

# ДИАЛОГ УНИВЕРСИТЕТОВ: ОПЫТ НИУ ВШЭ СПб

**А. В. Колмогорова<sup>1</sup>**

*Национальный исследовательский университет  
«Высшая школа экономики» (Санкт-Петербург)*

## ОТ РЕДАКТОРА: ЯЗЫКОВЫЕ ТЕХНОЛОГИИ В ПИТЕРСКОЙ ВЫШКЕ: ВЗГЛЯД МОЛОДЫХ ИССЛЕДОВАТЕЛЕЙ

Сочетание «языковые технологии», еще недавно непривычное в лингвистической среде, сегодня становится все более актуальным. Технологичный взгляд на язык, с одной стороны, позволяет увидеть как бы на расстоянии незаметные вблизи тренды, а с другой – дает инструменты для проверки самых смелых предположений и гипотез.

Языковые технологии, объединяя в себе корпусную и компьютерную лингвистику, дают импульс для разработки междисциплинарных проблем, соединяющих в себе семиотику, лингвоэмотиологию, лексикографию и многие другие.

Интересным и продуктивным, на наш взгляд, представляется то, что современные компьютерные технологии, ставшие доступными нам, лингвистам, позволяют с другой стороны посмотреть на классические постулаты языковедения, в некоторых случаях снижая градус их абсолютности и императивности, в других – подтверждая релевантность и применимость за пределами теоретической лингвистики.

Остановимся на некоторых из них.

В. И. Шаховский, основатель лингвоэмотиологии, отмечал, что «семантика всех слов ингерентно или адгерентно всегда эмоционально нагружена» [Шаховский, 2014, с. 14]. Бурное развитие в последние десять лет так называемой парадигмы аффективных вычислений в компьютерных науках и междисциплинарных исследованиях позволило разработать методы и технологии, известные в настоящее время как сентимент-анализ и эмоциональный анализ, которым подвергаются сегодня не только текстовые массивы различной жанровой принадлежности (посты в социальных сетях [Колмогорова, 2022], художественная [Sherstinova, 2023] и учебная литература [Solovyev, 2019]) но и интонация, мимика [Savchenko, 2022]. Результаты многочисленных исследований с использованием данных инструмен-

---

<sup>1</sup> Анастасия Владимировна Колмогорова – доктор филологических наук, профессор, профессор департамента филологии Школы гуманитарных наук и искусств Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург), заведующий лабораторией языковой конвергенции, [akolmogorova@hse.ru](mailto:akolmogorova@hse.ru).

тов подтверждают этот тезис: эмоциональность присутствует в большинстве типов текстов – она потенциально детектируема везде, но ее амплитуда неодинакова и зависит от жанра. Но, с другой стороны, есть основания для того, чтобы несколько уточнить тезис В. И. Шаховского: в естественных условиях устной коммуникации эмоциональная нагрузка свойственна всему коммуникативному поведению индивида, поэтому она динамически перераспределяется между разными его модальностями (словами, мимикой, жестами) в каждый момент развертывания коммуникации, то есть слова – далеко не единственные носители эмоциональности. В случае же письменной коммуникации эмоциональная нагрузка слов становится чрезвычайно изменчивой, мало привязанной к лексическому значению слова, но чрезвычайно зависимой от индивидуальных психологических и когнитивных характеристик реципиента, его знаний о мире и коммуникативного опыта.

Другой пример. Понятие темы хорошо изучено в лингвистической теории текста [Матвеева, 1990]: тема – это совокупность лексических единиц, отражающая некоторую денотативную ситуацию, являющуюся содержанием текста. Современные методы тематического моделирования (Topic Modeling) позволяют исследователям автоматически отбирать ряды слов, статистически неслучайно встречающихся вместе именно в этом тексте, а затем, экспертным образом соотнося их с некоторой денотативной ситуацией, формулировать метку темы (топика) и давать коммуникативно-прагматическую интерпретацию ее значимости для общей смысловой системы, например, художественного произведения. Несмотря на то, что последнее слово все равно остается за экспертом-лингвистом, компьютерные модели «помогают» увидеть возможных кандидатов на роль лексических единиц темы. Прodelать такую большую работу с колоссальными объемами данных лингвисту просто не под силу. Например, такой подход помог нам увидеть общие темы в текстах, составивших творческое наследие С. Довлатова, и текстах 36 произведений мировой литературы, которые писатель определял как оказавшие большое влияние на его писательский стиль [Колмогорова, 2023]

