

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНСТИТУТ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ РАН  
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ  
ИМ. А. И. ГЕРЦЕНА

---

ТРУДЫ  
МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ  
«КОРПУСНАЯ ЛИНГВИСТИКА — 2023»

**21-23 июня 2023 г.**  
**Санкт-Петербург**



ST. PETERSBURG STATE UNIVERSITY  
INSTITUTE FOR LINGUISTIC STUDIES (RAS)  
HERZEN STATE PEDAGOGICAL UNIVERSITY OF RUSSIA

PROCEEDINGS  
OF THE INTERNATIONAL CONFERENCE  
«CORPUS LINGUISTICS — 2023»

June 21–23, 2023, St. Petersburg

SAINT PETERSBURG UNIVERSITY PRESS  
2024



САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНСТИТУТ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ РАН  
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ  
ИМ. А. И. ГЕРЦЕНА

ТРУДЫ  
МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ  
«КОРПУСНАЯ ЛИНГВИСТИКА — 2023»

21–23 июня 2023 г., Санкт-Петербург



ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА

2024

ББК 81.1  
Т78

Ответственные редакторы:  
М. В. Хохлова, А. О. Гребенников, Т. Г. Скрещцова

*Рекомендовано к публикации Научной комиссией  
в области наук о языках и литературе  
Санкт-Петербургского государственного университета*

**Труды международной конференции «Корпусная лингвистика — 2023».** — СПб.: Изд-во С.-Петерб. ун-та, 2024. — 320 с.

Сборник содержит материалы докладов, представленных на Международной научной конференции «Корпусная лингвистика — 2023» 21–23 июня 2023 г. в Санкт-Петербурге.

Создание корпусов текстов является одним из приоритетных направлений в современной лингвистике. Материалы конференции по данной тематике знакомят ученых с современными разработками и новыми технологическими решениями в этой области, а также способствуют обобщению опыта научных исследований по корпусной лингвистике.

Сборник адресован специалистам по прикладной лингвистике и обработке естественного языка, а также может быть интересен широкому кругу лингвистов.

ББК 81.1

---

---

## ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ.....	9
<i>Н. Л. Аванесян, В. В. Зенькова, А. М. Чеповский</i> АНАЛИЗ КОРПУСОВ ТЕКСТОВ СООБЩЕСТВ В TELEGRAM.....	11
<i>В. А. Баранов</i> СЕМАНТИЧЕСКИЕ АНАЛОГИ СЛОВ <i>БРАНЬ</i> И <i>РАТЬ</i> (ПО ДАННЫМ ИСТОРИЧЕСКОГО КОРПУСА СРЕДНЕВЕКОВЫХ СЛАВЯНСКИХ РУКОПИСЕЙ).....	18
<i>А. А. Баркович</i> СЕМАНТИЧЕСКАЯ СТРУКТУРИЗАЦИЯ ЭКОЛОГИЧЕСКОГО ДИСКУРСА: ПОТЕНЦИАЛ КОРПУСНОЙ ИНТЕРПРЕТАЦИИ.....	28
<i>Н. В. Богданова-Бегларян</i> VASKSCHANNEL И ПЕРЕСПРОС В УСТНОМ ДИАЛОГЕ: СООТНОШЕНИЕ ПОНЯТИЙ И ВАРИАНТЫ РЕАЛИЗАЦИИ.....	35
<i>А. С. Васелюк, Д. В. Демидов, А. А. Потемкина</i> ФИЛЬТРАЦИЯ И БИНАРИЗАЦИЯ ИЗОБРАЖЕНИЙ ТЕКСТОВ СЛАВЯНСКИХ РУКОПИСНЫХ КНИГ.....	43
<i>А. А. Гончаров, Д. О. Добровольский, Анна А. Зализняк</i> БАЗА ДАННЫХ КОНСТРУКЦИЙ С НЕМЕЦКИМИ МОДАЛЬНЫМИ ГЛАГОЛАМИ И ИХ РУССКИХ СООТВЕТСТВИЙ.....	51
<i>М. А. Гриц, М. С. Коган</i> АЛГОРИТМ РАЗРАБОТКИ УЧЕБНОГО ГЛОССАРИЯ ПРЕДМЕТНОЙ ОБЛАСТИ (НА ПРИМЕРЕ ПРЕДМЕТНОЙ ОБЛАСТИ «ПРОМЫШЛЕННАЯ БЕЗОПАСНОСТЬ»).....	61
<i>Д. В. Демидов, А. Г. Кравецкий, А. А. Ларионов, А. А. Плетнёва</i> СОЗДАНИЕ КОРПУСА РУКОПИСНОГО НАСЛЕДИЯ ДРЕВНЕЙ РУСИ.....	72
<i>Д. В. Демидов, В. В. Миронова</i> ПРЕДОБРАБОТКА ИЗОБРАЖЕНИЙ БУКВИЦ В СЛАВЯНСКИХ РУКОПИСНЫХ КНИГАХ.....	82
<i>Д. О. Добровольский, И. Б. Левонтина</i> СОПОСТАВИТЕЛЬНОЕ КОРПУСНОЕ ИССЛЕДОВАНИЕ ГЛАГОЛА <i>ОСКОРБИТЬ</i> И ЕГО НЕМЕЦКИХ ВАРИАНТОВ ПЕРЕВОДА: НА МАТЕРИАЛЕ ПАРАЛЛЕЛЬНОГО КОРПУСА НКРЯ.....	89

