философии подсчета слов». Стиль концептуализируется как набор частот лексических единиц, прежде всего служебных слов. Инструментарий ограничен элементарной описательной статистикой. Результаты легко интерпретируемы: исследователь может указать конкретные слова, различающие авторов. Однако подход игнорирует порядок слов (парадигма «мешка слов») и семантические отношения.

Вторая парадигма (2010-е гг.) — семантическое моделирование — переносит акцент на понимание смысла. Стиль трактуется как конфигурация семантических связей в векторном пространстве. Применяются нейросетевые эмбединги (Word2Vec), теория графов, ансамблевое машинное обучение. Интерпретируемость снижается: вместо отдельных слов анализируются графовые метрики (центральность, плотность, веса ребер), смысл которых требует специальной экспертизы.

Третья парадигма (с 2018 г.) — гибридный интеллект — синтезирует семантические представления трансформерных моделей с классическими стилометрическими признаками. Философия данного этапа: стиль = смысл + форма. Используются предобученные модели со сотнями миллионов параметров (BERT) в сочетании с градиентным бустингом (LightGBM). Интерпретируемость минимальна: модель функционирует как черный ящик, хотя методы объяснимого ИИ (SHAP, LIME) частично компенсируют этот недостаток.

Выявленная закономерность инверсного соотношения точности и интерпретируемости не является специфичной для стилометрии — она отражает фундаментальное противоречие между предсказательной способностью модели и возможностью ее содержательного понимания, характерное для машинного обучения в целом. Однако именно стилометрия, с ее полуторавековой историей и богатым методологическим арсеналом, представляет собой идеальную лабораторию для изучения данной закономерности.

МАТЕРИАЛ И МЕТОДЫ ЭМПИРИЧЕСКОЙ ВЕРИФИКАЦИИ

Эмпирической базой верификации метамодели послужил корпус из 278 публицистических писем А. С. Суворина («Маленькие письма»), опубликованных в газете «Новое время» в 1904–1908 гг.¹⁸ Корпус объемом около 250 000 токенов распределен по годам следующим образом: 1904 г. — 55 документов, 1905 г. — 72, 1906 г. — 68, 1907 г. — 49, 1908 г. — 34 документа. Данный хронологический диапазон охватывает период Русско-японской войны и Первой русской революции, что создает предпосылки для исследования влияния социально-политических факторов на идиостиль автора.

Для каждой из трех парадигм метамодели реализована соответствующая вычислительная система. Система 1 (частотная парадигма) применяет метод Burrows' Delta с извлечением частот наиболее употребимых слов и иерархической кластеризацией. Система 2 (семантическая парадигма) использует модель Word2Vec с построением семантических графов по методологии Трипто и Али¹⁹, адаптированной для русского языка. Из графовой структуры извлекается 85–95 структурных признаков (центральности узлов, веса ребер, коэффициенты кластеризации), классификация осуществляется ансамблем VotingClassifier. Система 3 (гибридная парадигма) объединяет BERT-эмбединги (768 признаков, модель DeepPavlov/rubert-base-cased, стратегия Mean Pooling) с 200 отобранными стилометрическими признаками; классификатор — LightGBM²⁰.

Существенным методологическим решением стало упрощение 4-классовой задачи до бинарной: ранний период (1904–1905, 127 документов) против позднего периода (1906–1908, 151 документ). Данное решение обосновано как статистически (увеличение соотношения сигнал/шум), так и содержательно (события 1905 г. как исторический водораздел). Валидация проводилась методом 5-блочной стратифицированной кросс-валидации с ансамблированием (три запуска с различными random_state).

Эмпирические результаты подтверждают предсказания метамодели. Система 1 (Burrows' Delta) обеспечила базовую точность 65 % — существенно выше случайного уровня (50 %), что свидетельствует о наличии детектируемых стилистических различий между периодами. Система 2 (Word2Vec + Graph) достигла точности 72,3 % (прирост +7,3 п.п.), продемонстрировав преимущества семантического моделирования. Система 3 (BERT + LightGBM) показала пиковую точность 87 % (прирост +14,7 п.п. относительно Системы 2). Статистическая значимость различий между системами подтверждена тестом МакНемара: $\chi^2 = 23,4$, $p < 0,001$. Размер эффекта по шкале Коэна составил $d = 1,24$, что интерпретируется как большой эффект ($> 0,8$) и свидетельствует о практической значимости прироста точности. Доверительный интервал для Системы 3, полученный методом bootstrap (95 %), составляет [0,77; 0,81] при среднем $\mu = 0,81$.

