

Научная статья

УДК 81.33

doi: 10.17223/23062061/38/2

ЛИНГВИСТИЧЕСКАЯ СЛОЖНОСТЬ ТЕКСТОВ ЖАНРА «ВИРТУАЛЬНАЯ ЭКСКУРСИЯ ПО МУЗЕЮ» (НА МАТЕРИАЛЕ ВИРТУАЛЬНОГО ВИЗИТА В ГОСУДАРСТВЕННЫЙ ЭРМИТАЖ)

**Анастасия Владимировна Колмогорова¹,
Елизавета Романовна Куликова²,
Полина Алексеевна Колмогорова³**

^{1,2,3} *Национальный исследовательский университет
«Высшая школа экономики», Санкт-Петербург, Россия*

¹ *akolmogorova@hse.ru*

² *kulikova.e.r@hse.ru*

³ *kolmogorovapa@gmail.com*

Аннотация. Статья посвящена анализу метрик лингвистической сложности текстов Виртуального визита по Главному музейному комплексу Государственного Эрмитажа, размещенного на сайте музея. Цель исследования – определить, в какой мере лингвистические характеристики данных текстов коррелируют с признаками виртуальной экскурсии по музею как речевого жанра. В результате исследования сделан вывод о том, что базовые лингвистические метрики текстов Виртуального визита соответствуют большинству признаков жанра виртуальной экскурсии по музею, за исключением признака «аргументативность»; на уровне синтаксиса и морфологии данные тексты обнаруживают больше сходств с массмедийными текстами об искусстве, чем с научными искусствоведческими.

Ключевые слова: лингвистическая сложность текста, музей, научные искусствоведческие тексты, массмедийные тексты об искусстве, Виртуальный визит по Эрмитажу, речевой жанр, культурное наследие

Благодарности. Публикация подготовлена в результате проведения исследования по проекту № 24-00-033 «Экспериментальное изучение и моделирование когнитивных механизмов речевой деятельности» в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2025 г.

Для цитирования: Колмогорова А.В., Куликова Е.Р., Колмогорова П.А. Лингвистическая сложность текстов жанра «виртуальная экскурсия по музею» (на материале Виртуального визита в Государственный Эрмитаж) // Текст. Книга. Книгоиздание. 2025. № 38. С. 29–54. doi: 10.17223/23062061/38/2

Original article

THE LINGUISTIC COMPLEXITY OF TEXTS OF THE "VIRTUAL MUSEUM TOUR" GENRE (BASED ON THE MATERIAL OF A VIRTUAL VISIT TO THE STATE HERMITAGE MUSEUM)

Anastasia V. Kolmogorova¹, Elizaveta R. Kulikova²,
Polina A. Kolmogorova³

^{1, 2, 3} National Research University Higher School of Economics,
Saint Petersburg, Russian Federation

¹ akolmogorova@hse.ru

² kulikova.e.r@hse.ru

³ kolmogorovapa@gmail.com

Abstract. The article discusses the linguistic features of the texts of the virtual visit to the State Hermitage Museum, available on its official website. The aim of the study is to analyze the set of lexical, morphological, syntactic and discursive metrics of the linguistic complexity of these texts in comparison with the same metrics of scientific art history texts and mass media texts about art to answer the question of whether the level of complexity of the text content of a virtual excursion corresponds to the expectations that the frame of the speech genre "virtual excursion" forms – to be informative and, at the same time, attractive and uncomplicated. The research material consisted of 3 corpora: a corpus of virtual visit texts with a total volume of 169,223 tokens; a corpus of scientific articles on art history – 187,695 tokens, a corpus of mass media texts about art – 186,123 tokens. The research methodology included (1) description of the speech genre "virtual museum tour", its genre features, (2) calculation of the values of 132 metrics of linguistic complexity of texts from three subcorpora using a computer model, (3) identifying the specific features of "Hermitage" texts in comparison with the two other text subcorpora, (4) analysis of the correlation of specific features with the characteristics of the speech genre of virtual excursion. The analysis of text linguistic complexity metrics showed that the texts of the virtual visit to the Hermitage differ from both text corpora in comparison at the level of their basic metrics that are not closely related to one or another level of the language system: sentence length, indexes of lexical diversity, the presence of numeric symbols, high nominativity index, and low verballity index, which results in a number of grammatical text properties. The "Hermitage" texts approach mass media texts mainly at the level of morphology and syntax, and with scientific texts they only have in common a rather high value of the proportion of long words. Comparing the specifics identified at the level of linguistic complexity metrics with the characteristics of the "virtual excursion" speech genre, we determined that there is a correspondence between the metrics and such characteristics of the genre as descriptiveness, informativeness, linguistic simplicity, and non-banality of the communicative style. At the same

time, it was found that the texts practically do not contain verbal means contributing to the realization of such genre feature as "argumentativeness": the proportion of speech markers of persistence, deictic means of ensuring coherence of discourse, etc. is generally very low. The prospect of research is an experimental study of the difficulty of such texts for the reader's perception.

Keywords: linguistic complexity of text, museum, scientific art criticism, mass media texts about art, Virtual Tour of the Hermitage, speech genre, cultural heritage

Acknowledgments: The publication was prepared as a result of research on project No. 24-00-033 "Experimental study and modeling of cognitive mechanisms of speech activity" within the framework of the Program "Scientific Foundation of the National Research University Higher School of Economics (HSE University)" in 2025.

For citation: Kolmogorova, A.V., Kulikova, E.R. & Kolmogorova, P.A. (2025) The linguistic complexity of texts of the "virtual museum tour" genre (based on the material of a virtual visit to the State Hermitage Museum). *Tekst. Kniga. Knigoizdanie – Text. Book. Publishing.* 38. pp. 29–54. (In Russian). doi: 10.17223/23062061/38/2

Введение

Многие культурные институции сегодня ставят перед собой задачу популяризации искусства и повышения доступности культурного наследия для разных социальных групп. Одним из способов добиться подобного результата является цифровизация музеев и музейных экспозиций. Термин «клиенториентированность» стал уместным и в отношении музейных институций [1]: цифровизация музейных продуктов и услуг становится условием привлечения посетителей и обеспечения их комфортного и интересного пребывания в музее. Тот же вектор задает и государство. Так, в паспорте Национального проекта «Культура» [2] на период с 2018 по 2024 г. сформулирован целевой показатель «увеличение числа обращений к цифровым ресурсам в сфере культуры в 5 раз».

В этой связи актуальным становится вопрос о том, каких изменений в типичных практиках взаимодействия музея и посетителя потребует подобная трансформация. Ведь основная функция сотрудников музея, например, в сфере экскурсионно-образовательной деятельности – служить своеобразным медиатором между шедеврами мировой культуры и миром повседневности, из которого приходит обычный посетитель музейной экспозиции. Как отмечают С. Харрисон и П. Вуд, «устный или письменный процесс аналитического и вдумчивого рассуждения об искусстве можно сравнивать с надеванием одежды на привидение с той целью, чтобы другие могли его увидеть» [3. С. 1].

Если в классическом музее фигура такого медиатора вполне реальна – экскурсовод рассказывает харизматично, увлекательно, владея своим голосом, мимикой, речью, ему можно задавать вопросы – то в случае цифровой среды эстафету аттрактивности и адаптивности должен принять текст, например, на сайте музея, в приложении и т. д. Аттрактивность и несложность для восприятия и понимания становятся тем более важными чертами цифровой текстовой музейной среды в силу того, что пользователи привыкли потреблять именно такой цифровой контент. В музейной же практике каноническим текстовым жанром является так называемый искусствоведческий текст, сочетающий черты письменности, научности, включающий терминологию и сложную критическую рефлексия. Возникает некоторая коллизия: вливаясь в категорию цифрового контента и пытаясь сохранить свою медиативную функцию, текстовое сопровождение «новых» музейных практик должно отвечать требованиям лингвистической «несложности», однако оно зачастую попадает под инерционное влияние складывавшейся годами сложной стилистики научных искусствоведческих текстов.

