

АЛЕКСАНДР АЛЕКСАНДРОВИЧ ЛЕБЕДЕВ

кандидат филологических наук, старший преподаватель кафедры русского языка Института филологии Петрозаводский государственный университет (Петрозаводск, Российская Федерация)

ORCID 0000-0001-9939-9389; perevodchik88@yandex.ru

## АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ЧАСТЕЙ РЕЧИ И КАТЕГОРИЯ ИДИОСТИЛЯ

**Аннотация.** Рассматривается лингвистический подход к анализу идиостиля автора, предусматривающий учет морфологических особенностей текстов и последовательностей частей речи, представленных в таких текстах. Описывается цель рассмотрения последовательностей частей речи в контексте определения индивидуально-авторского стиля. Ставятся и конкретизируются проблемы, связанные с анализом последовательностей частей речи: множественность лингвистических концепций, вариативность глубины анализа, специфика выбора текстов и методов для их анализа. Описывается информационная система «Статистические методы анализа литературного текста» (СМАЛТ), которая может быть использована в том числе и для морфологического анализа публицистических текстов, обосновывается использование математического метода построения деревьев решений, перечисляются результаты, которые были достигнуты в ходе подобного анализа. Приводятся примеры сочетаний частей речи, отличающих тексты разных авторов. Делается вывод о влиянии морфологической структуры текста на его восприятие читателем, подчеркивается важность анализа последовательностей частей речи в контексте определения идиостиля автора.

**Ключевые слова:** идиостиль, части речи, последовательности частей речи, морфология, дерево решений, СМАЛТ

**Благодарности.** Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-012-90026 Достоевский («Проблема атрибуции анонимных и псевдонимных статей в журналах “Время”, “Эпоха” и еженедельнике “Гражданин”»).

**Для цитирования:** Лебедев А. А. Анализ последовательностей частей речи и категория идиостиля // Ученые записки Петрозаводского государственного университета. 2021. Т. 43, № 5. С. 29–34. DOI: 10.15393/uchz.art.2021.633

### ВВЕДЕНИЕ

Категория идиостиля автора, будучи одной из значимых и активно упоминающихся в современных филологических исследованиях, может изучаться в разных аспектах, как лингвистических, так и литературоведческих. Этому способствует в том числе и вариативность использования терминов «идиостиль» и смежного с ним «идиолект» (подробный обзор разных определений терминов «идиостиль» и «идиолект» представлен в работах [2], [16]). Н. А. Фатеева в своей работе «К вопросу об изучении идиостиля Ф. М. Достоевского» [14] выделяет четыре основных подхода к изучению данной категории:

1) лингвистический (анализ языковых единиц, характерных для писателя);

2) лексикографический (создание словарей языка писателя, в том числе и статистических);

3) когнитивный (формирование языковой картины мира, определение основных концептов);

4) художественно-изобразительный (анализ деталей и реалий внешнего мира).

При этом именно лингвистический подход зачастую позволяет отыскать «скрытые» для внешнего наблюдателя особенности и закономерности построения текста на разных его уровнях, которые могут сыграть фундаментальную роль в формировании индивидуально-авторского стиля. Для решения задачи, связанной с поиском закономерностей в построении текстов, имеет смысл углубиться в анализ грамматической структуры исследуемых материалов. Однако для корректной организации подобного рода лингвистических исследований следует учитывать ряд особенностей и нюансов, связанных со сбором и обработкой текстов, предназначенных для анализа, в противном случае достоверность сделанных выводов может быть поставлена под сомнение.

## ОСНОВНЫЕ ПРОБЛЕМЫ СИСТЕМНОГО АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ ЧАСТЕЙ РЕЧИ

Полный и исчерпывающий анализ всех без исключения особенностей текста того или иного автора представляет собой трудную в силу своей комплексности и многоаспектности задачу, особенно если учитывать, что в таких областях лингвистики, как семантика и прагматика, методология исследований еще до конца не сформировалась. Именно поэтому для решения вопросов, связанных с категорией идиостиля, имеет смысл в первую очередь обращаться к тем уровням языка, анализ которых опирается на лингвистическую традицию и достаточно последовательно описан в получивших признание и прошедших проверку временем научных трудах. Естественно, что полученные данные должны быть показательными с точки зрения решаемых вопросов определения индивидуально-авторского стиля. Все это позволяет определить одним из наиболее важных в изучении стилевых черт авторов именно морфологический аспект анализа текстов. Морфология как раздел грамматики, который изучает части речи и их категории, находится в центре лингвистических исследований уже несколько столетий, а потому сомневаться в значимости данного аспекта изучения текстов не приходится.

