

Инна Алексеевна Воронцова, к. филол. н.
Ярославский государственный педагогический университет им. К. Д. Ушинского,
Ярославль, Россия
э-почта: arinna1@yandex.ru

Inna Alekseevna Vorontsova, Cand. of Sc.
(Philology)
Yaroslavl State Pedagogical University named
after K. D. Ushinsky, Yaroslavl, Russia
e-mail: arinna1@yandex.ru

ЧЕТЫРЕ РЕВОЛЮЦИИ В НОВЕЙШЕЙ ЛЕКСИКОГРАФИИ: КАК МЕНЯЕТСЯ СЛОВАРНАЯ ПРАКТИКА

В статье освещаются основные трансформационные процессы, изменившие концепцию и особенности функционирования словарей в XX–XXI столетиях. Обсуждаются понятия описательной, корпусной, цифровой и нейросетевой революций в лексикографии.

Ключевые слова: лексикография; корпус; электронный словарь; цифровой словарь; нейросетевая революция; искусственный интеллект

FOUR REVOLUTIONS IN MODERN LEXICOGRAPHY: HOW DICTIONARY PRACTICES ARE CHANGING

The article highlights the main transformational processes that have shaped the concept and characteristics of dictionaries in the XX–XXI centuries. The phenomena of descriptive, corpus, digital and intellectual / neural network revolutions in lexicography are discussed.

Key words: lexicography; corpus; electronic dictionary; digital dictionary; neural network revolution; artificial intelligence

Трансформационные процессы, наблюдаемые в практике составления словарей и культуре их использования, вызванные широкой экспансией цифровых технологий и искусственного интеллекта, подводят к необходимости переосмысления не только формальных, содержательных, технологических и т. п. аспектов лексикографирования избранного объекта, но и сущностных основ лексикографии. Целью настоящего исследования является освещение основных трансформационных процессов, изменивших концепцию и особенности функционирования словарей в XX–XXI столетиях, преобразовавших лексикографическую парадигму. В ходе исследования были использованы следующие методы: метод теоретико-лингвистического анализа литературы, метод лексикографического анализа, разработанный отечественной школой лексикографии (Л. П. Ступин, О. М. Карпова и др.), метод критического анализа словарей (*dictionary criticism* – R. R. K. Hartmann).

По мнению С. А. Маник, за последние сто лет лексикография претерпела три крупных преобразования или «революции»: описательную (*descriptive revolution*) – на рубеже XIX–XX вв., корпусную (*corpus revolution*) – во второй половине XX в. и цифровую (*digital revolution*) – происходящую сейчас, в XXI в. [1, с. 520].

Описательная революция, происходившая в эпоху бумажных словарей (*Oxford English Dictionary (OED)*, *Deutsches Wörterbuch (DWD)* и др.), сформировала системный (системо- / лингвоцентрический) подход к изъяснению единиц в словаре. Г. М. Мандрикова выделяет такие свойства лингвоцентрического подхода, как восприятие языка как подлежащей данности, существование словаря для языка и при

языке, сфокусированность на фиксации, описании и оценке языковых фактов, ориентация словаря на традиционное отражение лингвистических свойств, безадресность словаря, лингвоцентричность единиц описания [2, с. 9].

Внедрение корпусных методов, или корпусная революция, затронувшая как бумажные, так и электронные словари и обеспечившая их основанность на массивах текстовых данных, принципиально изменила подход к работе над словарным производством, облегчив поиск и извлечение единиц в словник словаря, расширив его описательную базу, повысив эффективность редакторской работы. Это было в полной мере реализовано после появления аннотированных корпусов, создания программ-лемматизаторов и программных инструментов, способных осуществлять поиск заданных лексикографом N-грамм, на более поздних этапах – программного обеспечения для построения так называемых «эскизов слов» (*word sketches*), оцифровки значительного числа национальных языков, составления специализированных корпусов, дальнейшей разработки и систематизации методов лингвистической и метаразметки.

Корпусная революция изменила подход к описанию единиц в словаре, позволив составителям перейти к дефинициям, представленным полными предложениями, имеющими более высокий коммуникативный потенциал по сравнению с определениями-фразами, дефиниции приобрели и более разговорный характер. Возросшая роль разноплановых источников информации, усилившееся внимание лексикографов к речевой культуре общества, осознание особой миссии словаря в трансляции идей, культурных и ценностных смыслов определили движение словарей в сторону человека, обеспечив разворот словарной идеологии к антропоцентризму.

