



Актуальное состояние корпусной лингвистики: тенденции и перспективы

В. В. Изюмская

Московский государственный лингвистический университет, Москва, Россия

izyumskayaw@gmail.com

Аннотация. Автор статьи исследует актуальное состояние корпусной лингвистики для выявления существующих тенденций и перспективных траекторий развития. Для достижения цели был собран массив научных публикаций, опубликованных после 2019 г. и посвященных лингвистическим корпусам; далее была применена последовательная кластеризация, результаты которой подлежали статистическому анализу и последующей интерпретации.

Ключевые слова: корпусная лингвистика, прикладная лингвистика, лингвистический корпус, компьютерная лингвистика, кластеризация, статистический анализ

Благодарности: Исследование выполнено в рамках государственного задания ФГБОУ ВО МГЛУ № FSFU-2020-0020.

Для цитирования: Изюмская В. В. Актуальное состояние корпусной лингвистики: тенденции и перспективы // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2023. Вып. 8 (876). С. 27–33. DOI 10.52070/2542-2197_2023_8_876_27

Original article

Modern Corpus Linguistics: Trends and Prospects

Veronika V. Izyumskaya

Moscow State Linguistic University, Moscow, Russia

izyumskayaw@gmail.com

Abstract. The article contains an analysis of the state of modern corpus linguistics in order to identify current trends and prospects for development. For this purpose, an array of scientific publications (230 articles and 50 dissertations) published no earlier than 2019 and devoted to linguistic corpora was collected; then the method of sequential clustering was applied, the results of which were statistically analyzed and subsequently interpreted.

Keywords: corpus linguistics, applied linguistics, linguistic corpus, computational linguistics, clustering, statistical analysis

Acknowledgements: The research was carried out within the state assignment of Moscow State Linguistic University (theme No. FSFU-2020-0020).

For citation: Izyumskaya, V. V. (2023). Modern Corpus Linguistics: Trends and Prospects. Vestnik of Moscow State Linguistic University. Humanities, 8(876), 27–33. 10.52070/2542-2197_2023_8_876_27

ВВЕДЕНИЕ

Корпусная лингвистика как самостоятельная отрасль научного знания зародилась в 70-х гг. прошлого столетия, однако истоки ее базовых методов берут свое начало еще в XIII в., когда исследователи вручную составляли конкордансы к собраниям сочинений одного автора: как религиозным, так и литературным [Солнышкина, Гатиятуллина, 2020]. Наиболее показательным примером подобных трудов является конкорданс к латинскому переводу Библии «Вульгата», созданный около 1230 г. под руководством первого кардинала парижского монастыря Святого Иакова, который, согласно преданиям, привлек для достижения поставленной цели 500 монахов [Захаров, 2016]. Спустя два столетия лингвисты, как, например, Отто Есперсен (1909–1947), осознавая необходимость подкрепления исследований валидными примерами и придания им квантитативных характеристик, стали собирать массивы текстовых данных, зафиксированных на бумажных карточках, в так называемые «shoe box» корпусы или корпусы обувных коробок (так, Отто Есперсен был известен тем, что именно в них хранил собранные коллекции примеров); сам термин «корпус» в лингвистическом его значении тогда еще не существовал, однако то, как собирались подобные текстовые массивы и как применялись для дальнейших исследований, позволяет считать их предками современных лингвистических корпусов [Biber, Reppen, 2015]. На середину того же столетия приходится закат доэлектронной эпохи развития лингвистических корпусов, которая продлилась более семи столетий.

