

Научная статья

УДК 808.1

DOI: <https://doi.org/10.18721/JHSS.14108>



## О СМЕНЕ ПАРАДИГМЫ АВТОРСКОГО ИНВАРИАНТА

А.А. Рогов , Н.Д. Москин  , А.А. Лебедев 

Петрозаводский государственный университет,  
г. Петрозаводск, Российская Федерация

 [moskin@petrsu.ru](mailto:moskin@petrsu.ru)

**Аннотация.** Одной из актуальных задач в филологии является атрибуция текстов. Количественный показатель, с помощью которого можно различить произведения разных авторов, нужно назвать авторским инвариантом. В работе описывается ряд исследований (методика Г. Хетсо, методика оценки парной связи грамматических классов, методика «деревьев решений», Дельта-метод), результаты которых подтверждают, что изначальное определение авторского инварианта следует скорректировать. В частности, это касается временного промежутка, на котором параметр атрибуции должен сохранять «постоянное значение». Он не обязательно совпадает со всем периодом творчества писателя. Также из-за отсутствия универсального критерия, однозначно отличающего конкретного писателя от других, следует использовать совокупность характеристик авторских инвариантов на разных уровнях языка. Проведенный анализ показывает, что термин «авторский инвариант» следует разбить на две категории – «глобальный авторский инвариант» и «локальный авторский инвариант» – которые могут последовательно изучаться независимо друг от друга.

**Ключевые слова:** литературный текст, лингвостатистический параметр, авторский инвариант, классификация, интеллектуальный анализ данных.

**Для цитирования:** Рогов А.А., Москин Н.Д., Лебедев А.А. О смене парадигмы авторского инварианта // Terra Linguistica. 2023. Т. 14. № 1. С. 88–97. DOI: 10.18721/JHSS.14108



## ON THE PARADIGM SHIFT OF THE AUTHOR'S INVARIANT

A.A. Rogov , N.D. Moskin  , A.A. Lebedev 

Petrozavodsk State University,  
Petrozavodsk, Russian Federation

✉ [moskin@petsru.ru](mailto:moskin@petsru.ru)

**Abstract.** One of the urgent tasks in philology is the attribution of texts. The quantitative indicator by which one can distinguish between the works of different authors should be called the author's invariant. The paper describes a number of studies (the method of G. Kjetsaa, the method of evaluating the pair connection of grammatical classes, the method of "decision trees", the Delta method), the results of which confirm that the initial definition of the author's invariant should be corrected. In particular, this applies to the time interval during which the attribution parameter should keep "constant value". It does not necessarily coincide with the entire period of the writer's work. Also due to the lack of a universal criterion that uniquely distinguishes a particular writer from others, one should use a set of characteristics of author's invariants at different levels of the language. The performed analysis shows that the term "author's invariant" should be divided into two categories – "global author's invariant" and "local author's invariant" – which can be consistently studied independently of each other.

**Keywords:** literary text, linguostatistical parameter, author's invariant, classification, data mining.

**Citation:** A.A. Rogov, N.D. Moskin, A.A. Lebedev, On the paradigm shift of the author's invariant, *Terra Linguistica*, 14 (1) (2023) 88–97. DOI: 10.18721/JHSS.14108

### Введение

Вопрос об определении авторства текста – один из наиболее актуальных в филологии. Псевдонимные и анонимные произведения существовали еще в далеком прошлом; даже в эпоху античности появлялись и литературные подделки, и стилизации, и банальная путаница. Однако вопрос атрибуции подобных текстов не был устранен и после появления книгопечатания. В связи с этим важную роль могло бы сыграть формирование некоего метода, позволяющего при помощи современных информационных технологий решать данные вопросы атрибуции с высокой степенью достоверности.

Но хотя сегодня существует большое число различных способов, позволяющих определить авторство произведений, универсального метода, который был бы признан всеми исследователями, не выявлено. Очевидно, что основой изучения авторского стиля становится анализ языка конкретных литературных произведений определенного автора в контексте общезыковых закономерностей, присущих тому или иному этапу развития национального языка и конкретному литературному стилю. Однако закономерной является постановка вопроса о существовании подобного метода как такового – можно ли пронаблюдать некоторые лингвистические универсалии, которые прослеживаются на протяжении всего творчества автора и могут быть эффективным инструментом для определения индивидуально-авторского стиля?