Этап развития лексикографии в первой четверти XXI в. можно справедливо назвать цифровой революцией, трансформировавшей словарь в принципиально новый продукт. Электронные словари, появившиеся в конце XX столетия как компьютерные версии бумажных словарей, а затем – как интерактивные базы данных, постепенно преобразовались в гибридные, синергетические ресурсы, интегрирующие в своей структуре различные лексикографические формы, функции и инструменты, обеспечивающие решение различных пользовательских задач. Появилась новая бизнес-модель словаря, словарь стал рассматриваться как непрерывно оказываемая услуга, значительно возросла скорость извлечения информации из словаря, антропоцентрический характер цифрового словаря оказался «усилен» целым рядом новообразований в операциональном или утилитарном компоненте – организацией словарной статьи в виде гипертекста, профилизацией режимов работы со словарной статьей (принцип *Show more / Show less*) и др. Значимым для структурирования и содержательного наполнения современного электронного словаря стало вовлечение пользователей в процесс составления / обновления словника [1].

В текущий момент, на наш взгляд, лексикография вступает в новую фазу развития, становясь активным участником четвертой, «интеллектуальной», или «нейросетевой», революции. Лексикография постепенно сближается с таким направлением деятельности, как обработка естественного языка (*Natural language processing*), исследователи говорят о наступлении эпохи постлексикографии (*post-lexicography*), в обиход входит понятие GPT-логики, направления, исследующего данности и перспективы применения возможностей искусственного интеллекта для решения различных задач, в том числе в области лексикографии.

Технологии искусственного интеллекта задают как минимум три вектора развития словарной практики. Во-первых, нейросетевые инструменты встраиваются в структуру существующих ресурсов в качестве дополнительных опций, обеспечи-

вающих, например, машинный перевод, коррекцию, стилизацию текста и др. (интегрированные в один ресурс словарь *Linguee* и инструменты машинного перевода и письма *DeepL*, *DeepL Write*, *DeepL Voice*, *DeepL API*). Во-вторых, нейросети (чат-боты) способны самостоятельно генерировать словарные статьи и «эскизы слов», следуя традициям, устоявшимся в словарной практике, а также давать комментарии по частным аспектам семантики или прагматики лексической единицы. Оставляя за рамками дискуссии вопросы этики и качества выходных данных (известные «галлюцинации» чат-ботов), мы должны признать, что интеллектуальные системы, таким образом, принимают на себя роль лексикографа, а электронно-цифровой словарь конца первой четверти XXI в. теряет привычную «материализацию». В-третьих, возможности искусственного интеллекта, а точнее, возможность обращения к большим языковым моделям (*Large Language Model*, *LLM*), используются профессиональными лексикографами для сбора материала в словари, его обновления и уточнения. Функция составителя словаря, таким образом, трансформируется в «постредакторскую» (*post-editing lexicography*). Лексикограф решает задачи, связанные с созданием корректных промптов, верификацией выходных данных, валидацией их содержательной и формальной сторон на предмет соответствия ожиданиям предполагаемого пользователя, приведением данных к приемлемой лексикографической форме.

Таким образом, лексикография претерпела глубокие трансформации, обусловленные технологическим прогрессом и развитием методологий и практик научного познания. Современные лексикографические процессы отмечены интенсификацией применения цифровых технологий и искусственного интеллекта для генерирования нового и оптимизации существующего словарного материала, персонализации режимов работы с лексикографическими источниками, обеспечения «устойчивой лексикографии». Меняется субъектность лексикографа, обусловленная способностью искусственного интеллекта создавать лексикографический контент, что ведет, с одной стороны, к необходимости разграничения понятий «искусственный лексикограф» и «лексикограф-человек», с другой – к принятию преимуществ интеграции опыта традиционной и потенциала нейросетевой лексикографии. Соответственно, трансформируется и роль лексикографа-человека, все чаще выступающего в качестве валидатора и редактора данных, получаемых от искусственного интеллекта. Размываются традиционные формы словаря: словарь перестает быть статичным продуктом, а становится динамичной услугой, интегрированной в цифровые экосистемы, или «нейросетевым следом», оставленным в киберпространстве в ходе общения пользователя и нейросети. Основные вызовы текущего этапа развития лексикографии связаны с качеством генерируемого контента, этикой использования искусственного интеллекта и обеспечением научной достоверности словарных данных. Перспективы дальнейших исследований видятся в изучении влияния GPT-технологий на методологию лексикографии, разработке и анализе новых форматов словарей, оценке этических и практических аспектов взаимодействия человека и искусственного интеллекта при создании словарей.

ЛИТЕРАТУРА

1. Маник С. А. Современная лексикография в эпоху Chat GPT // Научно-исследовательская деятельность в классическом университете–2023: традиции и инновации : материалы Международного научно-практического фестиваля, Иваново, 10–28 апреля 2023 г. Иваново : Ивановский государственный университет, 2023. С. 519–524.
2. Мандрикова Г. М. Основные положения антропоцентрической лексикографии // МНКО. 2011. № 2. С. 8–12.