Рассвет же электронной эпохи ознаменовался созданием в 1960-х гг. первого электронного корпуса, «Брауновского корпуса» (The Brown Corpus), послужившего импульсом к открытию в 1970-х гг. первых исследовательских центров и лабораторий по всему миру, в которых лингвисты работали совместно с программистами не только над способами обработки текстовых массивов, но и над общими проблемами лингвистики [Солнышкина, Гатиятуллина, 2020], что привело к стремительному развитию данной области научного знания: начало разработки корпуса устной речи London-Lund (LLC) в 1975 г.; реализация Единого языка разметки текстов (Standard Generalized Markup Language, SGML) в начале 1980-х гг.; разработка Международной базы данных языка при Бирмингемском университете и компании Коллинз (Collins Birmingham University International Language Database, корпус COBUILD) объемом 20 млн словоупотреблений в конце 1980-х гг.; начало составления Британского национального корпуса (British National

Corpus, BNC) объемом 100 млн словоупотреблений в 1991 г.; начало разработки Национального корпуса русского языка (НКРЯ) в 2001 г.; разработка программ разметки видеозаписей на уровне жестов в середине 2000-х гг.; публикация Корпуса современной американской английской речи (The Corpus of Contemporary American English (COCA) объемом до 400 млн словоупотреблений в 2008 г. и многое другое за менее чем семь десятилетий.

Необходимо продолжать постоянный мониторинг тенденций корпусной лингвистики как динамично развивающегося направления исследований с целью выявления существующих лакун и прогноза дальнейших траекторий развития. В статье описывается актуальное состояние корпусной лингвистики.

Для достижения данной цели был собран массив научных публикаций (статей, диссертаций), посвященных лингвистическим корпусам, чтобы, применяя последовательную кластеризацию, выявить существующие тенденции и сформулировать потенциальные перспективы развития корпусной лингвистики. Собранный массив текстов включает 200 русскоязычных статей, 50 диссертаций и 30 англоязычных статей, опубликованных в открытом доступе начиная с 2019 г.

КЛАСТЕРЫ «КОРПУС КАК ОБЪЕКТ» И «КОРПУС КАК ИНСТРУМЕНТ»

Традиционно корпусная лингвистика рассматривается как «раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с использованием компьютерных технологий» [Захаров, 2005].

Вместе с тем результаты проведенной в рамках исследования первичной кластеризации позволяют предложить новый подход к описанию данной научной дисциплины.

В частности, собранный массив подразделяется на два основных кластера:

1. Труды, в которых лингвистический корпус является объектом исследования (например [Leedham, Lillis, Twiner, 2021]).
2. Труды, в которых лингвистический корпус является инструментом исследования (например [Орехов, Савчук, 2019]).

Так, соотношение данных кластеров в русскоязычных и англоязычных статьях представляется практически равным, с незначительным перевесом в пользу работ второго выделенного кластера (45 % на 55 % и 40 % на 60 % соответственно), в то время как в диссертациях это соотношение является критическим (4 % на 96 % соответственно).

Из полученных статистических данных следует, что сегодня лингвистические корпусы, являясь простым и общедоступным инструментом для проведения широкого спектра исследований и решения практических задач (например, для целей перевода, которому посвящены пять трудов из массива русскоязычных статей), используются не только корпусными лингвистами, но и другими специалистами гуманитарного профиля, которые также вносят свой вклад в развитие частных принципов применения лингвистических корпусов.

Следовательно, предлагается следующее актуализированное определение корпусной лингвистики: «раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов, а также описанием и унификацией существующих способов их применения».

КЛАСТЕРЫ «КОРПУСЫ В ЛИНГВИСТИКЕ» И «КОРПУСЫ В ДРУГИХ НАУКАХ»

Отечественный специалист в области теории познания И. Т. Касавин утверждает, что «междисциплинарное взаимодействие есть естественное состояние науки» [Касавин, 2006, с. 7]; современная корпусная лингвистика, согласно полученным в ходе кластеризации данным, является наглядным тому подтверждением. Так, 10 % русскоязычных статей, 20 % англоязычных статей и 26 % диссертаций, хотя и посвящены лингвистическим корпусам, относятся к другим областям научных знаний.

