7. *Ineke van H., Bína V.* Country profile the Netherlands // Council of Europe ERICarts: Compendium of Cultural Policies and Trends in Europe, 16th edition, 2015. URL: http://www.culturalpolicies.net/down/netherlands_072015.pdf

8. *Inkei P., Vaspál V.* Country Profile Hungary // Council of Europe ERICarts: Compendium of Cultural Policies and Trends in Europe, 16th edition, 2015. URL: http://www.culturalpolicies.net/down/hungary_122014.pdf

9. *Moss L.* Sheffield's cultural industries quarter 20 years on: What can be learned from a pioneering example? International Journal of Cultural Policy, 2002, 8: 2. P. 211–219.

10. *McIsaa P. M.* Public-Private support of the Arts and German Cultural Policy. International Journal of Cultural Policy, 2007, 13: 4. P. 371–391.

11. *Pyykkönen M., Simanainen N., Sokka S.* The Strengthening of International Influences // What about culture policy? Interdisciplinary Perspectives on Culture and Politics. — Helsinki, 2009. P. 2–17.

12. *Rindzeviciute E., Svensson J., Tomson K.* The International Transfer of Creative Industries as a Policy Idea. The International Journal of Cultural Policy, 2015, 1: 1, P. 1–17.

13. The report of the World Commission on Culture and Development: Our Creative Diversity. Paris. UNESCO Publishing. 1995.

14. *Quinn R. M.* Distance or intimacy? — The arm's length principle, the British government and the arts council of Great Britain. International Journal of Cultural Policy, 1997, 4: 1, P. 127–159.

15. *Vestheim G.* Instrumental cultural policy in Scandinavian countries. International Journal of Cultural Policy, 1994, 1: 1. P. 57–71.

А. Н. Ефремова

КОМПЛЕКСНЫЙ ЭЛЕКТРОННЫЙ СЛОВАРЬ: ОСОБЕННОСТИ ПРОЕКТИРОВАНИЯ

В статье рассмотрены особенности создания комплексного электронного словаря на основе существующих проверенных лексикографических источников и предложены основные этапы его проектирования.

Ключевые слова: проектирование словаря, комплексный электронный словарь, макроструктура словаря, микроструктура словаря, распознавание, фрейм.

A. Efremova

COMPLEX ELECTRONIC DICTIONARY: DETAILS OF CREATION

The paper considers the details of complex electronic dictionary creation process which is based on the existing dictionaries. The main steps of creation such a dictionary are suggested.

Keywords: dictionary creating, complex electronic dictionary, dictionary macrostructure, dictionary microstructure, recognition, frame.

Создание и ведение электронных словарей и терминологических баз данных является одной из основных задач современной компьютерной лексикографии. Такие лексикографические ресурсы необходимы в первую очередь переводчику, который в процессе своей деятельности тратит до 30% времени на собственно терминологическую работу [12].

В настоящее время создано множество сетевых терминологических баз данных, как государственных, так и корпоративных: Eurodicatoum [13], IATE (InterActive Terminology for Europe) [13], EurTermBank [14], TaaS (Terminology as a Service) [15], LexSite и др. Среди сетевых словарей наиболее популярными являются ABBYY Lingvo, Мультитран и Мультилекс [8].

Однако переводчику часто необходим собственный лексикографический ресурс, настроенный на конкретную предметную область; для решения этой задачи могут служить комплексные электронные словари (КЭС), которые включают в свою структуру словари различных типов: как общие, так и отраслевые, как переводные, так и энциклопедические и толковые, а также специализированные словари переводчика [9]. Словарная статья такого словаря предполагает наличие разноаспектной информации о лексических единицах, которая может быть собрана из уже существующих проверенных лексикографических ресурсов, объединение которых в КЭС целесообразно с точки зрения использования в работе переводчика. При таком подходе создание основы КЭС предполагает извлечение информации из различных лексикографических источников, бумажных в том числе, и объединение ее в форме единого ресурса. Рассмотрение особенностей создания КЭС представляет особый интерес и в теоретическом, и в практическом плане.

