

ФИЛОЛОГИЯ

УДК 81'374.2, 81'374.7

Н.А. Агапова, Н.Ф. Картофелева

О ПРИНЦИПАХ СОЗДАНИЯ ЭЛЕКТРОННОГО СЛОВАРЯ ЛИНГВОКУЛЬТУРОЛОГИЧЕСКОГО ТИПА: К ПОСТАНОВКЕ ПРОБЛЕМЫ

*Исследование выполнено в рамках государственного задания на выполнение НИР
«Изучение историко-культурного наследия России (сибирский аспект)» (код проекта 2059).*

Обосновывается актуальность создания электронных словарей различных типов; конкретизируются понятия «электронный словарь» и «электронный вариант словаря»; проводится сравнение информативных и навигационных возможностей традиционных бумажных и электронных лексикографических продуктов. Дается краткая характеристика существующих компьютерных словарей; рассматривается понятие «лингвокультурология»; намечаются перспективы создания словаря лингвокультурологического типа в электронном формате. Авторы выстраивают общую концепцию словаря, раскрывают содержание основных этапов работы над ним, презентуют базовый принцип организации материала в рамках создаваемого лексикографического продукта.

Ключевые слова: электронная лексикография; корпусная лингвистика; словарь компьютерного типа; лингвокультурология; этнолингвистика; лингвокультурология; народная примета; лингвокультурологический комментарий.

Лексикографическая практика является неотъемлемой и органичной частью науки о языке. При этом данные словаря могут быть использованы в качестве материалов исследования, и наоборот – сам словарь может выступать в качестве конечного результата, практического итога научной деятельности: «Словари обладают большими информативными возможностями и служат источником для теоретических построений и разработки других аспектов анализа. Существует и обратная связь “научная разработка – словарь”, при которой словарь представляет собой поле реализации той или иной научной концепции (например, теория мотивации, лингвоперсоналогия)» [1. С. 31–37].

Особую актуальность и востребованность в настоящее время приобрело такое направление, как компьютерная лексикография: сейчас разработка и создание электронных словарей, корпусов и баз данных являются наиболее продуктивными ее областями. Ценность данного направления заключается не только в разработке наглядных и удобных методов представления и классификации материала, но и в том, что электронные словари и компьютерные корпусы позволяют сохранить для дальнейшей обработки уже собранный (иногда – на протяжении очень долгого времени) материал, хранившийся ранее в менее удобной и неустойчивой к внешним воздействиям форме (бумажные картотеки, рукописные источники, аудио-, а также видеоносители, уязвимые перед внешними пагубными факторами, – аудиокассеты, видеокассеты, дискеты и т.д.).

Существует необходимость конкретизации терминов «компьютерная лексикография» и «электронный словарь»: как справедливо отмечает В.П. Селегей, «обычно подразумевается, что словарь на компьютере – это введенный в него бумажный словарь, снабженный удобными средствами поиска и отображения <...>. Компьютерная лексикография как область прикладной лингвистики, производящая такие словари, оказывается лишенной собственного языкового предмета. На ее долю оставляется только эффектная демон-

страция канонического содержания» [2]. Безусловно, необходимость перевода существующего массива словарей различных типов из бумажного формата в электронный назрела уже давно; данное направление сейчас достаточно успешно развивается в рамках отечественной лексикографии. Конечно, назвать это собственно научной деятельностью сложно, однако без предварительного проведенной технической работы невозможно получить готовый к исследованию материал – именно перевод информации в электронный формат является первым этапом работы с любым компьютерным корпусом. В отдельных случаях перевод данных в электронную форму сопровождается дополнительной и подчас весьма трудоемкой деятельностью. В качестве примера можно привести работу по переводу в электронный формат записей диалектной речи: это не только оцифровка самого материала (картотек, тетрадей), но и масштабная подготовительная работа, реализующаяся в виде разработки сложной системы разметки и ее применения в текстах, составляющих основу компьютерного диалектного корпуса [3].

