

Редакторская заметка

УДК 81'32, 81'33

DOI: <https://doi.org/10.18721/JHSS.14101>



ИНЖЕНЕРНЫЕ ЛИНГВИСТИЧЕСКИЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ ТЕКСТА

А.В. Колмогорова  

Национальный исследовательский университет «Высшая школа экономики»,
Санкт-Петербург, Российская Федерация

 nastiakol@mail.ru

Аннотация. Публикация посвящена анализу современного состояния инженерной лингвистики, ее основных направлений и исследовательских вызовов. Формулируется определение языковых технологий и их типология по критерию решаемых с их помощью задач. Отмечается, что отечественной школе инженерной лингвистики удастся сохранить баланс между технологичностью и лингвистичностью изысканий.

Ключевые слова: языковые технологии, инженерная лингвистика, компьютерная лингвистика, языковые модели.

Для цитирования: Колмогорова А.В. Инженерные лингвистические технологии в исследовании текста // Terra Linguistica. 2023. Т. 14. № 1. С. 7–10. DOI: 10.18721/JHSS.14101



Editorial note

DOI: <https://doi.org/10.18721/JHSS.14101>



ENGINEERING LINGUISTIC TECHNOLOGIES IN TEXT STUDIES

A.V. Kolmogorova 

National Research University Higher School of Economics,
St. Petersburg, Russian Federation

✉ nastiakol@mail.ru

Abstract. The publication is devoted to the analysis of the current state of engineering linguistics, its main directions and research challenges. The definition of language technologies and their typology are formulated according to the criterion of the tasks solved with their help. It is noted that the national school of engineering linguistics manages to maintain a balance between technological and linguistic research.

Keywords: linguistic technologies, engineering linguistics, computational linguistics, linguistics models.

Citation: A.V. Kolmogorova, Engineering linguistic technologies in text studies, Terra Linguistica, 14 (1) (2023) 7–10. DOI: 10.18721/JHSS.14101

Инженерные лингвистические технологии в исследовании текста: основные направления и результаты

Цифровой поворот во всех сферах жизни общества стал стимулом для активного роста исследований в области компьютерной лингвистики, которая имеет в отечественной исследовательской практике давние традиции.

60-е годы XX в. были отмечены бурным развитием первых компьютерных технологий и появлением в некотором смысле «идеологии технологичности»: роль искусственного интеллекта стала активно осмысливаться обществом. Устоявшиеся правила нормативной науки, будучи перенесены в новый контекст, требовали серьезных изменений. Лингвистика в этом смысле не стала исключением.

В эти годы преимущественно на Западе возникает направление компьютерной (вычислительной) лингвистики, а в СССР Раймондом Генриховичем Пиотровским разрабатывается концепция инженерной лингвистики, в фокусе которой оказываются «проблемы применения вероятностного моделирования для автоматической переработки текста» и разработки необходимых для этого лингвистических технологий [1].

Сегодня мы можем наблюдать, как вычислительные модели с успехом воплощаются в жизнь в виде языковых (лингвистических) технологий, позволяющих менять суть и свойства коммуникативных, информационных, этических, когнитивных и даже мировоззренческих процессов и категорий в нашей повседневной жизни.

В качестве одного из возможных определений лингвистических технологий можно было бы предложить следующее: это инженерные решения, имеющие воспроизводимый алгоритм применения, базирующиеся на вычислительных операциях обработки данных на естественном языке и позволяющие моделировать и воспроизводить широкий диапазон процессов, аналогичных тем, которые запускаются человеческим интеллектом в момент обработки сообщения на естественном языке.

По критерию решаемой задачи сегодняшние языковые технологии формируют несколько значимых кластеров:

1. Технологии автоматической обработки текста, которые делают текстовую ткань лингвистически транспарентной и доступной для дальнейшего машинного анализа: инструменты морфологического и синтаксического анализа и др.



2. Технологии и модели, моделирующие знаниевую компоненту текста: тезаурусы, онтологии, инструменты извлечения сущностей, тем и отношений.

3. Технологии, которые распознают и моделируют оценочную и эмоциональную составляющую текста (осуществляют аффективные вычисления): инструменты сентимент-анализа и эмоционального анализа.

4. Технологии, моделирующие процессы извлечения смысла из текстовых данных с возможностью его дальнейшей обработки, перекодирования и модификации: модели понимания и интерпретации естественного языка, инструменты парафразирования, симплификации и суммаризации, алгоритмы машинного перевода.

5. Технологии, моделирующие коммуникативное поведение человека, включающее не только вербальную составляющую, но и жестовую, мимическую и прагматическую: инструменты, позволяющие разрабатывать модели когнитивных ассистентов, чат-ботов, виртуальных дополненных личностей.

Приведенный список заведомо неполон — новые технологические решения постоянно пополняют имеющийся в распоряжении компьютерных и математических лингвистов арсенал.

Однако магия техники не должна мешать балансу делегирования — четкому пониманию соотношения той работы, которая выполняется технологическими средствами, и операций, которые остаются за человеком [2]. В нашем случае — за профессиональным лингвистом, который понимает, что именно находится «под капотом» у той или иной модели, алгоритма; как это работает; какие ограничения и искажения может давать применяемое технологическое решение; какая комбинация методов покажет наибольшую эффективность и, самое главное, как интерпретировать на языковедческом уровне те результаты, которые мы получаем «на выходе», или те сбои, которые мы наблюдаем, проводя вычислительные эксперименты.