В отечественной лингвистике подобного рода исследования, проводимые на стыке филологии и математики, имеют давнюю историю. Одним из первых авторов, которые в практике анализа русскоязычных текстов стали опираться на такие универсальные компоненты, был Н.А. Морозов [1]. Еще в начале XX века он в своей работе «Лингвистические спектры, как средство для отличения плагиатов от истинных произведений того или другого известного автора и для определения их эпохи» ввел понятие «лингвистический спектр» (график частоты встречаемости определенных слов или грамматических категорий в творчестве определенного автора), и на базе текстов А.С. Пушкина, Н.В. Гоголя, А.Н. Толстого, И.С. Тургенева, Н.М. Карамзина, М.Н. Загоскина



определил значимость служебных частиц в творчестве данных авторов. Н.А. Морозов считал, что такое исследование творчества позволяет «узнавать авторов», метафорически представляя свою идею следующим образом: «подобно тому, как каждый автор, всегда оставаясь человеком, имеет свою индивидуальную физиономию, так и его язык, все время оставаясь русским, или английским, или французским, обнаруживает свои особенные черты <...>» [1].

Однако более востребованным в работах на стыке математики и лингвистики стало понятие «авторский инвариант», которое было заявлено в работе В.П. Фоменко и Т.Г. Фоменко «Авторский инвариант русских литературных текстов» [2]. Исследователи под авторским инвариантом понимают числовую характеристику, соответствующую ряду требований:

- 1) это должен быть «бессознательный параметр», который укоренен столь глубоко, что автор о нем не задумывается, а если и пытается держать его под контролем, то не в состоянии проделывать это в течение долгого времени (вследствие чего тот возвращается в исходное состояние);
- 2) для текстов конкретного автора такой параметр должен быть постоянным (слабо отклоняться от среднего значения) на всех его произведениях;
- 3) с помощью данного параметра должны различаться разные группы авторов (т. е. он должен позволить различить писателей).

Только комбинация из трех вышеуказанных условий позволяет говорить о формировании авторского инварианта. Среди множества параметров в качестве авторского инварианта исследователями была выбрана частота употребления служебных слов, которая оказывается стабильным показателем на протяжении всего творчества писателя. Было отмечено, что именно частотность употребления предлогов, союзов и частиц соответствует всем вышеозначенным критериям (писатель, формирующий текст, специально не задумывается над служебными частями речи; параметр прослеживается во всех произведениях автора; параметр позволяет разграничивать авторов). На основании данного критерия авторами было поставлено под сомнение авторство «Тихого Дона» М.А. Шолохова.

Следует особо отметить, что вопрос об авторстве «Тихого Дона» является одним из наиболее сложных в аспекте изучаемой проблемы. Эту тему затрагивали и специалисты-литературоведы, и языковеды, в том числе и зарубежные, причем, как кажется, точка в этой дискуссии не поставлена до сих пор: существуют исследования, как указывающие на стилистическое единство всех частей «Тихого Дона», так и на значительные отличия (опираясь на язык романа, его хронологию, историческое содержание и т.п.). Не ставя перед собой задачу дать полноценный обзор данной конкретной проблемы, сошлемся на выводы, которые сделали В.П. Фоменко и Т.Г. Фоменко. Исследователи отмечают, что «количество служебных слов в его произведениях оказалось настолько неодинаковым, что появляется необходимость представить Шолохова в виде двух авторов, которых мы условно назвали: Шолохов I и подозреваемый Шолохов II» [2]. В соответствии с полученными данными, авторы статьи делают вывод о том, что «Статистические результаты, полученные в результате анализа авторского инварианта, подтверждают гипотезу, что части 1, 2, 3, 4, 5 и в значительной мере часть 6 романа «Тихий Дон» написаны не М.А. Шолоховым» [2].

Схожая методика была апробирована Г. Хетсо в работе «Принадлежность Достоевскому: к вопросу об атрибуции Ф.М. Достоевскому анонимных статей в журналах «Время» и «Эпоха» [3], ставшей одним из наиболее известных в аспекте атрибуции псевдонимных и анонимных статей, которые были опубликованы в журналах «Время» и «Эпоха», с опорой на математические методы. Исследование включало в себя 15 лингвостатистических параметров, 10 грамматических параметров, а также дополнительные критерии (например, лексический спектр текста на уровне словаря и на уровне текста, а также индекс разнообразия лексики). Подробное критическое исследование методики Г. Хетсо представлено во второй главе монографии «Проблема атрибуции в журналах «Время», «Эпоха» и еженедельнике «Гражданин» [4].



