

Компьютерная лингвистика и интеллектуальные технологии

По материалам ежегодной международной конференции
«Диалог» (2025)

Выпуск 23

Computational Linguistics and Intellectual Technologies

Papers from the Annual International Conference "Dialogue" (2025)

Issue 23

Редакционная
коллегия: *В. П. Селегей (главный редактор), В. И. Беликов, И. М. Богуславский,
Б. В. Добров, Д. О. Добровольский, И. М. Кобозева, Е. Б. Козеренко,
М. А. Кронгауз, Н. В. Лукашевич, Д. Маккарти, П. Наков, Й. Нивре,
В. Раскин, Э. Хови, Т. О. Шаврина, С. А. Шаров, Т. Е. Янко*

Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог». Вып. 23. 2025. С. 1–492.

Сборник включает 39 докладов международной конференции по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2025», представляющих широкий спектр теоретических и прикладных исследований в области описания естественного языка, моделирования языковых процессов, создания практически применимых компьютерных лингвистических технологий.

Для специалистов в области теоретической и прикладной лингвистики и интеллектуальных технологий.

Предисловие

23-й выпуск ежегодника «Компьютерная лингвистика и интеллектуальные технологии» содержит избранные материалы 30-й международной онлайн-конференции «Диалог». В 2025 году для публикации в основном томе сборника редколлегией были отобраны 39 докладов. Работы, представленные в сборнике, отражают те направления исследований в области компьютерного моделирования и анализа естественного языка, которые по традиции представляются на Диалоге:

- **Интеллектуальный анализ документов;**
- **Лингвистический анализ текста;**
- **Глубокое обучение в компьютерной лингвистике:** методики и примеры применения в лингвистических исследованиях, содержательная интерпретация работы LLM;
- **Компьютерные лингвистические ресурсы:** включая новые датасеты и новые сценарии и типы разметки, Evaluation Benchmarks;
- **Компьютерный анализ Social Media;**
- **Корпусная лингвистика и корпусометрия:** методики создания, использования и оценки корпусов;
- **Компьютерная семантика:** аналитические и дистрибуционные модели, связь между ними;
- **Лингвистические онтологии и автоматическое извлечение знаний;**
- **Мультимодальная коммуникация:** аналитические и нейронные модели речевого акта;
- **Модели общения и диалоговые агенты;**
- **Компьютерная лексикография;**
- **Многоязычие:** language transfer, новые технологии работы с малоресурсными языками.

В соответствии с традициями «Диалога» отбор работ основывается на представлении о важности соединения новых методов и технологий анализа языковых данных с полноценным лингвистическим анализом. Диалог является де-факто крупнейшим форумом по проблемам создания современных компьютерных ресурсов, моделей и технологий для русского языка, поэтому ключевым событием «Диалога» является подведение итогов технологических соревнований между разработчиками систем лингвистического анализа русскоязычных текстов — *Dialogue Evaluation*. В сборнике представлены статьи участников и организаторов 3-х соревнований:

- **CoBaLD Parsing:** Соревнование по автоматической лингвистической разметке;
- **RuOpinionNE-2024:** Соревнование по извлечению мнений из новостных текстов;
- **RuTermEval-2024:** Соревнование по выявлению терминов в научных статьях на русском языке.

Статьи в сборнике публикуются на русском и английском языках. При выборе языка публикации действует следующее правило:

- доклады по компьютерной лингвистике подаются на английском языке. Это расширяет их аудиторию и позволяет привлекать к рецензированию международных экспертов;
- доклады, посвященные лингвистическому анализу русского языка, предполагающие знание этого языка у читателя, подаются на русском языке (с обязательной аннотацией на английском).

Несмотря на традиционную широту тематики представленных на конференции и отобранных в сборник докладов, они не могут дать полной картины направлений «Диалога». Ее можно получить с помощью сайта конференции <https://dialogue-conf.org/>, на котором представлены обширные электронные архивы «Диалогов» последних лет и все результаты проведенных тестирований Dialogue Evaluation.

Мы обращаем внимание авторов и читателей сборника, что с 2018 года Редсовет отказался от печати сборника на бумаге. Все сборники размещаются на сайте конференции. С 2014 года основной том индексируется Scopus.