<i>О. В. Дони́на, Д. А. Фу́рсина, Н. С. Горбунов</i>	
СОЗДАНИЕ РЕГИОНАЛЬНОГО ПОДКОРПУСА: ОТ ИДЕИ ДО ВОПЛОЩЕНИЯ..	97
<i>И. А. Котюрова, Г. Р. Сафонов</i>	
МЕТАДАННЫЕ В ПЕТРОЗАВОДСКОМ АННОТИРОВАННОМ КОРПУСЕ ТЕКСТОВ .....	103
<i>Г. И. Кустова</i>	
КОРПУС КАК ИНСТРУМЕНТ ИССЛЕДОВАНИЙ: ВАРИАТИВНОСТЬ КОНСТРУКЦИЙ С СОЮЗОМ ЧТОБЫ .....	111
<i>А. Н. Лапошина, М. Ю. Лебедева</i>	
ПЕДАГОГИЧЕСКИЙ КОРПУС НЕПОДГОТОВЛЕННОЙ МОНОЛОГИЧЕСКОЙ УСТНОЙ РЕЧИ ИЗУЧАЮЩИХ РУССКИЙ ЯЗЫК КАК ИНОСТРАННЫЙ: ПИЛОТНОЕ ИССЛЕДОВАНИЕ .....	119
<i>М. Н. Лошанина</i>	
ПРОГРАММА MAHQDA КАК ИНСТРУМЕНТ КОРПУСНОГО АНАЛИЗА ДИСКУРСИВНЫХ МАРКЕРОВ ПРОТИВОРЕЧИЯ .....	126
<i>П. С. Макарова, Е. М. Гриднева, А. А. Иваненко, Н. С. Здорова, М. А. Грабовская</i>	
БАЗА РУССКИХ ИДИОМ (БРИ): ОПИСАНИЕ ПАРАМЕТРА ПРЕДСКАЗУЕМОСТИ И МЕСТА РАСПОЗНАВАНИЯ ИДИОМЫ .....	135
<i>О. А. Митрофанова</i>	
ПОИСК И РАНЖИРОВАНИЕ ТЕКСТОВ В СПЕЦИАЛЬНОМ КОРПУСЕ НА ОСНОВЕ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ .....	143
<i>А. Д. Москвина, М. А. Кирина</i>	
ОТ ЛЮБВИ ДО НЕНАВИСТИ: РАСПРЕДЕЛЕНИЕ ЭМОЦИОНАЛЬНОЙ ЛЕКСИКИ В РУССКОМ РАССКАЗЕ НАЧАЛА XX ВЕКА .....	156
<i>Х. Г. Нигматов, Б. Р. Менглиев, Ш. М. Хамраева</i>	
ЛИНГВИСТИЧЕСКИЕ ОСНОВЫ СОЗДАНИЯ МОРФОЛЕКСИКОНА, ИЛИ О МОДЕРНИЗИРОВАННОЙ КЛАССИФИКАЦИИ ЧАСТЕЙ РЕЧИ УЗБЕКСКОГО ЯЗЫКА .....	167
<i>И. С. Пименов, Н. В. Саломатина, А. С. Засыпкин</i>	
МАРКЕРНЫЙ ПОДХОД К АВТОМАТИЗАЦИИ ИЗВЛЕЧЕНИЯ АРГУМЕНТАТИВНЫХ СТРУКТУР ИЗ НАУЧНЫХ СТАТЕЙ .....	184
<i>Р. А. Родионов</i>	
СТИЛОМЕТРИЧЕСКИЙ АНАЛИЗ РУССКОГО РАССКАЗА НАЧАЛА XX ВЕКА: ВЫЯВЛЕНИЕ ГЕНДЕРНОГО СИГНАЛА .....	191