Однако даже базовый морфологический анализ (то есть разбор текстов определенного автора по частям речи) может быть сопряжен с некоторыми трудностями, как собственно лингвистическими, так и методологическими. К числу таких трудностей, с которыми может столкнуться исследователь, следует отнести:

1. Отсутствие среди лингвистов единого взгляда на систему частей речи в русском языке. Как общее количество частей речи, выделяемых в русском языке, так и необходимость выделения каких-то конкретных из них в грамматической системе языка может становиться предметом научного спора. Некоторые исследователи склонны полагать, что «установление частеречной принадлежности слова – вопрос из разряда “вечных” в языкоznании» [13: 30]. Приведем лишь несколько примеров дискуссий подобного типа:

а) «причастие и деепричастие – это самостоятельные части речи или формы глагола» [1], [10]?

б) следует ли выделять отдельную часть речи «категория состояния» [11]?

в) выделяются ли модальные слова как самостоятельная часть речи [6]?

В этом случае сначала разработчик системы, с помощью которой будет выполняться морфо-

логический анализ, а затем и специалист, выполняющий разбор текстов, должны заранее учесть все эти спорные вопросы. Оптимальной будет опора на какую-то из уже существующих морфологических систем, однако полностью избежать противоречий вряд ли удастся.

2. Глубина проводимого морфологического анализа. Знаменательные части речи обладают своей системой лексико-грамматических разрядов и грамматических категорий (например, у существительных определяются число, род и падеж; у прилагательных – степень сравнения, род, число и падеж и т. п.), которые тоже можно учитывать в ходе анализа идиостиля на уровне морфологии. Однако, во-первых, выделение тех или иных грамматических категорий и лексико-грамматических разрядов также зачастую является дискуссионным (например, один из проблемных для русской лингвистики – спор о категории вида как словоизменительной или словообразовательной [9]); во-вторых, усложнение морфологического разбора влечет за собой необходимость более трудоемкого и времязатратного анализа текстов. Как правило, перед исследователем в этом случае встает дилемма: либо выполнять углубленный анализ, но на меньшем объеме текстов, либо учитывать небольшое число параметров, но при этом иметь возможность охватить существенно больший текстовый материал.

3. Выбор текстов, которые будут подвергнуты морфологическому анализу. Нет сомнения в том, что на структуру текста (в том числе и на его морфологическое построение) оказывает прямое влияние не только индивидуально-авторский стиль, но и другие особенности, напрямую не связанные с личностью автора. В полной мере это касается принадлежности текста к тому или иному функциональному стилю, в случае с литературными текстами – спецификой жанра некоторых из текстов и т. п. При этом, как отмечается в работе М. П. Болотской «О стилистическом аспекте изучения грамматических категорий», «вопросы стилистического изучения грамматического строя художественного произведения разработаны недостаточно», а стилистическая многогранность частей речи «не изучена в должной мере» [3: 26], а это не всегда позволяет понять, что из морфологических особенностей текста относится именно к идиостилю. Поэтому исследователь в этом случае должен по возможности стремиться к однородности исследуемых текстов, особенно в тех случаях, когда ставится задача сравнить морфологическую структуру текстов двух разных авторов, иначе искажения неизбежны.

4. Выбор методов, позволяющих проанализировать морфологическую структуру текста. В последние десятилетия филология в этом аспекте тесно сблизилась с точными науками, что позволяет говорить об актуализации такого раздела языкознания, как квантитативная лингвистика. Однако даже в пределах квантитативной лингвистики предлагается большое разнообразие методов подсчета и анализа статистических данных для решения различных смежных задач, а потому выбор оптимального из них остается одной из важнейших задач для исследователя, занимающегося изучением идиостиля автора. В то же время полученные данные, которые связаны с последовательностями частей речи, могут быть продуктивно использованы в решении смежных с определением идиостиля вопросов, например в поиске неоднородностей в текстах (подробнее см. [12]).

### СИСТЕМА «СМАЛТ» И ЕЕ ИСПОЛЬЗОВАНИЕ В АНАЛИЗЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ЧАСТЕЙ РЕЧИ

В Петрозаводском государственном университете активно ведутся исследования, связанные с анализом публицистических текстов различных авторов. С 2018 года благодаря поддержке РФФИ разрабатывается веб-версия информационной системы «Статистические методы анализа литературного текста» (ИС «СМАЛТ», <http://smalt.karelia.ru/>), позволяющей выполнять в том числе и морфологический анализ текстов, представленных в дореформенной орфографии.

