

УДК 81`374:004

ОСНОВНЫЕ ПОНЯТИЯ ЭЛЕКТРОННОЙ ЛЕКСИКОГРАФИИ

А.В. Палкова

Тверской государственной университет, Тверь

Данная статья посвящена ключевым понятиям современной лексикографии, возникшим в условия активного развития информационно-коммуникационных технологий. Актуальность темы обусловлена терминологическим разнообразием в научной литературе.

Ключевые слова: *электронная лексикография, компьютерная лексикография, корпусная лексикография, электронный словарь, электронные лексикографические ресурсы.*

Развитие компьютерных технологий привело к выделению новой области лексикографии – *компьютерной* или *электронной* лексикографии (*computer-aided lexicography / electronic lexicography / e-lexicography*), в широком понимании занимающейся созданием электронных словарей.

За последние десятилетия в лексикографии произошло много событий, коренным образом изменивших облик современного словаря. По словам Е.Е. Голубковой, с внедрением новых технологий, Интернета и лингвистических корпусов лексикография кардинальным образом изменилась, столкнувшись с новыми возможностями, новыми потребностями пользователя, новой «средой обитания» и новыми трудностями. Новая информационная парадигма потребовала от составителей словарей серьезной переквалификации, переоценки стратегий составления лексикографических источников, содержания информации и способов взаимодействия с «новым» типом пользователя «digital native» – «уроженцем Интернета» [3]. Чтобы идти в ногу со временем, лексикография должна выйти далеко за пределы традиционных рамок и стать «наукой об информации» (*informology* или *information science*) [13; 15]. В современной лексикографии основное внимание уделяется трём взаимосвязанным аспектам: информации, доступу к ней и пользователю. Информация обеспечивается лингвистикой, профиль пользователя – социологией, доступ – информационными науками [3]. По мнению О.М. Карповой, компьютеризация лексикографической деятельности заключается прежде всего в создании специализированных языковых машинных фондов, в разработке методов их формирования и оперативном представлении информации пользователю [5].

В данной статье рассматриваются основные понятия современной лексикографии, возникшие в условия активного развития информационно-коммуникационных технологий. Актуальность темы обусловлена терминологическим разнообразием в научной литературе.

Анализируемое направление современной лексикографии называют по-разному. В научном дискурсе встречаются термины «вычислительная лексикография», «кибернетическая лексикография», «машинная лексикография», «автоматическая / автоматизированная лексикография», «электронная лексикография» (цит. по: [10]). Однако в последнее время наиболее употребительным является термин «компьютерная лексикография», отражающий

средство создания электронных словарей и инструмент, позволяющий читателю использовать их возможности.

В статье Е.Ю. Чепик приводится следующее определение: компьютерная лексикография – это прикладная научная дисциплина в языкознании, изучающая методы, технологию и отдельные приемы использования компьютерной техники в теории и практике составления словарей. Специальные программы – базы данных, компьютерные картотеки, программы обработки текста – позволяют в автоматическом режиме формировать словарные статьи, хранить словарную информацию и обрабатывать её [11].

В понимании О.П. Фесенко [10] компьютерная лексикография – это: 1) прикладная отрасль лексикографии, возникшая в последние десятилетия XX века и призванная осуществить переход от создания традиционных (бумажных) словарей к их электронным вариантам и модификациям с использованием информационных технологий – специальных компьютерных программ для составления электронных словарей, программ лингвистического анализа и обработки текстов, лексикографических баз данных и переводческих систем. Электронный словарь становится главным объектом компьютерной лексикографии; 2) вся совокупность словарей, доступных читателю в электронном формате. В.П. Селегей – специалист в области компьютерного моделирования Естественного Языка и электронной лексикографии, директор по лингвистическим исследованиям компании АВВУУ – придерживается точки зрения, согласно которой компьютерная лексикография является особым направлением в практической лексикографии со своими собственными подходами не только к отображению, но и к содержанию словаря [9].

На наш взгляд, более уместным является использование термина «электронная лексикография», так как в этом случае происходит смещение акцента с технического компонента на формат данных, что позволяет расширить понятие и уйти от ограничения процесса разработки и использования словаря только с помощью компьютера, поскольку сейчас многие лексикографические ресурсы устанавливаются не только на компьютеры, но и на другие электронные устройства (смартфоны, планшеты и т.п.).