В данной статье мы помещаем в зону нашего исследовательского интереса тексты Виртуального визита в Государственный Эрмитаж, размещенного на официальном сайте музея. Они будут рассмотрены на фоне двух жанровых полюсов своеобразной «шкалы» искусствоведческих текстов: научных искусствоведческих текстов и массмедийных текстов об искусстве. Цель исследования – проанализировав совокупность лексических, морфологических, синтаксических и дискурсивных метрик лингвистической сложности данных текстов в сравнении с метриками научных искусствоведческих текстов и массмедийных текстов об искусстве, ответить на вопрос о том, соответствует ли уровень сложности текстового контента виртуальной экскурсии тем ожиданиям, которые формирует рамка речевого жанра «виртуальная экскурсия» – быть информативным и, одновременно, привлекательным и несложным.

Понятие лингвистической сложности текста

Измерение сложности текста представляет собой междисциплинарную проблему как с точки зрения используемых методов (экспертный лингвистический анализ, статистические и компьютерные методы), так и с точки зрения применения полученных результатов. Исследования в данной области зачастую являются практико-ориентированными: запрос на измерение сложности текстов существует в юриспруденции [4], лингводидактике [5], медицине [6] и других сферах.

Стоит отметить, что ввиду многообразия подходов к данному вопросу существует неоднозначность в использовании трех понятий – трудность, сложность и читабельность текста. С опорой на мнения ряда отечественных исследователей [7, 8 и др.] дифференцировать эти понятия можно следующим образом.

Сложность текста рассматривается как объективная величина, которая может быть измерена на основе специально подобранных дескриптивных параметров. Такие параметры многочисленны. Наиболее базовыми из них являются количество слов в тексте, средняя длина предложения в словах, средняя длина слова в буквах и аналогичные, исключительно количественные критерии, которые были предложены на ранних этапах разработки проблемы. В ходе исследований на материале текстов разных стилей и жанров набор метрик был значительно расширен: в качестве предикторов сложности стали рассматриваться частотность используемой лексики [9], лексическая плотность [10], соотношение слов с конкретным и абстрактным значением, количество специальных терминов [11] и многие другие критерии. При оценке сложности часто прибегают к использованию формул читабельности (удобочитаемости). Это математические формулы, отражающие линейную зависимость сложности текста от его количественных характеристик. Значения индексов читабельности, как правило, интерпретируются в соответствии с возрастом читателя или количеством лет образования, необходимым для беспрепятственного чтения текста. Такие формулы языкоспецифичны: многие были созданы на материале английского языка, а затем адаптированы для других языков. Так, например, для русского языка формула Флеша–Кинкейда была скорректирована в исследовании И.В. Оборновой [12].

Таким образом, оценка сложности основывается на чисто лингвистическом подходе, при этом из поля нашего внимания исключается фактор языковой личности читателя как субъекта восприятия текста. Однако, как отмечает М.И. Солнышкина, «только включение читателя, а также процессов восприятия и понимания в объект исследования позволяет осуществить комплексную и всеобъемлющую оценку сложности текста» [13. С. 20]. В тех случаях, когда анализ осуществляется с учетом фоновых знаний, языковых навыков и когнитивных особенностей читателя или группы читателей, говорят об оценке трудности текста как более субъективной, чем сложность, категории.

Безусловно, данные два понятия взаимосвязаны: трудность текстового материала во многом определяется именно его сложностью, выраженной в показателях объективных параметров. Однако предопределять

трудность текста для разных групп читателей будут разные объективные текстовые характеристики. Так, для одной группы важным критерием сложности окажется терминологическая плотность, а для другой – длина предложения. В связи с этим измерение сложности и трудности текстов конкретного типа (учебных, юридических, медицинских, искусствоведческих) можно рассматривать как последовательные этапы: сначала выявляются общелингвистические, объективные предикторы сложности, которые затем уточняются в контексте триады «читатель–чтение–текст». С учетом такого подхода к обсуждаемым понятиям, данная работа посвящена изучению лингвистической сложности текстов, сопровождающих Виртуальный визит в Главный музейный комплекс Государственного Эрмитажа, с точки зрения соответствия метрик сложности характеристикам виртуальной экскурсии как жанра.

Жанровая специфика текстов Виртуального визита в Государственный Эрмитаж

Виртуальная экскурсия – это относительно новый жанр взаимодействия музея и посетителя. Л.Ю. Щипицина так определяет суть жанра: это любой вид представления экскурсионного объекта, опосредованный цифровыми устройствами и средствами связи и имеющий в качестве прагматической цели знакомство аудитории с новым для нее объектом показа, а также характеризующийся физической разобщенностью участников экскурсии, которые находятся на удалении от показываемого объекта и не могут воспринимать его собственными органами чувств [14. С. 236–238]. Тем не менее, виртуальные туры, несмотря на опосредованность восприятия цифровыми технологиями, позволяют посетителям музеев одновременно сформировать в своем когнитивном опыте полноценные симметричные когнитивный и аффективный образы культурных объектов, что не всегда достижимо в ситуации реального визита [15, 16]. В ходе виртуального визита посетитель «погружается» в новые локации, может, не торопясь и не стесняясь, обойти все интересные экспонаты, рассмотреть их, прочитать информацию о них, если нужно, то несколько раз. В ситуации реального визита данные эвристические действия (рассмотреть, узнать новое, вспомнить известную информацию, осмыслить полученное впечатление) могут быть затруднены большим скоплением людей, нехваткой времени, присутствием зрителей и т. д.

Совершить виртуальное путешествие по Государственному Эрмитажу посетитель официального сайта музея может, кликнув на вкладку

«Виртуальный визит» (<https://pano.hermitagemuseum.org/3d/html/pwoa/main>). Пользователь попадает на страницу, где представлены иконки для каждого составляющего Эрмитаж музейного комплекса, его галерей и выставочных центров, а также проектов – всего 6 зданий и 300 локаций. Кликнув на соответствующую иконку, виртуальный посетитель попадает на страницу, где представлен план музейного помещения, на котором обозначены основные его элементы: залы, переходы, лестницы. Нажав на название элемента музейного комплекса, посетитель получает на экране 3D-изображение помещения. При помощи специальных кнопок можно увеличивать и удалять изображение, двигаться по нему в том или ином направлении, а при нажатии значка «Информация» появляется всплывающее окно, которое содержит текст-описание помещения (рис. 1) или краткую характеристику экспоната (рис. 2). Материал нашего исследования составили только тексты первого типа.

Таким образом, если применить к данному типу текстовой продукции известную «анкету» речевого жанра Т.В. Шмелевой [17], то получается нижеследующая характеристика жанра «Виртуальный визит в музей»:

Коммуникативная цель – создание у посетителя привлекательного аффективного и насыщенного когнитивного образа объекта культурного наследия (экспозиции, музея).

Образ автора – специалист в области искусствоведения, выступающий в роли медиатора между высокой эстетикой культурного наследия и посетителем, имеющим, как правило, достаточно поверхностные, базовые знания об искусстве.