В то же время словарь, существующий в электронном формате, – это не просто электронная (отсканированная, оцифрованная) версия уже созданного бумажного словаря, это полностью самостоятельный продукт, и форма его бытования логично продолжает и дополняет содержательную составляющую. Актуальным представляется замечание о том, что «компьютерная лексикография является особым направлением в практической лексикографии со своими собственными подходами не только к отображению, но и к содержанию словаря» [2].

Безусловно, электронный словарь – не только «особый лексикографический объект, в котором могут быть реализованы и введены в обращение многие продуктивные идеи, невостребованные по разным причинам в бумажных словарях» [Там же], но также и эффективный исследовательский инструмент, использование которого может способствовать более продуктивным

накапливанию и обработке информации. Кроме того, сами электронные словари в настоящее время всё чаще становятся объектом исследований. Обращение лингвистов к словарям именно этого типа, как справедливо отмечает О.С. Рублева, «объясняется тем, что в последнее время создается значительное количество электронных словарей, развивается всемирная сеть Интернет, информация хранится в электронном виде, но нет комплексного лингвистического описания электронных словарей и возможностей их использования пользователями в разных целях» [4. С. 3]. Между тем именно исследование возможностей использования электронных словарей в настоящее время представляется одним из наиболее продуктивных направлений данной сферы.

В.П. Селегей противопоставляет электронные словари традиционным, опираясь на несколько основополагающих принципов: исследователь называет их «антиномиями бумажной лексикографии». Во-первых, как отмечает автор, большой объем бумажного словаря (являющийся следствием богатой лексической базы и полноты словарных статей) выступает серьезным препятствием для удобного пользования: многотомный (или же просто очень объемный) словарь становится буквально непригодным для быстрого поиска информации – не говоря уже о том, что он абсолютно не мобилен в процессе использования. Таким образом, высокое качество словаря на содержательном уровне является непосредственной причиной низкого качества удобства его использования; в качестве примера автор приводит 20-томный Оксфордский словарь.

Вторая антиномия касается актуальности представленной в словаре информации: необходимо отметить, однако, что данное замечание справедливо в первую очередь для двуязычных (переводных) словарей, а также словарей терминологических: «Чрезвычайно долгий цикл создания и модификации фундаментальных бумажных словарей приводит к тому, что образ мира, который они фиксируют в системе своих значений, примеров и переводов, уже заметно отличается от действительности. Многие словари, основной корпус статей которых сформировался в языковой атмосфере середины века, представляют собой лексикографические музеи (а то и терминологические кладбища, если говорить о специализированных словарях)» [2].

Третье противоречие строится на асимметрии формы и содержания: исследователь говорит о том, что чем интереснее собственно лексикографическая концепция словаря, чем интегральнее средства описания лексических значений, тем уже его лексическая база.

Еще один недостаток бумажных (традиционных) словарей – отсутствие удобной навигации: в данном вопросе наиболее объективной представляется позиция пользователя, работающего со словарем; при этом чем больше объем словаря, тем сильнее данный недостаток проявляется. Основным навигационным инструментом, используемым в словарях, является способ расположения материала: в алфавитном порядке или по тематическому принципу. Возможен и смешанный вариант данных способов подачи материала. Этим возможности бумажного словаря исчерпываются.

Возможности компьютерной лексикографии позволяют успешно преодолеть перечисленные трудности и

противоречия. Так, первая «антиномия» в рамках электронной лексикографии разрешается благодаря возможности поместить большой объем данных на легкий и мобильный носитель информации, а также разместить созданный словарь (корпус) в сети Интернет. Таким образом, к словарю можно будет обращаться практически в любое время и любом месте. Неблагоприятные последствия, связанные с долгим циклом создания словаря, можно нейтрализовать, если программа, в которой создается лексикографический продукт, допускает возможность внесения новой информации и корректировки уже имеющейся. Эта же опция позволяет пополнять базу словаря: таким образом, лексикографический объект перестает быть статичным, застывшим во времени и пространстве, и получает возможность обновляться, если таковая потребность возникает.