Иными словами, между компьютерным лингвистом и его технологическим инструментом выстраивается своеобразный эвристический диалог о языке [3]: интерпретация поведения вычислительной модели, обрабатывающей языковые данные (например, в ходе машинного или глубинного обучения), результатов применения математических методов к текстовому материалу дает лингвисту не прямые, но значимые ответы на вопросы о внутреннем устройстве языка, как это представлено, например, в публикации [4], функционировании идиолекта [5], языковых средств реализации художественного целого [6].

Подобный сочинительный принцип, соединяющий инженерию и теорию языка, был тонко подмечен еще в работах Р. Г. Пиотровского [7]. Ясно прослеживается он и в исследованиях, результаты которых представлены в тематическом выпуске журнала *«Инженерные лингвистические технологии в исследовании текста»*. Содержание номера, совмещающее в себе гносеологическую эталонность и творческий подход [8], позволит всем интересующимся лингвистикой и «доказательной лингвистикой» читателям познакомиться с современными достижениями инженерной лингвистики в России, даст импульс к новым экспериментам и эвристикам на пути познания языка.

СПИСОК ИСТОЧНИКОВ

1. **Беляева Л.Н., Богданов С.И., Горностай Т.** Инженерная лингвистика в контексте современной «Информации 4.0» // Proceedings. Сер. "CEUR-WS Workshop Proceedings" Том 2233. St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Herzen State Pedagogical University of Russia. 2018. С. 48–56.

2. **Вахштайн В.** Техника. СПб.: Издательство Европейского университета в Санкт-Петербурге, 2021. 156 с.



3. Колмогорова А.В., Калинин А.А., Маликова А.В., Вдовина Л.А. Методы компьютерной и корпусной лингвистики для решения задач эмоционального анализа интернет-текстов, М.: IPR Media, 2021.

4. Пиотровский Р. Г. Инженерная лингвистика и теория языка. Л.: Наука, 1979. 111 с.

5. Митрофанова О.А., Гаврилик Д.А. Эксперименты по автоматическому выделению ключевых выражений в стилистически разнородных корпусах русскоязычных текстов // Terra Linguistica. 2022. Т. 13. № 4. С. 22–40. DOI: 10.18721/JHSS.13402

6. Андреев В.С. Экспоненциальное распределение частей речи в стихотворном тексте: опыт стилометрического анализа // Общество. Коммуникация. Образование. 2021. Т. 12. № 4. С. 94–104. DOI: 10.18721/JHSS.12407

7. Се Линь, Загайнов А.И. Моделирование характеристик персонажей и их взаимосвязей в сюжете художественного произведения методами численного фрактального анализа // Terra Linguistica. 2022. Т. 13. № 3. С. 36–47. DOI: 10.18721/JHSS.13304

8. Беляева Л.Н., Чернявская В.Е. Доказательная лингвистика: метод в когнитивной парадигме // Вопросы когнитивной лингвистики. 2016. № 3 (48). С. 77–84. DOI: 10.20916/1812-3228-2016-3-77-84

REFERENCES

[1] L. Beliaeva, S. Bogdanov, T. Gornostay, Engineering in the Framework of Modern “Information 4.0”, Proceedings. Ser. “CEUR-WS Workshop Proceedings” Vol. 2233. St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Herzen State Pedagogical University of Russia. 2018. Pp. 48–56.

[2] V. Vakhshayn, Technics, Saint-Petersburg, Editions of European University in Saint-Petersburg, 2021.

[3] A.V. Kolmogorova, A.A. Kalinin, A.V. Malikova, L.A. Vdovina, Methods of computational and corpus linguistics in tasks of emotional analysis of the internet-texts, М.: IPR Media, 2021.

[4] R.G. Piotrovskiy, Inzhenernaya lingvistika i teoriya yazyka [Engineering linguistics and theory of language]. L.: Nauka, 1979. 111 p.

[5] O.A. Mitrofanova, D.A. Gavrilič, Experiments on automatic keyphrase extraction in stylistically heterogeneous corpus of Russian texts, Terra Linguistica, 13 (4) (2022) 22–40. DOI: 10.18721/JHSS.13402

[6] V.S. Andreev, Exponential distribution of parts of speech in verse text: experience in stylometric analysis, Society. Communication. Education, 12 (4) (2021) 94–104. DOI: 10.18721/JHSS.12407

[7] Xie Linyi, A.I. Zagaynov, Modeling of character characteristics and their relationships in a novel plot by methods of numerical fractal analysis, Terra Linguistica, 13 (3) (2022) 36–47. DOI: 10.18721/JHSS.13304

[8] L.N. Belyayeva, V.Ye. Chernyavskaya, Evidence-based linguistics: Methods in cognitive paradigm, Issues of Cognitive Linguistics. 3 (48) (2016) 77–84. DOI: 10.20916/1812-3228-2016-3-77-84

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Колмогорова Анастасия Владимировна

Anastasia V. Kolmogorova

E-mail: nastiakol@mail.ru

ORCID: <https://orcid.org/0000-0002-6425-2050>

Поступила: 14.02.2023; Одобрена: 17.03.2023; Принята: 17.03.2023.

Submitted: 14.02.2023; Approved: 17.03.2023; Accepted: 17.03.2023.