Дело в том, что не существует детального универсального набора этапов проектирования словаря, процесс создания зависит от множества факторов: стартовые условия создания, тип словаря, его первоначальный замысел и др. [11, с. 50]. Однако возможно создание моделей работы над словарем, имеющих практическую ценность [11, с. 51]. Подобные модели создания традиционных бумажных словарей были предложены многими исследователями [1, с. 93; 5, с. 11; 6, с. 370–373].

Рассмотрим особенности создания комплексного электронного словаря на основе существующих проверенных лексикографических источников. При создании любого электронного словаря, в том числе и комплексного, необходимо учитывать, что в его составе принято различать собственно базу данных (коллекцию словарных статей) и систему программ, осуществляющую работу с этой базой [2, с. 15]. Словари, на основе которых строится база данных КЭС, в дальнейшем будем называть исходными. Определим основные этапы проектирования КЭС.

1. Выбор исходных лексикографических ресурсов

Состав исходных лексикографических ресурсов в первую очередь зависит от тематики текстов, для перевода которых создается словарь. Кроме того, при выборе исходных словарей необходимо учитывать одну из печальных тенденций в современной лексикографии — некоторые издания могут претендовать на статус словаря только по формальным признакам [10, с. 6].

2. Разработка макро- и микроструктуры КЭС

Макроструктура словаря реализуется на общих принципах организации структуры лексикографического произведения, принципах расположения языковых единиц в словаре и т. п. [6, с. 60–61]. Микроструктурой называют собственно структуру словарной статьи [3, с. 16]. Для электронного словаря выбор микроструктуры [1, с. 84–90; 3; 6, с. 56–60] представляет собой более сложную проблему, чем выбор его макроструктуры, поскольку требует анализа и выявления всего комплекса потенциально возможных и представленных в

словарях разных типов видов информации. Статья комплексного словаря должна максимально полно описывать лексическую единицу; формат словарной статьи, в свою очередь, должен обеспечивать возможность создания такого описания и его модификации и/или пополнения.

3. Проектирование базы данных КЭС

База данных КЭС предназначена для хранения информации о лексических единицах и должна наглядно представлять словарную информацию и обеспечивать быстрый поиск в ней.

Для представления словарной информации в базе данных КЭС может быть использована фреймовая модель знаний, основанная на идеях М. Минского [7] и его последователей. Согласно этой модели фрейм представляет собой иерархически организованную структуру данных, репрезентирующую знания о некоторой стереотипной ситуации или классе ситуаций. Верхние уровни структуры содержат данные, всегда справедливые для анализируемой ситуации, а нижние — пустые узлы (слоты), заполняемые конкретными данными соответствующей ситуации [7].

Основным преимуществом фреймов как модели представления знаний является их способность отражать наши представления о концептуальной организации памяти человека, а также их гибкость. Фреймовая модель позволяет достаточно наглядно представлять информацию, что облегчает как ее восприятие человеком, так и дальнейшую компьютерную обработку. Семантический фрейм, необходимый для описания значения лексической единицы в словаре, можно рассматривать как концептуальную структуру, описывающую событие, отношение или объект и его участников (см., например, проект FrameNet).

При таком подходе каждая словарная статья организуется как система фреймов; структура такой системы основана на разработанном формате статьи комплексного словаря. Соответственно, система включает фреймы разного уровня описания (например, основной фрейм статьи, фрейм зоны заголовка, фрейм зоны перевода и т. д.). Имена слотов фреймов соответствуют наименованиям зон или компонентов зон (например, в зоне «Пример» можно выделить подзоны «Заголовок примера» и «Перевод примера»). Значением слота является контент соответствующей зоны.

4. Распознавание макро- и микроструктуры исходных словарей и заполнение базы данных КЭС

Автоматизация данного этапа позволит значительно ускорить процесс создания КЭС.

Для автоматического создания основы КЭС необходимо разработать универсальный метод распознавания структуры исходных словарей. Бумажные словари при этом должны быть предварительно преобразованы в электронный формат путем сканирования с помощью современных средств распознавания текста [2, с. 21].

Распознавание структуры каждого исходного словаря предполагает распознавание его макро- и микроструктуры.