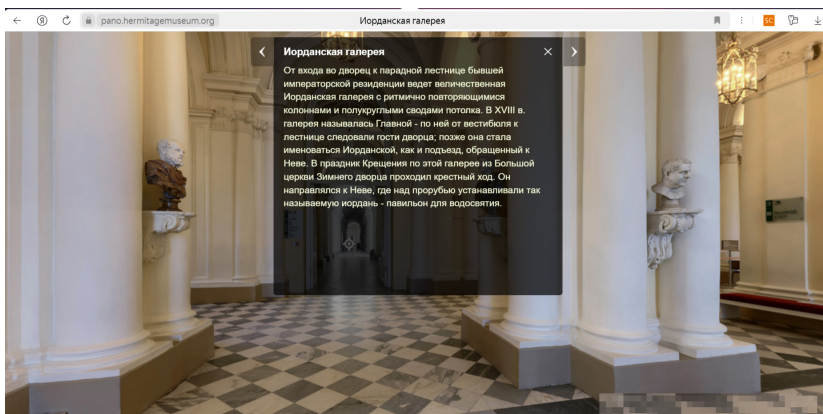


Рис. 1. Скриншот экрана Виртуального визита по Государственному Эрмитажу с всплывающим окном – текстовым описанием музейной локации



Рис. 2. Скриншот экрана Виртуального визита по Государственному Эрмитажу с всплывающим окном – текстовым описанием музейного экспоната

Образ адресата – человек, не имеющий специальных искусствоведческих знаний, погруженный в собственную повседневность, но заинтересованный данным музейным комплексом и имеющий некоторое, как правило, небольшое, количество свободного времени, доступ к электронному устройству (компьютеру или смартфону) и сети Интернет.

Образ будущего складывается из того, какое решение примет виртуальный посетитель музея – купить билет и прийти в локации в реальном времени и пространстве или не делать этого.

Образ прошлого формируется под влиянием двух факторов: опыт посещения музейных объектов в реальной жизни посетителя и опыт его взаимодействия с цифровой средой в целом, в том числе характер цифрового контента, ставшего привычным: социальные сети, рекомендательные платформы, маркетплейсы, компьютерные игры.

Диктумное содержание жанра очерчивается схемой описания того, что видит посетитель, находясь в локации.

Отметим, что указанные выше характеристики жанра сформулированы нами как результат обсуждений с коллегами из Государственного Эрмитажа, в частности, работающих с Цифровой коллекцией музея, относительно того, какими они видят цель виртуальной экскурсии, ее потенциального посетителя и других условий бытования жанра. Иными словами, это своего рода проект жанровых ожиданий музейных работ-

ников в том числе. При этом проверить, насколько текстовая реальность близка к данным ожиданиям, было поручено нам, лингвистам.

Исходя из вышеуказанных жанровых характеристик, мы можем сформулировать некоторые ожидания: какими лингвистическими характеристиками должен обладать текст, обрамленный в такую жанровую рамку. Перечислим их:

- дескриптивность, наличие дейктиков и других способов реализации категории референтности (диктумное содержание – описание локации);

- информативность (коммуникативная цель – создать насыщенный когнитивный образ культурного объекта);

- черты скрытого аргументатива: сравнительные конструкции, лексические единицы с семантикой оценочности, средства выражения модальных значений, маркеры персуазивности (образ будущего – создаются условия для прихода посетителя в музей; образ адресата – человек, заинтересовавшийся искусством, мотивацию которого нужно поддерживать);

- лингвистическая несложность: недлинные предложения и слова, отсутствие сложных синтаксических конструкций и перечислений (образ прошлого: посетитель – опытный потребитель цифрового контента);

- небанальность коммуникативного стиля, его некоторая «эксклюзивность»: использование редких лексических единиц, отсутствие разговорной лексики и сленгизмов (образ адресанта – медиатор между областью высокого искусства и повседневностью);

- эмоциональность: наличие эмотивов, прецедентных феноменов, средств создания образности (коммуникативная цель – создать аффективно-насыщенный образ музейного объекта).

Мы постараемся проверить, соответствует ли корпус текстов, взятых со страницы Виртуального визита, данным лингвистическим ожиданиям, обусловленным их жанром. Для этого мы посчитаем значения 132 лингвистических метрик, используемых для оценки лингвистической сложности текста, в трех подкорпусах: текстах со страницы Виртуального визита, научных искусствоведческих текстах, массмедийных текстах об искусстве. Сопоставление значений метрик поможет нам лучше увидеть специфику текстов виртуальной экскурсии и сделать выводы о том, насколько они соответствуют с точки зрения выбора языковых средств поставленной жанровой задаче.

Материал и методы

Материал исследования составили три корпуса. Первый корпус, анализ которого является основной зоной нашего интереса, – это собранная нами коллекция текстов, сопровождающих Виртуальный визит по Главному музейному комплексу Государственного Эрмитажа [18]. Корпус включает 886 текстов общим объемом 169,223 токена. Два других корпуса используются для сравнительно-сопоставительного анализа: это корпус научных статей по искусствоведению и корпус массмедийных текстов об искусстве. Научный корпус сформирован нами из находящихся в открытом доступе 50 статей из разделов «Теория и история искусства», «Виды искусства», «Декоративно-прикладное искусство» из журналов «Манускрипт» и «Культура и искусство». Объем корпуса – 187,695 токенов. Набор массмедийных текстов сформирован из текстов, опубликованных на сайте газеты «Новости искусства» в рубриках «Новости», «Выставки» и «Музей будущего». Объем корпуса – 186,123 токена.

В выборе методики оценки сложности мы опираемся на исследование О.В. Блиновой и Н.А. Тарасова, проведенное на материале юридических текстов [19]. В наших текстах из трех подкорпусов оценивались 132 лингвистические метрики: их значения извлекались автоматически при помощи компьютерной модели, предложенной упомянутыми выше авторами [20], затем доработанной А.В. Марголиной [21]. В код модели нами были внесены дополнительные изменения, обусловленные объектом нашего исследования. Поскольку исходная модель О.В. Блиновой и Н.А. Тарасова была разработана для текстов юридического характера, в ней заложено измерение доли юридических терминов, что нерелевантно для нашей задачи, поэтому для оценки терминологической плотности текстов взамен словаря юридических терминов нами был использован словарь искусствоведческих терминов Российской академии художеств [22]. Таким образом, тексты каждого из трех подкорпусов подавались на вход модели, которая выдавала значения метрик.

Все оцениваемые параметры можно разделить на несколько групп (ниже приводим список групп с несколькими примерами):

- 1) базовые метрики: количество лемм, словоформ, букв, средняя длина предложений в словах;
- 2) формулы читабельности: адаптированный для русского языка индекс Флеша–Кинкейда, адаптированный SMOG;

3) частеречные показатели: доля глаголов, существительных, функциональных слов;

4) биграммные и триграммные метрики: доля биграмм, включающих различные комбинации частей речи;

5) частотность слов: доля слов, входящих в каждую из 9 групп лексем, разделенных на ранги на основе закона Ципфа;

6) морфологические параметры: доля словоформ в родительном падеже, доля существительных среднего рода;

7) лексические и семантические параметры: доля архаичных слов и выражений, доля абстрактных лемм;

8) синтаксические метрики: доля аппозитивных конструкций, доля составных (неоднословных) выражений;

9) метрики связности: количество повторов существительных в соседних предложениях, количество повторов граммем времени и вида у глаголов в личной форме (в соседних предложениях).

Для анализа полученных значений метрик использовался сравнительно-сопоставительный метод.

Результаты

После получения значений всех метрик для каждого текста в каждом корпусе мы нормализовали абсолютные величины для тех метрик, которые не предусматривают нормализацию в рамках своей математической формулы, для того чтобы учесть разницу в количестве токенов в каждом из корпусов, и затем получили средние значения для всех метрик.