<i>С. О. Савчук</i>	
НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА НАКАНУНЕ ЮБИЛЕЯ: КРАТКИЙ ОБЗОР ИННОВАЦИЙ .....	198
<i>Е. Г. Соколова</i>	
СОСТАВЛЯЕТ ЛИ МОРФОЛОГИЧЕСКИЙ ИМПЕРФЕКТИВ ВИДОВУЮ ПАРУ С ИСХОДНЫМ ГЛАГОЛОМ ИЗ ОДНОЙ ГРУППЫ ЗА-ИНХОАТИВОВ? .....	217
<i>А. И. Фокина, А. М. Чеповский</i>	
АНАЛИЗ КОРПУСА ПОЭТИЧЕСКИХ ТЕКСТОВ НА ПЛАТФОРМЕ TXM .....	224
<i>Т. Ю. Шерстинова, А. Д. Москвина, М. А. Кирина, А. С. Карышева, Е. О. Колпащикова, П. И. Максименко, Р. А. Родионов, А. Р. Сейнова</i>	
НЕСЧАСТЛИВЫ ПО-СВОЕМУ: КАК ИЗМЕРИТЬ ТОНАЛЬНОСТЬ ЛИТЕРАТУРНОГО ТЕКСТА? .....	232
<i>О. I. Babina, A. Y. Zinovyeva, E. D. Nerucheva</i>	
CONCEPTUAL TAGGING OF E-NEWS CORPUS WITH FINE-TUNED PRE-TRAINED ENCODER MODELS .....	241
<i>A. V. Belyi, O. A. Mitrofanova, N. A. Dubinina</i>	
DISTRIBUTIVE SEMANTIC MODELS IN LANGUAGE LEARNING: AUTOMATIC GENERATION OF LEXICAL-GRAMMATICAL TESTS FOR RUSSIAN AS A FOREIGN LANGUAGE .....	252
<i>A. O. Grebennikov, T. G. Skrebtsova</i>	
RUSSIAN SHORT STORIES: CHANGES IN WORD GROUP FREQUENCIES OVER THE LAST CENTURY .....	260
<i>М. I. Kiose, S. D. Makeev, A. A. Rzheshhevskaya, A. I. Izmalkova</i>	
ARCHITECTURE AND SEMANTIC PROTOCOL OF GAZE BEHAVIOR CORPUS MULTICORTEXT .....	267
<i>М. V. Koryshev, L. A. Kulikova, K. V. Mazin, A. V. Mushulova, D. A. Starodub, M. V. Khokhlova</i>	
THE IDEOLOGICAL SCENERY OF MODERN GERMANY: A CORPUS ANALYSIS OF PRINT MEDIA .....	276