В отличие от бумажной, электронная лексикография обладает весьма обширными навигационными возможностями. Помимо традиционных способов организации материала – в алфавитном порядке и по тематическому принципу – лексикограф получает возможность выстроить максимально удобные для пользователя словаря связи между объектами: лексикографируемыми единицами, статьями, комментариями, списками источников и литературы, списком составителей и т.д.

О.С. Рублева продолжает мысль В.П. Селегея и создает развернутую классификацию достоинств (плюсов) электронных словарей (обширный объем словника, удобство и скорость поиска информации, наличие звуковой и графической иллюстрации слова, возможность самостоятельного заказа структуры словарной статьи через заданный путь поиска, возможность разностороннего раскрытия значения слова и др.) [4].

Наибольшее количество электронных лексикографических продуктов в настоящее время создается в области переводческой деятельности. Существовать переводные словари могут как в самостоятельном формате – в виде скачиваемой и устанавливаемой отдельной программы, так и в формате интерактивном, т.е. интегрированными в какой-либо электронный ресурс: иначе говоря, пользоваться таким словарем можно в режиме on-line, посетив нужный сайт в сети Интернет. Как правило, способы представления материала в таких словарях однотипны и достаточно просты: пользователю нужно в максимально короткие сроки получить точный перевод отдельной лексемы или выражения, это, в свою очередь, диктует специфику навигации: в таком словаре всегда есть поле для ввода переводимого слова и поле, в котором появляются перевод либо варианты перевода. Практически все двуязычные словари, существующие в электронном формате, подразумевают подобный способ работы с материалом. Специально для создания словарей переводческого типа компанией АВВУУ – одним из наиболее известных в России разработчиков, занимающихся созданием электронных словарей, – был придуман язык DSL (Dictionary Specification Language) [5]. Широко известных аналогов данного языка в рамках российской электронной лексикографии в настоящий момент не применяется: связано это, прежде всего, с тем, что большинство компьютерных приложений для использования в

разных областях – и электронная лексикография не является исключением – до сих создается за рубежом.

Использование языка DSL возможно и для создания толковых словарей: способ представления материала по принципу «слово – толкование» аналогичен первоначальной схеме «слово – перевод»; меняется только наполнение словаря.

Однако давно назрела необходимость создания электронных словарей не только переводного и толкового, но и других типов: перечисленные ранее «антиномии бумажной лексикографии» актуальны для всех лексикографических областей; не являются исключением словари лингвокультурологического и этнолингвистического типов. Закономерным выглядит предположение о том, что организация материала в словарях этого типа предполагает принципиально иной подход, отличающийся от подхода к организации материала толковых и двуязычных словарей; следовательно, и создание компьютерных приложений для них будет строиться иных на принципах.

Некоторыми учеными предлагается объединять словари лингвокультурологического и этнолингвистического типов одним понятием – «лингвокультурология». В толковании этого термина мы идем вслед за О.К. Ансимовой, говорящей о том, что «объектом лингвокультурологии являются языковые и неязыковые (например, мимика, жесты) единицы, содержащие культурную информацию, знание которой необходимо для успешной коммуникации в рамках определенного лингвокультурного сообщества; предметом – параметры описания данных единиц в соответствующих словарях, т.е. их лексикографическая интерпретация» [6]. Словари такого типа диктуют более сложную – по сравнению с другими типами словарей – форму организации материала: связано это, прежде всего, с тем, что в поле зрения лексикографа наряду с собственно языковыми попадают неязыковые явления (обычаи, обряды, жесты, костюмы...), требующие особого подхода к возможности их представления в рамках словарной статьи. Это актуально для традиционного – бумажного – формата словаря и является не менее актуальным для электронной лингвокультурологии.

