

# ЕСТЕСТВЕННЫЙ ЯЗЫК И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: КОГНИТИВНЫЙ АСПЕКТ

**Н.К. Рябцева**

Институт языкознания РАН

[Nadia\\_riabceva@mail.ru](mailto:Nadia_riabceva@mail.ru)

В своем приветствии участникам организованного Андреем Александровичем Кибриком в 2020 г. Лингвофорума «Язык и искусственный интеллект» он совершенно точно обозначил проблему «Язык и искусственный интеллект» как «общественно значимую».

Как ни парадоксально, естественный язык кажется естественным только своему носителю. Так, носители языка в разных странах, даже специалисты, поначалу (60 лет назад и позже) были уверены, что любая обработка естественного языка — **Natural Language Processing (NLP)** дело довольно простое и выполнимое. На практике же оказалось, что огромное количество лингвистических свойств почти не поддается формализации. И именно подобные свойства имеются в виду, когда подчеркивается важность изучения когнитивных процессов, проявляющихся в использовании языка. В статье Кибрик 2020: 59 в связи с этим отмечается: ««Язык и речевая деятельность представляют собой одно из основных проявлений невидимой когнитивной системы. Каждый представитель *Homo sapiens* постоянно и массово порождает материал, который может использоваться в качестве данных для исследования общекогнитивных процессов». К особо важным с точки зрения NLP когнитивным явлениям относятся такие важнейшие свойства языка, как многозначность, омонимия, недискретность семантики, аналоговость (Кзаковская, Онипенко 2020), имплицитность, инферентность, коннотативность, неоднозначность, синкретизм, «нелинейность» выражения содержания и др. Так, в статье Борисова 2020: 47 отмечено в связи с этим: «оттенок значения» — это «промежуточное между двумя значениями». Поэтому, как отмечается в статье Александрова 2020: 37: «ИИ не может соперничать с человеком во многих областях», не говоря уже о (машинном) переводе, который, хотя и значительно эволюционировал в положительную сторону, еще далеко не везде может быть использован; подробнее см. в статье Рябцева 2023. В результате современные лингвистические исследования в этом отношении связаны с выявлением когнитивных структур и процессов, лежащих в основе

языковой, речевой и невербальной деятельности человека. Особое внимание при этом уделяется типам и видам знания, их использованию, в том числе в цифровых технологиях, особенно энциклопедическим, фоновым (Виссон 2007, Cozma 2012) и специальным знаниям.

В прикладном отношении соответствующие проблемы моделирования когнитивной сферы человека решаются сегодня (1) посредством совершенствования лингвистического обеспечения при создании лексикографических цифровых ресурсов принципиально нового «когнитивного» типа (Katić, Novakov 2022); (2) посредством выявления постоянно возникающих новых лингвистических средств представления специальных знаний, особенно связанных с интенсивно развивающимися цифровыми технологиями. В первом случае передним краем создания лексикографических цифровых ресурсов нового «когнитивного» поколения выступает их интеграция с энциклопедическими и др. данными. Так, в когнитивном отношении принципиально новым шагом в создании цифровых лексикографических ресурсов (нового, третьего) поколения стало сочетание энциклопедической и лексикографической информации в толковании лексических единиц и устойчивых словосочетаний как общеупотребительного, так и терминологического типа, что подчеркивается, в частности, в статье Fuertes-Olivera, Esandi-Baztan 2020: 91: «Precise definitions, similar to terminological (i.e. encyclopaedic) definitions, for most lemmas increase the tool's functions. Such definitions offer a very different picture of current monolingual Spanish and bilingual Spanish–English dictionaries». Предполагается, что подобные универсальные ресурсы могут быть интегрированы в любые цифровые лингвистические продукты.

В области развития систем **автоматического перевода (АП)**, в первую очередь научных и технических текстов, решается наиболее актуальная в этой области проблема — установление межъязыковых соответствий при переводе с английского языка на родной и обратно многокомпонентных терминологических словосочетаний, возникающих исходно в настоящее время в английском языке во всех научных дисциплинах в результате стремительного развития и внедрения «в практику» цифровых технологий. Такие инновационные словосочетания становятся главными средствами передачи новых дисциплинарных знаний и в подавляющем большинстве случаев носят принципиально лингвоспецифичный характер и к тому же не успевают попасть в словарь.

Так, в английском языке для такой современной области, как разработка новых цифровых лексикографических ресурсов (одновременно) дидактического, профессионального и двуязычного предназначения, весьма характерно использование (и создание) развернутых терминологических лингвоспецифических словосочетаний типа: *context-aware (electronic) dictionaries, task-*

*based (electronic) dictionary training, user-friendly (electronic) search routes, in no user-unfriendly (electronic) way, lexicography-based (electronic) applications, a(n) (electronic) product-based approach, commercially driven integrated (electronic) dictionary portals, a test-driven development of an (electronic) lexicographic product* (Huang, Tarp 2021: 68–92) и др. Причем самые лучшие современные автоматические переводчики типа Google Translate, Yandex Translate и др. пока не в состоянии корректно перевести на русский язык подобные типичные для английского специального текста развернутые терминологические лингвоспецифические конструкции. Ср. выделенные *курсивом* лингвоспецифические конструкции в аутентичных терминологических словосочетаниях английского языка, их некорректно переведенные системами АП элементы «под вопросом» (Google Translate) и их правильные межъязыковые эквиваленты: context-aware dictionaries — контекстно-?зависимые словари vs контекстологические словари; task-based dictionary training — обучение словарю ?на основе задач vs компьютерное обучение работы со словарем; in no user-unfriendly way — ?неудобным для пользователя способом vs неудобно; lexicography-based applications — приложения ?на основе лексикографии vs лексикографические приложения; commercially driven integrated dictionary portals — ?коммерчески управляемые интегрированные словарные порталы vs коммерческие интегрированные словарные порталы и т. д.