В целом, идея авторского инварианта ставилась под сомнение более поздними исследователями. Из числа наиболее значимых работ отметим статью Т.В. Батуры «Формальные методы определения авторства текстов», где указывается, в частности, что значимым ограничением данного метода выступает «очень низкая разделительная способность оценки в случае большого числа авторов (потенциально метод может разделять лишь 10 авторских стилей)» [5]. Критические замечания, связанные с категорией инварианта, отмечаются и в статье Е.Б. Трофимовой «Инвариант: реальность или фикция» [6] (неясно, насколько обязательно присутствие инварианта во всей совокупности его вариантов; отсутствует строго формальная процедуры идентификации отношений такого рода), а также в работе А.А. Боронина «Инвариант в текстовой проекции (интерпретация персонажных субтекстов в условиях эксперимента)», где отмечается, что «инвариант как лингвистическая абстракция условен и не претендует на абсолютную точность» [7]. Особо этот вопрос рассматривается не только в аспекте анализа уже имеющихся произведений, но и с точки зрения генерации и моделирования новых текстов. В частности, в статье А.В. Мордвинова «Системный подход в моделировании текста» [8] отмечается, что при воплощении текста в конкретной языковой форме «происходит потеря информации, которая выражается, в том числе, и в потере части авторских инвариантов», и, как следствие, происходит частичное искажение интегрального понятия «авторский инвариант».

В соответствии с проведенными исследованиями, мы можем предположить, что само понятие авторского инварианта как некоей постоянной величины, сохраняющейся во всех произведениях автора, может быть подвергнуто сомнению. Стиль автора может меняться со временем; более того, даже в пределах одного временного промежутка на отдельные аспекты построения структуры текста могут быть наложены определенные ограничения. Как отмечает О.В. Карелова, «своеобразие индивидуального слога обусловлено как индивидуальными особенностями автора, так и историческими событиями, происходящими в тот или иной период времени и влияющими на творчество писателя» [9]. К таким глобальным историческим событиям, к примеру, может быть отнесена цензура, характерная для определенных временных промежутков (в частности, И. С. Чирскова указывает на то, что «в борьбе с цензурой русская литература оттачивала язык общения с читателем, научившись выражаться емко, иносказательно и многозначительно» [10], что не могло не отразиться на отдельных особенностях авторского идиостиля, реализуемого в тексте произведения). Таким образом, говорить об авторском инварианте в аспекте индивидуально-авторского стиля нужно с большой аккуратностью, учитывая стилистические особенности подвергаемых анализу произведений.

### Методика исследования

Однако проводимые в последние десятилетия исследования указывают на отсутствие подобных инвариантов. Используемые характеристики авторских инвариантов противоречат второму или третьему пункту основных свойств, заявленных в изначальном определении. Рассмотрим несколько примеров:

1) Проанализируем в качестве возможного авторского инварианта **частоту употребления служебных слов**. Заметим, что описание проведенных экспериментов в [2] не позволяет сделать вывод о том, что найденный параметр «количество служебных слов» действительно является авторским инвариантом. Как отмечают сами авторы, он ведет себя по-разному на больших и малых текстах. Для его стабилизации требуется анализ фрагментов объемом 16000 слов. При этом, как указано в [2], «близкие значения инварианта отнюдь не означают, что исследуемые произведения написаны одним автором. Как мы отмечали, встречаются разные писатели с близкими значениями инварианта. Например, Леонов и Фадеев, у которых эти числа равны соответственно 23,08 и 23,40». Это утверждение явно противоречит третьему пункту основных свойств авторского инварианта.



2) **Методика Г. Хетсо**, заявленная в работе «Принадлежность Достоевскому: к вопросу об атрибуции Ф.М. Достоевскому анонимных статей в журналах «Время» и «Эпоха» [3]. Проведенные исследования, описанные в монографии [4] показывают, что ни один из 15 признаков и методик их оценки нельзя использовать как авторский инвариант самостоятельно или в комбинации с другими. Главное отличие описанного эксперимента от оригинального исследования, проведенного Г. Хетсо, заключается в том, что была исследована вторая группа текстов, отсутствующая в исследованиях Г. Хетсо. Это было связано с выявлением наличия ошибки второго рода: необходимо было сравнить тексты Достоевского со статьями, автором которых точно не является писатель. Поэтому в исследование вошли тексты М.М. Достоевского, Н.Н. Страхова и А.А. Григорьева, где редакторское вмешательство Ф.М. Достоевского исключено. По всем параметрам тексты Григорьева, Страхова и М.М. Достоевского попали в группу статей, автором которых является Ф.М. Достоевский. Наиболее эффективными оказались параметры, связанные с лексикой, то есть признаки № 15 – индекс разнообразия лексики (точность составила всего 61%, не отвергнутыми оказались 7 текстов из 18 проанализированных) и № 14 – лексический спектр текста на уровне текста (точность 56%). У остальных признаков точность оказалась еще ниже. Такие результаты явно противоречат третьему пункту основных свойств авторского инварианта.