¹⁸ Суворин А. С. Русско-японская война и русская революция. Маленькие письма (1904–1908). М.: Алгоритм, 2005.

¹⁹ Tripto N. I. Op. cit.

²⁰ Ke G. LightGBM: a highly efficient gradient boosting decision tree / G. Ke, Q. Meng, T. Finley et al // Advances in Neural Information Processing Systems 30. 2017. P. 3146–3154.

Анализ важности признаков в Системе 2 выявил ключевые графовые метрики: средний вес ребер (`avg_edge_weight`), стандартное отклонение весов (`std_edge_weight`), количество множественных узлов (`multi_nodes_count`) — маркеры публичного стиля; максимальный вес ребра (`max_edge_weight`), плотность графа (`density`) — маркеры личного стиля. В Системе 3 среди информативных признаков: `function_word_ratio`, `short_sentence_ratio`, символьные N-граммы (`char_2gram_to`, `char_3gram_ств`), морфологические признаки (`pos_ratio_adjs`, `pos_ratio_numr`).

Визуализация временной динамики стилометрических показателей методом PCA выявила кластерную структуру: тексты 1904–1905 гг. образуют компактный кластер, тексты 1906–1907 гг. демонстрируют более рассеянное распределение, тексты 1908 г. формируют отдельный кластер. В 1906 г. зафиксирован переломный момент: резкое снижение средней длины предложения и коэффициента лексического разнообразия (`Type Token Ratio`).

Полученные результаты эмпирически верифицируют предложенную метамодель эволюции стилометрических методов. Закономерность инверсного соотношения точности и интерпретируемости прослеживается на всех трех этапах: от полностью прозрачного метода Delta (каждый признак — конкретное слово) через частично интерпретируемые графовые метрики к «черному ящику» трансформерной архитектуры.

Данная закономерность вписывается в историческую логику развития стилометрии. Лингвистические спектры Морозова были наглядны и понятны современникам, однако академик Марков справедливо критиковал их статистическую ненадежность²¹. Метод Mosteller–Wallace (1964) повысил строгость за счет байесовского анализа, но потребовал специальной математической подготовки для интерпретации результатов. Современные нейросетевые модели достигают беспрецедентной точности, однако механизм их работы остается непрозрачным даже для специалистов.

Важным методологическим выводом является неэффективность прямого дообучения (`fine-tuning`) трансформерных моделей на малых корпусах. Эксперименты с дообучением BERT продемонстрировали классическую дилемму: при полной разморозке слоев — точность на валидации падает до 36 %, при усиленной регуляризации — недообучение (29 %). Оптимальной стратегией оказалось использование BERT в режиме экстрактора признаков с последующей классификацией градиентным бустингом. Этот результат согласуется с рекомендациями для работы с малыми данными в NLP.

Содержательная интерпретация позволяет связать стилистические изменения с социально-политическим контекстом эпохи. Упрощение синтаксиса (сокращение длины предложений), снижение лексического разнообразия после 1905 г. могут отражать адаптацию А. С. Суворина к изменившимся условиям общественной коммуникации в период политической турбулентности. Эта гипотеза требует дальнейшей верификации на сопоставимых корпусах других авторов.

²¹ Марков А. А. Об одном применении статистического метода.

ВЫВОДЫ

В настоящем исследовании предложена и эмпирически верифицирована метамодель эволюции стилометрических методов, фиксирующая инвариантную закономерность: повышение точности классификации сопровождается снижением интерпретируемости результатов. Метамодель охватывает три парадигмальных этапа развития дисциплины — от частотных методов XIX века до современных нейросетевых архитектур.

Основные результаты исследования сводятся к следующим положениям. Во-первых, подтверждена возможность фиксации измеримых диахронических изменений в идиостиле А. С. Суворина: точность бинарной классификации достигла 87 % при статистически значимом превышении над случайным уровнем ($p < 0,001$). Во-вторых, выявлена закономерность инверсного соотношения точности и интерпретируемости, характерная для всех трех парадигм стилометрии. В-третьих, установлена корреляция стилистических изменений с событиями 1905 г., что подтверждает чувствительность идиостиля к внешним социально-политическим факторам.

