

# ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ТЕХНОЛОГИИ СОЗДАНИЯ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

---

УДК 81`374:811.161.1`93

## СИСТЕМАТИЗАЦИЯ ВИРТУАЛЬНОГО ЛЕКСИЧЕСКОГО ПРОСТРАНСТВА: ЛЕКСИКОГРАФИЧЕСКИЙ АСПЕКТ

Л.Е. Бессонова

(Таврический национальный университет им. В.И. Вернадского,  
г. Симферополь, АРК, Украина)

*В статье анализируются гипертекстовые технологии в компьютерной лексикографии, описывается структура разрабатываемого электронного словаря политической лексики.*

**Ключевые слова:** компьютерная лексикография, лингвистика, семантика, слово

*У статті аналізуються гіпертекстові технології в комп'ютерній лексикографії, описується структура розроблюваного електронного словника політичної лексики.*

**Ключові слова:** комп'ютерна лексикографія, лінгвістика, семантика, слово

*The hypertext technologies in the computer lexicography are analyzed in the article, the structure of investigated electronic dictionary of political vocabulary is described.*

**Keywords:** computer lexicography, linguistics, semantics, word

Интенсивное развитие когнитивной лингвистики обуславливает необходимость нового осмысления традиционных объектов языкознания, в том числе семантики слова. Отражение семантического и концептуального содержания слова в лексикографическом аспекте с позиции современных научных парадигм представляет собой одну из актуальных задач исследовательских словарных проектов. Лексикографы сегодня активно обсуждают такие вопросы, как структура и объем электронных словарей, построение базы данных и принципы системной организации лексики, систематизация и параметризация материала, лингвистическую модель и программную оболочку текстового пространства словаря.

Основные аспекты этих и многих других научных проектов связаны не только с теоретическим исследованием естественного языка с позиций системно-квантитативного анализа, но и, что очень важно, с поиском новых возможностей использования компьютерных технологий при описании языковых единиц. Это и определило актуальность обозначенной проблемы, связанной, на наш взгляд, с организацией гипертекстового пространства в лексикографической практике и построением коммуникативно-дискурсивной модели словаря политического языка.

Опыт разработки текстовых корпусов как новой технологии прикладной лингвистики достаточно полно представлен в лингвистической литературе [3; 6; 8]. В научном обзоре четко обозначается круг проблем, одна из которых связана с отсутствием единого формата представления кор-

пусов текстов, а также с принципами их организации. Описывая эти проблемы, А.Н. Баранов справедливо замечает, что в корпусной лингвистике сегодня необходимо ввести ясное разграничение между «исследовательскими» и «иллюстративными» корпусами. Это снимет сложности построения лингвистических корпусов при той ситуации, когда «и материал слишком велик, и он не всегда отражает реальный узус». Здесь же лингвист приводит удачную, на наш взгляд, метафору У. Фрэнсиса, сравнившего процесс формирования корпуса с попыткой вычерпывания ведром океана [3, с. 113]. Фактор «безграничности» корпусов в современных научных проектах может сниматься за счет динамической модели корпуса, более гибкой по сравнению с традиционной. Так, созданный корпус текстов по современной русской публицистике с учетом выбранных параметров рассматривается как модель функционирования языка в политическом дискурсе [7].

Основной параметр организации любого корпуса – аннотация, или разметка, является главной характеристикой корпуса. Она-то и отличает его от коллекций, библиотек, энциклопедий, широко представленных в современном Интернете [4]. Как известно, чем богаче и разнообразнее разметка, тем выше научная и учебная ценность корпуса. В Национальном корпусе русского языка сегодня разработаны пять типов разметки: метатекстовая, морфологическая, акцентная, синтаксическая и семантическая. Так определяются важные параметры научного описания языка со структурно-семантической и функциональной позиций. Пространство электронных текстовых корпусов дает возможность результативного использования *электронных конкордансов*, которые открывают перспективы моделирования языковой картины мира. Традиционное понимание конкорданса как специализированной лингвистической прикладной программы помогает решить проблему, во-первых, формирования корпуса текстов и, во-вторых, его использования в конкретных задачах (например, при автоматической выборке заданных языковых единиц) [6, с. 73].

Безусловно, корпусная лингвистика ярко и четко репрезентирует себя в одном из перспективных направлений – *компьютерной лексикографии*. Электронные словари и энциклопедии разрабатываются сегодня как автономные и сетевые программные продукты. Не вызывает сомнения тот факт, что электронные словари предоставляют пользователю множество дополнительных возможностей по сравнению с печатными аналогами:

1. хранение большого объема информации за счет гиперссылок: электронный словарь предполагает включение в свою структуру нескольких словарей разных типов и жанров (словаря сочетаемости, терминологии, грамматических норм и др.);
2. доступность словаря за счет эффективной системы поиска (полнотекстовый поиск, одновременный поиск в нескольких словарях, скорость поиска);
3. применение средств мультимедиа для семантизации лексики: озвучивание заголовочных слов, введение иллюстративного материала с фотографиями, анимацией, видеофрагментами;
4. использование словарей в локальной и глобальной сетях, причем работа со словарями может проводиться многими пользователями одновременно;
5. экономия времени и материальных затрат при создании компьютерных словарей.