Программный комитет конференции «Диалог»
Редколлегия сборника «Компьютерная лингвистика и интеллектуальные технологии»

Рецензенты

Баюк Александра Михайловна
Баюк Илья Сергеевич
Беликов Владимир Иванович
Богданова-Бегларян Наталья Викторовна
Богуславский Игорь Михайлович
Бухаров Ян Михайлович
Васильев Виталий Геннадьевич
Галицкий Борис Васильевич
Добров Борис Викторович
Добровольский Дмитрий Олегович
Жарков Андрей
Ильвовский Дмитрий Алексеевич
Инькова Ольга Юрьевна
Киосе Мария Ивановна
Киселева Ксения Львовна
Клышинский Эдуард Станиславович
Клячко Елена Леонидовна
Князев Сергей Владимирович
Кобозева Ирина Михайловна
Козеренко Елена Борисовна
Коротаев Николай Алексеевич
Котельников Евгений Вячеславович
Котов Артемий Александрович
Кутузов Андрей Борисович
Лапошина Антонина Николаевна
Лепехин Михаил

Левонтина Ирина Борисовна
Лобанов Борис Мефодьевич
Лукашевич Наталья Валентиновна
Ляшевская Ольга Николаевна
Мамонтова Ангелина
Митрофанова Ольга Александровна
Мичурина Мария Александровна
Никишина Ирина Юрьевна
Пазельская Анна Германовна
Переверзева Светлана Игоревна
Петрова Мария Владимировна
Пиперски Александр Чедович
Подлеская Вера Исааковна
Рыгаев Иван Петрович
Селегей Владимир Павлович
Смирнов Иван Валентинович
Смуров Иван Михайлович
Татевосов Сергей Георгиевич
Урысон Елена Владимировна
Федорова Ольга Викторовна
Феногенова Алена Сергеевна
Хохлова Мария Владимировна
Циммерлинг Антон Владимирович
Шаврина Татьяна Олеговна
Шаров Сергей Александрович
Янко Татьяна Евгеньевна

Contents¹

Akhmadeeva I., Kononenko I., Sidorova E., Shestakov V.

Using Rhetorical Structures to Analyze Argumentation in Scientific Communication Texts	1
---	---

Alibekov A., Migal A., Matenkov A., Muryshv A., Bolshakov V., Kozachenko A., Mukhtarova G., Mikhaylovskiy N.

RuTaR—A Dataset in Russian for Reasoning about Taxes	12
---	----

Baiuk I., Baiuk A., Petrova M.

CoBaLD Parser: Joint Morphosyntactic and Semantic Annotation	28
---	----

Баранов А. Н.

На отскребись: вложенная цитация как фактор идиоматичности	38
---	----

Богуславский И. М.

Совместность и раздельность	45
------------------------------------	----

Бонч-Осмоловская А. А., Гладиллин С. А., Козеренко А. Д., Ляшевская О. Н., Морозов Д. А., Кузнецова Ю. Н.,
Махова А. А., Пискунова С. В., Буйлова Н. Н., Бородин Д. Г., Виноградова И. И., Сизов В. Г., Дьяченко П. В.,
Казенников А. О., Власова Н. А., Глазкова А. В., Столяров С. С., Гарипов Т. А., Смаль И. А., Губарькова Я. Н.

Национальный корпус русского языка 2.0: корпусная платформа, инструменты анализа, нейросетевые модели разметки данных	57
--	----

Буров Д. А., Панич М. Б., Садковский Ф. А., Федорова О. В., Шевелев С. М.

Задача лексического решения: модальность имеет значение	74
--	----

Chuikova O. Iu., Gorbova E. V.

Derivation of Aktionsarten from (non-)directional verbs of motion: posing restrictions on restrictions	86
---	----

Galitsky B., Ilvovsky D., Morkovkin A.

Enhancing RAG and Knowledge Graphs with Discourse	103
--	-----

Gromenko E., Kalacheva D., Klokova K., Krongauz M., Moroz O., Shulginov V., Yudina T.

Cultural Evaluation of LLMs in Russian: Catchphrases and Cultural Types	117
--	-----

Инькова О. Ю.

Неоднозначность коннекторов: некоторые уточнения	129
---	-----

Ivleva I., Solovyev V.

An Algorithm for Genre Imbalance Correction in the Russian Subcorpus of the Google Books Ngram Corpus	137
--	-----

Khomenko A., Komratova A., Isakov D., Balba D., Shishkovskaya T., Khudyakova M.

