

УДК 81'374=161.1:004.415

КОМПЬЮТЕРНАЯ ЛЕКСИКОГРАФИЯ КАК ОДНО ИЗ НАПРАВЛЕНИЙ СОВРЕМЕННОЙ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

Чепик Е.Ю.

В современной лингвистике развитие лексикографии определяется как «степень моделирования языковой системы» [14, с.27]. В рамках данной статьи мы рассматриваем одно из *актуальных* направлений компьютерной лингвистики – лексикографическое моделирование, целью которого является систематизация и классификация семантического пространства политического дискурса.

Постановка проблемы. Очень важным для такого моделирования является определение важнейших понятийных категорий компьютерной лингвистики «фрейм» и «сценарий», так как «жизненный контекст дискурса моделируется в форме «фреймов» (типовых ситуаций) или «сценариев» (делающих акцент на развитии ситуаций)» [2, с.137].

Что понимается под фреймом при лексикографическом подходе? Большинство лингвистов рассматривает фрейм как категорию мыслительную, единицу памяти, особым образом организованную структуру представления знаний о том или ином явлении (М. Минский, Дж. Лакофф, Ч. Филлмор, Е.С. Кубрякова, В.З. Демьянков и др. [11]). Знание о явлении представляется в виде важнейших признаков, которые составляют суть этого явления. Эти представления существуют в сознании человека и образуют понятийные (концептуальные) «блоки», определяющие наше виденье действительности, наше восприятие и членение окружающего мира, а также обуславливающие национальную специфику стереотипов как поведенческих в целом, так и коммуникативных – в частности. Важным для определения фрейма является понимание того, как происходит структурирование нашего знания о том или ином явлении действительности и как проявляются понятийные блоки в языковом сознании. Процесс концептуализации связан, на наш взгляд, с ассоциативно-вербальными сетями [8], слотами [12], набор которых формирует в сознании человека фрейм-структуру. При восприятии и понимании произнесенной языковой структуры, например слова, в сознании языковой личности возникают ассоциации, которые помогают ей идентифицировать услышанное слово с уже известными ключевыми словами, хранящимися в сознании и дающими имя слоту. Через понимание слота или нескольких слотов, поскольку в сознании происходит цепочка подобных ассоциаций, реципиент определяет фрейм, который имелся в виду в данном высказывании.

Таким образом, под фреймом понимается особым образом организованная структура знаний о типизированном объекте или стереотипной ситуации.

Фрейм фиксирует прежде всего те факты, которые характерны и типичны для данной культуры. Он закрепляет стандартные, стереотипные способы действия субъекта в какой-либо типичной ситуации и тем самым регулирует поведение человека (в том числе и речевое). Механизм регулирования М. Минский представляет так: человек, оказавшись в определенной ситуации, извлекает из памяти соответствующую структуру

знаний о ней – фрейм, которую при необходимости корректирует применительно к реальности [12, с.250].

В отличие от фрейма, сценарий – это процедурный способ представления стереотипного знания, который формулируется в терминах алгоритма или инструкции. Сценарий вырабатывается в результате интерпретации текста, в процессе создания ключевыми словами и идеями текста тематических (сценарных) структур, извлекаемых из памяти на основе стандартных, стереотипных значений.

