

Коган Марина Самуиловна

кандидат технических наук, доцент
Санкт-Петербургский
политехнический университет
г. Санкт-Петербург, Россия

Marina Kogan

PhD in technology, Associate professor
Peter the Great Saint Petersburg
Polytechnic University
Saint Petersburg, Russia
m_kogan@inbox.ru

Коростелев Денис Александрович

магистр лингвистики
Санкт-Петербургский
политехнический университет
г. Санкт-Петербург, Россия

Denis Korostelev

MA in Linguistics
Peter the Great Saint Petersburg
Polytechnic University
Saint Petersburg, Russia
korostelev.da@spbstu.ru

Дмитриев Александр Владиславович

кандидат филологических наук, доцент
Санкт-Петербургский
политехнический университет
г. Санкт-Петербург, Россия

Alexandr Dmitriev

PhD in philology, Associate professor
Peter the Great Saint Petersburg
Polytechnic University
Saint Petersburg, Russia
dmitriev_av@spbstu.ru

МЕТАДААННЫЕ ПОДДЕРЖАННЫХ ЗАЯВОК НА ГРАНТЫ РФ
В ОБЛАСТИ КОМПЬЮТЕРНЫХ НАУК
КАК ОБЪЕКТ ЛИНГВИСТИЧЕСКОГО АНАЛИЗА

В современном мире грантовая поддержка от разных фондов государственных/негосударственных/частных/корпоративных является важным источником финансирования научных исследований. Благодаря общей тенденции на открытость науки победители конкурсов делают доступными результаты своих исследований, а на сайтах фондов размещается краткая информация о поддержанных заявках, содержащая название проекта, информацию о руководителе и главных исполнителях, аннотацию, ключевые слова и др.

В проведенном исследовании сделана попытка выявить типичные лингвистические характеристики заявок в области компьютерных наук, поддержанных РФ (Российским научным Фондом) в 2020–2023 гг. Собранный вручную корпус, содержащий 350 заявок, был предобработан с помощью известных инструментов автоматической обработки текста. После этого был проведен анализ употребления выделенных авторами ключевых слов, в названиях проектов и в аннотациях; заголовки и аннотации исследовались на удобочитаемость и длину. 12 заявок в области компьютерной лингвистики были проанализированы более детально. Проведенный анализ, не позволяет сформулировать четких рекомендаций по написанию заявок для молодых исследований, т.к. в решении поддержать заявку или нет на уровне РФ важную роль играют не только лингвистические, но и экстралингвистические факторы, рассмотрение которых выходило за рамки проведенного исследования. Однако полученные результаты можно использовать в некоторых курсах магистратуры с целью развития необходимого для начинающих исследователей навыка убедительно рассказать о своем исследовании, чтобы обеспечить его финансовую поддержку.

Ключевые слова: заявки на грант; РФФ; метаданные заявки; лингвистический анализ; автоматическая обработка текста; удобочитаемость; корпусный менеджер AntCont.

METADATA OF RSF'S SUPPORTED GRANT APPLICATIONS IN INFORMATION TECHNOLOGY DOMAIN AS THE OBJECT OF LINGUISTIC ANALYSIS

In the modern world grant funding from different funds – governmental, non-governmental, private, corporate – has become a cornerstone of scientific research. Thanks to the general trend towards open science, grant receivers make the outcomes of their research available while the funds' websites preserve brief information about supported applications including the Title of the project, biodata about the research team leader, main executors, summary and expected results, keywords, etc.

This study attempts to reveal typical linguistic characteristics of applications in IT domain supported by the Russian Science Foundation (RSF) in 2020–2023. The mutually collected corpus of 350 applications was pre-processed with available NLP tools. This was followed by the analysis of the titles and summaries' length and readability, titles' structure, and the keywords use in the titles and summaries. 12 applications in Computer linguistics domain have been studied in more detail. Though the analysis hasn't resulted in clear recommendations for young researchers on linguistic orchestrating their grant applications, yet findings could be implemented into a number of master courses to develop in young researchers a skill of clearly and convincingly presenting their research to provide its financial support.