Наиболее частотными в отобранном материале являются научные тексты по таким специальностям, как теоретические основы информатики, а также математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей (на них приходится 66 % от русскоязычных нелингвистических статей, 69 % от нелингвистических диссертаций и 83 % от англоязычных нелингвистических статей). Главное отличие от работ по компьютерной лингвистике заключается в том, что в них лингвистический корпус является не объектом исследования, а лишь преобразованной базой данных естественного языка. Например, И. А. Шилин в кандидатской диссертации «Метод и алгоритмы интерпретации неполных высказываний пользователя для управления устройствами Интернета вещей на основе онтологического подхода» (ВАК РФ 05.13.17) описывает создание экспериментального корпуса пользовательских высказываний, который служит основой архитектуры программной системы для взаимодействия с устройствами Интернета вещей [Шилин, 2019]. Примечательно, что в англоязычных статьях встречается такое понятие, как

псевдокорпус (*pseudo corpus*), под которым понимается массив данных на естественном языке, не подразумевающий аннотирования [Yamanishi et al., 2022].

Вторым по частотности направлением является информационная безопасность, на которую приходится 10 % от русскоязычных нелингвистических статей и 7 % от нелингвистических диссертаций, где лингвистический корпус также выступает в качестве базы данных естественного языка, как, например, в исследовании А. М. Лаврентьева, И. В. Смирнова, Ф. Н. Соловьёва и др. «Анализ корпусов текстов террористической и антиправовой направленности» [Лаврентьев и др., 2019].

Остальной же массив текстов нелингвистической направленности включает в себя труды из таких областей, как литературоведение, философия религии, управление в социальных и экономических системах и этнография.

КЛАСТЕРЫ «НАЦИОНАЛЬНЫЕ КОРПУСЫ» И «СПЕЦИАЛЬНЫЕ КОРПУСЫ»

Национальные корпусы уже более полувека являются объектом и инструментом лингвистических исследований, поскольку с момента их создания первоочередной задачей являлось описание актуального состояния естественного языка в целом. Статистический анализ исходного массива показал, что национальные корпусы всё еще представляют интерес для исследователей: наиболее востребованными национальными корпусами являются Национальный корпус русского языка, НКРЯ (40 трудов), British National Corpus, BNC (7 трудов) и Corpus of Contemporary American English, COCA (4 труда); при этом продолжают развиваться и новые национальные лингвистические корпусы, примером чему может послужить исследование «О создании национальных корпусов тюркских языков» [Сиразитдинов, Садыков, 2021].

Однако полученные в ходе кластеризации статистические данные также показали, что на настоящий момент исследователи больше заинтересованы не в национальных, а в специальных корпусах. Так, из всего массива русскоязычных статей только 34 посвящены национальным корпусам, в то время как специальным практически в два раза больше (60 статей); в диссертациях 18 трудов посвящены национальным корпусам и 25 специальным; в англоязычных статьях количественный разрыв между национальными и специальными корпусами еще более наглядный: 1 к 19 соответственно. Подобное соотношение говорит нам о том, что сегодня исследователи заинтересованы в проведении более узкоспециализированных исследований,

которые не могут быть реализованы на базе национального корпуса или, например, на базе основного корпуса в случае НКРЯ.

Таким образом, возрастает интерес, во-первых, к созданию собственных исследовательских корпусов, которые составляют 17 % от числа русскоязычных статей, 64 % от числа диссертаций и 100 % от числа англоязычных статей, посвященных специальным корпусам, во-вторых, к использованию подкорпусов НКРЯ, из которых наиболее востребованными являются параллельные (9 публикация), МУРКО (4 публикации) и акцентологические (2 публикации), в-третьих, к созданию новых специальных корпусов, которые бы не только решали поставленную в рамках конкретного исследования задачу, но и давали бы материал для целых новых направлений исследований, как например, в работе «Корпус русского языка повседневного общения 'Один речевой день' (ОРД): текущее состояние и перспективы» Н. В. Богдановой-Бегларян, О. В. Блиновой, Г. Я. Мартыненко и Т. Ю. Шерстиновой, в которой идет речь о создании корпуса повседневной русской речи с уникальной методологической основой, заключающейся в осуществлении звукозаписей посредством информантов-добровольцев, «готовых прожить день с диктофоном на шее» [Богданова-Бегларян и др., 2019].