Специфические черты текстов Виртуального визита в Государственный Эрмитаж

Рассмотрим группу базовых метрик. В коллекции «эрмитажных» текстов отмечено наибольшее количество числовых символов (метрика N на рис. 3). Это связано с тем, что тексты, сопровождающие Виртуальный визит по Эрмитажу, знакомя посетителя с музейным объектом, фиксируют его внимание на датах создания объекта и годах жизни его автора, например:

1) *В витринах у окна представлены французский фаянс второй половины XVIII века в стиле рококо и предметы из десертного сервиза со сценами по гравюрам Жана Батиста Пиллемана (1728–1808)* (Виртуальный визит в Государственный Эрмитаж: Главный музейный комплекс).

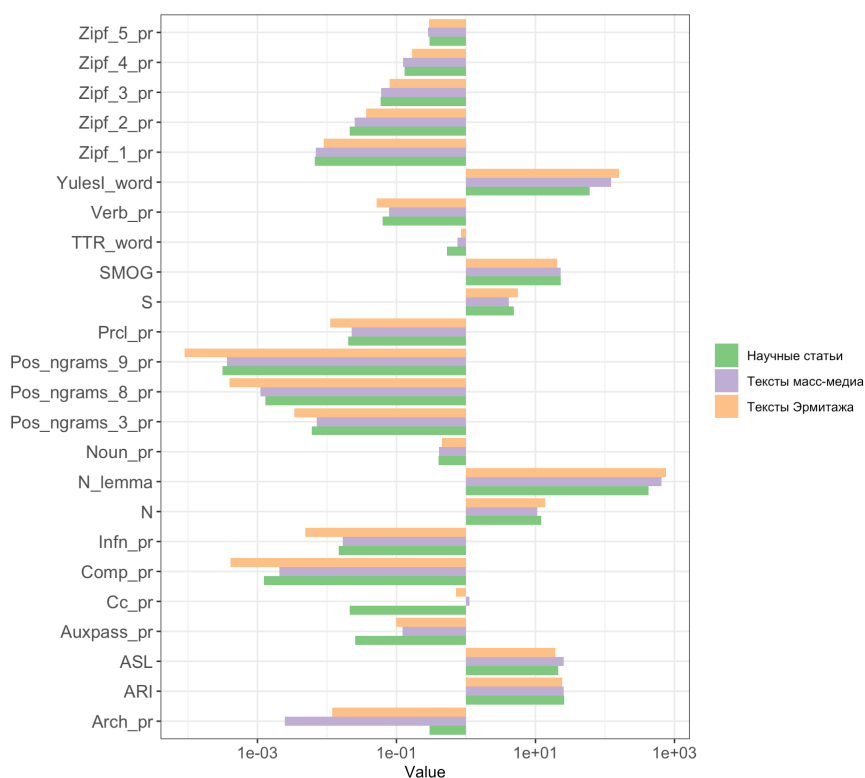


Рис. 3. Логарифмированная шкала значений метрик лингвистической сложности, специфичных для текстов Виртуального визита в Государственный Эрмитаж на фоне научных и искусствоведческих текстов

В приведенном примере (пр. 1) мы видим и порядковые, и количественные числительные, записанные как буквами, так и римскими и арабскими цифрами.

Тексты сайта Эрмитажа оказались самыми разнообразными с точки зрения используемого вокабуляра: если наибольшее количество словоформ имеют научные искусствоведческие тексты, то наибольшее количество лемм – тексты Виртуального визита (N_lemma : 750 против 420 у научных текстов, рис. 3), т. е. в «эрмитажных» текстах используется больше разных слов, чем словоформ одного и того же слова. Данное наблюдение коррелирует со значениями других метрик (см. ниже).

Что касается метрик удобочитаемости, то отдельного внимания заслуживает индекс Юла. Yule's I (Индекс Юла Ай) – это статистический показатель, который измеряет долю уникальных слов среди всех слов текста. Индекс Юла часто используется в анализе текстов для оценки лексического разнообразия [23]. Показатель этого индекса (YulesI_word, рис. 3) у «эрмитажных» текстов значительно превышает аналогичные показатели для словоформ и лемм у фоновых текстов: 158,6 (словоформы) / 82,4 (леммы) – у «эрмитажных»; 60,3 (словоформы) / 18,3 (леммы) – у «научных»; 123,3 (словоформы) / 56,8 (леммы) – у массмедийных. Подтверждает данную тенденцию и коэффициент лексического разнообразия TTR (Type Token Ratio) (рис. 3), представляющий собой отношение количества уникальных слов к общему количеству слов в тексте [24]: для «эрмитажных» текстов его значение, например, для словоформ составляет 0,85, в то время как для научных – 0,52, а для массмедийных – 0,74.

Данные метрики позволяют нам сделать вывод о том, что текст виртуальной экскурсии менее клиширован, чем текст массмедийный или научный, поскольку включает в себя номинации разнообразных предметов и характеристик коллекции, имена их авторов, названия используемых техник и материалов, локаций, архитектурных деталей. Ср. в нижеследующем предложении все леммы уникальны (пр. 2):

2) От входа во дворец к парадной лестнице бывшей императорской резиденции ведет величественная Иорданская галерея с ритмично повторяющимися колоннами и полукруглыми сводами потолка (Виртуальный визит в Государственный Эрмитаж: Главный музейный комплекс).

В метриках, предопределяющих удобочитаемость, например ASL (Average Sentence Length in Words) и S (Number of sentences per 100 word forms), хорошо заметно (рис. 3), что у текстов Виртуального визита при наибольшем значении количества предложений на 100 слов – 5 на фоне 4 в массмедийном и научном подкорпусах – наблюдается наименьшая длина предложения в словах (19, на фоне – 21 и 25 слов в фоновых подкорпусах). Иными словами, в целом в «эрмитажных» текстах используются более короткие предложения. Кроме того, тексты Виртуального визита имеют самые низкие из трех подкорпусов значения метрик удобочитаемости ARI и SMOG (рис. 3) – 20 на фоне 23 и 24 в других подкорпусах, что говорит о том, что они менее сложны для чтения.

Говоря о частеречных признаках, стоит отметить, что у «эрмитажных» текстов самые высокие значения индекса номинативности

(Noun_pr), пропорции полных причастий (Prtf_pr) и самые низкие значения индекса вербальности (Vrb_pr), пропорции инфинитивов (Inf_pr) (табл. 1). Представляется, что это связано с диктумной частью данного жанрового типа текстов – описание локаций и объектов, которое не требует частого использования глаголов.

Низкие же значения пропорций местоимений (Npro_pr), сравнительных конструкций (Comp_pr) и частиц (Prcl_pr) (табл. 1) обусловлены, по-видимому, слабой выраженностью в текстах виртуальной экскурсии категории персуазивности. Этим же фактом объясняется и то, что таким текстам в наименьшей степени свойственна модальность деонтической возможности и необходимости (Deont_pr): 0,004 на фоне 0,006 и 0,005 у научных и газетных текстов соответственно.

При анализе биграмм было обнаружено, что в текстах сайта Эрмитажа в наименьшей степени представлены биграммы типа «наречие + глагол» (ADVB + VERB – 0,0034), в то время как для публицистических и научных статей показатели данной метрики составляют 0,0060 и 0,0070 соответственно, а также биграмм типа «деепричастие + существительное» (GRND + NOUN) и «наречие + деепричастие» (ADVB + GRND). Данные метрики подтверждают наблюдения о том, что глагольность в целом мало выражена в текстах виртуальных экскурсий.