Key words: *grant applications; RSF; applications' metadata; linguistic analysis; natural language processing; readability; AntConc software.*

В современном мире грантовая поддержка от разных государственных/негосударственных/частных/корпоративных фондов является важным источником финансирования научных исследований. Благодаря общей тенденции на открытость науки победители конкурсов делают доступными результаты своих исследований, а на сайтах фондов размещается краткая информация о поддержанных заявках, содержащая название поддержанного проекта, информацию о руководителе и главных исполнителях, аннотацию, ключевые слова и др.

Большие гранты обычно получают исследовательские коллективы, в состав которых входят известные в своей области ученые, талантливые студенты или предприниматели [1]. На сегодняшний день гранты считаются одним из наиболее распространенных и эффективных способов поддержки научных исследований и имеют важное статусное значение, принося их обладателям признание научного сообщества и способствуя росту профессиональной репутации [2; 3]. В России большие научные проекты получают поддержку из наиболее известной некоммерческой организации – Российского научного фонда (РНФ), созданного по инициативе Президента России в ноябре 2013 года.

Факторы, влияющие на успех грантовой заявки, можно условно разделить на 2 категории: нелингвистические и лингвистические.

Нелингвистические факторы, которые зависят от предмета и типа конкурса, в совокупности гарантируют, что заявка удовлетворяет основным требованиям фонда. К ним относятся новизна, актуальность, цели, задачи, методология и план проведения исследований, задел по теме исследования, бюджет проекта и др., большинство из которых описываются в отдельном разделе заявки под соответствующей рубрикой.

Заявки на грант образуют отдельный жанр среди письменных текстов научного стиля [4].

Лингвистические факторы касаются того, как заявка написана, структурирована, насколько она удобочитаема за счет длины слов и предложений, использованных синтаксических конструкций, насколько она способствует автоматическому профилированию информации в условиях 4-й промышленной революции [5].

Из-за доступности в открытом доступе только метаданных поддержанных заявок на гранты РФ наше исследование ограничилось проведением лингвистического анализа заголовков, аннотаций и ключевых слов заявок в сфере ИТ за 2020–2023 годы.

Цель исследования состояла в поиске ответа на вопрос о наличии лингвистических характеристик типичных для успешных грантов в данной области, в которой наблюдается острая конкуренция между разными проектами за получение финансовой поддержки в условиях «цифрового поворота» практически во всех сферах деятельности и отраслях промышленности.

Материал и процедура исследования

Материалом для исследования стал корпус из 350 метаданных поддержанных заявок в области ИТ (Отрасль знания 01 «Математика, информатика и науки о системах» по классификации РФ) за 2020–2023 гг. собранный вручную с сайта РФ (<https://www.rscf.ru/>). Метаданные заявок были сохранены в формате .xlsx как один документ в виде набора данных, который содержит 350 строк и 8 атрибутов, включающих ID, заголовок, ключевые слова, код и название отрасли и подотрасли, текст аннотации и в некоторых случаях информацию о месте работы руководителя гранта и размере исследовательского коллектива.

Лингвистический анализ включал исследование заголовков (длина, структура, удобочитаемость), исследование аннотаций (длина, удобочитаемость) и ключевых слов, выделенных авторами, для проверки соответствия некоторым условиям автоматического профилирования информации. Эти условия связаны с авторской разработкой структурированного контента с разбивкой на тематические разделы, которые достаточно полно описываются ключевыми терминами. В настоящее время выбор ключевых слов рассматривается исследователями как основополагающая задача формулирования полученного результата, важный параметр для дальнейшего использования текста специалистами и распространения результатов [5]. Рекомендацию по использованию ключевых слов в заголовках научных статей содержат некоторые пособия по академическому письму [6], а гипотеза о том, что ключевые слова должны включаться в аннотации статей, выдвигается некоторыми исследователями [5, с. 58].

