

RU

Онлайн-ресурс лингвистических характеристик произведения художественной литературы: ключевые параметры и технология разработки

Горожанов А. И.

Аннотация. Цель настоящего исследования – описать технологию составления справочного онлайн-ресурса о лингвистических особенностях языка произведения художественной литературы на основе корпусных данных. Научная новизна в широком смысле обусловлена тем, что впервые описывается технология автоматизированного составления универсального справочного ресурса, полученного на основании результатов работы программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер», а в узком смысле – тем, что подобный онлайн-ресурс впервые создается для оригинальных текстов писателя Э. М. Ремарка. В ходе работы были сформулированы теоретические принципы составления онлайн-ресурса, его технологические параметры, а также осуществлена апробация в виде экспериментальной разработки и тестирования готового программного продукта, размещенного в открытом доступе по интернет-адресу <https://idiostil.ecovle.ru/index.php>. В результате было создано готовое программное решение, обладающее параметром масштабирования и пригодное в качестве источника справочных данных для проведения профессиональных лингвистических исследований авторского идиостиля.

EN

Online resource of linguistic features in a fictional work: Key parameters and development methodologies

A. I. Gorozhanov

Abstract. The purpose of this study is to describe the technology of compiling an online reference resource on the linguistic features of a work of fiction based on corpus data. Scientific novelty, in a broad sense, is due to the fact that for the first time the technology of automated compilation of a universal reference resource obtained on the basis of the output of the software package “Balanced Linguistic Corpus Generator and Corpus Manager” is described, and in a narrow sense, by the fact that such an online resource is created for the first time for the original texts of the writer E. M. Remarque. In the course of the work, the theoretical principles of compiling the online resource, its technological parameters were formulated, and testing was carried out in the form of experimental development and testing of the ready-to-use software product, hosted in the public domain at the Internet address <https://idiostil.ecovle.ru/index.php>. As a result, a ready-made software solution, having a scaling parameter and suitable for use as a source of reference data for professional linguistic research of author’s idiosyncrasy, was created.

Введение

В предисловии к объемному Словарю языка Достоевского Ю. Н. Караулов отмечает, что «изучение писательского идиолекта может осуществляться разными путями. Исследователь может отправляться непосредственно от текстов... Другой подход предполагает представление языка писателя через построение его словаря, причем лексикографический путь решения задачи обладает известными преимуществами по сравнению с иными подходами» (Словарь языка Достоевского, 2008, с. IX).

В другом предисловии, к Словарю рифм М. Ю. Лермонтова, который входит в состав Лермонтовской энциклопедии (1981), составители указывают на то, что «словари языка писателей – редкий вид справочных изданий». Уже тогда «при подготовке Словаря широко использовалась электронная вычислительная техника» (<http://m-y-lermontov.ru/lermontov/enc/in/fo/info/st010.shtml>), а говоря современным языком – специализированное программное обеспечение.

В настоящей работе предпринимается попытка составить пусть пока еще не такой масштабный и всеобъемлющий справочник языка отдельного писателя, но все же полноценный, значительный по объему и представленный в Интернете ресурс, посвященный лингвистической составляющей идиостиля автора и структурированный по отдельным произведениям, созданный автоматизированным способом, т. е. автоматически (на основе современных программных средств с участием технологичного искусственного интеллекта) и одновременно с некоторой долей ручного труда исследователя (эксперта-лингвиста).

Понятие *идиостиля* не привлекается нами в качестве образующего, однако этот термин объединяет многие исследования, в фокусе которых находятся особенности индивидуального языка того или иного писателя, отсюда мы должны соотносить нашу работу с работами ученых, разрабатывающих проблему идиостиля.

В этой связи важно основополагающее определение идиостиля В. В. Виноградова как «системы индивидуально-эстетического использования свойственных данному периоду развития художественной литературы средств словесного выражения» (1963, с. 126), поскольку в нем изначально заложена лингвистическая (языковая, словесная) составляющая идиостиля.