КЛАСТЕР «КОРПУСЫ В ИССЛЕДОВАНИЯХ ИНОСТРАННЫХ ЯЗЫКОВ»

Постоянно растущий уровень глобализации¹ затрагивает и современные корпусные исследования, что подтвердили результаты статического анализа собранного массива: 20 % всех англоязычных статей, 22 % всех русскоязычных статей и 54 % всех диссертаций были связаны с исследованием иностранных языков с помощью лингвистических корпусов. Анализ частотности обращений к тому или иному иностранному языку показал следующее их соотношение:

- русскоязычные статьи: английский (38,64 %), немецкий (18,18 %), испанский (9,09 %), китайский, японский (по 6,82 %), итальянский, казахский, таджикский (по 4,55 %), чешский, киргизский и иврит (по 2,27 %);
- диссертации: английский (46,15 %), китайский (23,08 %), немецкий (19,23 %), французский (11,54 %), испанский (7,69 %), казахский и таджикский (по 3,85 %);
- англоязычные статьи: арабский, урду, китайский, французский, итальянский (по 20 %).

¹ URL: <https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html>

Таким образом, наиболее частотными иностранными языками в корпусных исследованиях являются английский, китайский и немецкий. Примечательно, что все публикации, связанные с иностранными языками, являются лингвистическими исследованиями, в которых в большинстве случаев корпус выступает в качестве инструмента, а не рассматривается как объект анализа (в русскоязычных статьях: 38 % и 62 % соответственно; в англоязычных статьях: 20 % и 80 % соответственно; в диссертациях все публикации данного кластера посвящены использованию корпуса в качестве инструмента).

Наиболее востребованными для проведения подобных исследований являются национальные и специальные одноязычные корпуса (75,56 % русскоязычных статей, 80 % англоязычных статей и 92,31 % диссертаций), в остальных случаях используются параллельные корпуса, которые могут служить материалом для исследования не только иностранного, но и родного языка, как, например, в исследовании Д. О. Добровольского и А. А. Зализняка «Русские конструкции с потенциально модальным значением по данным параллельных корпусов» [Добровольский, Зализняк, 2020].

КЛАСТЕРЫ «КОРПУСЫ И МИНОРИТАРНЫЕ ЯЗЫКИ», «КОРПУСЫ И ДИАЛЕКТЫ»

Согласно М. П. Безеновой и Г. Л. Григорьеву, чье исследование стало частью отобранного массива, «в последнее время созданию корпусов текстов на языках народов России также уделяется особое внимание» [Безенова, Григорьев, 2020], что подтверждается результатами проведенной кластеризации – исследованию миноритарных языков посвящено 8 % русскоязычных статей, 6 % диссертаций и 3,33 % англоязычных статей всего собранного материала.

В частности, русскоязычные статьи посвящены таким миноритарным языкам, как бурятский, башкирский, удмуртский, шорский, эрзянский, мокшанский, тувинский, калмыцкий, хакасский, якутский, татарский, карельский, вепсский; в диссертациях исследования затрагивают татарский и цыганский языки; в англоязычных статьях были выявлены труды, посвященные валлийскому языку.

Диапазон создаваемых и используемых корпусов в данном направлении исследований достаточно обширен: национальные корпуса, параллельные, диахронические, фонетические и речевые.

Результат кластеризации позволил выделить еще одно близкое к исследованиям миноритарных языков направление – исследование диалектов, которым посвящено 9 % от всего массива русскоязычных статей. Содержащие в данном кластере

исследования посвящены диалектам русского, башкирского, бурятского, якутского языков, а также диалектам Северного Приангарья, Среднего Прииртышья и российских немцев.

Наиболее востребованным для исследований русских диалектов является Томский диалектный корпус (38 % от данного кластера), в то время как во всем собранном массиве обнаружено только одно исследование на материале диалектного подкорпуса НКРЯ. В остальных же случаях исследователи отдают предпочтение созданию собственного корпуса, как, например, в труде «Региолекты русского жестового языка: мультимодальный электронный корпус (на материале коммуникативного пространства Восточной Сибири)» [Куликова, Шахотина, 2021]; примечательно, что только это исследование во всем отобранном массиве научных текстов было посвящено мультимодальному корпусу и корпусу РЖЯ.