Отдельной областью и новым направлением современной лексикографии является корпусная лексикография (corpus-oriented lexicography). «Корпусной лексикографией называют область изучения языка на основе текстовых или акустических корпусов, при постоянном использовании компьютера в определённых фазах хранения, извлечения или анализа данных» [5]. Значение корпусной лексикографии для развития лингвистической науки огромно, так как многие лингвистические исследования проводятся сегодня с опорой на корпусные данные; научные описания языковой системы, а также авторитетные академические словари составляются на основе корпусов этих языков. Так, популярная серия словарей Collins базируется на известном языковом корпусе The Bank of English, лексикографические произведения серии Oxford University Press базируются на British National Corpus, словари серии Лонгман – на Longman Mini Concordances. В основе немецкоязычных словарей издательства Duden находится собственный корпус текстов, словарь DWDS (Das Digitale Wörterbuch der deutschen Sprache) функционирует в тесной взаимосвязи с 15 корпусами немецкого языка. По словам В.А. Букиной, языковые банки

данных обладают большой ценностью, так как сотни примеров, содержащихся в них, представляют собой точные данные для понимания того, как именно, в какой коннотации и с какими употребительными ограничениями слова английского языка используются его носителями. Таким образом, они могут способствовать максимальному приближению к образу мышления британцев, а также пониманию мира их языка [2].

Другим направлением лексикографии, связанным с использованием компьютера, является, по словам О.М. Карповой, кибернетическая лексикография, или киберлексикография (по аналогии Интернет – киберпространство), получившая в конце XX века широкое распространение. М. Карр понимает термин «киберлексикография» (cyberlexicography) как «использование Интернет для составления или создания словаря» (цит. по: [5]).

Объектом исследования для выше описанных областей современной лексикографии является так называемый электронный словарь. Самому понятию «электронный словарь» дать чёткое определение достаточно сложно. Для словарей, создаваемых с помощью компьютерных технологий, в научной литературе используются различные термины: «машинный словарь», «автоматический словарь», «автоматизированный словарь», «компьютерный словарь», «электронный словарь» (цит. по: [12]). При этом различные электронные лексикографические ресурсы могут сильно отличаться друг от друга не только содержанием, но и структурой, назначением, техническим инструментарием, так что возникает вопрос о правомерности их отнесения к одной категории «электронных словарей».

В широком понимании электронный словарь представляет собой особый лексикографический объект, обладающий рядом специфических структурных особенностей. Л.П. Нелюбин приводит такое определение: электронный словарь – это любой упорядоченный, относительно конечный массив лингвистической информации, представленный в виде списка, таблицы или перечня, удобного для размещения в памяти ЭВМ и снабжённого программами автоматической обработки и пополнения [6]. По мнению Х. Неси, термин «электронный словарь» может быть использован для обозначения любого справочного материала, хранящегося в электронном виде и предоставляющего информацию о написании, значении или использовании слов [14]. В понимании Я.А. Перванова, электронные словари – это действующие продукты, созданные в электронном виде, существующие на электронном носителе, предназначенные для электронного чтения и имеющие перспективы электронного развития [7].

В диссертационном исследовании И.И. Саженина «Корпусные методы в лексикографии: опыт создания модели Словарного корпуса» [8] был проведен анализ научной литературы, в которой поднимаются вопросы, касающиеся проблемы определения понятия «электронный словарь». Автор исследования отмечает, что спектр лексикографических объектов в электронной форме настолько широк, что сложно подобрать единый термин для их описания, а существующие определения недостаточны для их исчерпывающей характеристики. При этом И.И. Саженину удалось выявить ряд признаков, характеризующих данные объекты. С позиции содержания электронные лексикографические ресурсы делятся на две категории: те, что используют материалы уже существующих словарей, и те, что создаются с новым содержанием. При