Нами отмечена интересная особенность текстов виртуальных экскурсий по Эрмитажу: значение метрики «пропорция слов текстового дейксиса, обеспечивающих связность», для текстов Эрмитажа – 0,0007 на фоне 0,141 и 0,001 для научных и газетных текстов соответственно. Аналогично и для метрики «доля вхождений элементов, служащих для сегментации дискурса и обеспечения связности»: для «эрмитажных» текстов она составляет 0,003 на фоне научных – 0,14 и массмедийных – 0,017.

Таблица 1

Значения частеречных метрик лингвистической сложности

Подкорпус/ метрика	Vrb_pr	Noun_pr	Prtf_pr	Npro_pr	Comp_pr	Inf_pr	Prcl_pr
Виртуальный визит	0,051	0,453	0,028	0,013	0,0004	0,004	0,011
Научные тексты	0,063	0,400	0,022	0,021	0,0012	0,014	0,020
Массмедиа	0,078	0,406	0,016	0,028	0,0020	0,016	0,022

Как мы увидим далее, отсутствие дейктиков в текстах виртуальной экскурсии компенсируется таким средством лексической связности, как повторы лексем в рамках сверхфразового единства, что сближает «эрмитажные» тексты и массмедийные тексты. Отсутствие дейктиков, возможно, объясняется и наличием информативного визуального компонента в текстах Виртуального визита.

Обратим внимание на характеристики, связанные с частотностью слов: для текстов оценивалась доля слов, входящих в один из девяти рангов, выделенных на основе закона Ципфа (где 0 – самые редкие, 8 – самые частотные слова языка). Так, значения пропорции слов низкой и средней частотности Zipf_1–Zipf_5 (рис. 3) для «эрмитажных» текстов оказались более высокими, чем в других подкорпусах, даже в научном. Представляется, что данная характеристика связана с уникальностью экспонатов коллекции, описание которых предполагает употребление редких лексем, номинирующих их детали, элементы.

Тексты Виртуального визита ожидаемо содержат наибольшую долю терминов из области искусствоведения: 0,249 на фоне 0,052 для научных и 0,196 для массмедийных текстов. Например, в нижеследующем отрывке (пр. 3) используются такие термины, как *десюдепорт* (декоративная композиция в интерьере, расположенная над дверным проёмом), *медальон*:

3) *В двери цветного дерева, над которыми располагаются живописные десюдепорты, вмонтированы расписные фарфоровые медальоны, в мраморные каминные вставлены мозаичные композиции* (Виртуальный визит в Государственный Эрмитаж: Главный музейный комплекс).

Что касается синтаксических метрик, то для «эрмитажных» текстов характерны самые высокие значения метрики «доля вхождений пассивного подлежащего главной или зависимой клаузы»: 0,438 на фоне 0,0256 для научных и 0,284 для массмедийных текстов. Подобная специфика текстов обусловлена тем, что экскурсии Эрмитажа строятся на описании непосредственно «от музейного объекта»: где он расположен, кем был создан, чем украшен (пр. 4):

4) У окна **расположен** беломраморный камин. На каминной полке – бронзовые часы с группой «Амур и Психея» (Санкт-Петербург, фабрика Феликса Шопена, XIX). Стол перед камином **создан** во Флоренции в 1873. Его подстолье **украшено** бронзовыми барельефами, столешница с изображением колесницы бога Аполлона и атрибутов девяти Муз **выполнена** в технике флорентийской мозаики. Две мозаичные картины «Храм Сивиллы в Тиволи и водопад» и «Грот в Тиволи» **исполнены** мас-

тером Дж. Раффаэлли в 1817 в технике римской мозаики (Виртуальный визит в Государственный Эрмитаж: Главный музейный комплекс).

Характеристики «эрмитажных» текстов, приближающие их к массмедийным текстам об искусстве

Остановимся на некоторых характеристиках, приближающих тексты Виртуального визита в Эрмитаж к массмедийным текстам. Более подробно схожие метрики будут представлены в табл. 2.

В целом массмедийные и тексты с сайта Эрмитажа близки друг к другу по большинству синтаксических метрик: оба корпуса не отягощены составными (неоднословными) выражениями (Compound_pr, рис. 4) – их доля имеет значение 0,00051 и 0,00012 соответственно; в то время как для научных текстов этот параметр более чем в 1000 раз больше – 0,1318. В обоих случаях также имеем долю клауз с элементами, трактуемыми как связочные (Cor_pr, рис. 4), ниже чем 0,08, в то время как та же метрика у научных текстов была оценена моделью в 0,2.

Паратаксис и парафраз измеряются метрикой Parataxis_pr. Для научных текстов эта метрика оказалась значительно ниже (0,0596), чем для «эрмитажных» и публицистических текстов (0,3913 и 0,512 соответственно).

Среди синтаксических метрик обращает на себя внимание метрика «доля союзов, связанных с конъюнктами синтаксическим отношением "сс" (координация)» (Cc_pr, рис. 3): наибольшие значения метрика демонстрирует в текстах массмедийного подкорпуса – 1,13, «эрмитажные» тексты имеют близкое значение – 0,7208, в то время как для научных текстов сочинительная связь оказалась нехарактерна – 0,0211. Полагаем, что подобное распределение значений метрик связано с большей дескриптивностью массмедийных и «эрмитажных» текстов по сравнению с научными, поскольку у последних преобладает аргументативная составляющая.

В обоих типах текстов преобладают словоформы дательного (datv, рис. 4) и именительного падежа (nomn, рис. 4), которые призваны отобразить типичную ситуацию социального взаимодействия, где есть соответственно бенефициар и агент (напр. *император заказал эту картину известному художнику*), а также глаголы несовершенного вида (Impf_pr, рис. 4). Последнее наблюдение довольно неожиданно, но обращение непосредственно к текстовым данным показало, что в массмедийных материалах об искусстве нередко подчеркивается ритуаль-

ность определенных художественных практик (*художник всегда ставил на своих картинах...*), а в «эрмитажных» текстах – типичность использованных техник изготовления чего-либо или способов использования интерьеров (*для балов лестницы украшали цветами...*). Одновременно глаголы в форме будущего времени (Futr_pr, рис. 4) совершенно нехарактерны – «эрмитажные» тексты описывают музейную коллекцию и ее историю, а тексты медиа рассказывают о недавних событиях в мире культуры и новых именах. В текстах обоих жанров относительно редко используются формы настоящего времени (в отличие от текстов научных): доля данных глагольных форм (Pres_pr, рис. 4) составляет 0,03 в обоих подкорпусах; для сравнения в научных текстах по искусствоведческой проблематике – 0,14.

Интересно, что и в публицистических, и в «эрмитажных» текстах велика доля абстрактных лемм (0,04). Так, абстрактные леммы включают в себя слова, которые обозначают идеи, концепции, чувства, состояния, названия категорий объектов, их типов и жанров. Сравним два предложения – из текста виртуальной экскурсии и из статьи электронных медиа:

5) *В начале XVII века в Италии возникло и получило **развитие направление**, известное как болонский **академизм*** (Виртуальный визит в Государственный Эрмитаж: Главный музейный комплекс);

6) *Они были **модернистами по самоощущению** и работали вместе, «чтобы создать новый художественный **текст, язык со своим специфическим грамматическим синтаксисом**»* (Новости искусства, «Премьера авангарда из Узбекистана на мировой сцене проходит во Флоренции и Венеции»).

Полужирным шрифтом (пр. 5, 6) выделены лексические единицы абстрактной семантики: названия типов объектов искусства, направлений в искусстве, эпох и традиций. Как видим, фрагмент публицистического текста также изобилует лексемами абстрактного характера, обозначающими различные временные, модальные, категориальные отношения, типы явлений, действий и решений.