3) **Методика оценки парной связи грамматических классов**. Она достаточно подробно описана в монографии [11]. Методика основана на гипотезе, в соответствии с которой «...стиль автора проявляется в «пристрастии» к определенным грамматическим связям, частота появления которых в тексте высока. Остальные связи соответствует слабым, несущественным статистическим связям на уровне грамматических классов; частоты этих связей малы ..., а их появление в сильной мере случайно ...». В указанной монографии описан эксперимент успешного анализа с помощью данной методики разных списков «Повести временных лет». В монографии [4] описан эксперимент о возможности использования данной методики для разделения произведений следующих писателей: Ф.М. Достоевского, М.М. Достоевского, В.П. Мещерского, К.П. Победоносцева, Я.П. Полонского, А.А. Григорьева, Н.Н. Страхова, П.К. Щербальского. Оказалось, что даже полный перебор настроечных параметров методики не позволяет последовательно разделить тексты этих авторов. Вновь наблюдается явное противоречие третьему пункту основных свойств авторского инварианта. Заметим, что проведенные эксперименты (после нахождения соответствующих настроечных параметров) позволяли корректно разграничить произведения двух и трех авторов.

4) **Методика «деревья решений»**. Данная методика и ее обобщение под названием «Случайный лес» успешно используется в машинном обучении. В частности, в статье [12] деревья решений анализируются как один из подходов к автоматической классификации текстов с опорой на понятие авторского стиля и с высокой точностью различения таких индивидуальных стилей. Эксперименты, описанные в монографии [4], позволяют сделать следующее заключение: точность данной методики достигала 95%. Точность отделения Ф.М. Достоевского от других авторов оказывалась ниже, чем при попытке разделить двух конкретных авторов. Для решения задачи по различению произведений двух разных авторов использовались разные деревья, что вновь противоречит третьему пункту основных свойств авторского инварианта.

5) **Дельта-метод**, основывающийся на расстоянии Дельта-Бёрроуза. Метод был предложен в 2001 году австралийским филологом Джоном Бёрроузом, и используется во многих исследованиях, большая часть которых посвящена установлению авторства различных произведений.

Метод работает следующим образом: для каждого текста строятся лексические спектры самых частотных слов, а затем рассматривается расстояние между этими спектрами. Эмпирическая закономерность заключается в том, что расстояние для текстов, написанных одним автором, оказывается меньше, чем для любой пары текстов, написанных разными людьми. Для сравнения расстояний между текстами и разбиения их на кластеры используются различные вариации



иерархического кластерного анализа, чаще всего – метод Варда. Визуализации результатов кластеризации представляется в виде дендрограммы.

Исследование, опубликованное в работе [13], показывает, что расстояние между произведениями многих авторов (например, между М. Булгаковым и Н. Островским, между Л. Леоновым и А. Фадеевым и т.д.) значительно меньше, чем между произведениями М. Шолохова «Тихий Дон» и «Поднятая целина». Из этого можно сделать вывод: авторский инвариант, определяемый с помощью данного метода, явно противоречит второму пункту основных свойств (его устойчивость на протяжении жизни автора может быть поставлена под сомнение).

В работе [14] описано применение различных модификаций данного метода, разного набора частотного словаря и преобразования исходных данных для определения авторства романов «Двенадцать стульев» и «Золотого тельца». Несмотря на в целом интересные полученные результаты в данной работе, можно заметить неустойчивость данной методики в области определения авторского стиля. Аналогичные результаты получены в работе [14] при сравнении стилей публицистических статей Ф.М. Достоевского, М.М. Достоевского, В.П. Мещерского, К.П. Победоносцева, Я.П. Полонского, А.А. Григорьева, Н.Н. Страхова, П.К. Щербальского. Из работы видно, что атрибутировать тексты В.П. Мещерского и К.П. Победоносцева от других авторов удалось успешно – их кластер собрался в результате разных экспериментов. Для А.А. Григорьева и М.М. Достоевского тексты группировались в кластер за исключением нескольких текстов – выбросов. Произведения Ф.М. Достоевского и П.К. Щербальского разбивались на несколько однородных относительно расстояния дельта-Берроуза кластеров.