Заметим здесь, что сочетание «словесное выражение», редко используемое в наше время, закрепилось в русской литературе как обозначающее нечто *материальное, формальное, измеримое*, противоположное *мысли или неизмеримому*. Приведем в качестве доказательства несколько цитат.

«Есть мучительное несоответствие между моей мыслью и ее словесным выражением» (Бердяев Н. А. Самопознание. 1949 год).

«Но помощь эта касается единственно словесного выражения мыслей» (Добролюбов Н. А. Забытые люди. 1861 год).

«И что не оттого ли в первые времена не было ереси, что христиане судили друг друга не по словесным выражениям, но по сердцу и делам» (Толстой Л. Н. Царство Божие внутри вас, или Христианство не как мистическое учение, а как новое жизнепонимание. 1893 год).

Последующие определения также подчеркивают важность лингвистического компонента идиостиля, например, А. Н. Баранов и Д. О. Добровольский подразумевают под идиостилем «совокупность ментальных и языковых структур художественного мира писателя» (1997, с. 11). Стилистический энциклопедический словарь русского языка (2006 год издания) приводит следующее определение: «...идиостиль имеет комплексный характер, разноаспектно выражает социально-историческую сущность, национальные, индивидуально-психологические и нравственно-этические особенности человека. В идиостиле проявляется его мировоззрение и знание о мире (концептуальная картина мира и тезаурус), общая и языковая культура в их текстовом воплощении. Идиостиль, таким образом, – это стиль личности во всем многообразии ее многоуровневых текстовых проявлений (в структуре, семантике и прагматике текста)» (https://stylistics.academic.ru/58/Коммуникативная_стилистика_художественного_текста).

Таким образом, ядром идиостиля является текстовая, лингвистическая составляющая, которая подлежит в настоящее время некоторой оценке с помощью формальных измерительных процедур, в то время как абстрактные (ментальные, психологические, нравственно-этические) компоненты все еще требуют, по нашему мнению, глубокой мысленной интерпретации, осуществляемой коллективами экспертов.

Перечислим задачи нашего исследования:

- 1) сформулировать теоретические принципы составления онлайн-ресурса о лингвистических особенностях языка произведения художественной литературы на основе корпусных данных;
- 2) разработать технические требования к такому онлайн-ресурсу;
- 3) провести апробацию предложенной технологии на материале оригинального текста романа Э. М. Ремарка «*Liebe deinen Nächsten*» («Возлюби ближнего своего»).

В ходе решения первой задачи применяются методы *анализа* данных, которые могут быть получены в ходе автоматической обработки текста произведения художественной литературы, и *описания* функциональных особенностей онлайн-ресурса, т. е. его потенциальных возможностей для оператора-исследователя.

Вторая задача решается с помощью метода *проектирования* архитектуры онлайн-ресурса, включая описание технических параметров серверной и клиентской частей.

Для решения третьей задачи используется метод *эксперимента* (т. е. разработки экспериментального онлайн-ресурса и размещения его на веб-сервере) с последующей *верификацией* полученного результата в части его соответствия сформулированным ранее теоретическим принципам и *валидацией* размещенных данных (т. е. оценкой их пригодности для установления особенностей языка художественного произведения).

Кроме того, в ходе решения задач исследования мы руководствуемся общенаучными принципами *объективности* (в процессе обоснования актуальности работы и экспликации теоретических положений), *доказательности* (путем эксперимента), *систематичности* и *повторяемости результатов* (предлагая универсальную технологию, которая может быть масштабирована на другой языковой материал).

Теоретическая база работы сформирована трудами советских и российских ученых в области исследования идиостиля (Виноградов, 1963; Григорьев, 1983; Болотнов, 2016; Баранов, Добровольский, 2022), творчества Э. М. Ремарка (Омелаенко, 2023; Киселева, Жилиякова, 2024; Нечай, Линке, 2024), корпусной лингвистики (Голубкова, Бухарова, 2023; Степанова, 2023; Степанова, Яновская, 2024).