рассмотрении адресованности также можно выделить две категории: ресурсы, адресованные человеку как конечному пользователю, а также ресурсы, адресованные машине. Л.Н. Беляева отмечает, что принято различать *автоматизированные словари*, предназначенные для конечного пользователя-человека, и *автоматические словари*, предназначенные для программной обработки текста. В такой формулировке под автоматическим словарем подразумевается автоматический переводной словарь (АПС), предназначенный в основном для обработки научно-технических текстов. Такие словари, являясь упорядоченным массивом лингвистической информации, регистрируют и хранят лексические единицы (слова и словосочетания) с их морфологическими, синтаксическими и семантическими характеристиками, необходимыми для синтаксического анализа (парсинга) и автоматического перевода текста [1]. А.А. Залевская также различает словари по данному критерию, однако обозначает их с помощью иных терминов: *электронные словари* – для конечного пользователя-человека, *компьютерные словари* – для компьютерных программ, работающих с текстами на естественных языках, в том числе включённых в системы машинного перевода, автоматического реферирования, информационного поиска и т.д. [4: 16]. Е.Ю. Чепик оба типа словарей называет *автоматическими*: «автоматический словарь, представляет собой словарь в особом машинном формате, который предназначен как для работы с ним непосредственно пользователем, так и самой ЭВМ. Таким образом, различают автоматические словари для пользователя-человека и автоматические словари для обработки текста специальной программой. Словари для пользователя по своей структуре, внешнему виду (интерфейсу) и ряду других моментов отличаются от тех, которые входят в системы машинного перевода, системы автоматического реферирования, информационного поиска и т.д.» [11].

Для ряда исследователей критерий формы существования лексикографического ресурса (электронная или бумажная) является достаточным для отнесения словаря в категорию электронных (автоматических, автоматизированных), для других исследователей данный критерий не является определяющим. Так, некоторые лингвисты предпринимают попытку переосмыслить существующие подходы к созданию электронных лексикографических ресурсов, отрицая путь воспроизводства словарного и структурного содержания бумажного словаря в его электронной версии [7; 9]. По мнению В.П. Селегея, *электронный словарь* является продуктом, который необходимо создавать, а не воспроизводить, при этом создавать его на совершенно новых принципах. В нём могут быть реализованы и введены в обращение многие продуктивные идеи, не востребованные по разным причинам в бумажных словарях [9]. Качественное отличие электронного словаря от бумажного видится не столько в электронной разметке и гипертексте, в карточках и окнах на месте бумажных страниц и словарных статей, сколько в новой манере лексикографической интерпретации языковой семантики [7].

Основной проблемой при создании электронных лексикографических ресурсов является то, что машина не способна в полной мере работать с текстом на естественном языке для репрезентации пользователю информации, соответствующей возможному спектру его запросов. Данная проблема была

решена специалистами, разрабатывающими корпусы текстов посредством использования такого инструмента, как разметка [9].

По словам Е.Е. Голубковой, современные электронные словари представляют собой гибкую интерактивную систему, обеспечивающую доступ к информации самого разного характера, выходящей далеко за пределы сугубо лингвистических знаний. Это скорее мультимедийные базы данных, всё больше расширяющие возможности и объём электронных источников информации и приближающие их к депоzitариям информации тезаурусного типа, максимально приближенных к способу хранения информации о словах в ментальном лексиконе человека. При этом речь идёт о переходе от словаря-книги к его современному последователю – *lexicographic tool* [3].

На наш взгляд, в современных условиях следует говорить не об отдельных электронных словарях, а об электронных лексикографических системах, которые представляют собой комплексные лексикографические ресурсы, обеспечивающие взаимосвязь с различными источниками информации. Это могут быть системы, обращающиеся только к «своим» словарям, т.е. словарям определённого издательства или системы (например, Duden, DWDS, OWID), либо к словарям различных издательств (например, лексикографическая система «YourDictionary.com», «Slovar-Vocab.com», раздел «Словари» на портале «Грамота.ру», раздел «Энциклопедиум» на портале «Библиоклуб.ру», энциклопедический ресурс «Рубикон»). В любом случае, автоматизированность и гипертекстуальность позволяют современным электронным словарям работать в тесной взаимосвязи друг с другом; пользователь сможет самостоятельно выбрать электронные издания, по которым будет осуществляться поиск информации.