В то же время специфика материала, отображаемого в словаре лингвокультурологического типа, соединившись с навигационными и техническими возможностями компьютерной лексикографии, способна дать продуктивные результаты в виде не только удобно выстроенной навигации между материалом и иллюстрациями [7]. Использование медиа-приложений (видео-, аудиозаписей, анимированных приложений) является эффективным инструментом для иллюстрирования статей лингвокультурологического характера.

Представляется справедливым утверждение о том, что в рамках отечественной культурологии пока не создано полноценных компьютерных словарей. Однако электронные словари на материале русского языка создаются сейчас достаточно активно; определенным разнообразием отличаются типы создаваемых электронных лексикографических продуктов. Достаточно подробный список существующих электронных словарей размещен в рамках ресурса «Грамота.ру» [8]: так, в разделе «Словари русского языка» представлены, в

основном, словари толкового типа. Причем на равных с новыми – недавно созданными – словарями фигурируют и уже ставшие классическими словари, воплощенные в электронном варианте («Толковый словарь живого великорусского языка В.И. Даля» [9]; «Толковый словарь русского языка» под ред. Д.Н. Ушакова [10]). Отдельную группу формируют энциклопедические, терминологические словари и издания [8]; иноязычные словари и переводчики также достаточно широко представлены в рамках обозначенного электронного ресурса. Компьютерных словарей лингвокультурологического и этнолингвистического типов среди перечисленных на сайте словарей и справочников не встречается.

Авторы настоящей статьи видят своей целью разработку электронного словаря именно лингвокультурологического типа: в качестве примера такового было принято решение разработать и создать электронный словарь народных примет, призванный в компьютерном формате объединить корпус народных примет и лингвокультурологический комментарий, сопровождающий собранный материал. Воплощение этого проекта в жизнь предполагает разработку общей концепции словаря, структуры (модели) словарной статьи и принципов организации представленного в словаре материала, а также разработку компьютерного приложения, посредством которого данным словарем можно будет эффективно пользоваться как составителям словаря, так и читателям.

Уникальность заявленного проекта словаря заключается, с одной стороны, в его формальном воплощении – подобные типы словарей пока не представлены в рамках лингвокультурологических (как, впрочем, и филологических в целом) работ. Создание электронного словаря подобного типа позволит выработать и апробировать ряд универсальных принципов лексикографической работы с материалом, обладающим лингвокультурологической ценностью, прежде всего в области репрезентации такого материала в формате компьютерного словаря. С другой стороны, актуальность и научная ценность исследования заключаются в содержательной части словаря. Это связано, в свою очередь, с тем, что содержательное наполнение заявленного лексикографического продукта базируется на уникальном материале – корпусе народных примет, фиксируемых в рамках работы с разнообразными источниками, обладающих несомненной культурологической ценностью, создаваемом в течение нескольких лет и пополняемом до сих пор.

Работа над представленным лексикографическим продуктом включает в себя несколько базовых этапов.

Первый этап. Сбор материала, который выступит в качестве основы словаря; написание словарных статей, репрезентирующих лингвокультурологический комментарий. В настоящий момент этот этап в целом завершен [11].

Второй этап. Разработка общей концепции словаря; выбор формы существования словаря; составление программистом технического задания в соответствии с разработанной концепцией.

Третий этап. Выявление функциональных и нефункциональных требований, детализация технического задания на основе выявленных требований; написание словарных статей для электронного словаря.

Четвертый этап. Проектирование архитектуры компьютерного приложения с несколькими вариантами поиска и группировки материала, выбор средств разработки.

Пятый этап. Реализация электронного лингвокультурологического словаря народных примет на основе спроектированной архитектуры приложения. Разработка программной платформы словаря, позволяющей осуществлять поиск по заданным параметрам, проводить статистическую обработку материала, сохранение результатов поиска. Завершение разработки электронного ресурса: создание интуитивно понятного интерфейса и удобной навигации.

Шестой этап. Апробация созданного электронного словаря в рамках научных исследований, проводимых в сфере лингвокультурологии и смежных с ней научных областей, а также в сфере лексикографической практики (на базе кафедры русского языка и лаборатории общей и сибирской лексикографии филологиче-

ского факультета Национального исследовательского Томского государственного университета).