Однако сопоставление статистических данных показало, что миноритарным языкам и диалектам уделяется существенно меньше внимания по сравнению с иностранными языками (процент исследований, посвященных миноритарным языкам и диалектам, по отношению ко всему массиву не доходит и до 10 %, в то время как иностранным языкам посвящено не менее 20 % исследований).

КЛАСТЕР «КОРПУСЫ В ПРЕПОДАВАНИИ»

Цифровизация представляется одной из главных тенденций современной экономики¹, что не может обойти стороной и сферу образования². Анализ собранного массива показал, что лингвистические корпуса становятся одним из инструментов цифровизации преподавания.

Так, данной теме посвящено 15 % всех русскоязычных статей, 20 % англоязычных и 3 диссертации. Выделяется четыре основных направления:

1. Корпусы в преподавании иностранных языков;
2. Корпусы в преподавании родного языка;
3. Корпусы в преподавании родного языка как иностранного;
4. Корпусы в преподавании перевода.

В русскоязычных статьях первому направлению уделяется 50 % всего кластера корпусов в преподавании, второму и третьему – чуть больше 15 % каждому, четвертому – 20 %.

В англоязычных статьях первому направлению уделяется 33 % всего кластера корпусов в преподавании, второму – чуть больше 15 %, третьему – 50 %, четвертому – отсутствует вовсе в указанном массиве.

Таким образом, использование лингвистических корпусов наиболее описано в трудах, посвященных преподаванию иностранных языков; вторым по степени проработанности темы представляется преподавание родного (русского и английского) как иностранного; преподавание родного языка и преподавание переводу являются наименее представленными в отобранном массиве.

Методики применения лингвистических корпусов в преподавании, описанные в исследуемом массиве, направлены на использование не только в рамках общего практического курса преподавания языков, но и отдельных его аспектов, например, грамматики и аудирования (как, например, в исследовании «К вопросу о модели обучения аудированию на материалах национального корпуса русского языка» [Фидчук, 2021]).

Для целей преподавания исследователи предлагают использовать национальные корпуса, параллельные корпуса, учебные корпуса, а также корпуса профессиональных текстов и звучащей речи.

ЗАКЛЮЧЕНИЕ

Проведенная в рамках исследования кластеризация, а также статистический анализ полученных результатов позволили сформулировать следующие существующие тенденции корпусной лингвистики:

- лингвистические корпуса являются общедоступным инструментом для проведения исследований широкого спектра, зачастую выходящих за рамки лингвистики;
- лингвистические корпуса также активно используются в таких направлениях, как теоретические основы информатики, однако в этой области они нередко редуцируются до псевдокорпусов;
- исследователи продолжают обращаться к национальным корпусам, тем не менее в связи с потребностью в проведении более узкоспециализированных исследований возрастает интерес к специальным корпусам;
- лингвистические корпуса активно используются для исследований иностранных языков, в то время как миноритарным языкам и диалектам уделяется существенно меньше внимания;
- лингвистические корпуса становятся инструментом цифровизации преподавания, в частности, для повышения навыков грамматики и аудирования.

¹ URL: https://digital.gov.ru/ru/activity/directions/858/?utm_referrer=https%3a%2f%2fyandex.ru%2f

² URL: <https://aiedu.hse.ru/mirror/pubs/share/308201188>

В то же время проведенное исследование позволяет сформулировать потенциально перспективные направления для дальнейшего развития корпусной лингвистики:

- описание и унификация существующих способов применения лингвистических корпусов;
- создание узкоспециальных корпусов для лингвистических целей;
- создание лингвистических корпусов для не-лингвистических целей;
- исследование параллельных корпусов для расширения представлений о родном языке;
- введение практики использования лингвистических корпусов в рамках языковой политики по исследованию и сохранению миноритарных языков и диалектов;
- применение лингвистических корпусов для преподавания родного языка;
- создание мультимодальных корпусов и корпусов русского жестового языка.