Актуальность исследования заключается в необходимости разработки и внедрения эффективных онлайн-платформ, которые не только обеспечивали бы доступ к статистическим и аналитическим данным литературных произведений, но и являлись бы масштабируемым образцом для создания подобных программных продуктов, применимых к произведениям других авторов, в том числе и опубликованных на различных языках.

Кроме того, об актуальности темы говорят неослабевающий интерес к изучению идиостилей отдельных писателей и наличие изданных в различные годы и популярных в настоящее время справочников, таких как «Словарь языка Пушкина» под редакцией В. В. Виноградова (Словарь языка Пушкина: в 4 т. / отв. ред. В. В. Виноградов М.: Азбуковник, 2000), «Словарь языка Достоевского» (2008) авторского коллектива под руководством Ю. Н. Караулова, «Лермонтовская энциклопедия» коллектива авторов под редакцией В. А. Мануйлова (Лермонтовская энциклопедия / гл. ред. В. А. Мануйлов. М.: Советская энциклопедия, 1981), «Словарь диалектной лексики в произведениях Н. А. Некрасова» (Мальшева И. Ю. Словарь диалектной лексики в произведениях Н. А. Некрасова // Электронный научно-методический журнал Костромского областного института развития образования. 2018. № 46 (1)).

Лингвистическим материалом исследования является сбалансированный лингвистический корпус оригинального текста романа Э. М. Ремарка “*Liebe deinen Nächsten*” («Возлюби ближнего своего»), полученный в результате работы программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер», разработанного в лаборатории фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета (Свидетельство о государственной регистрации программы для ЭВМ № 2023683209 Российская Федерация. «Генератор сбалансированного лингвистического корпуса и корпусный менеджер»: № 2023682269: заявл. 25.10.2023: опубл. 03.11.2023 / А. И. Горожанов; заявитель федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный лингвистический университет»).

В практическом отношении результаты исследования могут рассматриваться как руководство к созданию сходных онлайн-ресурсов, которые, в свою очередь, могут быть полезны широкому кругу профессиональных исследователей для анализа идиостиля того или иного автора или определения характерных языковых черт того или иного произведения художественной литературы, а также для широкого ряда сопоставительных исследований.

Обсуждение и результаты

Мы исходим из того, что наполнение онлайн-ресурса должно производиться данными, полученными в ходе работы программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» без их последующей ручной обработки (по крайней мере, на текущем этапе исследования). Логично предположить, что характер данных напрямую зависит от возможностей указанного программного обеспечения, которые были подробно описаны нами ранее (Горожанов, 2024а; Gorozhanov, Guseynova, Stepanova, 2024). Кратко суммируем, что типовыми результатами выполнения запросов к корпусному менеджеру, задействованными нами здесь, являются частотные списки по частям речи, а также списки контекстов (как правило, в пределах одного предложения), содержащие указанные оператором языковые явления (например, модальные глаголы или подчинительные союзы).

Говоря о теоретических принципах составления онлайн-ресурса, сформулировать которые необходимо в рамках выполнения первой задачи исследования, мы назовем принцип *быстрого обновления содержания*, принцип *превалирования содержания над формой*, а также принцип *учета возможной последующей программной обработки данных*.

Быстрота обновления содержания понимается нами как возможность замены отдельной категории данных (например, частотного списка прилагательных) путем исполнения минимального набора операций. Несмотря на то, что сам текст художественного произведения не изменяется во времени, технологии искусственного интеллекта программного комплекса могут совершенствоваться в сторону уменьшения погрешности вывода (англ. ‘output’), поэтому через какое-то время может быть целесообразна замена данных на более актуальные. Учитывая в перспективе наличие множества онлайн-ресурсов для отдельных авторов и произведений, такая замена должна восприниматься как рутинная механическая процедура.