Computer tools in mental disorders diagnostics by oral speech	147
--	-----

Киосе М., Ржешевская А., Петров А.

Локальная структура дискурса описания в речи взрослых, изучающих русский язык как иностранный	158
--	-----

* The reports of each section are ordered by the surname of the first author in compliance with the English alphabet.

Коротаев Н. А. Дискурсивные контексты наложений в диалоге на русском языке	169
Kotov A. A., Nosovets Z. A., Filatov A. A., Arinkin N. A. The development of a natural language reasoning system for a companion robot	183
Kovalev G., Tikhomirov M., Kozhevnikov E., Kornilov M., Loukachevitch N. Building Russian Benchmark for Evaluation of Information Retrieval Models	195
Lee G. F., Loukachevitch N. V., Khokhlov A. R. Generating Encyclopedic Articles Based on a Collection of Scientific Publications	206
Lepekhnin M., Sharoff S. Causal Models and Adversarial Training: Selecting the right properties for robust non-topical text classification	224
Loukachevitch N. V., Tkachenko N. M., Lapanitsyna A. M., Tikhomirov M. M., Rusnachenko N. L. RuOpinionNE-2024: Extraction of Opinion Tuples from Russian News Texts	234
Mamontova A., Ischenko R., Vorontsov K. RuTermEval-2024: Cross-domain Automatic Term Extraction and Classification in Russian scientific texts	245
Morozov D., Glazkova A., Garipov T. BERT-like Models for Automatic Morpheme Segmentation of the Russian Language	257
Подлеская В. И. «Ой, нет! Не в среду! В четверг!»: корпусное исследование междометия в контексте речевого сбоя	267
Prokofyeva O., Kiose M., Leonteva A., Smirnova E. Intensity and its manifestations in speech and recurrent gestures in spontaneous dialogue	281
Rossyaykin P. Structured sentiment analysis using few-shot prompting of an ensemble of LLMs	291
Rozhkov I. S., Loukachevitch N. V. Methods for Recognizing Nested Terms	299
Satdarov K., Kharlamova D., Yakubov A., Lezina A. Faroese Corpus Development: Strategies for Low-Resource Languages Corpora	312
Semak V. V., Bolshakova E. I. Comparing Transformer-Based Approaches for Term Recognition in Russian texts	323
Шерстинова Т. Ю., Мельник А. Г., Петрова И. А., Азаревич К. И., Мелкозерова В. И., Чеповецкая С. В. «Оки доки, вот вам без рофла!»: Повседневная речь молодежи в корпусном представлении (структура и назначение звукового корпуса КУРС/ESC)	331
Shulginov V., Şimşek H. B., Kudriashov S., Randautsova R., Shevela S. A. Evaluating the Pragmatic Competence of Large Language Models in Detecting Mitigated and Unmitigated Types of Disagreement	345
Sidorova E. A., Ivanov A. I., Ilina D. V., Ovchinnikova K. A., Osmushkin N. M., Sery A. S. An Approach to Information Extraction from Texts of a Limited Subject Domain Based on a Chain of Large Language Models	361

Sorokin A., Nasyrova R. LORuGEC: the Linguistically Oriented Rule-annotated corpus for Grammatical Error Correction of Russian	374
Studenikina K. A., Lyutikova E. A., Gerasimova A. A. Gradual Acceptability Judgments with LLMs: Evidence from Agreement Variation in Russian	384
Татевосов С. Г., Киселева К. Л. Видовые пары глагольных идиом в русском языке: условия и механизмы аспектуальной редукции	393
Тимошенко С. П., Сердобольская Н. В., Кобозева И. М. Статистические метрики как критерии отбора составных коннекторов (на материале базы Рускон)	407
Vatolin A. Structured Sentiment Analysis with Large Language Models: A Winning Solution for RuOpinionNE-2024	416
Vatolin A., Gerasimenko N., Loukachevitch N., Ianina A., Vorontsov K. ruSciFact: Open Benchmark for Verifying Scientific Facts in Russian	435
Янко Т. Е. Русский upstep: просодическая модель переспроса к плану выражения	460
Zimmerling A. V., Baiuk A. M. Operator words versus MT systems and LLMs	471
Abstracts	481
Авторский указатель	489
Author Index	490