Опять замечаем явное противоречие третьему пункту основных свойств авторского инварианта.

б) **Нейросетевые методы.** В последние десятилетия при анализе литературных текстов активно используют методы машинного обучения. Особое внимание в этом аспекте уделяется нейронным сетям, которые показали свою эффективность в машинном переводе, в решении задачи построения ответа на задаваемые вопросы, а также в рассмотрении некоторых других лингвистических проблем. Как следствие, возникает идея попробовать отыскать с их помощью авторский инвариант. В работе [4] описывается сравнение следующих методов для атрибуции текстов: дерево решений, рекуррентные сети без модификаций и с LSTM ячейками и трансформер. Лучшие результаты показал трансформер. Заметим, что в нейросетевых методах, как правило, отсутствует визуализация обоснования принятия решения. Это существенно затрудняет их использование в научных целях. Однако, проведенные эксперименты показали, что они тоже не могут достигнуть 100% точности на тестовых данных. Значит, тот набор признаков, который используют данные методы, не удовлетворяет третьему пункту основных свойств авторского инварианта.

Данный перечень может быть продолжен с опорой на другие признаки и методики.

Заметим, что существует целый ряд публикаций, где успешно решаются отдельные задачи по кластеризации текстов, например, [15, 16, 17]. Однако примененные там признаки и методы их оценки нельзя рассматривать как авторский инвариант, поскольку в данных работах не проводится исследование на наличие ошибок второго рода, устойчивость признаков за пределами рассматриваемых текстов и т. д. Следует заметить, что применение ансамблевых методов при атрибуции текстов заранее предполагает неиспользование авторского инварианта в описанном ранее смысле.

### Заключение

Проведенный анализ показывает, что термин «авторский инвариант» следует разбить на две категории – «глобальный авторский инвариант» и «локальный авторский инвариант» – которые могут последовательно изучаться независимо друг от друга, а выбор между ними осуществляется с учетом тех конкретных задач, которые ставит перед собой исследователь.



Под глобальным авторским инвариантом понимается количественный показатель (группа количественных показателей), который удовлетворяет первому и второму свойству исходного определения. К его недостаткам следует отнести возможную существенную ошибку второго рода (принять гипотезу о принадлежности текста автору, если она не верна). Однако, за счет выбора критического параметра, его можно настроить на минимальную ошибку первого рода (отвергнуть верную гипотезу). Поэтому его надо использовать для опровержения гипотезы об авторстве, а не для ее принятия. Для нахождения глобального авторского инварианта нужно четко определить условия проведения эксперимента. Прежде всего, должна быть возможность его повторения, ясно определены признаки, размеры фрагментов, на которых они рассчитываются, круг текстов и т. д. Заметим, что размеры текстов в этом случае играют существенную роль, что было отмечено целым рядом исследователей (после проведения многочисленных экспериментов возникла своеобразная поговорка – «если что-то идет не так – смотри размер текста»). К глобальному авторскому инварианту можно отнести количественные показатели, рассмотренные в работах Н.А. Морозова, В.П. Фоменко, Т.Г. Фоменко, Г. Хетсо и пр.

Под локальным авторским инвариантом следует понимать количественный показатель, с помощью которого можно различить произведения разных авторов, который удовлетворяет следующим свойствам:

1) первое свойство, связанное со слабой контролируемостью такого анализируемого параметра автором, остается неизменным. В свою очередь, два других свойства авторского инварианта следует изменить:

2) искомый параметр, входящий в авторский инвариант, должен сохранять «постоянное значение» (слабо колебаться) для стилистически однородных произведений данного автора *на протяжении определенного периода времени*. Данный временной промежуток не обязательно совпадает со всем периодом творчества писателя.

3) в решении задачи по разграничению разных авторов мы должны использовать различные характеристики авторских инвариантов.

Такое изменение связано с отсутствием универсального критерия, однозначно отличающего одного конкретного писателя от бесчисленного множества других авторов, пишущих на том же языке. Однако при попарном или групповом сравнении грамматических и лексических характеристик, присущих текстам определенных писателей, мы можем выявить существенные различия на разных уровнях языка, которые могут сыграть важную роль в решении вопросов, связанных с определением авторства текста. К локальному авторскому инварианту можно отнести количественные показатели, рассмотренные в большинстве работ, связанных с анализом текстов, например, в работах А.А. Рогова, М.А. Марусенко, А.В. Седова и пр.