Наконец, последний кейс, на котором хотелось бы остановиться. В связи с бумом генеративных моделей, которые после появления на рынке в 2022 году ChatGPT создали иллюзию ненужности естественного, т.е. исходящего от человека, процесса текstopорождения, стал активно обсуждаться вопрос о том, чем отличаются тексты, порожденные искусственным интеллектом (ИИ) и человеком. Такая постановка вопроса вдохнула новую жизнь в психолингвистические и когнитивные исследования процессов текстовой обработки человеком. Предварительный вывод, который активно обсуждается сегодня в литературе [Li, 2021], состоит в том, что если ИИ предсказывает следующее слово в предложении на основе статистических паттернов, выученных на больших объемах данных, то естественный генератор текстов (человек), продуцируя или понимая текст, совершает итерации между языковыми и когнитивными структурами, которые уже имеются в его опыте (модель мира) или структурируются на основе уже имеющегося знания в режиме «здесь-и-сейчас» в форме, например, ментальных пространств [Fauconnier, 2006]. Как следствие такого вывода, опирающегося на наработки когнитивной лингвистики, фокус

современных разработок направлен на создание у больших языковых моделей своеобразных моделей мира, системы знаний о мире, на которые они могли бы опираться, генерируя тексты.

В рамках концепции нашей магистерской программы мы ставим задачу объединения лингвистической школы и базовых навыков в ИТ, чтобы все это помогло нашим выпускникам видеть и понимать механику языковых процессов, запускаемых как искусственным, так и естественным интеллектом, и уметь перекинуть между ними своеобразный эвристический «мостик».

В предлагаемых статьях, написанных молодыми исследователями магистерской программы «Языковые технологии в бизнесе и образовании» НИУ «Высшая школа экономики» в Санкт-Петербурге в соавторстве с их наставниками, формулируются достаточно смелые гипотезы, которые проверяются с помощью инструментария компьютерной лингвистики.

Работа Е. Г. Кузиной и Т. А. Литвиновой носит практический характер и направлена на разработку англо-русского словаря терминологии NLP. Не секрет, что большинство терминов данной профессиональной области бытуют и в русскоязычном сообществе в своем первоначальном виде как не адаптированные заимствования. Однако это нарушает принципы лингвоэкологии и не соответствует современному стремлению российской общественности к поддержанию системности и однородности родного языка. В статье представлены структура будущего словаря, принципы отбора словника и технические решения, которые будут применяться.

Статья А. Д. Егоровой и Н. В. Станкевич посвящена междисциплинарной проблематике – в ней обсуждаются точки пересечения лингвистики и математики. Рассматривается нетривиальное понятие сентиментной траектории текста, которую можно построить, откладывая по оси абсцисс порядковый номер предложения, а по оси ординат – значения сентимента предложения. Получаемые при соединении вершин полученных графов траектории и являются сентиментными траекториями, показывающими динамику эмоциональной тональности в анализируемом тексте. Такой подход позволяет сравнивать разные тексты одного и того же автора, а также переводы одного и того же текста на предмет сохранения или степени модификации эмоциональной тональности.

В публикации М. О. Сергеевой, А. В. Мигалья и А. В. Колмогоровой также обсуждается проблематика анализа эмоциональной тональности, но не текста, а жеста. Но поскольку словарей эмоциональной тональности жестов пока не существует, для получения эмоциональной метки текста используются результаты сентимент-анализа тех слов, которые описывают его значения в аннотации, предлагаемой Мультимедийным корпусом Национального корпуса русского языка. При сопоставлении эмоциональной метки и активного органа, используемого для исполнения жеста, предпринимается попытка сделать вывод о том, есть ли у жестов, исполняемых разными частями тела, своя эмоциональная нагрузка, которую приписывают им носители того или иного культурного соматического кода, например, актеры советского кино.

Представленные статьи приглашают к дискуссии, предлагают идеи, прошедшие только первичную апробацию на авторских тестовых выборках или представляющие собой просpekt будущего продукта, но при этом показывающие возможности компьютерных инструментов и, в определенной степени, их идеологию.

#### СПИСОК ЛИТЕРАТУРЫ

**1. Колмогорова, А. В.** Компьютерное моделирование как инструмент анализа художественного текста / А. В. Колмогорова, Е. Д. Залевская // Филологический класс. – 2023. – Т. 28. – № 2. – С. 22–33.

**2. Колмогорова, А. В.** Эмоциональный анализ постов ВКонтакте: классификатор или регрессор / А. В. Колмогорова, А. А. Калинин // Компьютерная лингвистика и интеллектуальные технологии : По материалам ежегодной международной конференции «Диалог». – Вып. 21. – Москва: РГТУ, 2022. – С. 311–322.