СПИСОК ИСТОЧНИКОВ

1. Солнышкина М. И., Гатиятуллина Г. М. История развития корпусной лингвистики (на примере англоязычных корпусов) // Вестник Томского государственного университета. Филология. 2020. № 63. С. 132–160.
2. Захаров В.П. Прологомены к корпусной лингвистике // Вопросы психолингвистики. 2016. № 28. С. 150–161.
3. Biber D., Reppen R. The Cambridge Handbook of English Corpus Linguistics. Cambridge: Cambridge University Press, 2015.
4. Захаров В. П. Корпусная лингвистика: учебно-методическое пособие. СПб., 2005.
5. Leedham M., Lillis T., Twiner A. Creating a corpus of sensitive and hard-to-access texts: Methodological challenges and ethical concerns in the building of the WiSP Corpus // Applied corpus linguistics. 2021. Vol. 1. № 3. DOI <https://doi.org/10.1016/j.acorp.2021.100011>
6. Орехов Б. В., Савчук С. О. Акцентологический корпус как инструмент для исследования русского ударения // Труды Института русского языка им. В. В. Виноградова. 2019. № 21. С. 61–63.
7. Касавин И. Т. Дискурс-анализ как междисциплинарный метод гуманитарных наук // Эпистемология и философия науки. 2006. Т. 10. № 4. С. 5–16.
8. Шилин И. А. Метод и алгоритмы интерпретации неполных высказываний пользователя для управления устройствами Интернета вещей на основе онтологического подхода: дис. ... канд. техн. наук. СПб., 2019.
9. Yamanishi R. et al. Applying Existing Datasets as a Pseudo Corpus for Sentiment Representation on Social Media // Procedia Computer Science. 2022. Vol. 207. P. 335–342.
10. Лаврентьев А. М. [и др.]. Анализ корпусов текстов террористической и антиправовой направленности // Вопросы кибербезопасности. 2019. № 4 (32). С. 54–60.
11. Сиразитдинов З. А., Садыков Т. С. О создании национальных корпусов тюркских языков // ALATOO ACADEMIC STUDIES. 2021. № 2. С. 245–255.
12. Богданова-Бегларян Н. В. [и др.]. Корпус русского языка повседневного общения «Один речевой день» (ОРД): текущее состояние и перспективы // Труды Института русского языка им. В. В. Виноградова. 2019. № 21. С. 100–110.
13. Добровольский Д. О., Зализняк А. А. Русские конструкции с потенциально модальным значением по данным параллельных корпусов // Труды Института русского языка им. В. В. Виноградова. 2020. № 3 (25). С. 35–49.
14. Безенова М. П., Григорьев Г. Л. Роль программы проверки удмуртской орфографии в пополнении национального корпуса удмуртского языка // Ежегодник финно-угорских исследований. 2020. V. 14. № 3. С. 549–556.
15. Куликова Л. В., Шахотина С. А. Региолекты русского жестового языка: мультимодальный электронный корпус (на материале коммуникативного пространства Восточной Сибири) // Вестник Санкт-Петербургского университета. Язык и литература. 2021. Т. 18. № 4. С. 750–759.
16. Фидчук М. Ю. К вопросу о модели обучения аудированию на материалах национального корпуса русского языка // Высшее образование сегодня. 2021. № 2. С. 59–62.

REFERENCES

1. Solnyshkina, M. I., Gatiyatullina, G. M. (2020). The History of Corpus Linguistics (on the example of the English language corpora). Tomsk State University Journal of Philology, 63, 132–160. (In Russ.)