«Эрмитажные» и массмедийные тексты также близки по метрике, оценивающей количество повторов существительных в соседних предложениях (0,3913 и 0,3826 соответственно; Cohes_1, рис. 4). Иными словами, лексическая связность в текстах Эрмитажа выше, чем в научных искусствоведческих текстах. В нижеследующем фрагменте хорошо прослеживаются лексические повторы (пр. 7: *зал, концертный зал, ярус, колонны, ярус*).

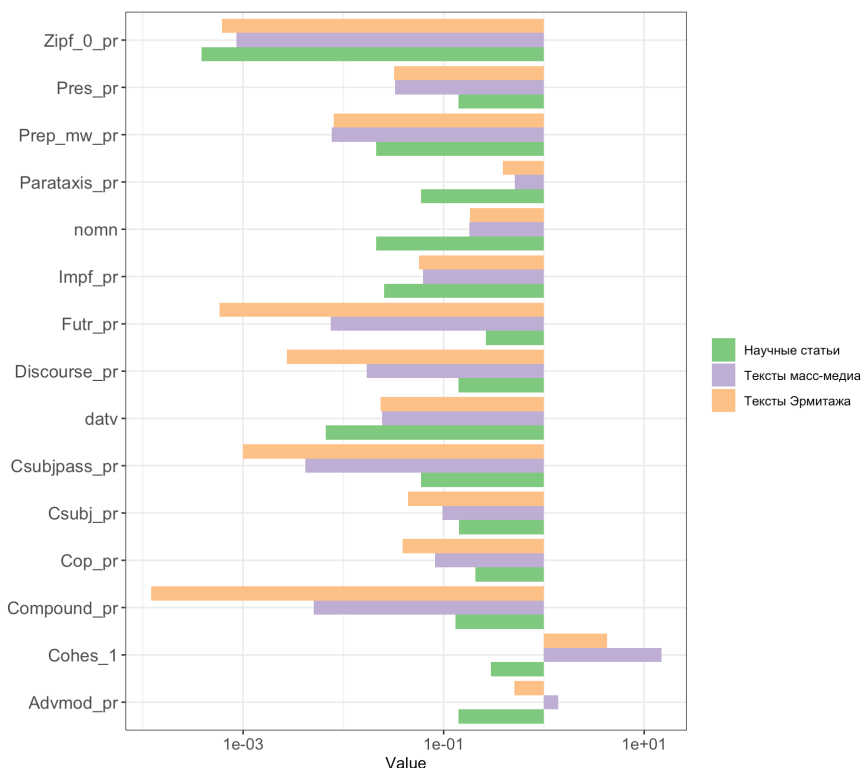


Рис. 4. Логарифмированная шкала значений метрик лингвистической сложности, значения которых близки у «эрмитажных» и массмедийных текстов

7) Ансамбль Невской анфилады замыкает **Концертный зал**. Он был создан на месте растреллиевского Аванзала. После пожара П. Стасов восстановил это помещение почти без изменений. Антаблемент с сильно вынесенным карнизом делит **зал** по высоте на два **яруса**. Стены первого **яруса** украшают трехчетвертные **колонны** коринфского ордера, увенчанные статуями муз. Живопись на падугах и потолке, имитирующая лепку, с изображениями атрибутов искусства и аллегорических фигур, близка по тематике скульптурному оформлению **зала**. Эффект убранства интерьера строится на контрасте фактур белого цвета – матовых стен и полированных **колонн**, выполненных из искусственного мрамора. **Концертный зал** предназначался для придворных **концертов**

и балов (Виртуальный визит в Государственный Эрмитаж: Главный музейный комплекс).

Тексты Виртуального визита в Эрмитаж приближаются к массмедийным статьям и по метрике «доля длинных слов» (4 и более слога): в «эрмитажных» текстах и массмедийных статьях доля таких слов несколько меньше, чем в научных – 43,5 и 40,1 соответственно, против 48,9.

Характеристики «эрмитажных» текстов, схожие с характеристиками научных текстов по искусствоведению

Единственной метрикой, которая некоторым образом сближает «эрмитажные» тексты с научными оказалась метрика, фиксирующая долю архаичных слов: 0,297 – у научных текстов, 0,012 – у «эрмитажных» и 0,002 – у массмедийных. Вторая позиция в данном импровизированном рейтинге у «эрмитажных» текстов вполне ожидаема, поскольку в описаниях интерьеров появляется немало названий объектов, которые не встречаются более в рутинных практиках жизни домохозяйств и общества в целом (например, *складень, трон* и т. д.).

Обсуждение

В табл. 2 мы постарались систематизировать: 1) лингвистические метрики, составляющие специфику текстов Виртуального визита по Государственному Эрмитажу, т. е. отличающие их от научных искусствоведческих и массмедийных в области искусства одновременно; 2) метрики, близкие у «эрмитажных» и массмедийных текстов журнала «Новости искусства» (<https://www.theartnewspaper.ru/>); 3) метрики, близкие у «эрмитажных» и научных искусствоведческих текстов.

Анализируя табл. 2, мы видим, что у текстов с сайта Государственного Эрмитажа определенно есть свои особенности, которые проявляются преимущественно на уровне неспецифичных метрик, т. е. не связанных тесно с тем или иным уровнем языковой системы. Это длина предложения, индексы лексического разнообразия, присутствие числовых символов, общая тенденция к номинативности, а не вербальности, следствием которой становится ряд грамматических метрик. «Эрмитажные» тексты сближаются с массмедийными в основном на уровне морфологии и синтаксиса. Сходство «эрмитажных» текстов с научными фрагментарно – общим является только довольно высокое значение доли архаичных слов.

Таблица 2

**Результаты анализа метрик лингвистической сложности текстов
Виртуального визита в Эрмитаж на фоне двух других подкорпусов**

Метрики	Категория метрик
Метрики, имеющие значения, специфические для текстов на сайте Эрмитажа	
Количество числовых символов, N_lemma, Yule's I, TTR	Базовые метрики
ASL, S, ARI, SMOG	Метрики удобочитаемости
Пропорция инфинитивов, Пропорция частиц, Пропорция сравнительных форм, Индекс вербальности, Индекс номинативности, Пропорция полных причастий, Пропорция местоимений	Частеречные метрики
Доля искусствоведческих терминов	Лексические и семантические метрики
Пропорция ADVB + VERB, GRND + NOUN, ADVB + GRND	Метрики биграмм и триграмм
Цифф 1-5	Частотность слов
Доля вхождений пассивного подлежащего главной или зависимой клаузы	Синтаксические метрики
Метрики, имеющие значения, приближающие «эрмитажный» текст к массмедийному	
Длина слова	Базовые метрики
Цифф 0 (самые редкие слова)	Частотность слов
Доля словоформ в дательном падеже, в именительном падеже, глаголов несовершенного вида, в форме будущего времени, в форме настоящего времени	Морфологические метрики (словообразовательные и граммемные)
Доля неоднословных предлогов	Лексические и семантические метрики

Метрики	Категория метрик
Доля наречных модификаторов предиката (наречий или наречных групп), составных (неоднословных) выражений, клауз с элементами, трактуемыми как связочные, конструкций «с сентенциальным субъектом», пассивных конструкций «с сентенциальным субъектом», вхождений элементов, служащих для сегментации дискурса и обеспечения связности, элементов, связанных паратактическим отношением с другими элементами (discourse-like equivalent of coordination), а также парафраз; доля союзов, связанных с конъюнктами синтаксическим отношением «с» (координация)	Синтаксические метрики
Количество повторов существительных в соседних предложениях	Метрики связности
Метрики, имеющие значения, приближающие «эрмитажный» текст к научному	
Доля архаичных слов и выражений (Arch_pr)	Лексические и семантические метрики

Если мы вернемся к жанровым чертам текстов виртуальной экскурсии по музею, сформулированным нами *ad hoc* в самом начале, то увидим, что проанализированные нами лингвистические метрики данных текстов коррелируют с такими чертами, как дескриптивность (средства лексической когезии, частотность сочинительной синтаксической связи, тяготение к пассиву), информативность (частотность числовых символов), лингвистическая несложность (тенденция к использованию недлинных предложений, достаточно низкие значения индексов удобочитаемости, но при этом далеко не базовый вокабуляр), небанальность коммуникативного стиля (высокие индексы лексического разнообразия, наличие архаизмов, редких и низкочастотных слов).