Грамматическая разметка текстов в системе «СМАЛТ» сформирована с опорой на концепцию В. В. Виноградова [5] и включает 14 частей речи, каждая из которых имеет свое числовое обозначение (существительное – 0, прилагательное – 1, числительное – 2, местоимение – 3, глагол – 4, причастие – 5, деепричастие – 6, наречие – 7, категория состояния – 8, частица – 9, предлог – 10, союз – 11, модальное слово – 12, междометие – 13). Также слова в анализируемых текстах можно помечать как цитаты, иностранные слова, вводные слова, сокращенные слова и неязыковые символы. Таким образом, любое предложение (и шире – любой текст) можно представить в виде числовой последовательности, например:

*Послѣдніе два года были особенно тяжелы для славянофильского кружка*

1 – 2 – 0 – 4 – 7 – 1 – 10 – 1 – 0

Одним из наиболее продуктивных математических методов в решении вопросов классификации стало построение деревьев решений (математический аспект данного метода подроб-

но описан в работах [8] и [17]). В основе метода построения деревьев решений лежит поиск таких последовательностей частей речи, которые будут характерны для одного исследуемого автора и не будут характерны для другого (то есть будут отличать их тексты и служить одним из элементов формирования идиостиля). Подобного рода сопоставление авторов на базе последовательностей частей речи играет важную роль в решении вопросов атрибуции текстов (в частности, оно упомянуто в одной из наиболее значимых статей, посвященных атрибуции текстов [19]), когда один и тот же текст потенциально могли написать два автора. В этом случае именно поиск таких частотных морфологических структур может помочь в определении авторства.

Материалами для исследования послужили тексты 60–70-х годов XIX века (преимущественно это журналы «Время», «Эпоха» и еженедельник «Гражданин»). Такой выбор не случаен: с одной стороны, обеспечивается единство времени написания текстов; с другой – появляется возможность проанализировать идиостиль разных авторов (это не только Ф. М. Достоевский, но и М. М. Достоевский, А. А. Григорьев, Н. Н. Страхов, В. П. Мещерский, Я. П. Полонский и многие другие), а также привлечь к исследованию анонимные и псевдонимные тексты.

Ряд исследований, осуществленных в рамках проекта РФФИ «Проблема атрибуции анонимных и псевдонимных статей в журналах “Время”, “Эпоха” и еженедельнике “Гражданин”», включал в себя подобный морфологический анализ текстов. В частности, методика анализа последовательностей частей речи была продуктивно использована в статье Д. Д. Бучневой «Кто автор редакционной статьи “Желание” в первом номере “Гражданина” за 1873 год?» [4], где в том числе и с опорой на последовательности частей речи «Прилагательное + Существительное» и «Глагол + Существительное» был сделан вывод о признании В. П. Мещерского автором спорной статьи «Желание».

Другим объектом исследования в работе «Text Attribution in Case of Sampling Imbalance by the Method of Constructing an Ensemble of Classifiers Based on Decision Trees» [18] стала статья «Стихотворения Хомякова», которую в течение долгого времени приписывали Аполлону Григорьеву, однако данное суждение было поставлено под сомнение (подробно литературоведческий аспект данного спора описан в [7]). При помощи математических методов была выделена значимая для эталонных текстов А. Григорьева последовательность частей речи «Частица +

Прилагательное»; низкая частотность данной последовательности в спорном тексте стала одним из доводов того, что данная статья не принадлежит А. Григорьеву.

Следует отметить, что подобного рода анализ последовательностей частей речи может быть выполнен как на материале всего текста, так и с опорой на отдельные его составляющие. В частности, в работе Г. Хетсо [15] было заявлено 15 параметров определения индивидуально-авторского стиля, связанных в том числе и с анализом сочетаний частей речи в первых и последних позициях предложения.