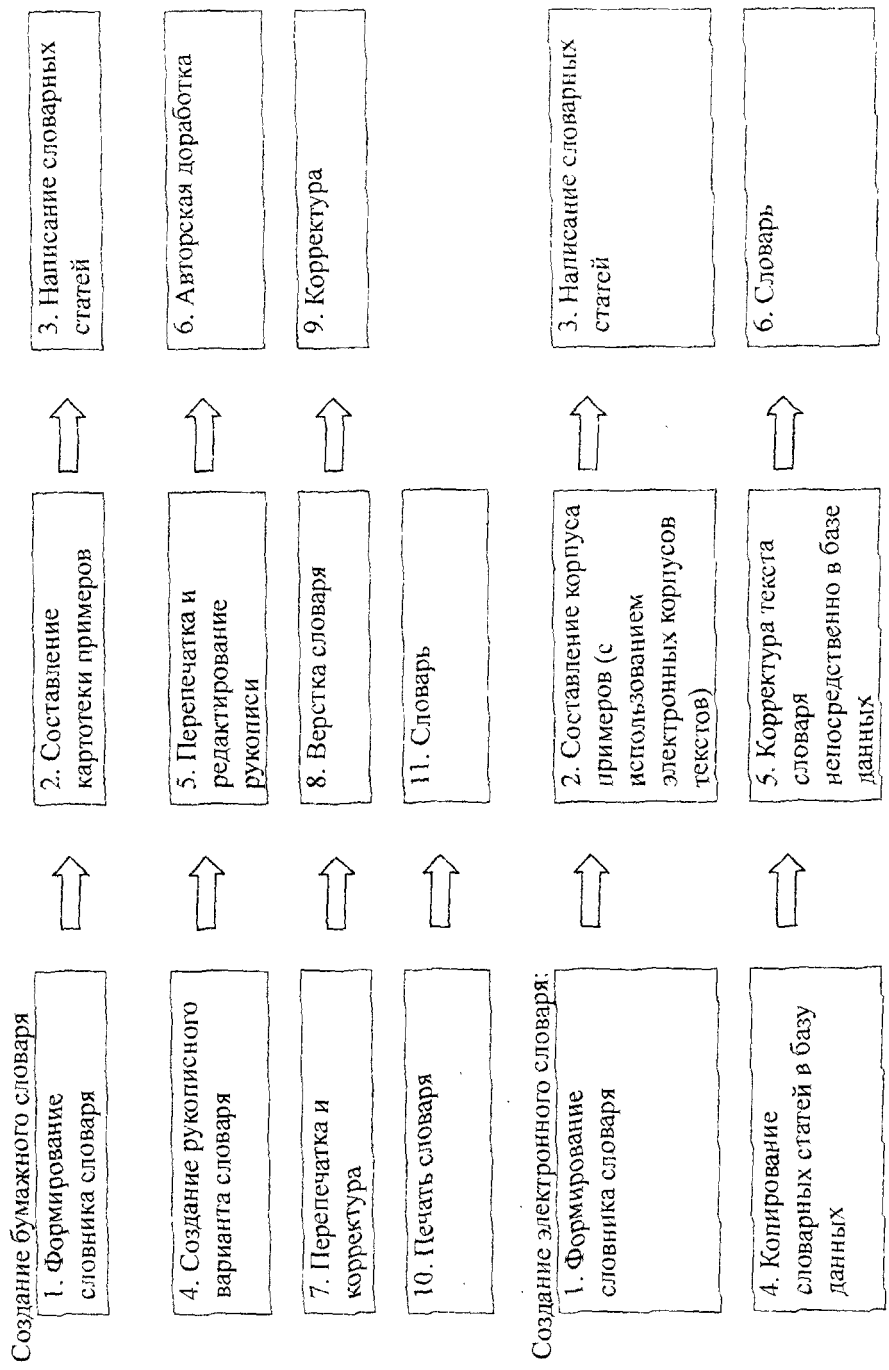
Определенным образом организованный набор структур знаний формирует «модель мира» когнитивной системы и ее компьютерной модели. В системах искусственного интеллекта модель мира образует особый блок, в который в зависимости от выбранной архитектуры могут входить общие знания о мире (в виде простых пропозиций типа «летом жарко» или в виде правил продукций «если на улице холодно, то надо надеть теплые вещи»), некоторые специфические факты («Самое соленое море в мире – Красное»), а также ценности и их иерархии.

В инструментальном аспекте знание может быть воплощено в совокупности дескрипций (описаний), в наборе данных, с одной стороны, и в алгоритме, инструкции, которую выполняет компьютерная или какая-либо другая модель когнитивной системы, – с другой [17].

В рамках компьютерной лексикографии разрабатываются компьютерные технологии составления и эксплуатации словарей. Специальные программы – базы данных, компьютерные картотеки, программы обработки текста – позволяют в автоматическом режиме формировать словарные статьи, хранить словарную информацию и обрабатывать ее. Иллюстрацией особенностей создания электронного словаря по сравнению с бумажным может служить следующая схема:

В отечественной лексикографической практике различные компьютерные лексикографические программы разделяются на две большие группы: с одной стороны, программы поддержки лексикографических работ и, с другой стороны, автоматические словари различных типов, включающие лексикографические базы данных. Автоматический словарь – это словарь в специальном машинном формате, предназначенный для использования на ЭВМ пользователем или компьютерной программой обработки текста. Иными словами, различаются автоматические словари конечного пользователя-человека и автоматические словари для программ обработки текста. Автоматические словари, предназначенные для конечного пользователя, по интерфейсу и структуре словарной статьи существенно отличаются от автоматических словарей, включенных в системы машинного перевода, системы автоматического реферирования, информационного поиска и т.д. Такие словари создаются как самостоятельные лексикографические продукты (так, например, БРУМС (Большой русский морфемный словарь) [10], словари рифм русского языка [9], частотные словари словоформ и др.) Но чаще всего они являются компьютерными версиями хорошо известных обычных словарей. Так, в сфере программного обеспечения имеются компьютерные аналоги толковых словарей английского языка (автоматический Webster, автоматический вариант Нового большого англо-русского словаря под ред. Ю.Д. Апресяна и Э.М. Медниковой), существует и компьютерная версия словарей С.И. Ожегова, Н.Д. Ушакова, В.И. Даля и т.д.