Однако мы не обнаружили среди метрик лингвистических коррелятов такой предполагаемой жанровой черты виртуальной экскурсии по сайту музея, как скрытая аргументативность: для данных текстов оказалось характерно только незначительное количество сравнительных конструкций, средств выражения деонтической модальности, а также местоимений. Эмоциональность как жанровую черту оценить пока не удалось, поскольку среди метрик нет ни доли эмоциональных слов, ни

эмотивов. Однако на следующем этапе мы планируем провести сентимент-анализ подобных текстов.

Выводы

Проведенный анализ текстов Виртуального визита на официальном сайте Государственного Эрмитажа позволяет заключить, что их лингвистические характеристики в целом соответствуют большинству признаков (черт) жанра «виртуальная экскурсия по музею». При этом хорошо заметно, что, например, по морфологическим и синтаксическим характеристикам тексты виртуальной экскурсии ближе к массмедийным текстам об искусстве, чем к научным. Это свидетельствует о том, что продуценты текстов виртуальных экскурсий видят и понимают специфичность задачи таких текстов, их несводимость к простому трансферу из коллекций научных искусствоведческих текстов. Однако пока наиболее слабо выраженной является такая жанровая черта текстов виртуальных экскурсий, как персуазивность, скрытая аргументативность, коррелирующая с образом будущего в данном жанре – виртуальный посетитель получает такой заряд мотивации, который «приводит» его в реальные локации музея и подпитывает его дальнейший интерес к искусству.

На следующем этапе исследования мы планируем оценить эмоциональность текстов Виртуального визита на сайте Государственного Эрмитажа, а также их трудность для восприятия. Для решения последней задачи планируется использование инструментов фиксации глазодвигательной активности.

Список источников

1. Сизова И.А., Гордин В.Э. Цифровизация музеев: трудности, успехи, перспективы // Информационное общество. 2022. № 4. С. 35–44.
2. Национальный проект «Культура». URL: <https://culture.gov.ru/about/national-project/about-project/> (дата обращения: 07.05.2024).
3. Harrison C., Wood P. Art in Theory: 1900–2000. Oxford : Blackwell Publishing, 2002. 1214 p.
4. Белов С.А., Гулида В.Б. Язык юридических документов: сложности понимания // Acta Linguistica Petropolitana. 2019. № 15. С. 56–103.
5. Solnyshkina M.I., Vishnyakova O.D., Gafiyatova E.V., Gabitov A.I. English textbooks for Russian students: Problems and specific features // Journal of Social Studies Education Research. 2017. № 3 (8). P. 215–226.
6. Oliffe M., Thompson E., Johnston J., Freeman D., Bagga H., Wong P.K.K. Assessing the readability and patient comprehension of rheumatology medicine information

- sheets: A cross-sectional Health Literacy Study // *BMJ Open*. 2019. № 9 (2). doi: 10.1136/bmjopen-2018-024582.
7. Blinova O., Tarasov N. Complexity metrics of Russian legal texts: selection, use, initial efficiency evaluation // *Computational Linguistics and Intellectual Technologies: Proc. International Conference “Dialogue 2022”* (June 15–18, 2022). 2022. P. 1017–1028. doi: 10.28995/2075-7182-2022-21-1017-1028
 8. Оборнева И.В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров : дис. ... канд. пед. наук. М., 2006. 165 с.
 9. Leroy G., Kauchak D. The effect of word familiarity on actual and perceived text difficulty // *Journal of the American Medical Informatics Association*. 2014. V. 21 (e1). P. 169–172.
 10. Gatiyatullina G.M., Solnyshkina M.I., Kupriyanov R.V., Ziganshina C.R. Lexical density as a complexity predictor: the case of Science and Social Studies textbooks // *Research Result. Theoretical and Applied Linguistics*. 2022. № 9 (1). P. 11–26. doi: 10.18413/2313-8912-2023-9-1-0-2
 11. Monakhov S.I., Turchanenko V.V., Cherdakov D.N. Terminology use in school textbooks: corpus analysis // *Research Result. Theoretical and Applied Linguistics*. 2023. № 9 (1). P. 27–49. doi: 10.18413/2313-8912-2023-9-1-0-3
 12. Оборнева И.В. Автоматизация оценки качества восприятия текста // *Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования*. 2005. № 5. С. 86–91.
 13. Солнышкина М.И., Соловьев В.Д., Гафиятова Э.В., Мартынова Е.В. Сложность текста как междисциплинарная проблема // *Вопросы когнитивной лингвистики*. 2022. № 1. С. 18–39.
 14. Щипицина Л.Ю. Жанровая характеристика виртуальной экскурсии // *Актуальные проблемы филологии и педагогической лингвистики*. 2022. № 3. С. 234–248.
 15. Hyun M.Y., O’Keefe R.M. Virtual destination image: Testing a telepresence model // *Journal of Business Research*. 2012. № 65 (1). P. 29–35. <https://doi.org/10.1016/j.jbusres.2011.07.011>
 16. Zhu Ch., Hoc Nang Fong L., Hall C.M., She C., Koupaei S.N. Understanding a virtual heritage site through the lens of telepresence and virtual destination image // *Journal of Heritage Tourism*. 2023. doi: 10.1080/1743873X.2023.2289949
 17. Шмелева Т.В. Модель речевого жанра // *Жанры речи*. 1997. № 1. С. 88–98.
 18. Государственный Эрмитаж: виртуальный визит. URL: https://hermitagemuseum.org/wps/portal/hermitage/panorama/virtual_visit/panoramas-m-1/?lng=ru (дата обращения: 01.04.2024).
 19. Blinova O.V., Tarasov N.A. Language complexity across sub-styles and genres in legal Russian // *Research Result. Theoretical and Applied Linguistics*. 2023. V. 9 (2). P. 73–96. doi: 10.18413/2313-8912-2023-9-2-0-5
 20. Blinova O., Tarasov N. A hybrid model of complexity estimation: Evidence from Russian legal texts // *Front. Artif. Intell*. 2022. V. 5. P. 1008530. doi: 10.3389/frai.2022.1008530
 21. Github – репозиторий Generated VS Actual Movie Reviews in Russian. URL: <https://github.com/z00logist/generated-vs-actual-reviews/tree/main> (01.04.2024).
 22. Словарь терминов Российской академии художеств. URL: <https://rah.ru/science/glossary/?ID=19830&let=%D0%9F> (дата обращения: 01.04.2024).

23. Yule G. The statistical study of literary vocabulary. Cambridge : Cambridge Univ. Press, 1944.
24. Казачкова М.Б., Галимова Х.Н. Лексическое разнообразие текста как параметр сложности текстов // Вестник Марийского государственного университета. 2021. Т. 15, № 3. С. 384–390.