Этот метод анализа начал предложений как сильных позиций текста, значимых с точки зрения формирования идиостиля, в рамках выполненного проектного исследования был усовершенствован (анализировалось распределение трех частей речи в начале предложения и трех частей речи в конце предложения; предложения, размер которых составлял менее трех слов, пропускались). Анализу были подвергнуты 54 эталонных текста четырех авторов (Ф. М. Достоевского, А. А. Григорьева, Н. Н. Страхова, В. П. Мещерского). Наиболее значимыми для начала предложений Ф. М. Достоевского оказалась комбинация «Местоимение + Частичца»; среди трех последних слов в предложении для Ф. М. Достоевского наиболее типичным вариантом является завершение предложения глаголом. После анализа таких контекстов было обнаружено, что подобный завершающий глагол у Достоевского часто комбинируется с вопросительной формой предложений, причем небольших. Приведем несколько примеров из введения к «Ряду статей о русской литературе»<sup>1</sup>:

- Но неужели жъ разувѣрять?
- Отчего же не говорить? Отчего же именно непремѣнно молчать?
- Неужели же для того, чтобъ ничего не дѣлать?
- Гдѣ же ваша способность самоосужденія, гдѣ вашъ трезвый взглядъ, которыми вы такъ хвалились?

- Да гдѣ жъ они были, спрашивали мы, гдѣ жъ они до сихъ поръ прятались?
- Что имъ дѣлать?
- Ну, чтожъ имъ отвѣтить?
- Чего бояться?

Подобного рода статистические результаты могут быть интересны исследователям, анализирующими фигуры речи, связанные с расположением слов в предложении и повторами (анафоры, эпифоры, синтаксические повторы, риторические вопросы и т. п.), и влияние этих фигур речи на читателей или слушателей текста.

## ВЫВОДЫ

Морфологическая структура текста содержит в себе отпечаток индивидуально-авторского стиля, воплощаемого в использовании определенных частей речи и их сочетаний. Применение тех или иных комбинаций может быть как бессознательным, так и вполне осознанным (особенно в тех случаях, когда автор текста ставит перед собой соответствующую задачу воздействия на читателя при помощи тропов и фигур речи, которые зачастую тяготеют к началу или концу предложения). Но и в том, и в другом случае именно морфологический анализ частеречных последовательностей, выполненный на большом массиве текстов, становится одним из эффективных инструментов, отличающих идиостили разных авторов, а также помогающих более обоснованно определять авторство анонимных и псевдонимных текстов. Проделанные в ходе работы над проектом «Проблема атрибуции анонимных и псевдонимных статей в журналах “Время”, “Эпоха” и еженедельнике “Гражданин”» исследования, связанные с выделением значимых для создателя текста последовательностей частей речи в тексте, способны сыграть важную роль в решении вопросов определения идиостиля автора.

## ПРИМЕЧАНИЕ

<sup>1</sup> Указанные контексты цитируются по тексту «Ряд статей о русской литературе. Введение» в базе данных «СМАЛТ» (<http://smalt.karelia.ru/shower/?id=42>).

## СПИСОК ЛИТЕРАТУРЫ

1. Абдулхакова Л. Р. Русское деепричастие: часть речи или форма глагола? // Вестник Волгоградского государственного университета. Серия 2 (12). Языкоznание. 2010. № 2. С. 67–72.
2. Боданова Е. В. О некоторых аспектах изучения термина *идиолект* в отечественной и западной лингвистике // Вестник Ленинградского государственного университета им. А. С. Пушкина. 2011. Т. 1. № 4. С. 100–108.
3. Болотская М. П. О стилистическом аспекте изучения грамматических категорий // Известия Пензенского государственного педагогического университета им. В. Г. Белинского. 2009. № 15. С. 25–29.
4. Бучнева Д. Д. Кто автор редакционной статьи «Желание» в первом номере «Гражданина» за 1873 год? // Неизвестный Достоевский. 2020. № 2. С. 142–161. DOI: 10.15393/j10.art.2020.4721