В качестве базового принципа организации материала в словаре решено использовать *принцип ключевых слов*, ставший уже традиционным при создании лексикографических продуктов различной направленности. Основное отличие данного принципа от способов организации материала в словарях, написанных на языке DSL, заключается в том, что он позволит выстроить максимально удобную навигацию между материалом (корпусом примет) и лингвокультурологическим комментарием, сопровождающим данный материал. Данный способ организации материала, структура словарной статьи, а также специфика технического воплощения представленного лексикографического продукта – особенности проектирования архитектуры приложения и выбранных средств разработки – будут подробным образом описаны в последующих работах авторов.

ЛИТЕРАТУРА

1. Демешкина Т.А. Лексикографическое направление в Томской диалектологической школе. Итоги и перспективы // Вестник Томского государственного университета. Филология. 2011. № 3. С. 31–37.
2. Селегей В.П. Электронные словари и компьютерная лексикография. URL: http://www.lingvoda.ru/transforum/articles/selegey_a1.asp
3. Юрина Е.А. Томский диалектный корпус: в начале пути // Вестник Томского государственного университета. Филология. 2011. № 2. С. 58–63.
4. Рублева О.С. Слово в электронном словаре (с позиции пользователя электронными ресурсами) : автореф. дис. ... канд. филол. наук. Тверь : Тверской государственный университет, 2010. 20 с.
5. Ассоциация лексикографов Lingvo. URL: <http://www.lingvoda.ru/dictionaries/create.asp#p2>
6. Анисимова О.К. Лингвокультурология как отдельная филиация общей лексикографии. URL: http://www.sociosfera.com/publication/conference/2012/119/lingvokulturografiya_kak_otdelnaya_filiaciya_obwej_leksikografii/
7. Евсюкова Т.В. Лингвокультурологическая концепция словаря культуры : дис. ... д-ра филол. наук. Нальчик, 2002. 304 с. : ил.
8. Грамота.ру. URL: <http://www.gramota.ru/slovari/online>
9. Даль В.И. Толковый словарь живого великорусского языка. URL: <http://vidahl.agava.ru>
10. Ушаков Д.Н. Толковый словарь русского языка. URL: <http://ushdict.narod.ru>
11. Агапова Н.А. Лингвокультурологический потенциал ключевого слова народной приметы : дис. ... канд. филол. наук. Томск, 2013. 184 с.

Статья представлена научной редакцией «Филология» 27 марта 2014 г.

ON THE PRINCIPLES OF CREATING AN ELECTRONIC DICTIONARY OF THE LINGUOCULTUROLOGICAL TYPE: RAISING THE PROBLEM

Tomsk State University Journal. No. 382 (2014), 6-10.

Agapova Nina A., Kartofeleva Natalia F. Tomsk State University (Tomsk, Russian Federation). E-mail: NinaAgapowa@yandex.ru, natconnor.neo@gmail.com

Keywords: computational lexicography; corpus linguistics; electronic dictionary; linguoculturology; ethnolinguistics; linguoculturography; national sign; linguoculturological commentary.