References

1. Sizova, I.A. & Gordin, V.E. (2020). Tsifrovizatsiya muzeev: trudnosti, uspekhi, perspektivy [Digitalization of museums: Difficulties, successes, prospects]. *Informatsionnoe obshchestvo*. 4. pp. 35–44.
2. Russian Federation. (n.d.) *Natsional'nyy proekt "Kul'tura"* [National Project "Culture"]. [Online] Available from: <https://culture.gov.ru/about/national-project/about-project/> (Accessed: 7th May 2024).
3. Harrison, C. & Wood, P. (2002) *Art in Theory: 1900–2000*. Oxford: Blackwell Publishing.
4. Belov, S.A. & Gulida, V.B. (2019) Yazyk yuridicheskikh dokumentov: slozhnosti ponimaniya [The language of legal documents: Difficulties of understanding]. *Acta Linguistica Petropolitana*. 15. pp. 56–103.
5. Solnyshkina, M.I., Vishnyakova, O.D., Gafiyatova, E.V. & Gabitov, A.I. (2017) English textbooks for Russian students: Problems and specific features. *Journal of Social Studies Education Research*. 3(8). pp. 215–226.
6. Oliffe, M., Thompson, E., Johnston, J., Freeman, D., Bagga H. & Wong, P.K.K. (2019) Assessing the readability and patient comprehension of rheumatology medicine information sheets: A cross-sectional Health Literacy Study. *BMJ Open*. 9(2). DOI: 10.1136/bmjopen-2018-024582
7. Blinova, O. & Tarasov, N. (2022) Complexity metrics of Russian legal texts: Selection, use, initial efficiency evaluation. *Computational Linguistics and Intellectual Technologies*. Proc. International Conference "Dialogue 2022", June 15–18, 2022. pp. 1017–1028. DOI: 10.28995/2075-7182-2022-21-1017-1028
8. Osborneva, I.V. (2006) *Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov* [Automated assessment of the complexity of educational texts based on statistical parameters]. Pedagogy Cand. Diss. Moscow.
9. Leroy, G. & Kauchak, D. (2014). The effect of word familiarity on actual and perceived text difficulty. *Journal of the American Medical Informatics Association*. 21. pp. 169–172.
10. Gatiyatullina, G.M., Solnyshkina, M.I., Kupriyanov, R.V. & Ziganshina, C.R. (2022) Lexical density as a complexity predictor: The case of Science and Social Studies textbooks. *Research Result. Theoretical and Applied Linguistics*. 9(1). pp. 11–26. DOI: 10.18413/2313-8912-2023-9-1-0-2
11. Monakhov, S.I., Turchanenko, V.V. & Cherdakov, D.N. (2023) Terminology use in school textbooks: Corpus analysis. *Research Result. Theoretical and Applied Linguistics*. 9(1). pp. 27–49. DOI: 10.18413/2313-8912-2023-9-1-0-3
12. Osborneva, I.V. (2005) *Avtomatizatsiya otsenki kachestva vospriyatiya teksta* [Automation of Text Perception Quality Assessment]. *Vestnik Moskovskogo gorodskogo*

- pedagogicheskogo universiteta. Seriya: Informatika i informatizatsiya obrazovaniya.* 5. pp. 86–91.
13. Solnyshkina, M.I., Soloviev, V.D., Gafiyatova, E.V. & Martynova, E.V. (2022) Slozhnost' teksta kak mezhdistsiplinarnaya problema [Text complexity as an interdisciplinary problem]. *Voprosy kognitivnoy lingvistiki.* 1. pp. 18–39.
 14. Shchipsitsina, L.Yu. (2022) Zhanrovaya kharakteristika virtual'noy ekskursii [Genre characteristics of a virtual excursion]. *Aktual'nye problemy filologii i pedagogicheskoy lingvistiki.* 3. pp. 234–248.
 15. Hyun, M.Y. & O'Keefe, R.M. (2012) Virtual destination image: Testing a telepresence model. *Journal of Business Research.* 65(1). pp. 29–35. DOI: 10.1016/j.jbusres.2011.07.011
 16. Zhu, Ch., Hoc Nang Fong, L., Hall, C.M., She, C. & Koupaei, S.N. (2023) Understanding a virtual heritage site through the lens of telepresence and virtual destination image. *Journal of Heritage Tourism.* 19(5). pp. 682–695. DOI: 10.1080/1743873X.2023.2289949
 17. Shmeleva, T.V. (1997) Model' rechevogo zhanra [Speech genre model]. *Zhany rechi.* 1. pp. 88–98.
 18. *The State Hermitage Museum: A virtual visit.* [Online] Available from: https://hermitagemuseum.org/wps/portal/hermitage/panorama/virtual_visit/panoramas-m-1/?lng=ru (Accessed: 1st April 2024).
 19. Blinova, O.V. & Tarasov, N.A. (2023) Language complexity across sub-styles and genres in legal Russian, *Research Result. Theoretical and Applied Linguistics.* 9(2). pp. 73–96. DOI: 10.18413/2313-8912-2023-9-2-0-5
 20. Blinova, O. & Tarasov, N. (2022) A hybrid model of complexity estimation: Evidence from Russian legal texts. *Front. Artif. Intell.* 5. p. 1008530. DOI: 10.3389/frai.2022.1008530
 21. *Github – Repository Generated VS Actual Movie Reviews in Russian.* [Online] Available from: <https://github.com/z00logist/generated-vs-actual-reviews/tree/main> (Accessed: 1st April 2024).
 22. The Russian Academy of Arts. (n.d.) *Slovar' terminov Rossiyskoy akademii khudozhestv* [Dictionary of Terms of the Russian Academy of Arts]. [Online] Available from: <https://rah.ru/science/glossary/?ID=19830&let=%D0%9F> (Accessed: 1st April 2024).
 23. Yule, G. (1944) *The Statistical Study of Literary Vocabulary.* Cambridge University Press.
 24. Kazachkova, M.B. & Galimova, Kh.N. (2021) Leksicheskoe raznoobrazie teksta kak parametr slozhnosti tekstov [Text lexical diversity as a text complexity parameter]. *Vestnik Mariyskogo gosudarstvennogo universiteta.* 15(3). pp. 384–390.

Сведения об авторах:

Колмогорова Анастасия Владимировна – доктор филологических наук, заведующий лабораторией языковой конвергенции, профессор департамента филологии Школы гуманитарных наук и искусств НИУ ВШЭ – Санкт-Петербург (Санкт-Петербург, Россия). E-mail: akolmogorova@hse.ru

Куликова Елизавета Романовна – младший научный сотрудник лаборатории языковой конвергенции НИУ ВШЭ – Санкт-Петербург (Санкт-Петербург, Россия).

E-mail: kulikova.e.r@hse.ru

Колмогорова Полина Алексеевна – аспирант НИУ ВШЭ – Санкт-Петербург (Санкт-Петербург, Россия). E-mail: kolmogorovapa@gmail.com

Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Anastasia A. Kolmogorova, Dr. Sci. (Philology), head of the Laboratory of Language Convergence, professor, National Research University Higher School of Economics (HSE University) (Saint Petersburg, Russian Federation). E-mail: akolmogorova@hse.ru

Elizaveta R. Kulikova, junior research assistant, Laboratory of Language Convergence, National Research University Higher School of Economics (HSE University) (Saint Petersburg, Russian Federation). E-mail: kulikova.e.r@hse.ru

Polina V. Kolmogorova, PhD student, National Research University Higher School of Economics (HSE University) (Saint Petersburg, Russian Federation). E-mail: kolmogorovapa@gmail.com

The authors declare no conflicts of interests.

*Статья поступила в редакцию 09.05.2024;
одобрена после рецензирования 29.08.2024; принята к публикации 29.08.2024*

*The article was submitted 09.05.2024;
approved after reviewing 29.08.2024; accepted for publication 29.08.2024*