5. Виноградов В. В. Русский язык. М.: Высш. шк., 1972. 616 с.
6. Гусева Л. А. Модальные слова как часть речи // Филологические науки. Вопросы теории и практики. 2017. № 1–2 (67). С. 109–113.
7. Захаров В. Н. Вопрос о А. С. Хомякове // Имя автора – Достоевский. Очерк творчества. М.: Индрик, 2013. С. 231–247.
8. Кафтанников И. Л., Парасич А. В. Особенности применения деревьев решений в задачах классификации // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2015. Т. 15. № 3. С. 26–32.
9. Коваленко Б. Н. О спорных вопросах теории вида русского глагола // Инновационная наука. 2015. № 6–2. С. 158–162.
10. Кунавина И. И. Грамматический статус причастия в современном русском языке // Актуальные проблемы филологии и педагогической лингвистики. 2016. № 2 (22). С. 68–73.
11. Мельник А. Д. Проблема слов категории состояния в современной лингвистической литературе // Современные проблемы науки и образования. 2015. № 2–2 [Электронный ресурс]. Режим доступа: <http://www.science-education.ru/ru/article/view?id=23004> (дата обращения 09.02.2021).
12. Седов А. В., Рогов А. А. Анализ неоднородностей в тексте на основе последовательностей частей речи // Современные проблемы науки и образования. 2013. № 1 [Электронный ресурс]. Режим доступа: <http://www.science-education.ru/ru/article/view?id=8339> (дата обращения 09.02.2021).
13. Тукова Т. В. Части речи в курсе грамматики // Ученые записки Крымского федерального университета имени В. И. Вернадского. Филологические науки. 2014. № 1–1. С. 30–34.
14. Фатеева Н. А. К вопросу об изучении идиостиля Ф. М. Достоевского // Григорьевские чтения (2020): Тезисы. Сайт Института русского языка имени В. В. Виноградова Российской академии наук [Электронный ресурс]. Режим доступа: <http://www.ruslang.ru/doc/grigoriev2020/Fateeva.pdf> (дата обращения 09.02.2021).
15. Хетсо Г. Приналежность Достоевскому: к вопросу об атрибуции Ф. М. Достоевскому анонимных статей в журналах «Время» и «Эпоха». Осло: SOLUM FORLAG A. S., 1986. 86 с.
16. Чернышева Т. А. Идиостиль: лингвистические контуры изучения // Вестник Череповецкого государственного университета. 2010. № 1. С. 30–34.
17. Breiman L., Friedman J. H., Olshen R. A., Stone C. J. Classification and regression trees. Wadsworth, Belmont, Ca, 1984. 368 p.
18. Rogov A. A., Moskin N. D., Abramov R. V., Kulakov K. A. Text attribution in case of sampling imbalance by the method of constructing an ensemble of classifiers based on decision trees // Data Analytics and Management in Data Intensive Domains: XXII International Conference DAMDID/RCDL'2020 (October 13–16, 2020, Voronezh, Russia): Extended abstracts of the conference. Voronezh: Voronezh State University, 2020. P. 185–188 [Электронный ресурс]. Режим доступа: [http://damdid2020.cs.vsu.ru/DAMDID\\_2020\\_Extended\\_Abstracts.pdf](http://damdid2020.cs.vsu.ru/DAMDID_2020_Extended_Abstracts.pdf) (дата обращения 09.02.2021).
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.

*Поступила в редакцию 01.03.2021; принята к публикации 17.05.2021*

---

Original article

Alexander A. Lebedev, Cand. Sc. (Philology), Petrozavodsk State University (Petrozavodsk, Russian Federation)  
ORCID 0000-0001-9939-9389; [perevodchik88@yandex.ru](mailto:perevodchik88@yandex.ru)

## PARTS OF SPEECH SEQUENCES ANALYSIS AND THE CATEGORY OF IDIOSTYLE

**A b s t r a c t.** The article addresses the linguistic approach to the analysis of the author's idiosyncrasy, which takes into account the morphological features of texts and the parts of speech sequences presented therein. The author outlines the purpose of analyzing the parts of speech sequences in the context of determining the individual author's style. He also reveals specific problems related to such analysis – multiplicity of linguistic concepts, variability of the analysis depth, and specificity of the selection of texts and methods for their analysis. The paper describes the Statistical Methods of Literary Text Analysis (SMALT) information system, which can be used, inter alia, for the morphological analysis of publicistic texts, substantiates the usage of a mathematical method for constructing decision trees, and provides the results of the research. The author gives particular examples of the parts of speech combinations that distinguish texts written by different authors. Eventually the conclusion is made about the impact of the text morphological structure on its perception by readers, as well as about the importance of analyzing the parts of speech sequences for determining the author's idiosyncrasy.

**K e y w o r d s :** idiosyncrasy, parts of speech, parts of speech sequences, morphology, decision tree, SMALT

**A c k n o w l e d g e m e n t s .** The reported study was funded by the Russian Foundation for Basic Research (RFBR), project No 18-012-90026 Dostoevsky ("Attribution of anonymous and pseudonymous articles in *Time*, *Epoch* and *The Citizen* periodicals").