Схема 1. Сравнительный анализ этапов создания бумажного и электронного словаря



В настоящее время в компьютерной лексикографии под электронным словарем принято понимать «словник, процедуры укладання якого здійснює комп'ютер» [7, с.54].

Для электронной лексикографии особое значение имеет идея гипертекста как основного способа организации виртуального текстового пространства. А.Н. Баранов рассматривает гипертекст с нескольких позиций: «С одной стороны, это особый способ представления, организации текста, с другой – новый вид текста, противопоставленный по многим своим свойствам обычному тексту, сформированному в гутенберговской традиции книгопечатания. И, наконец, это новый способ, инструмент и новая технология понимания текста» [3, с.31-32]. Принимая позицию Р.К. Потаповой, под гипертекстом мы понимаем «соединение смысловой структуры, структуры внутренних связей некоего содержания и технической среды, технических средств, дающих человеку возможность осваивать структуру смысловых связей, осуществлять переходы между взаимосвязанными элементами» [13]. Надо отметить, что практическая ценность гипертекста в компьютерной лексикографии состоит в том, что он описывает тип интерактивной среды с возможностями переходов по ссылкам. Ссылки, которыми являются слова, фразы или рисунки, позволяют пользователю выбирать тот или иной текст или рисунок и выводить на экран связанные с ним сведения и материалы мультимедиа. Нелинейный характер гипертекста дает возможность формирования информации в виде разветвленной структуры, что позволяет в значительной степени расширить рамки словарной статьи и переосмыслить традиционное понимание термина «композиция». В контексте электронного словаря композицией можно было бы назвать его пользовательский интерфейс. Но это лишь «внешнее» представление композиции, т.е. то, что находится на поверхности. В действительности пользователь может создавать свою собственную композицию, активизируя ту или иную информацию в соответствии с его запросом. Надо отметить, что наличие определенных композиционных компонентов является весьма существенным для стадии формирования электронного словаря. Так, для описания структуры электронного словаря политической лексики обратимся к макроструктуре и микроструктуре словаря, где макроструктура – общая структура словаря, т.е. содержание и связь его частей, а микроструктура – это формат словарных статей и параметры его заполнения.

Макроструктура электронного словаря в линейной последовательности во многом повторяет макроструктуру традиционного бумажного словаря. Составные части макроструктуры электронного лексикографического издания легко просматриваются: предисловие, правила пользования, список используемых сокращений, корпус, грамматический очерк, различные списки (например, списки источников).

Но если бумажный словарь подразумевает упорядоченность этих блоков по приведенному нами списку, то в электронном словаре благодаря гипертекстовому устройству каждый из них может быть активизирован в любой очередности в зависимости от желания пользователя.

Рассмотрим специфику микроструктуры глоссария. Электронный словарь повторяет структуру словарной статьи своего «бумажного» аналога. Важно отметить, что взаимоотношения между частями словарной статьи не являются линейными и структурирование в словаре такого типа подчиняется законам гипертекста. Словарная статья имеет четкую логическую структуру с иерархическими связями между элементами. Каждая информационная категория занимает здесь строго фиксированное место, или зону [15]. Пользователь, проявляя интерес к той или иной информации,

запрашивает определенный параметр и получает доступ к отдельным фрагментам статьи. В соответствии с запросом активизируется лишь отдельно взятая зона, поэтому нет необходимости просматривать всю статью. Другими словами, электронный словарь берет на себя функцию вычленения необходимой информации и ее преобразования в наиболее удобную для потребителя форму. Следовательно, создатели электронного словаря могут предусмотреть довольно большое количество словарных входов, позволяющее пользователю легко и быстро получать любую необходимую ему информацию. Так, например, электронный словарь политической лексики [16] обеспечивает двухканальный вход в словарную статью – через меню прокрутки, где представлены все лексемы словаря, а также через поле поиска.