В чем состоит преимущество разделения авторского инварианта на глобальный и локальный? Такой подход позволяет подбирать инструменты (как математические, так и собственно филологические) сообразно имеющейся проблеме, минимизируя вероятность появления ошибок и формирования ложных выводов. К примеру, если мы ставим перед собой задачу анализа специфики построения текстов определенного автора в отдельно взятом стихотворном сборнике, произведения для которого были написаны в пределах небольшого временного диапазона, нет никакого смысла обращаться к глобальному авторскому инварианту – достаточно его локальной версии, позволяющей противопоставить анализируемые тексты сходным (и по стилю, и по времени написания) текстам других авторов. В то же время, если филолог задается глобальным, комплексным, многоаспектным вопросом выделения неких индивидуально-авторских особенностей, проявляющихся на протяжении абсолютно всего творчества автора и выделяющих его на фоне не только отдельной эпохи, но и всего литературного наследия, включающего в себя многообразие текстов, написанных на определенном языке – то в этом случае должна быть задействована категория «глобальный авторский инвариант». Разумеется, делать это следует с учетом ранее



упомянутых оговорок и требований, предъявляемых к подобного рода исследованиям, поскольку без них достоверность выполненной работы может быть поставлена под сомнение.

При этом анализ любого из типов авторских инвариантов (как глобального, так и локального) возможен только с учетом комбинированных комплексных подходов, предусматривающих проведение атрибуции по нескольким параметрам сразу. Только многоаспектный разносторонний взгляд на проблему, комбинирующий в себе лингвостатистику, современные математические варианты обработки информации и традиционные методы филологии, позволит добиться необходимых результатов в контексте анализа индивидуально-авторского стиля и решения вопросов, связанных с атрибуцией текстов.

Материалы статьи были представлены на IV Международной конференции по инженерной и прикладной лингвистике «Пиотровские Чтения – 2022», посвященной 100-летию со дня рождения профессора Р.Г. Пиотровского в РГПУ им. А.И. Герцена 22 ноября 2022 г.

### СПИСОК ИСТОЧНИКОВ

1. **Морозов Н.А.** Лингвистические спектры: средство для отличия плагиатов от истинных произведений того или иного известного автора. Стилеметрический этюд // Известия отделения русского языка и словесности Императорской Академии наук. 1915. Т. 20, № 4. С. 93–134.
2. **Фоменко В.П., Фоменко Т.Г.** Авторский инвариант русских литературных текстов. Предисловие А.Т. Фоменко // Фоменко А.Т. Новая хронология Греции: Античность в средневековье. М.: Изд-во МГУ, 1996. Т. 2. С. 768–820. URL: [https://chronologia.org/seven2\\_2/add3.html](https://chronologia.org/seven2_2/add3.html) (дата обращения: 25.12.2022)
3. **Kjetsaa G.** Attributed to Dostoevsky: the problem of attributing to Dostoevsky anonymous articles in Time and Epoch. Oslo: Solum Forlag A. S., 1986. 82 p.
4. **Рогов А.А., Абрамов Р.В., Бучнева Д.Д., Захарова О.В., Кулаков К.А., Лебедев А.А., Москин А.А., Отливанчик А.В., Савинов Е.Д., Сидоров Ю.В.** Проблема атрибуции в журналах «Время», «Эпоха» и еженедельнике «Гражданин». Петрозаводск: Издательство «Острова», 2021. 400 с.
5. **Батура Т.В.** Формальные методы определения авторства текстов // Вестник НГУ. Серия: Информационные технологии. 2012. Т. 10, № 4. С. 81–94.
6. **Трофимова Е.Б.** Инвариант: реальность или фикция? // Языковое бытие человека и этноса. 2004. № 7. С. 167–171.
7. **Боронин А.А.** Инвариант в текстовой проекции (интерпретация персонажных субтекстов в условиях эксперимента) // Вопросы психолингвистики. 2011. № 14. С. 28–35.
8. **Мордвинов А.В.** Системный подход в моделировании текста // Вестник Нижегородского университета им. Н.И. Лобачевского. 2010. № 2-1. С. 185–190.
9. **Карелова О.В.** К вопросу изучения индивидуального стиля автора // Известия РГПУ им. А.И. Герцена. 2006. Т. 3, № 20. С. 24–29.
10. **Чирскова И.М.** Цензура как историко-культурный феномен в России XIX века // Вестник РГГУ. Серия: История. Филология. Культурология. Востоковедение. 2008. №10. С. 115–125.
11. **Милов Л.В., Бородкин Л.И., Иванова Т.В., Неберекутина Е.В., Полянская И.В., Романкова Н.В., Саркисова Г.И.** От Нестора до Фонвизина: Новые методы определения авторства. М.: Прогресс, 1994. 445 с.
12. **Шевелев О.Г., Петраков А.В.** Классификация текстов с помощью деревьев решений и нейронных сетей прямого распространения // Вестник Томского государственного университета. 2006. № 290. С. 300–307.
13. **Великанова Н.П., Орехов Б.В.** Цифровая текстология: атрибуция текста на примере романа М.А. Шолохова «Тихий Дон» // Мир Шолохова. 2019. Т. 1, № 11. С. 70–82.
14. **Масаева О.С.** Метод Дельта Бёрроуза // Научно-исследовательская работа обучающихся и молодых учёных: материалы 74-й Всероссийской (с международным участием) научной конференции обучающихся и молодых учёных. Петрозаводск, 2022. С. 262–266.
15. **Марусенко М.А.** Атрибуция анонимных и псевдонимных текстов как типичная задача распознавания образов // Историография и источниковедение отечественной истории. Санкт-Петербург, 2003. Вып. 3. С. 116–135.