Для подготовки и анализа данных использовались инструменты автоматической обработки текста, а именно: библиотека NLTK, pandas, библиотеки ruTS для русского языка и популярная программа AntConc – свободно распространяемый корпусный менеджер, разработанный Л. Антони [7].

Предобработка текстовых данных осуществлялась с помощью библиотек NLTK и pandas для Python и включала токенизацию, удаление стоп-слов и знаков препинания, изменение регистра на строчный и сохранение обработанного текста в новом файле “cleared.xlsx” для текстов аннотаций, заголовков и ключевых слов. Длина заголовков и аннотаций определялась с помощью стандартной программы, написанной на языке программирования Python. Для определения удобочитаемости использовалась специализированная библиотека ruTS для расчета показателей удобочитаемости текстов на русском языке. Библиотека ruTS представляет собой пакет Python, который содержит набор показателей удобочитаемости, включая Flesch-Kincaid Reading Ease (FRE), Gunning FOG, Simple Measure of Gobbledygook (SMOG) и индекс Coleman-Liau (CLI) [8].

Обнаруженные в корпусе 12 заявок в области компьютерной лингвистики были проанализированы более детально.

Результаты и обсуждение

Длина заголовков и аннотаций

На этапе определения длины заголовков названий проектов и аннотаций были получены результаты, представленные в табл. 1.

Т а б л и ц а 1

Результаты расчета длины заголовков и аннотаций

Метрика	Значение
Средняя длина заголовка	13,63
Средняя длина аннотации	747,83
Диапазон слов в заголовке	4–31
Диапазон слов в аннотации	139–3838
Среднеквадратичное отклонение длины заголовка	4,8
Среднеквадратичное отклонение длины аннотации	421,69

Результаты расчета показывают, что средняя длина заголовка составляет 13,6 слова, а средняя длина аннотации – 747,8 слова. Наиболее длинные заголовки характерны для заявок в отраслях «Информационные технологии интеллектуальной поддержки принятия решений» и «Математическое моделирование технических систем». Короткие названия проектов встречались в таких отраслях, как «Искусственный интеллект и принятие решений»,

«Интеллектуальный анализ данных и распознавание образов», «Вычислительная математика», «Обработка и анализ изображений и сигналов», «Системы и технологии интеллектуального анализа данных и распознавания образов».

Наиболее длинные аннотации заявок были обнаружены в тех же отраслях, в которых были найдены самые длинные названия. Это отрасли, связанные с математическим моделированием, искусственным интеллектом и интеллектуальным анализом. В табл. 2 представлена количественная информация о самых длинных и коротких заголовках и аннотациях. К коротким мы относили заголовки и аннотации, отличающиеся от среднего на число, большее чем одно стандартное отклонение. (В нашем случае это заголовки, содержащие от 4 до 8 слов, и аннотации от 139 до 326 слов). К длинным мы относили заголовки и аннотации, отличающиеся от среднего на число, большее чем одно стандартное отклонение. (В нашем случае это заголовки, содержащие от 18 до 31 слова, и аннотации от 1 170 до 3 838 слов).

Т а б л и ц а 2

Количество самых длинных и самых коротких аннотаций и заголовков

Атрибут	Диапазон слов	Количество заявок
Длинные заголовки	18–31	35
Короткие заголовки	4–8	47
Длинные аннотации	1 170–3 838	35
Короткие аннотации	139–326	28

Таким образом, в нашем корпусе доля длинных аннотаций и заголовков составила 10 %; коротких аннотаций оказалось 8 % и 13,4 % коротких заголовков.

Структура заголовков проектов

При проведении эксперимента был произведен анализ наиболее типовых конструкций и лексики заголовков грантовых заявок, использованных авторами с опорой на типы заголовков, выделенные в исследовании Н. К. Рябцевой [9]. По результатам анализа заголовков в корпусе можно сделать следующие выводы.