Lexicographic practice is an essential and organic part of linguistics. Presently such a branch as computational lexicography has taken on special topicality and importance: development and implementation of electronic dictionaries, corpuses and databases are its most productive areas today. Undoubtedly, the need for converting the existing number of dictionaries of different kinds from paper to an electronic form appeared long ago: it is impossible to get some material prepared for work without any preliminary technical work – converting information to an electronic format is the first step of any activity with computer corpora. In particular cases data conversion may be accompanied by additional – and sometimes quite laborious – activities. Converting dialect speech records into an electronic format can be used for illustrative purposes: there is not only the digitization of the material itself (card indexes, notebooks, etc) but also huge preparation work realized with the development of a complex meta markup system and its adaptation in texts that form the basis of a computer dialect corpus. A dictionary existing in an electronic format is not just an electronic (scanned, digitized) version of a paper dictionary, it is a completely independent product, and its form of existence logically continues and supplements its content. At the present time, the greatest number of electronic lexicographic products is created in the area of translation industry. However, the need for creating Russian-language dictionaries of not only translational or explanatory types but also other types is obvious: linguoculturological and ethnolinguistic dictionaries are no exception. The authors of this paper aim at the development of an electronic dictionary of an indeed linguoculturological type: it was decided to develop and implement an electronic dictionary of national signs as an example which should unite the corpus of national signs and the linguoculturological commentary of the collected material in the electronic format. The realization of the project proposes the elaboration of the general concept of the dictionary, the structure (model) of a dictionary entry, and the principles of organisation of the material presented in the dictionary, and also it proposes the development of software to be conveniently and effectively used by dictionary compilers and readers. The originality of the project of the dictionary consists in, first of all, its physical implementation. The creation of an electronic dictionary of such a type allows elaborating and testing a series of universal principles for lexicographic work with the material of linguoculturological value, including the representation of such a material in the computer dictionary format. On the other hand, the topicality and scientific value of the research consist in the

physical part of the dictionary as well. It, in its turn, concerns with the fact that the content of the announced lexicographic product is based on the unique material – the corpus of national signs fixed in the process of work with different sources, it has certain culturological value and was created during several years and it is still supplemented.

REFERENCES

1. Demeshkina T.A. Lexicographic studies of Tomsk Dialect School. Results and prospects. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*, 2011, no. 3 (15), pp. 31-37. (In Russian).
2. Selegey V.P. *Elektronnye slovari i komp'yuternaya leksikografiya* [Electronic dictionaries and computer lexicography]. Available at: http://www.lingvoda.ru/transforum/articles/selegey_a1.asp.
3. Yurina Ye.A. Tomsk dialectal corpora: the starting point. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*, 2011, no. 2 (14), pp. 58-63. (In Russian).
4. Rubleva O.S. *Slovo v elektronnom slovare (s pozitsii pol'zovatelya elektronnyimi resursami)*. Avtoref. dis. kand. filol. Nauk [Word in an electronic dictionary (by the user of electronic resources). Abstract of Philology Cand. Diss.]. Tver': Tver' State University Publ., 2010. 20 p.
5. Assotsiatsiya leksikografov Lingvo [Lingvo Lexicographer Association]. Available at: <http://www.lingvoda.ru/dictionaries/create.asp#p2>.
6. Ansimova O.K. *Lingvokul'turografiya kak otdel'naya filiatsiya obshchey leksikografii* [Linguoculturography as a separate filiation of general lexicography]. Available at: http://www.sociosfera.com/publication/conference/2012/119/lingvokulturografiya_kak_otdelnaya_filiatsiya_obwej_leksikografii/.
7. Evsyukova T.V. *Lingvokul'turologicheskaya kontseptsiya slovarya kul'tury*. Dis. doktora filologicheskikh nauk [Linguoculturological concept of a culture dictionary. Philology Dr. Diss.]. Nal'chik, 2002. 304 p.
8. Gramota.ru. Available at: <http://www.gramota.ru/slovari/online/>. (In Russian).
9. Dal' V.I. *Tolkovyy slovar' zhivogo velikorusskogo yazyka* [The Explanatory Dictionary of the Living Great Russian Language]. Available at: <http://www.vidahl.agava.ru/>.
10. Ushakov D.N. *Tolkovyy slovar' russkogo yazyka* [The Explanatory Dictionary of the Russian Language]. Available at: <http://www.usdict.narod.ru/>.
11. Agapova N.A. *Lingvokul'turologicheskyy potentsial klyuchevogo slova narodnoy primety*. Dis. kandidata filologicheskikh nauk [Linguoculturological potential of the keyword of the folk sign. Philology Cand. Diss.]. Tomsk, 2013. 184 p.

Received: March 27, 2014