Подобные инновационные многокомпонентные терминологические словосочетания английского языка отличаются целым комплексом взаимосвязанных свойств, требующих учета в автоматических системах перевода. Так, 1) они являются средством передачи особо прецизионной профессионально значимой информации, носителями нового специального знания, и потому нуждаются в точном, полном и качественном переводе на другой язык; 2) они постоянно обновляются, уточняются и «развиваются» — в связи с интенсивным развитием цифровых технологий, и часто носят междисциплинарный и динамичный характер; 3) для них в высшей степени характерны варьирование, модификация, контекстуализация и интеграция с аналогичными текстуальными единицами; 4) их учет и фиксирование в специальных словарях и ресурсах значительно отстает от скорости их появления и использования в специальных текстах, а их автоматический перевод носит пословный характер и оказывается неприемлемым; 5) Основным препятствием их качественного автоматического перевода выступает их лингвоспецифичность: образование по характерным для современной научной терминологии на английском языке активным моделям (Рябцева 2022).

Учет при этом наиболее активных и динамичных свойств типичных для английского специального текста развернутых терминологических лингвоспецифических конструкций: их многокомпонентность, междисциплинарный

характер, способность к модификации, варьированию, расширению и др., посредством сочетания их лексикографического представления, лингвистического анализа и его интеграции с энциклопедическими и фоновыми знаниями (ср. Кибрик 2019: 238), а также их алгоритмического моделирования позволяет избежать самой типичной и, строго говоря, недопустимой ошибки в их автоматическом переводе — дословности, пословности, буквальности и потому некорректности их перевода.

Таким образом, лингвистическое и программное обеспечение цифровых систем обработки естественного языка (от обучающих иностранному языку до автоматического перевода специальных текстов) посредством моделирования наиболее значимых когнитивно-коммуникативных процессов передачи прецизионной информации позволит значительно повысить их качество и эффективность и совместить лексикографически продвинутое технологии с энциклопедическими, фоновыми и специальными знаниями.

### Л и т е р а т у р а

- Александрова Н.Ш. 2020. Понимание речи человеком: двойственность организации. Под ред. А.А. Кибрика, В.Ю. Гусева, Д.А. Залманова. *Лингвистический форум 2020: Язык и искусственный интеллект*. Тезисы докл. междунар. конф. Институт языкознания РАН, Москва. 12–14 ноября 2020 г. М.: ИЯз РАН, 37–38.
- Борисова Е.Г. 2020. Модель для обработки недискретных данных в лингвистическом описании. Под ред. А.А. Кибрика, В.Ю. Гусева, Д.А. Залманова. *Лингвистический форум 2020: Язык и искусственный интеллект*. Тезисы докл. междунар. конф. Институт языкознания РАН, Москва. 12–14 ноября 2020 г. М.: ИЯз РАН, 47–48.
- Под ред. А.А. Кибрика, В.Ю. Гусева, Д.А. Залманова. *Лингвистический форум 2020: Язык и искусственный интеллект*. Тезисы докл. междунар. конф. Институт языкознания РАН, Москва. 12–14 ноября 2020 г. М.: ИЯз РАН, 96–97.
- Кибрик А.А. 2019. Теория речевой деятельности: вызовы современности. *Вопросы психолингвистики*, № 3 (41), 238.
- Кибрик А.А. 2020. Язык и когнитивная система. В кн.: *I Национальный конгресс по когнитивным исследованиям, искусственному интеллекту и нейроинформатике. Сборник пленарных и индустриальных докладов. Доклады симпозиумов*. М.: Общероссийская общественная организация Российская академия естественных наук, 59–61.
- Рябцева Н.К. 2022. Современные терминологические проблемы в межъязыковом аспекте. *Научный диалог*, т. 11, № 6, 123–139.
- Рябцева Н.К. 2023. Естественный и искусственный интеллект в процессах перевода. В кн.: А.В. Вдовиченко (ред.) *Язык и искусственный интеллект*. М.: ЯСК, 237–261.
- Виссон Л. 2007. *Синхронный перевод с русского на английский*. М.: Р. Валент.

- Cozma M. 2012. Perceptions of cultural competence: The trainees perspective. *Professional Communication and Translation Studies*, n. 5(1-2), 67–74.
- Fuertes-Olivera P.A., EsandiBaztan M.A. 2020. Integrating Terminological Resources in Dictionary Portals: The Case of the Diccionarios Valladolid-UVa. *Lexikos*, n. 30, 90–110.
- Huang F., Tarp S. 2021. Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvement. *Lexikos*, n. 31, 68–92.
- Katić M., Novakov P. 2022. The Translation Stage in LSP Lexicography: A Mixed Translation Model for LSP Bilingual Dictionary Terms. *International Journal of Lexicography*, v. 35, Issue 4, December, 427–448.