Для грантовых заявок наиболее распространенным паттерном заголовков является именная группа с большим количеством зависимых членов, которые всегда в постпозиции (примеры: «Разработка и исследование методов и алгоритмов визуального внимания на основе нейронных сетей на графах и данных с ограниченной аннотацией», «Разработка модели компьютерного зрения для интеллектуальной навигации робототехнических систем, основанной на построении трехмерных сцен по картам глубин»).

Продуктивным является использование генитивных (неличных) глагольных форм, отглагольных существительных. В 45 % заголовков был обнаружен подобный паттерн, часто в виде сочинительной конструкции с союзом «и» (примеры: «Автоматические методы построения *и* пополнения баз зна-

ний на основе кросс-языковых технологий», «Разработка и имплементация некоторых быстрых алгоритмов в полиномиальных кольцах в библиотеке Rings для задач компьютерной алгебры и криптографии»).

Как и в случае научных публикаций, в заголовках грантовых заявок присутствует большое количество абстрактной лексики в сочетании с отглагольными существительными, самым частотным из которых было слово «метод» (анализа, обработки, машинного обучения), которое употребляется в 48,6 % заголовков.

Типовым было использование металингвистических конструкций в виде существительного с предлогом (*на основе, на примере, с помощью*), которые поясняли и дополняли первую часть заголовка. Примеры: «Исследование внутренней структуры неоднородных материалов при нагружении *с помощью* компьютерного томографа», «Методы создания и адаптации интеллектуальных систем *на основе* физически информированных нейронных сетей».

Имели место случаи (всего 9 из 350) использования двухчастных конструкций с двоеточием (примеры: «Мировое развитие и «пределы роста» в 21 веке: моделирование и прогноз», «Математическое моделирование динамики сплошных сред: разработка параллельных алгоритмов и программного обеспечения для суперкомпьютеров с GPUs»).

Конструкции с союзом *как* (2 %) и предлогом *для* (33 %). Примеры: «Методы повышения точности позиционирования с использованием технологии дополненной реальности *как* способа навигации в хирургическом вмешательстве», «Методы предварительного обучения графовых нейронных сетей *для* задач моделирования исходного кода программ».

Определение удобочитаемости аннотаций и заголовков

Экспериментальное определение удобочитаемости аннотаций показало, что по каждой метрике удобочитаемости все тексты получили самую высокую категорию сложности, так что особенных паттернов для определенных метрик выявить не удалось. Это ожидаемый результат, т.к. целевой аудиторией заявок на гранты являются авторитетные эксперты предметной области. Несмотря на разные критерии, заложенные в разные формулы удобочитаемости, каждая аннотация попала в категорию текстов наивысшей сложности по каждому индексу. Как представляется, проблема кроется в том, что формулы создавались для и проверялись на других категориях текстов, в первую очередь для оценки удобочитаемости учебных текстов для обучающихся разных возрастных категорий (от младшей школы до окончания университета) и плохо приспособлены для оценки сложности текстов, расположенных на границе и за границей категории текстов самой высокой сложности, ориентированных на экспертов предметной области, а не просто носителей языка с высшим образованием.

Анализ встречаемости ключевых слов в заголовках и аннотациях

Анализ встречаемости ключевых слов, выделенных авторами заявок на гранты, в названиях проектов и аннотациях проводился с помощью корпусного менеджера AntConc (версия 4.2.4), доступного на сайте разработчика.