Превалирование содержания над формой трактуется в данном случае как отказ от излишних элементов управления и оформления веб-страниц, что придает онлайн-ресурсу строго профессиональный характер. Интерфейс должен быть удобен, прост и понятен. Здесь же упомянем тот факт, что материалы онлайн-ресурса должны позволять исследователю рассматривать нужные ему явления в широком авторском контексте, т. е. быть достаточными для формулирования собственных выводов, но в то же время онлайн-ресурс не должен предлагать какие-либо готовые шаблоны для решения проблемы построения идиостиля.

Учет возможной последующей программной обработки данных важен, поскольку материалы онлайн-ресурса также рассматриваются нами и как промежуточный вариант представления данных, которые могут быть далее обработаны каким-либо образом. В этой связи списки не должны содержать лишние элементы, но в то же время должны легко копироваться и иметь типовые формальные разделители в каждой строке. Поскольку корпусный менеджер программного комплекса создавался именно с учетом того, что выводы запросов максимально структурированы для последующей программной обработки, наша задача сводится к переносу этих данных в онлайн-ресурс с минимальными изменениями, что полностью отвечает первому принципу.

Для каждого произведения художественной литературы предлагается следующий набор статистических данных:

- полный нумерованный список контекстов;
- частотный список глаголов;

- частотный список существительных;
- частотный список прилагательных;
- частотный список числительных;
- частотный список имен собственных;
- связанные элементы для первых 10 имен собственных;
- сочетания «прилагательное + существительное»;
- сочетания «прилагательное + прилагательное + существительное»;
- контексты употребления модальных глаголов;
- контексты употребления форм конъюнктива;
- контексты с сочинительными союзами;
- контексты с подчинительными союзами;
- контексты употребления основных колоронимов.

Заметим, что обозначенный список имеет демонстрационный характер, обусловлен текущими возможностями корпусного менеджера и впоследствии может быть изменен с учетом результатов долговременной апробации онлайн-ресурса.

Такой набор данных может быть достаточным для оператора-исследователя, чтобы качественно проанализировать, например, такие особенности языка произведения, как употребление числительных, имен собственных, эпитетов, сложность синтаксических конструкций и мн. др.

Далее было необходимо разработать технические требования к онлайн-ресурсу. Поскольку материалы должны быть доступны через Интернет, проектируемый программный продукт должен иметь так называемые серверную и клиентскую части. Для написания первой был выбран язык программирования PHP, в написании второй, как это обычно бывает в подобных случаях, была применена «классическая триада» «HTML + JavaScript + CSS».

На текущем этапе работы было принято решение не закрывать ресурс, создавая учетные записи, что дало возможность отказать пока от использования баз данных (например, MySQL).

Архитектура онлайн-ресурса представляется достаточно простой (Схема 1).

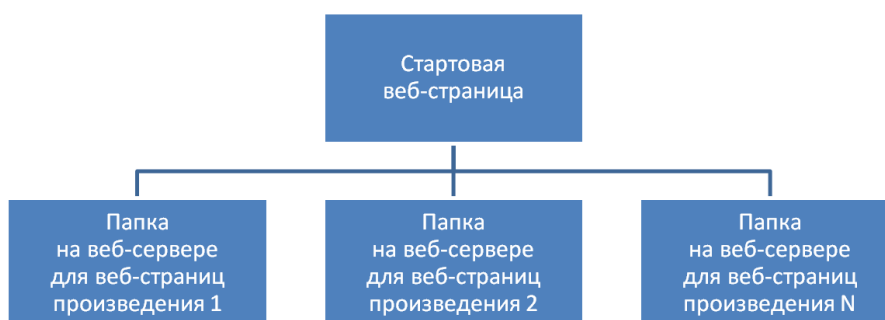


Схема 1. Упрощенная модель архитектуры серверной части онлайн-ресурса

Прочие технические детали являются типовыми для любого веб-сайта и поэтому не нуждаются в освещении.