## **СТУДЕНЧЕСКАЯ СЕКЦИЯ**

<i>О. У. Серова</i>	
ИССЛЕДОВАНИЕ ФЕНОМЕНА TRANSLATIONESE (НА МАТЕРИАЛЕ МАШИННОГО ПЕРЕВОДА НАУЧНЫХ ТЕКСТОВ С АНГЛИЙСКОГО НА РУССКИЙ ЯЗЫК).....	286

<i>Д. Д. Сухан, В. С. Павликова</i>	
ОПЫТ РЕАЛИЗАЦИИ СИСТЕМЫ РАЗРЕШЕНИЯ СИНТАКСИЧЕСКОЙ НЕОДНОЗНАЧНОСТИ В ПАРСЕРЕ NLTK4RUSSIAN.....	294
<i>А. В. Трошина</i>	
КРАТКОЕ СОДЕРЖАНИЕ ХУДОЖЕСТВЕННОГО ПРОИЗВЕДЕНИЯ КАК КОНТРОЛЬНЫЙ МАТЕРИАЛ ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ РАБОТЫ АЛГОРИТМОВ ПО ВЫДЕЛЕНИЮ КЛЮЧЕВЫХ СЛОВ .....	304
<i>С. В. Чеповецкая</i>	
«ПРОБЛЕМА СОСТОИТ» VS. «ПРОБЛЕМА ЗАКЛЮЧАЕТСЯ» В ТЕКСТАХ НАУЧНОГО СТИЛЯ (ПО КОРПУСНЫМ ДАННЫМ) .....	312

## ПРЕДИСЛОВИЕ

Настоящий сборник содержит статьи, представленные в рамках 11-й международной конференции «Корпусная лингвистика 2023 (Corpora 2023)», которая прошла с 21 по 23 июня 2023 года в Санкт-Петербургском государственном университете, и его публикация в нынешнем 2024 году является определенно важным событием, поскольку год ознаменован празднованием 300-летия Санкт-Петербургского государственного университета.

Первая конференция, заложившая основу многолетней традиции, носила название «Корпусная лингвистика и лингвистические базы данных» и состоялась в СПбГУ далеком 2002 году. Направление, связанное с созданием и использованием корпусов, тогда еще только зарождалось. Прошло более 20 лет, и мы видим, как стремительно развиваются корпусы текстов благодаря современным технологиям. Можно сказать, что на наших глазах появилось множество проектов, которые впервые были представлены на этих конференциях. Корпусная лингвистика не стоит на месте: по-прежнему большое внимание уделяется вопросу использования и обработке корпусных данных в лингвистических исследованиях, а также обсуждаются перспективы и результаты применения нейросетевых технологий. 10-я конференция по корпусной лингвистике, которая проводилась в 2021 году, должна была стать по-настоящему юбилейной, однако пандемия внесла свои коррективы, и мероприятие прошло онлайн, не позволив нам встретиться лично в Санкт-Петербурге в уже ставшее традиционным для нас время в конце июня, в период белых ночей.

Нам хотелось бы вспомнить наших учителей и вдохновителей конференции — почетного профессора и заведующего кафедрой математической лингвистики СПбГУ Александра Сергеевича Герда и доцента кафедры математической лингвистики СПбГУ Виктора Павловича

Захарова, которые сделали так много для становления конференции, и вклад которых сложно переоценить. Благодаря их организаторским усилиям конференция стала площадкой для встреч ученых и обмена опытом с коллегами из разных городов Российской Федерации, ближнего и дальнего зарубежья. Некоторые из участников стали добрыми друзьями конференции, с которыми мы с большой радостью встречаемся каждые два года. От лица программного комитета хотим поблагодарить наших многочисленных авторов, которые принимали участие в конференциях прошлых лет, а также отдельно авторов настоящего сборника, которые терпеливо дождались его выхода в свет. Мы надеемся на продолжение сотрудничества и будем рады увидиться в следующем году.

*Ответственные редакторы*