Алгоритм использования данной программы для решения поставленной задачи следующий: 1) создаются корпуса (файлы) в формате txt с кодировкой Юникод UTF-8, которые анализируются на наличие определенных лексических единиц; 2) корпус ключевых слов сравниваются с корпусами заголовков и аннотаций при помощи инструмента *Word*, который позволяет соотносить слова в разных корпусах; 3) далее данные сортируются по параметру *Range*, который показывает, в скольких корпусах содержатся интересующие исследователя слова. В ряде задач этот параметр позволяет отсеять узкоспециальную низкочастотную лексику, представленную с непропорционально высокой частотой только в одном корпусе, тем самым уменьшая систематические искажения в списке частотной лексики. Для этой цели задают количество корпусов, в которых исследуемая лексика должна присутствовать.

В нашем случае, если $Range=3$, мы получим список слов, которые встречаются во всех 3 корпусах – ключевых словах, заголовках и аннотациях. При $Range=2$, можно определить количество слов, встречающихся в 2 из 3 корпусов. Если $Range=1$, то будут доступны все слова, встречающиеся в каждом корпусе с заданной частотой. (В нашем случае – хотя бы 1).

По результатам анализа, из 2399 ключевых слов в корпусе только 650 (или 27 %) были использованы в аннотациях и в заголовках. Большая часть ключевых слов (1749 слов, или 73 %) не встретилась ни в заголовках, ни в аннотациях. Последующее сравнение каждого из списков слов позволило установить, что в заголовках использовано значительно больше ключевых слов, чем в аннотациях (488 против 162), и что 472 ключевых слова употреблялись как минимум дважды и в аннотациях, и в названиях.

Что касается предметной области, то заявки, в названии которых присутствовали ключевые слова, соответствовали таким отраслям, как «Системы и технологии интеллектуального анализа данных и распознавания образов», «Системы семантического моделирования», «Искусственный интеллект и принятие решений». В аннотациях, напротив, не наблюдалось какой-либо определенной закономерности использования или неиспользования ключевых слов в зависимости от отрасли.

Функция *Range* в программе *AntConc* позволяет автоматически находить и сравнивать количество одинаковых токенов, представленных униграммами, т.е. состоящих из одного слова, но не позволяет получить информацию о том, использовались ли двух- и многокомпонентные словосочетания, выделенные авторами в качестве ключевых слов, в аннотациях и /или заголовках в исходной или измененной форме.

Для решения этой проблемы был написан скрипт на языке Python с использованием библиотеки *pandas*. Принцип работы скрипта заключается в следующем: программа разбивает ключевые слова по запятой (,) и читает их целиком, сравнивая с текстом заголовков и аннотаций; в случае хотя бы однократного использования ключевого слова в метаданных оно записывается в отдельный столбец «Found Keywords», образуя таким образом подкорпус аннотаций и заголовков, содержащих ключевые слова (фрагмент скрипта

представлен на рис.1). По результатам анализа в корпусе аннотаций хотя бы одно ключевое слово было обнаружено в 201 аннотации и в 42 заголовках в диапазонах от 1 до 9 слов и 1 до 3 слов соответственно.

Нужно отметить, что у данного метода есть ограничения в обнаружении ключевых слов на русском языке ввиду его синтетического строя. Он обнаруживает иноязычные термины, например, *BERT*, *GPT* в заявках в области компьютерной лингвистики, и одно- и многокомпонентные словосочетания из списка ключевых слов, если в заголовках и аннотациях они также используются в именительном падеже. Однако если ключевое слово употреблено в другом падеже (например, [для] «генерации текстов» в заголовке, а не «генерация текстов» как в ключевых словах), оно не будет идентифицировано. Это означает, что полученные данные представляют нижний порог оценки употребления ключевых слов, выделенных автором, в аннотациях или заголовках заявок, и при использовании более сложного алгоритма, осуществляющего стемминг или лемматизацию, размер этого подкорпуса увеличится.