Для экспериментального апробирования, как уже упоминалось ранее, был выбран сбалансированный лингвистический корпус текста романа Э. М. Ремарка «Возлюби ближнего своего» объемом 136 926 токенов или 14 068 контекстов (предложений). В этой связи мы говорим о «младшем» и «старшем» токене соответственно (Горожанов, 2024а, с. 26). Таким образом, создаваемый онлайн-ресурс был наполнен материалами только одного произведения художественной литературы, что является рациональным решением на этапе проведения эксперимента. Для онлайн-ресурса был выбран интернет-адрес <https://idiostil.ecovle.ru/index.php>, где *index.php* – стартовая веб-страница. Предполагается, что данный интернет-адрес будет активным в среднесрочной перспективе.

Каждый из упомянутых выше наборов данных представляет собой отдельную веб-страницу, связанную со стартовой веб-страницей гиперссылкой (Рисунок 1).

В верхней части веб-страницы располагается указание на автора и навигационное меню, в котором размещены два пункта: «ГЛАВНАЯ» и ‘Liebe Deinen Nächsten’ (в представленной версии прописные буквы выставляются автоматически для каждого слова наименования произведения).

Рассмотрим далее форматы представления некоторых веб-страниц. Например, частотный список глаголов выглядит следующим образом (Рисунок 2).

Инфинитивы глаголов выведены в виде пронумерованного списка – от самого частотного к самому редкому. Сообразуясь с принципом *учета возможной последующей программной обработки данных*, каждая строка вывода имеет структуру «числительное + точка + пробел + инфинитив глагола + пробел + двоеточие + числительное». Сходным образом выглядят все представленные частотные списки.

Приведем пример из раздела «Дополнительные материалы». В нем сейчас расположена единственная веб-страница «Связанные элементы для первых 10 имен собственных» (Рисунок 3).

E. M. Remarque

ГЛАВНАЯ Liebe Deinen Nächsten

Liebe deinen Nächsten

[Полный нумерованный список контекстов](#)

[Частотный список глаголов](#)

[Частотный список существительных](#)

[Частотный список прилагательных](#)

[Частотный список числительных](#)

[Частотный список имен собственных](#)

[Открыть доп. материалы](#)

[Сочетания "прилагательное + существительное"](#)

[Сочетания "прилагательное + прилагательное + существительное"](#)

Рисунок 1. Фрагмент стартовой веб-страницы для материалов к роману «Возлюби ближнего своего»

Частотный список глаголов

1. sagen : 914
2. sehen : 569
3. gehen : 468
4. kommen : 417
5. haben : 357
6. fragen : 345
7. stehen : 278
8. geben : 237
9. machen : 222
10. wissen : 191
11. nehmen : 187
12. bleiben : 161
13. erwidern : 153
14. denken : 141
15. lassen : 131
16. weiß : 128
17. hören : 124
18. sitzen : 123
19. tun : 110
20. glauben : 106

Рисунок 2. Фрагмент веб-страницы частотного списка глаголов

Связанные элементы для перв

1. Steiner : 632
2. Ruth : 342
3. Marill : 130
4. Moritz : 72
5. Lilo : 54
6. Ludwig : 52
7. Ammers : 52
8. Binder : 47
9. Kern : 44
10. Schweiz : 43

Steiner

1. ich : 146
2. sagen : 112
3. sich : 55

Рисунок 3. Фрагмент веб-страницы связанных элементов

Здесь даются члены предложения, для которых указанные имена собственные являются непосредственно главным или зависимым словом. Эти данные могут быть полезны для получения характеристики конкретного персонажа, а также его речевого портрета (Горожанов, 2024b).

В следующем примере покажем раздел «Сочетания 'прилагательное + прилагательное + существительное'» (Рисунок 4).