16. **Седов А.В., Рогов А.А.** Анализ неоднородностей в тексте на основе последовательностей частей речи // *Материалы VI Международной научно-практической конференции «Информационная среда вуза XXI века»*. 2012. С. 135–139.

17. **Щеголева Л.В., Лебедев А.А., Москин Н.Д.** Методы анализа данных в задаче разграничения фольклорных и авторских текстов // *Вопросы языкознания*. 2020. № 2. С. 61–74. DOI: 10.31857/S0373658X0008823-4

## REFERENCES

[1] **N.A. Morozov**, Lingvisticheskiye spektry: sredstvo dlya otlicheniya plagiatov ot istinnykh proizvedeniy togo ili inogo izvestnogo avtora. Stilemetricheskii etyud [Linguistic spectra: a tool for distinguishing plagiarisms from the true works of one or another famous author. Stylometric study], *Izvestiya otdeleniya russkogo yazyka i slovestnosti Imperatorskoy Akademii nauk* [Proceedings of the Department of the Russian Language and Literature of the Imperial Academy of Sciences]. 20 (4) (1995) 93–134.

[2] **V.P. Fomenko, T.G. Fomenko**, Avtorskiy invariant russkikh literaturnykh tekстов. Predisloviye A.T. Fomenko [Author's invariant of Russian literary texts. Foreword by A.T. Fomenko], *Fomenko A.T. Novaya khronologiya Gretsii: Antichnost v srednevekovye* [New chronology of Greece: Antiquity in the Middle Ages]. M.: Izd-vo MGU [Publishing House of Moscow State University], 2 (1996) 768–820. URL: [https://chronologia.org/seven2\\_2/add3.html](https://chronologia.org/seven2_2/add3.html) (data obrashcheniya: 25.12.2022)

[3] **G. Kjetsaa**, Attributed to Dostoevsky: the problem of attributing to Dostoevsky anonymous articles in Time and Epoch. Oslo: Solum Forlag A. S., 1986. 82 p.

[4] **A.A. Rogov, R.V. Abramov, D.D. Buchneva, O.V. Zakharova, K.A. Kulakov, A.A. Lebedev, A.A. Moskin, A.V. Otlivanchik, Ye.D. Savinov, Yu.V. Sidorov**, Problema atributsii v zhurnalakh «Vremya», «Epokha» i yezhenedelnike «Grazhdanin» [The problem of attribution in the magazines “Time”, “Epoch” and the weekly “Citizen”]. Petrozavodsk: Izdatelstvo «Ostrova» [Islands Publishing House], 2021. 400 p.

[5] **T.V. Batura**, Formalnyye metody opredeleniya avtorstva tekстов [Formal methods for determining the authorship of texts], *Vestnik NGU. Seriya: Informatsionnyye tekhnologii* [Vestnik NSU. Series: Information Technologies]. 10 (4) (2012) 81–94.

[6] **Ye.B. Trofimova**, Invariant: realnost ili fiktsiya? [Invariant: reality or fiction?], *Yazykovoye bytiye cheloveka i etnosa* [The linguistic existence of a person and an ethnic group]. 7 (2004) 167–171.

[7] **A.A. Boronin**, Invariant v tekstovoy proyektzii (interpretatsiya personazhnykh subtekstov v usloviyakh eksperimenta) [Invariant in text projection (interpretation of character subtexts in experimental conditions)], *Voprosy psikholingvistiki* [Questions of psycholinguistics]. 14 (2011) 28–35.