Это свидетельствует о том, что метаданные большинства поддержанных заявок соответствуют принципам повышения эффективности продвижения полученного результата, выдвигая тематику и результаты исследования на первый план.

```
# Создаём пустой список для найденных слов
found_keywords = []
# Проходим через каждую строку в датафрейме
for index, row in df.iterrows():
    keywords = row['Ключевые слова'].split(",") #Разделяем
    ключевые слова по запятой
    title = row['Заголовок']
# Проверяем, есть ли ключевые слова в заголовке
    found = [keyword.strip() for keyword in keywords if
    keyword.strip() in title]
# Добавляем найденные слова в список через запятую
    found_keywords.append(", ".join(found))
```

Рис. 1. Фрагмент скрипта для поиска ключевых слов

Краткая характеристика поддержанных заявок в области компьютерной лингвистики

Исследования в этой группе охватывают вопросы измерения сложности и способы упрощения текстов, анализ и определение тональности текстов русскоязычного сегмента Интернета и отзывов на продукцию фармацевтической промышленности, совершенствование вопросно-ответных систем на основе усвоения знаний из неоднородных источников, управляемой генерации текстов, извлечения структурированной аргументации и др. Большинство поддержанных проектов в области компьютерной лингвистики выполнялось в малых группах. Руководители исследовательских коллективов аффилированы с Казанским федеральным университетом, московскими университетами, включая МГУ, ИТМО-университетом в Санкт-Петербурге,

Вятским государственным университетом, Институтом систем информатики им. А. П. Ершова СО РАН и др. В некоторых коллективах участвуют исследователи из разных университетов. В этой группе университетские исследовательские коллективы преобладают по сравнению с исследовательскими коллективами из исследовательских институтов. Это тенденция, которую отмечают многие исследователи, занимающиеся грантовой проблематикой (см. [3]).

Внутри этой выборки средняя длина заголовков равняется 13,25 слова, средняя длина аннотации равняется 610,83 слова, среднеквадратичное отклонение равно 3,46 и 193,93 слова соответственно.

Если говорить про ключевые слова, используемые в данной выборке, то по результатам анализа среднее количество ключевых слов на заявку равнялось 6 в диапазоне от 5 до 9 терминов. В среднем ключевом термине было использовано 2,2 слова. Автоматическое распознавание ключевых слов показало их полное отсутствие в заголовках, однако ввиду понимания ограничений подобного метода в отношении русскоязычного текста был проведен дополнительный анализ вручную. Несмотря на отсутствие ключевых слов в именительном падеже, выделенные автором термины в родительном падеже в составе именной группы и иногда во множественном числе встречались, в среднем, в количестве 1 слова или словосочетания на заголовок (Примеры «нечеткий граф» – «...на основе периодических нечетких графов»). Также можно отметить случай использования синонимичного выражения в заголовке («экстракция содержательной информации» – «Разработка комплекса нейросетевых алгоритмов выделения содержательной информации...»)

Выводы

Проведенный лингвистический анализ собранного вручную и предобработанного с помощью стандартных библиотек корпуса из 350 поддержанных заявок на получение грантов РФФИ за период 2020–2023 гг. в сфере ИТ позволяет сделать следующие выводы.

Доля поддержанных заявок в области компьютерной лингвистики сравнительно невелика (3,4 %).

Лингвистические характеристики, такие как длина заголовка и аннотации, варьируются достаточно сильно. Выявить какие-то закономерности от подобласти/отрасли не удалось.

Большее единообразие обнаружено в структуре заголовков: в 45 % заголовков встречаются неличные формы глагола и отглагольные существительные, часто соединенные союзом *и*; другой особенностью является использование абстрактной лексики (самое частотное слово *метод* встретилось в 48,6 % заголовков) в сочетании с отглагольными существительными.

Низкие индексы удобочитаемости текстов аннотаций и заголовков с одной стороны, ожидаемы, т. к. целевой аудиторией этих текстов являются эксперты в предметной области, с другой – говорят об отсутствии адекватных инструментов/способов выявления различий в сложности специальных текстов, которые попадают в категорию самых сложных текстов по всем индексам удобочитаемости, адаптированных для русского языка.