Возможности книжного словаря, касающиеся количества входов в словарь, ограничены. Так, абсолютное большинство словарей использует алфавитный принцип организации словарного материала, и это обеспечивает единственный вариант входа в словарь. Поскольку внутренняя структура электронного словаря невидима пользователю, она может быть совсем не алфавитной, пользователь видит лишь конечный результат – не обязательно считанную, а, скорее всего, синтезированную в момент ответа на запрос словарную статью [1]. Поэтому в компьютерной лексикографии традиционные проблемы алфавитного расположения словарных статей отсутствуют.

В словаре все информационные блоки, заложенные в словарную статью, самодостаточны, но в то же время могут функционировать в качестве ценного дополнения по отношению друг к другу. Таким образом, в электронном словаре становится возможным вариант лексикографического моделирования, при котором появление на экране той или иной зоны словарной статьи осуществляется по выбору пользователя.

Выводы.

1. Компьютерная лексикография представляет собой одно из интересных и актуальных направлений прикладной лингвистики, формирующееся в последние десятилетия на базе лингвистического обеспечения информационных систем разных типов.

2. В лексикографическом моделировании необходимым условием является организация набора структур знаний, формирующая концептуальную «модель мира».

3. Основными принципами создания виртуального лексического пространства являются стратегии гипертекста, что позволяет по-новому рассматривать методы семантизации лексики.

Список литературы

1. Андрущенко В.М. Автоматизация в лексикографии. Современное состояние и новые возможности // Советская лексикография. – М.: Русский язык, 1988. – С.201-224.
2. Арутюнова Н.Д. Дискурс// Лингвистический энциклопедический словарь / Под ред. В.Н.Ярцевой. – М.: Советская энциклопедия, 1990. – С. 136-137.
3. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – изд. 2-е, исправленное. – М.: Едиториал УРСС, 2003. – 360 с.
4. Бовтенко М.А. Компьютерная лингводидактика: Учебное пособие / М.А.Бовтенко. – М.: Флинта: Наука, 2005. – 216 с.
5. Богуславский И.М. Аннотированный корпус русских текстов: концепция, инструменты разметки, типы информации// Труды Международного семинара по компьютерной лингвистике и ее приложениям «Диалог-2000». – Протвино, 2000.

6. Зубов А.В. Информационные технологии в лингвистике: Учеб. пособие для студ. лингв. фак-тов высш. учеб. заведений/ А.В.Зубов, И.И.Зубова – М.: Издательский центр «Академия», 2004. – 208 с.
7. Карпіловська Є.А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: Підручник. – Донецьк: ТОВ «Юго-Восток, Лтд», 2006. – 188 с.
8. Красных В.В. К вопросу о лингвокогнитивных аспектах коммуникативного акта // Науч.докл.филол.ф-та МГУ. – М., 1998. – Вып. 2. –374с.
9. Кретов А.А., Воронина И.Е., Кунаковская О.В.. Программа для формирования словаря рифм русского языка. Программное средство. Зарегистрировано в Государственном фонде алгоритмов и программ Российской Федерации (регистрационный номер 50990000172 от 6.12.99)
10. Кретов А.А. Теоретические и практические аспекты создания морфемного словаря // Вестник ВГУ. Серия «Лингвистика и межкультурная коммуникация».– 2002. – № 2.
11. Кубрякова Е.С., Демьянков В.З., Панкрац Ю.Г., Лузина Л.Г. Краткий словарь когнитивных терминов/ Под ред. Е.С. Кубряковой. – М.,1996. –245 с.
12. Минский М. Структура для представления знания // Психология машинного зрения. – М., 1978. – 403 с.
13. Потапова Р.К. Новые информационные технологии и лингвистика. – М.: МГЛУ, 2002. – 575 с.
14. Проблемы построения лингвистического обеспечения образовательных систем высшей школы: Сборник научных статей. / Сост. В.В. Дубичинский. К.А. Метешкин, Л.А. Федорченко. – Харьков.: МСУ, 2005. – 106 с.
15. Убин И.И. Автоматический словарь как средство автоматизации лексикографических работ// Теория и практика научно-технической лексикографии: Сборник статей. – М.: Русский язык, 1988. – С. 234-240.
16. Чепик Е.Ю. Политическое слово в структуре электронного словаря // Культура народов Причерноморья. – 2005. – №69. – С.205-210.
17. <http://www.krugosvet.ru/articles/92/1009220/print.htm> – проверено 2.04.2006

Поступила в редакцию 14.02.2006 г.