**3. Матвеева, Г. В.** Функциональные стили в аспекте текстовых категорий: синхронно-сопоставительный очерк / Г. В. Матвеева. – Свердловск: изд-во УрГУ, 1990. – 182 с.

**4. Шаховский, В. И.** Эмоции: долигвистика, лингвистика, лингвокультурология / В. И. Шаховский. – Москва: Либроком, 2010. – 128 с.

**5. Fauconnier, G.** Mental spaces: conceptual integration networks / G. Fauconnier, M. Turner // Cognitive linguistics: basic readings / edited by Dirk Geeraerts. Walter de Gruyter GmbH & Co. KG, 2006 – P. 303–371.

**6. Li, B. Z.** Implicit Representations of Meaning in Neural Language Models / B. Z. Li, M. Nye, J. Andreas. – arXiv: 2106.00737. – 2021. – URL: [arxiv.org/abs/2106.00737](https://arxiv.org/abs/2106.00737) (дата обращения: 01.04.2024).

**7. Savchenko, A.** HSEmotion: High-speed emotion recognition library / A. Savchenko // Software Impacts. – 2022. – №14. – P.100433.

**8. Sherstinova, T.** Sentiment Analysis of Literary Texts vs. Reader's Emotional Responses / T. Sherstinova, A. Moskvina, M. Kirina, A. Karysheva, E. Kolpashchikova, P. Maksimenko, A. Seinova, R. Rodionov // Proceedings of the 33rd Conference of Open Innovations Association (FRUCT), 2023. – P. 243–249.

**9. Solovyev, V.** Sentiment in Academic Texts / V. Solovyev, M. Solnyshkina, E. Gafiyatova, D. McNamara, V. Ivanov // Proceedings of the 24th Conference of Open Innovations Association (FRUCT), 2019. – P. 408–414.

#### REFERENCES

**1. Fauconnier, G.** Mental spaces: conceptual integration networks / G. Fauconnier, M. Turner // Cognitive linguistics: basic readings / edited by Dirk Geeraerts. Walter de Gruyter GmbH & Co. KG, 2006 – P. 303–371.

**2. Kolmogorova, A. V.** Komp'yuternoe modelirovanie kak instrument analiza khudozhestvennogo teksta / A. V. Kolmogorova, E. D. Zalevskaya // Filologicheskii klass. – 2023. – Т. 28. – № 2. – С. 22–33.

**3. Kolmogorova, A. V.** Emotsional'nyy analiz postov VKontakte: klassifikator ili regressor / A. V. Kolmogorova, A. A. Kalinin // *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoy mezhdunarodnoy konferentsii «Dialog»*. – Vyp. 21. – Moskva: RGTU, 2022. – С. 311–322.

**4. Matveeva, G. V.** Funktsional'nye stili v aspekte tekstovykh kategoriy: sinkhronno-sopostavitel'nyy ocherk / G. V. Matveeva. – Sverdlovsk: izd-vo UrGU, 1990. – 182 s.

**5. Shakhovskiy, V. I.** Emotsii: dolingvistika, lingvistika, lingvokul'turologiya / V.I. Shakhovskiy. – Moskva: Librokom, 2010. – 128 s.

**6. Li, B. Z.** Implicit Representations of Meaning in Neural Language Models / B. Z. Li, M. Nye, J. Andreas. – arXiv: 2106.00737. – 2021. – URL: [arxiv.org/abs/2106.00737](https://arxiv.org/abs/2106.00737) (data obrashcheniya: 01.04.2024).

**7. Savchenko, A.** HSEmotion: High-speed emotion recognition library / A. Savchenko // *Software Impacts*. – 2022. – №14. – P.100433.

**8. Sherstinova, T.** Sentiment Analysis of Literary Texts vs. Reader's Emotional Responses / T. Sherstinova, A. Moskvina, M. Kirina, A. Karysheva, E. Kolpashchikova, P. Maksimenko, A. Seinova, R. Rodionov // *Proceedings of the 33rd Conference of Open Innovations Association (FRUCT)*, 2023. – P. 243–249.

**9. Solovyev, V.** Sentiment in Academic Texts / V. Solovyev, M. Solnyshkina, E. Gafiyatova, D. McNamara, V. Ivanov // *Proceedings of the 24th Conference of Open Innovations Association (FRUCT)*, 2019. – P. 408–414.