Ограничением проведенного исследования является невозможность провести лингвистический анализ неподдержанных заявок из-за отсутствия доступа к мета-данным таких заявок. Можно предположить, что роль экстра-лингвистических факторов при принятии решения о финансовой поддержке проекта из фонда РНФ очень велика, и, возможно, превалирует над лингвистическими.

Тем не менее мы считаем, что выявленные закономерности или их отсутствие могут стать определенным ориентиром для начинающих исследователей при подготовке их первых заявок на гранты и на конкурсы более низкого уровня, например, региональные, при оценке которых лингвистические факторы, характеризующие логичность, связность, структурированность материала и ясность изложения имеют более высокий вес при принятии решения конкурсной комиссией. (В Санкт-Петербурге таким авторитетным конкурсом является конкурс, проводимый Комитетом по науке и высшей школе при правительстве Санкт-Петербурга). Одной такой рекомендацией, безусловно, может стать рекомендация по более ответственному выделению ключевых слов/терминов в своей заявке и использованию их в названии проекта и аннотации. Соответствие ключевых слов структуре организации информации в заявке, их наличие в заголовке и аннотации способствует продвижению полученных научных результатов в условиях 4-й промышленной революции.

ЛИТЕРАТУРА

1. Боброва Н. А. Политика предоставления научных грантов в России и ее перспективы [Электронный ресурс] // Правоприменение. 2022. Т. 6, № 3. С. 262–268. DOI: [https://doi.org/10.52468/2542-1514.2022.6\(3\).262-268](https://doi.org/10.52468/2542-1514.2022.6(3).262-268).

2. Белявский О. В. Эффективность системы грантовой поддержки научных исследований [Электронный ресурс] // Пробелы в российском законодательстве. 2018. № 4. С. 395–399. URL: <https://cyberleninka.ru/article/n/effektivnost-sistemy-grantovoy-podderzhki-nauchnyh-issledovaniy> (дата обращения: 11.06.2024).

3. Эбзеева Ю. Н. Грантовая поддержка как фактор развития научных исследований в российских университетах [Электронный ресурс] // Вестник РУДН. Серия: Психология и педагогика. 2022. Т. 19, № 1. С. 146–157. DOI <https://doi.org/10.22363/2313-1683-2022-19-1-146-157>.

4. Федяева Н. Д., Мецевич И. В. Жанр «Заявка на грант»: становление жанровой нормы // Филологические науки. Вопросы теории и практики. 2017. Т. 78, № 12, Ч. 2. С. 163–166.

5. Беляева Л. Н., Чернявская В. Е. Научный и технический текст и Информация 4.0: ключевые задачи при создании структурированного контента [Электронный ресурс] // Научно-технические ведомости СПбГПУ. Гуманитарные и общественные науки. 2019. Т. 10, № 2. С. 53–63. DOI: <https://doi.org/10.18721/JHSS.10205>.

6. Armer T. Cambridge English for scientists. Cambridge : Cambridge University Press, 2014. 128 p.

7. Anthony L. AntConc (Version 4.3.0) [Computer Software] [Electronic resource]. Tokyo : Waseda University, 2024. URL: <https://www.laurenceanthony.net/software> (accessed: 10.05.2024).

9. Gómez P. C., Sánchez-Lafuente Á. A. Readability indices for the assessment of textbooks: a feasibility study in the context of EFL [Electronic resource] // Vigo Intern. J. of Appl. Linguistics (VIAL). 2019. Vol. 16. P. 31–52. DOI: <https://doi.org/10.35869/vial.v0i16.92>.

10. Рябцева Н. К. Название как доминантный компонент научного текста: русско-английские межъязыковые «несоответствия» [Электронный ресурс] // Вестник Волгоградского государственного университета. Серия 2, Языкознание. 2018. Т. 17, № 2. С. 33–43. DOI: <https://doi.org/10.15688/jvolsu2.2018.2.4>.