[8] **A.V. Mordvinov**, Sistemnyy podkhod v modelirovanii teksta [System approach in text modeling], *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo* [Vestnik of Lobachevsky University of Nizhni Novgorod]. 2-1 (2010) 185–190.

[9] **O.V. Karelova**, K voprosu izucheniya individualnogo stilya avtora [On the issue of studying the author's individual style], *Izvestiya RGPU im. A.I. Gertsena* [Izvestia: Herzen University journal of Humanities & Sciences]. 3 (20) (2006) 24–29.

[10] **I.M. Chirskova**, Tsenzura kak istoriko-kulturnyy fenomen v Rossii XIX veka [Censorship as a historical and cultural phenomenon in Russia in the 19th century], *Vestnik RGGU. Seriya: Istoriya. Filologiya. Kulturologiya. Vostokovedeniye* [RGGU Bulletin. Series: History. Philology. Cultural studies. Oriental studies]. 10 (2008) 115–125.

[11] **L.V. Milov, L.I. Borodkin, T.V. Ivanova, Ye.V. Neberekutina, I.V. Polyanskaya, N.V. Romankova, G.I. Sarkisova**, Ot Nestora do Fonvizina: Novyye metody opredeleniya avtorstva [From Nestor to Fonvizin: New methods authorship definitions]. M.: Progress, 1994. 445 p.

[12] **O.G. Shevelev, A.V. Petrakov**, Klassifikatsiya tekстов s pomoshchyu derevyev resheniy i neyronnykh setey pryamogo rasprostraneniya [Text classification with decision trees and feed-forward neural networks], *Vestnik Tomskogo gosudarstvennogo universiteta* [Tomsk State University Journal]. 290 (2006) 300–307.

[13] **N.P. Velikanova, B.V. Orekhov**, Tsifrovaya tekstologiya: atributsiya teksta na primere romana M.A. Sholokhova «Tikhyy Don» [Digital textology: text attribution on the example of the novel by M. A. Sholokhov “Quiet Flows the Don”], *Mir Sholokhova* [World of Sholokhov]. 1 (11) (2019) 70–82.



[14] **O.S. Masayeva**, Metod Delta Berrouza [The Burrow's Delta method], Nauchno-issledovatel'skaya rabota obuchayushchikhsya i molodykh uchenykh: materialy 74-y Vserossiyskoy (s mezhdunarodnym uchastiyem) nauchnoy konferentsii obuchayushchikhsya i molodykh uchenykh [Research work of students and young scientists: materials of the 74<sup>th</sup> All-Russian (with international participation) scientific conference of students and young scientists]. Petrozavodsk, 2022. Pp. 262–266.

[15] **M.A. Marusenko**, Atributsiya anonimnykh i psevdonimnykh tekstov kak tipichnaya zadacha raspoznavaniya obrazov [Attribution of anonymous and pseudonymous texts as a typical task of pattern recognition], Istoriografiya i istochnikovedeniye otechestvennoy istorii [Historiography and source study of national history]. St. Petersburg, 3 (2003) 116–135.

[16] **A.V. Sedov, A.A. Rogov**, Analiz neodnorodnostey v tekste na osnove posledovatelnostey chastey rechi [Analysis of heterogeneities in the text based on sequences of parts of speech], Proceedings of the VI International Scientific and Practical Conference “Information Environment of the XXI Century University”. 2012. Pp. 135–139.

[17] **L.V. Shchegoleva, A.A. Lebedev, N.D. Moskin**, Metody analiza dannykh v zadache razgraniicheniya folklornykh i avtorskikh tekstov [Methods of data analysis in the task of distinguishing folklore and author's texts], Voprosy yazykoznaniiya [Questions of Linguistics]. 2 (2020) 61–74. DOI: 10.31857/S0373658X0008823-4

#### **СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS**

**Рогов Александр Александрович**

**Alexander A. Rogov**

E-mail: rogov@petsru.ru

ORCID: <https://orcid.org/0000-0002-8815-7920>

**Москин Николай Дмитриевич**

**Nikolai D. Moskin**

E-mail: moskin@petsru.ru

ORCID: <https://orcid.org/0000-0001-5556-5349>

**Лебедев Александр Александрович**

**Alexander A. Lebedev**

E-mail: perevodchik88@yandex.ru

ORCID: <https://orcid.org/0000-0001-9939-9389>

*Поступила: 30.12.2022; Одобрена: 29.01.2023; Принята: 17.03.2023.*

*Submitted: 30.12.2022; Approved: 29.01.2023; Accepted: 17.03.2023.*