Распознавание макроструктуры словаря основано на выделении границ словарных статей в исходном тексте. Решение этой задачи требует определения признаков словарной статьи и зависит от способа представления статей в словаре — алфавитного, гнездового, алфавитно-гнездового, тематического и т. д.

Распознавание микроструктуры словаря заключается в выделении границ зон внутри словарных статей и их классификации, что может устанавливаться на основе признаков

каждой зоны. К признакам границ зон и их типов относятся, в частности, различные специальные символы, отделяющие одну зону от другой, позиционные характеристики расположения конкретной зоны относительно других зон словаря, изменение языка, используемого для описания информации в зоне (для переводных словарей), регистр букв (верхний/нижний) и т. д. На этом этапе важно учитывать, что в разных словарях признаки зон, содержащих информацию одного типа, могут отличаться кардинально.

Универсальная процедура распознавания структуры исходных словарей должна быть основана на полном наборе признаков, характерных для различных словарей. Установление такого набора требует специального исследования макро- и микроструктур большого количества словарей различных типов.

Следует иметь в виду, что при построении универсальной процедуры распознавания невозможно полностью учесть все особенности любого привлекаемого словаря, поэтому необходима либо ее постоянная доработка, либо дальнейшая «ручная» проверка распознанных данных.

При заполнении базы данных сложной проблемой может оказаться также синтез информации из нескольких словарей: необходимо разработать алгоритм, определяющий словарные статьи, описывающие одно и то же понятие [4, с. 27].

5. Разработка программного обеспечения для работы с базой данных КЭС и его тестирование

Этап включает в себя разработку алгоритма поиска по базе данных словаря, проектирование пользовательского интерфейса и пр.

Рассмотрение особенностей создания КЭС позволило выявить основные направления исследования для его проектирования, а также определить этапы в процессе создания подобных словарей. Однако следует отметить, что предложенная модель содержит только основные этапы процесса проектирования КЭС, для создания словаря по такой схеме необходима более детальная проработка каждого этапа.

СПИСОК ЛИТЕРАТУРЫ

1. *Баранов А. Н.* Введение в прикладную лингвистику: Учебное пособие. М.: Эдиториал УРСС, 2001. 360 с.
2. *Беляева Л. Н.* Автоматизированная лексикография: гуманитарные технологии. СПб.: Изд-во РГПУ им. А. И. Герцена, 2011. 75 с.
3. *Берков В. П.* Двуязычная лексикография: учебник. 2-е изд., перераб. и доп. М.: ООО «Издательство Астрель»: ООО «Издательство АСТ»; ООО «Транзиткнига», 2004. 236, [4] с.
4. *Власов Д. Ю., Пальчунов Д. Е., Степанов П. А.* Автоматизация извлечения отношений между понятиями из текстов естественного языка // Вестник НГУ. Серия: Информационные технологии. 2010. Т. 8. Вып. 3. С. 23–33.
5. *Городецкий Б. Ю.* Проблемы и методы современной лексикографии // Новое в зарубежной лингвистике. М.: Прогресс, 1983. Вып. 14. С. 5–22.
6. *Дубичинский В. В.* Лексикография русского языка: Учеб. пособие. М.: Наука: Флинта, 2008. 432 с.
7. *Минский М.* Фреймы для представления знаний: пер. с англ. М.: Энергия, 1979. 152 с., ил.
8. *Кит М. С.* О стратегии построения высокоэффективных сетевых словарей (на базе разработки словаря LexSite) // Вестник РГГУ. 2010. № 9. С. 149–160.

9. *Климзо Б. Н.* Ремесло технического переводчика: об английском языке, переводе и переводчиках научно-технической литературы. 2-е изд., перераб. и доп. М.: Р. Валент, 2006. 508 с.

10. *Козырев В. А., Черняк В. Д.* Современные ориентации отечественной лексикографии // Вопросы лексикографии / Томский государственный университет. Томск. 2014. № 1 (5). С. 5–15.

11. *Кудашев И. С.* Проектирование переводческих словарей специальной лексики. Helsinki: Helsinki University Print, 2007. 443 с.

12. *Gornostay T.* Terminology management in real use // Proceedings of the 5th International Conference "Applied Linguistics in Science and Education". Saint-Petersburg, 2010. P. 25–26.