Список литературы

1. Беляева Л.Н. Потенциал автоматизированной лексикографии и прикладная лингвистика [Электронный ресурс] // Известия Российского государственного педагогического университета им. А.И. Герцена. 2010. № 134. URL: <http://cyberleninka.ru/article/n/potencial-avtomatizirovannoy-leksikografii-i-prikladnaya-lingvistika> (дата обращения: 25.09.2015).
2. Букина В.А. Современная английская лексикография [Электронный ресурс] // Вестник Ленинградского государственного университета им. А.С. Пушкина. 2014. Выпуск № 4. Том 1. URL: <http://cyberleninka.ru/article/n/sovremennaya-angliyskaya-leksikografiya#ixzz3IPndWxTi> (дата обращения: 25.09.2015).
3. Голубкова Е.Е. Новый облик современной лексикографии: будет ли существовать словарь через 20 лет? [Электронный ресурс] // Вестник Московского государственного лингвистического университета. 2014. № 20. С. 73-80. URL: <http://elibrary.ru/item.asp?id=22796673> (дата обращения: 25.09.2015).
4. Залевская А.А. Информационные технологии в лингвистике: практикум для студентов 1 курса. 3-е изд., испр. Тверь: Твер. гос. ун-т, 2015. 28 с.
5. Карпова О.М. Методические рекомендации к спецкурсу «Проблемы современной лексикографии». Иваново, 2002. [Электронный ресурс]. URL: <http://w3.ivanovo.ac.ru/win1251/proect/karпова/elslovari.pdf> (дата обращения: 25.09.2015).
6. Нелюбин Л.П. Перевод и прикладная лингвистика. М.: Высшая школа, 1983. 208 с.
7. Перванов Я.А. Заметки по электронной лексикографии. [Электронный ресурс]. URL: <http://arxiv.org/pdf/1107.1753.pdf> (дата обращения: 25.09.2015).

8. Саженин И.И. Корпусные методы в лексикографии: опыт создания модели словарного корпуса: автореф. дис. ... канд. филол. наук. Новосибирск, 2013. 24 с.
9. Селегей В.П. Электронные словари и компьютерная лексикография [Электронный ресурс]. URL: <http://goo.gl/ujMnG> (дата обращения: 25.09.2015).
10. Фесенко О.П. Компьютерная лексикография: программные средства и продукт научной отрасли [Электронный ресурс] // Инновационное образование и экономика. 2014. № 15. С. 28–30. URL: <http://elibrary.ru/item.asp?id=21950987> (дата обращения: 25.09.2015).
11. Чепик Е.Ю. Компьютерная лексикография как одно из направлений современной прикладной лингвистики [Электронный ресурс] // Ученые записки ТНУ. Т.19 (58). №2: Филология. Симферополь. 2006. С. 274–280. URL: http://www.science.cri-me.a.edu/zapiski/2006/filologiya/uch_19_2fn/chepik_51.pdf (дата обращения: 25.09.2015).
12. Чумарина Г.Р. Классификация электронных словарей в современной лексикографии и лексикологии и особенности их использования [Электронный ресурс] // Балтийский гуманитарный журнал. 2013. № 4. С. 123–126. URL: <http://elibrary.ru/item.asp?id=21142170> (дата обращения: 25.09.2015).
13. Bergenholtz H., Theo J. D. Bothma. Needs-adapted Data Presentation in e-Information Tools. [Electronic resource]. URL: <http://books9.org/h/henning-bergenholtz-s45/> (accessed at 25.09.2015).
14. Nesi H. Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art // Proceedings of the Ninth EURALEX International Congress, EURALEX 2000: Stuttgart, Germany, August 8th-12th. 2000. P. 839–847. [Electronic resource]. URL: http://www.euralex.org/elx_proceedings/Euralex2000/099_Hilary%20NESI_Electronic%20Dictionaries%20in%20Second%20Language%20Vocabulary%20Comprehension%20and%20Acquisition_the%20State%20of%20the%20Art.pdf (accessed at 25.09.2015).
15. Tarp S. Beyond Lexicography: New Visions and Challenges in the Information Age // H. Bergenholtz, S. Nielsen & S. Tarp (Eds.) Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow. Bern: Peter Lang, 2009. Pp. 17–32.

BASIC DEFINITIONS OF ELECTRONIC LEXICOGRAPHY

A.V. Palkova

Tver State University, Tver

This article is devoted to the basic terms of modern electronic lexicography, which come into existence in the context of dynamic development of information and communication technology. The relevance of the subject is determined by terminological variety in the research literature.

Keywords: *computer-aided lexicography, electronic lexicography, e-lexicography, corpus-oriented lexicography, electronic dictionary.*

Об авторе:

ПАЛКОВА Анна Викторовна – кандидат филологических наук, доцент кафедры немецкого языка Тверского государственного университета, e-mail: anna_sap@rambler.ru.