Сочетания "прилагательное + прилагательное + существительное"

- >Tacoma<. Portugiesischer Frachtdampfer : 13892 : Mit der >Tacoma<. Por
- alten eigenen Sachen : 3905 : Mit unseren alten eigenen Sachen.
- alten griechischen Helden : 6898 : Die alten griechischen Helden weinte
- alten jüdischen Friedhofs : 2639 : Kern saß auf der Mauer des alten jüc
- alten schönen Zeiten : 12267 : Gott ja, Kinder, die alten schönen Zeite
- altes jiddisches Lied : 12853 : Dann richtete sie sich ein wenig auf ur
- behendes kleines Männchen : 5204 : Direktor Potzloch war ein behendes k
- billigen französischen Rotwein : 12232 : Fast alle den billigen französ
- blauen gestrickten Jäckchen : 3423 : Auf dem Boden verstreut lagen die
- buschigen schwarzen Haar : 10794 : Sehen Sie den Mann mit dem buschiger
- dicke blonde Wirtin : 336 : Nur die dicke blonde Wirtin lamentierte un
- dicken kurzen Fingern : 7421 : Er war ein gutgelaunter, kräftiger Mann
- dritte gesunde Tochter!< : 4283 : Kommt da heute morgen die Hebamme her
- dunklem russischem Brot : 5654 : Die dunkle Mutter der Lebensfreude!< l
- dürftigen älteren Mann : 3794 : DEN dürftigen älteren Mann auf den erst
- Edelstes deutsches Barock : 7443 : Edelstes deutsches Barock!

Рисунок 4. Фрагмент веб-страницы сочетаний нескольких частей речи

Здесь в алфавитном порядке представлены контексты, имеющие в своем составе относительно редкое сочетание «два прилагательных + существительное», что может, по нашему мнению, являться чертой идиостиля писателя. Каждая строка показывает само сочетание, далее следуют пробел, разделитель-двоеточие, пробел, номер контекста в корпусе, пробел, разделитель-двоеточие, пробел и сам контекст. Номер контекста важен для потенциального получения более полного текстового окружения в разделе «Полный нумерованный список контекстов».

Таким образом, был создан и размещен на веб-сервере экспериментальный онлайн-ресурс, который не показал технических ошибок, прошел процедуры верификации (соответствие установленным параметрам) и валидации (оценка пригодности размещенных материалов для профессиональных исследователей-лингвистов).

С онлайн-ресурсом возможно работать как со стационарного компьютера, так и с планшета или смартфона.

Заключение

Поставленная в ходе исследования цель была достигнута. Во-первых, были сформулированы теоретические принципы составления онлайн-ресурса о лингвистических особенностях языка произведения художественной литературы на основе корпусных данных: принципы быстрого обновления содержания, превалирования содержания над формой и учета возможной последующей программной обработки данных.

Во-вторых, мы разработали технические требования к такому онлайн-ресурсу как к открытому веб-сайту, определена его архитектура.

В-третьих, была проведена апробация полученной технологии на материале оригинального текста романа Э. М. Ремарка «Возлюби ближнего своего», создан хорошо функционирующий программный продукт на языке программирования PHP, размещенный по интернет-адресу <https://idiostil.ecovle.ru/index.php>.

Отметим также масштабируемость предложенного технического решения, которая становится возможной благодаря тому, что списки контекстов и частотные списки размещаются в виде необработанного результата запроса к корпусу, пусть и с некоторой долей погрешности.

В качестве перспективы исследования обозначим, прежде всего, дополнение представленного онлайн-ресурса данными, автоматически извлеченными из лингвистического корпуса. Далее необходимо расширить набор произведений одного автора до нескольких, провести апробацию материалов в рамках исследовательских работ различного уровня (от студенческих до диссертационных) и по ее итогам уточнить набор представляемых данных. В более далекой перспективе кажется целесообразным расширить функционал онлайн-ресурса возможностью входа в личный кабинет и формирования индивидуальных подборок лингвистического материала для пользователей.