13. *Johnson I., Macphail A.* IATE — Inter-Agency Terminology Exchange: Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union [Электронный ресурс]. URL: http://www.mt-archive.info/LREC-2000-Johnson.pdf (дата обращения: 27.09.15).

14. *Maslias R.* Combining EU Terminology with Communication and Ontology Research // Terminology and Knowledge Engineering 2014: Proceedings of the Conference, 19–21 Jun 2014. Berlin, 2014. P. 48–56.

15. *Vasiljevs A., Pinnis M., Gornostay T.* Service model for semi-automatic generation of multilingual terminology resources // Terminology and Knowledge Engineering 2014: Proceedings of the Conference, 19–21 Jun 2014. Berlin, 2014. P. 67–76.

REFERENCES

1. *Baranov A. N.* Vvedenie v prikladnuju lingvistiku: Uchebnoe posobie. M.: Jeditorial URSS, 2001. 360 s.

2. *Beljaeva L. N.* Avtomatizirovannaja leksikografija: gumanitarnye tehnologii. SPb.: Izd-vo RGPU im. A. I. Gercena, 2011. 75 s.

3. *Berkov V. P.* Dvujazychnaja leksikografija: uchebnik. 2-e izd., pererab. i dop. M.: OOO «Izdatel'stvo Astrel'«: OOO «Izdatel'stvo AST»; OOO «Tranzitkniga», 2004. 236, [4] s.

4. *Vlasov D. Ju., Pal'chunov D. E., Stepanov P. A.* Avtomatizacija izvlechenija otnoshenij mezhdu ponjatijami iz tekstov estestvennogo jazyka // Vestnik NGU. Serija: Informacionnye tehnologii. 2010. T. 8. Vyp. 3. S. 23–33.

5. *Gorodeckij B. Ju.* Problemy i metody sovremennoj leksikografii // Novoe v zarubezhnoj lingvistike. M.: Progress, 1983. Vyp. 14. S. 5–22.

6. *Dubichinskij V. V.* Leksikografija russkogo jazyka: Ucheb. posobie. M.: Nauka; Flinta, 2008. 432 s.

7. *Minskij M.* Frejmy dlja predstavlenija znanij: per. s angl. M.: Jenergija, 1979. 152 s., il.

8. *Kit M. S.* O strategii postroenija vysokojeffektivnyh setevyh slovarej (na baze razrabotki slovarja LexSite) // Vestnik RGGU, 2010. № 9. C. 149–160.

9. *Klimzo B. N.* Remeslo tehnicheskogo perevodchika: ob anglijskom jazyke, perevode i perevodchikah nauchno-tehnicheskoj literatury. 2-e izd., pererab. i dop. M.: R. Valent, 2006. 508 s.

10. *Kozyrev V. A., Chernjak V. D.* Sovremennye orientacii otechestvennoj leksikografii // Voprosy leksikografii / Tomskij gosudarstvennyj universitet. Tomsk. 2014. № 1 (5). S. 5–15.

11. *Kudashev I. S.* Proektirovanie perevodcheskih slovarej special'noj leksiki. Helsinki: Helsinki University Print, 2007. 443 s.

12. *Gornostay T.* Terminology management in real use // Proceedings of the 5th International Conference "Applied Linguistics in Science and Education". Saint-Petersburg, 2010. P. 25–26.

13. *Johnson I., Macphail A.* IATE — Inter-Agency Terminology Exchange: Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union [Jelektronnyj resurs]. URL: http://www.mt-archive.info/LREC-2000-Johnson.pdf (data obrashhenija: 27.09.15).

14. *Maslias R.* Combining EU Terminology with Communication and Ontology Research // Terminology and Knowledge Engineering 2014: Proceedings of the Conference, 19–21 Jun 2014. Berlin, 2014. P. 48–56.

15. *Vasiljevs A., Pinnis M., Gornostay T.* Service model for semi-automatic generation of multilingual terminology resources // Terminology and Knowledge Engineering 2014: Proceedings of the Conference, 19–21 Jun 2014. Berlin, 2014. P. 67–76.