Теоретическая значимость исследования заключается в систематизации методов стилометрии в единой метамодели, позволяющей прогнозировать характеристики новых методов. Практическая значимость определяется разработкой методологии, применимой к задачам атрибуции авторства и диахронического анализа текстов. Программный код проекта доступен в открытом репозитории: https://github.com/ipk-2025/Suvorin_project-2025.

Перспективы дальнейших исследований связаны с расширением корпуса, применением методов объяснимого ИИ (SHAP, LIME) для повышения интерпретируемости гибридных моделей, а также сравнительным анализом стилистической эволюции разных авторов рассматриваемого периода.

ЛИТЕРАТУРА

1. *Белый А.* Символизм. М. : Мусагет, 1910.
2. *Марков А. А.* Об одном применении статистического метода // Известия Имп. Академии наук. СПб., 1913. Сер. 6. Т. 7. С. 239–242.
3. *Мартыненко Г. Я.* Стилометрия: возникновение и становление в контексте междисциплинарного взаимодействия. Ч. 1. Первые шаги: XIX век // Структурная и прикладная лингвистика. Вып. 10. СПб. : Изд-во СПбГУ, 2014. С. 3–22; Ч. 2. Первая половина XX в. // Структурная и прикладная лингвистика. Вып. 11. СПб. : Изд-во СПбГУ, 2015. С. 9–28.
4. *Морозов Н. А.* Лингвистические спектры: Средство для отличия плагиатов от истинных произведений того или другого известного автора // Известия Отд. рус. языка и словесности Имп. Академии наук. СПб., 1915. Т. XX. Кн. 4. С. 93–134.
5. *Суворин А. С.* Русско-японская война и русская революция. Маленькие письма (1904–1908). М. : Алгоритм, 2005.
6. *Тынянов Ю. Н.* Архаисты и новаторы. Л. : Прибой, 1929.

7. *Чупров А. А.* Очерки по теории статистики. М., 1910.
8. *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. DOI: 10.1093/lc/17.3.267.
9. *Campbell L.* The sophistes and politicus of plato. Oxford : Clarendon Press, 1867.
10. *Devlin J.* BERT: Pre-training of deep bidirectional transformers for language understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // *Proceedings of NAACL-HLT*. 2019. P. 4171–4186.
11. *Dittenberger W.* Sprachliche kriterien fur die chronologie der platonischen Ddaloge // *Hermes*. 1881. Bd. 16. S. 321–345.
12. *Grzybek P.* The emergence of stylometry // *Glottometrics*. 2014. Vol. 27. P. 60–83.
13. *Ke G.* LightGBM: a highly efficient gradient boosting decision tree / G. Ke, Q. Meng, T. Finley et al // *Advances in Neural Information Processing Systems* 30. 2017. P. 3146–3154.
14. *Kurатов Y.* Adaptation of deep bidirectional multilingual transformers for Russian language / Y. Kuratov, M. Arkhipov // *Computational Linguistics and Intellectual Technologies: Dialogue*. 2019. Vol. 18. P. 333–340.
15. *Lutoslawski W.* Principes de stylometrie appliques a la chronologie des oeuvres de Platon // *Revue des etudes grecques*. 1897. Vol. 10. P. 61–81.
16. *Mendenhall T. C.* The characteristic curves of composition // *Science*. 1887. Vol. 9, № 216. P. 237–249.
17. *Mikolov T.* Efficient estimation of word representations in vector space / T. Mikolov, K. Chen, G. Corrado, J. Dean // *Proceedings of ICLR*. 2013. arXiv:1301.3781.
18. *Mosteller F.* Inference and disputed authorship: The Federalist. Reading / F. Mosteller, D. L. Wallace. MA : Addison-Wesley, 1964.
19. *Stamatatos E.* A survey of modern authorship attribution methods // *Journal of the American Society for Information Science and Technology*. 2009. Vol. 60, № 3. P. 538–556.
20. *Tripto N. I.* The Word2vec graph model for author attribution and genre detection in literary analysis / N. I. Tripto, M. E. Ali // *arXiv preprint*. 2023. arXiv:2310.16972.

Поступила в редакцию 23 марта 2026 г.;
одобрена после рецензирования 29 марта 2026 г.;
принята к публикации 31 марта 2026 г.