Источники | References

1. Баранов А. Н., Добровольский Д. О. Понятие корпусной модели идиостиля // Труды института русского языка им. В. В. Виноградова. 2022. № 1. <https://doi.org/10.31912/pvrl-2022.1.2>
2. Баранов А. Н., Добровольский Д. О. Постулаты когнитивной семантики // Известия Российской академии наук. Серия литературы и языка. 1997. Т. 56. № 1.

3. Болотнов А. В. Идиостиль информационно-медийной языковой личности: коммуникативно-когнитивные аспекты исследования: дисс. ... д. филол. н. Томск, 2016.
4. Виноградов В. В. Стилистика. Теория поэтической речи. Поэтика. М.: АН СССР, 1963.
5. Голубкова Е. Е., Бухарова А. А. Загадки конструкции XYZ(W): корпусный подход // От слова к дискурсу: взаимодействие форм и (не)предсказуемость смыслов: тезисы международной научной конференции (г. Минск, 12-13 мая 2023 года). Мн.: МГЛУ, 2023.
6. Горожанов А. И. Архитектура сбалансированного лингвистического корпуса, полученного автоматическим путем (опыт Московского государственного лингвистического университета) // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2024а. № 11 (892).
7. Горожанов А. И. Метод программного анализа контекста лексической единицы // Актуальные проблемы филологии и педагогической лингвистики. 2024b. № 3. <https://doi.org/10.29025/2079-6021-2024-3-178-190>
8. Григорьев В. П. Грамматика идиостиля: В. Хлебников. М.: Наука, 1983.
9. Киселева Н. М., Жилиякова С. А. Полифункциональность пейзажных описаний в романе Э. М. Ремарка «Черный обелиск» // Современное педагогическое образование. 2024. № 11.
10. Нечай Ю. П., Линке Н. О. Экспрессивность как семантико-прагматическая категория высказывания в художественной прозе Э. М. Ремарка // Гуманитарные и социальные науки. 2024. Т. 104. № 3.
11. Омеланко Н. В. Концепт «счастье» в романе Э. М. Ремарка «Жизнь взаимы, или У неба любимчиков нет» // Гуманитарные и социальные науки. 2023. Т. 98. № 3.
12. Словарь языка Достоевского. Идиоглоссарий / Российская академия наук, Институт русского языка им. В. В. Виноградова; гл. ред. чл.-корр. РАН Ю. Н. Караулов. М.: Азбуковник, 2008.
13. Степанова Д. В. Программный комплекс для генерации динамического корпуса текстов СМИ // Вестник Минского государственного лингвистического университета. Серия 1. Филология. 2023. № 6 (127).
14. Степанова Д. В., Яновская А. С. Языковые средства описания внешности литературного героя: сопоставительный анализ на базе корпуса параллельных текстов // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2024. № 13 (894).
15. Gorozhanov A. I., Guseynova I. A., Stepanova D. V. Natural Language Processing and Fiction Text: Basis for Corpus Research // RUDN Journal of Language Studies, Semiotics and Semantics. 2024. Vol. 15. № 1. <https://doi.org/10.22363/2313-2299-2024-15-1-195-210>

Информация об авторах | Author information



Горожанов Алексей Иванович¹, д. филол. н., доц.

¹ Московский государственный лингвистический университет



Alexey Ivanovich Gorozhanov¹, Dr

¹ Moscow State Linguistic University

¹ a_gorozhanov@mail.ru

Информация о статье | About this article

Дата поступления рукописи (received): 03.12.2024; опубликовано online (published online): 20.01.2025.

Ключевые слова (keywords): справочный лингвистический онлайн-ресурс; произведение художественной литературы; авторский идиостиль; корпусная лингвистика; обработка естественного языка; online linguistic reference resource; work of fiction; author's idiosyncrasy; corpus linguistics; natural language processing.