2. Zakharov, V. P. (2016). Prolegomena to Corpus Linguistics. Issues of psycholinguistics, 28, 150–161. (In Russ.)
3. Biber, D., Reppen, R. (2016). The Cambridge Handbook of English Corpus Linguistics. Cambridge: Cambridge University Press.
4. Zakharov, V. P. (2005). Korpusnaya lingvistika: Uchebno-metod. posobiye = Corpus Linguistics: Textbook. St. Petersburg. (In Russ.)
5. Leedham, M., Lillis, T., Twiner, A. (2021). Creating a corpus of sensitive and hard-to-access texts: Methodological challenges and ethical concerns in the building of the WiSP Corpus. Applied corpus linguistics, 1(3). <https://doi.org/10.1016/j.acorp.2021.100011>
6. Orekhov, B. V., Savchuk, S. O. (2019). Accentological corpus as a tool for studying Russian word stress. Proceedings of the V. V. Vinogradov Russian Language Institute, 21, 61–63. (In Russ.)
7. Kasavin, I. T. (2006). Diskurs-analiz kak mezhdistsiplinarnyy metod gumanitarnykh nauk = Discourse-analysis as an interdisciplinary method of the humanities. Epistemology and Philosophy of Science, 10(4), 5–16. (In Russ.)
8. Shilin, I. A. (2019). Metod i algoritmy interpretatsii nepolnykh vyskazyvanij pol'zovatelya dlya upravleniya ustroystvami Interneta veshchej na osnove ontologicheskogo podhoda = Method and algorithms for interpreting incomplete user statements to control Internet of Things devices based on an ontological approach: PhD in Technical Sciences. St. Petersburg. (In Russ.)
9. Yamanishi, R., et al. (2022). Applying Existing Datasets as a Pseudo Corpus for Sentiment Representation on Social Media. Procedia Computer Science, 207, 335–342.
10. Lavrentiev, A. M. et al. (2019). Analysis of corpus of extremist texts and unlawful texts. Cybersecurity issues, 4(32), 54–60. (In Russ.)
11. Sirazitdinov, Z. A., Sadykov, T. S. (2021). O sozdanii natsional'nykh korpusov tyurkskikh yazykov = On the creation of national corpus of Turkic languages. ALATOO ACADEMIC STUDIES, 2, 245–255. (In Russ.)
12. Bogdanova-Beglarian, N. V. et al. (2019). Russian Everyday Speech Corpus “One Day of Speech”: Current State and Perspectives. Proceedings of the V. V. Vinogradov Russian Language Institute, 21, 100–110. (In Russ.)
13. Dobrovol'skij, D. O., Zalizniak, A. A. (2020). Russian constructions with potentially modal meanings: an analysis based on parallel corpus data. Proceedings of the V. V. Vinogradov Russian Language Institute, 3(25), 35–49. (In Russ.)
14. Bezenova, M. P., Grigoriev, G. L. (2020). The Role of the Udmurt Spell Checker in Replenishment of the Udmurt National Corpus. Yearbook of Finno-Ugric Studies, 14(3), 549–556. (In Russ.)
15. Kulikova, L. V., Shakhovina, S. A. (2021). Regiolects of the Russian Sign Language: A Multimodal Electronic Corpus (based on the communicative space of Eastern Siberia). Vestnik of Saint Petersburg University. Language and Literature, 18(4), 750–759. (In Russ.)
16. Fidchuk, M. Y. (2021). Revising the question of the model to teach listening comprehension skills based on the Russian National Corpus materials. Higher Education Today, 2, 59–62. (In Russ.)

ИНФОРМАЦИЯ ОБ АВТОРЕ

Изюмская Вероника Викторовна

аспирантка кафедры прикладной и экспериментальной лингвистики
младший научный сотрудник лаборатории корпусной лингвистики
Московского государственного лингвистического университета.

INFORMATION ABOUT THE AUTHOR

Izyumskaya Veronika Viktorovna

Postgraduate student at Applied and Experimental Linguistics Department,
Junior Researcher at the Laboratory of Corpus Linguistics, Moscow State Linguistic University

Статья поступила в редакцию
одобрена после рецензирования
принята к публикации

15.05.2023
21.06.2023
22.06.2023

The article was submitted
approved after reviewing
accepted for publication