For citation: Lebedev, A. A. Parts of speech sequences analysis and the category of idiom. *Proceedings of Petrozavodsk State University*. 2021;43(5):29–34. DOI: 10.15393/uchz.art.2021.633

#### REFERENCES

1. Abdulkhakova, L. R. Russian adverbial participles: part of speech or verb form? *Science Journal of Volgograd State University. Linguistics*. 2010;2(12):67–72. (In Russ.)
2. Bogdanova, E. V. Some aspects of the study of the term “idiolect” in Russian and Western linguistics. *Pushkin Leningrad State University Journal*. 2011;1(4):100–108. (In Russ.)
3. Bolotskaya, M. P. Stylistic aspect of studying grammatical categories. *Izvestia of Belinsky State Pedagogical University*. 2009;15:25–29. (In Russ.)
4. Buchneva, D. D. Who is the author of the editorial “Zhelanie” (“Desire”) in the first issue of *Grazhdanin (Citizen)* for 1873? *The Unknown Dostoevsky*. 2020;2:142–161. DOI: 10.15393/j10.art.2020.4721 (In Russ.)
5. Vinogradov, V. V. Russian language. Moscow, 1972. 616 p. (In Russ.)
6. Guseva, L. A. Modal words as part of speech. *Philological Sciences. Issues of Theory and Practice*. 2017;1–2(67):109–113. (In Russ.)
7. Zakharov, V. N. The question about A. S. Khomyakov. *The author's name is Dostoevsky. An essay on the creative work*. Moscow, 2013. P. 231–247. (In Russ.)
8. Kaftannikov, I. L., Parasich, A. V. Decision trees' features of application in classification problems. *Bulletin of the South Ural State University. Series: Computer Technology, Automatic Control & Radioelectronics*. 2015;15(3):26–32. (In Russ.)
9. Kovalenko, B. N. Some controversial issues of the theory of the Russian verb aspect. *Innovative Science*. 2015;6–2:158–162. (In Russ.)
10. Kunavina, I. I. The grammatical status of participle in modern Russian. *Current Issues in Philology and Pedagogical Linguistics*, 2016;2(22):68–73. (In Russ.)
11. Mel'nik, A. D. The problem of the state category words in the modern linguistic literature. *Modern Problems of Science and Education*. 2015;2(2). Available at: <http://www.science-education.ru/ru/article/view?id=23004> (accessed 09.02.2021). (In Russ.)
12. Sedov, A. V., Rogov, A. A. Analysis of text heterogeneities based on the parts of speech sequences. *Modern Problems of Science and Education*. 2013;1. Available at: <http://www.science-education.ru/ru/article/view?id=8339> (accessed 09.02.2021). (In Russ.)
13. Tukova, T. V. Parts of speech in grammar course. *Scientific Notes of V. I. Vernadsky Crimean Federal University. Philological sciences*. 2014;1(1):30–34. (In Russ.)
14. Fateeva, N. A. The study of Dostoevsky's idiom. *Proceedings of the Grigoryev Readings (2020)*. Website of the V. V. Vinogradov Russian Language Institute of the Russian Academy of Sciences. Available at: <http://www.ruslang.ru/doc/grigoriev2020/Fateeva.pdf> (accessed 09.02.2021). (In Russ.)
15. Kjetsaa, G. Attributed to Dostoevsky: the problem of attributing to Dostoevsky anonymous articles in *Time* and *Epoch*. Oslo, 1986. 86 p. (In Russ.)
16. Chernysheva, T. A. Idiom: linguistic aspect. *Cherepovets State University Bulletin*. 2010;1:30–34. (In Russ.)
17. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. Classification and regression trees. Wadsworth, Belmont, Ca, 1984. 368 p.
18. Rogov, A. A., Moskin, N. D., Abramov, R. V., Kulakov, K. A. Text attribution in case of sampling imbalance by the method of constructing an ensemble of classifiers based on decision trees. *Data Analytics and Management in Data Intensive Domains: XXII International Conference DAMDID/RCDL'2020 (October 13–16, 2020, Voronezh, Russia): Extended abstracts of the conference*. Voronezh, 2020. P. 185–188. Available at: [http://damdid2020.cs.vsu.ru/DAMDID\\_2020\\_Extended\\_Abstracts.pdf](http://damdid2020.cs.vsu.ru/DAMDID_2020_Extended_Abstracts.pdf) (accessed 09.02.2021).
19. Stamatatos, E. A Survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*. 2009;60(3):538–556.

*Received: 1 March, 2021; accepted: 17 May, 2021*