Значительный опыт научных проектов в компьютерной лексикографии позволяет расширить возможности описания смысловой структуры политического слова. В этом исследовательском контексте формулируется основная цель словаря – дать лексикографическую интерпретацию массиву политически релевантных лексических единиц русского языка, определив их коммуникативный статус. Не останавливаясь на уточнении терминов «политическое слово», «идеологема», «советизм», «политема» а также на обосновании их выбора, определим специфику макроструктуры электронного словаря политических слов.

- Словарь представляет собой многоуровневое лексикографическое произведение, в котором нашли отражение языковые факты политической жизни конца XX – начала XXI века.
- Словарь построен по гипертекстовой структуре и включает информацию в виде лексических баз данных реляционного типа, когда смысловое содержание единицы распределяется по полям, причем главное внимание уделяется системам представления информации, характерной для целого класса лексем.
- Словарь организует гипертекстовое пространство, которое является отражением когнитивного пространства человека.

- Словарь оформлен по алфавитному принципу и принадлежит к толково-справочным изданиям с элементами энциклопедического типа. Наряду с собственно лингвистической информацией в нем даются сведения по истории слова, его функционированию в различных сферах.
- Ориентированность на словарное произведение малого объема определило включение в словарь единиц, принадлежащих ядру современного политического лексикона – актуальной и употребительной политической лексики и фразеологии. Состав и объем словника, как известно, является одним из важных вопросов, стоящих перед лексикографами. Поэтому при разработке электронного словаря не следует ограничиваться специальной политической терминологией, необходимо учитывать и тот коммуникативно-прагматический потенциал лексического значения, который выводит слово с идеологизированным компонентом в политический контекст.
- Материал словаря обладает коммуникативной направленностью.

Эти параметры могут быть реализованы, на наш взгляд, в следующей структуре словаря, предлагаемой для обсуждения.

Слово в электронном варианте реализуется в семантическом макрополе, которое состоит из пяти полей, выстроенных в определенную систему:

1. Поле индексации служит для удобного поиска единицы в базе данных, которая создается в системе управления Access.
2. Поле заголовочного слова характеризуется традиционной, исходной, формой подачи и включает в себя несколько блоков.
  - 2.1. Блок помет о произношении слова или его форм: информация об особенностях произношения помещается непосредственно после входного слова и представляет собой указание на его нормативное произношение или допустимый вариант. Слово может быть озвучено средствами мультимедиа.
  - 2.2. Блок грамматической информации, прагматическая нагрузка которой связана с конкретным грамматическим статусом слова в текстах. Блок содержит основные формы заголовочного слова, его грамматические характеристики, что соответствует правилам отечественной лексикографии.
  - 2.3. Блок этимологических справок: слово в лексикографическом описании любого типа апеллирует не только к синхронии, но и к диахронии.
  - 2.4. Блок толкования строится не по всей семантической парадигме слова, а лишь по ее основной части, тематически связанной с политической сферой употребления. Эта часть микроструктуры словаря сопровождается примерами, предложениями.
  - 2.5. Блок фразеологии.
3. Поле иллюстраций работает для верификации дефиниции, тем самым выявляет семантику политического слова в определенном контекстуальном окружении. Важно отметить, что иллюстрации из политических текстов различных жанров (материалы Корпуса газетных текстов русского языка конца 20-го века [6], Интернет-ресурсов, СМИ) используются и в традиционном виде как оправдательный контекст (это особенно важно для метафорических значений) и как лексикографический прием, помогающий создать пространственную, временную и социально-историческую перспективы. Репрезентативный текстовый материал, организованный с помощью электронного конкурданса, позволяет моделировать политическую картину мира и ее составляющие.
4. Поле толковых словарей как микроструктура электронного словаря включает дефиниции толковых и толково-справочных словарей, основной материал которых составляет общественно-политическая лексика. Таким образом определяется репрезентативность лексической единицы и ее роль в формировании политического языкового пространства.
5. Поле энциклопедических сведений отражает основные этапы становления данной дефиниции, эволюцию ее доминантных признаков, а также дает дополнительные сведения об истории слова.

При лексикографической реализации семантики слова раскрывается тот прагматический потенциал, который составляет эмпирическое поле словаря. Что же понимать под прагматической ценностью словарного материала? Вопрос, поставленный в лингвистической семантике в 80-90-х годах [1], до сих пор не получил однозначного решения, однако обозначил новый вектор в широкой научной дискуссии. В статье «Два принципа и два понятия системной лексикографии» (2005) Ю.Д. Апресян, развивая свою концепцию интегральности лексикографического описания лексемы, выдвинутую в начале 90-х гг. [2, с. 270–271], выделяет семантико-прагматическое представление языковой единицы как особый тип лексикографической информации в словарных статьях интегрального словаря.

Такое широкое понимание прагматического компонента, а именно: отношение к действительности, к содержанию сообщения и к адресату – позволяет делать вывод о коммуникативно-прагматическом потенциале единиц лексикографического описания. Г.Н. Складневская, автор больших научных проектов, при лексикографировании слов выходит за рамки традиционных способов описания, выделяя в словарных статьях коммуникативно-ситуационный, когнитивный, эмоциональный и идеологический компоненты как маркеры речевой ситуации. Как отмечает исследовательница, при формировании структурированной электронной картотеки целесообразно выделить особую прагматическую зону значения, которая позволит дать более полное и адекватное языковой действительности представление о содержании лексического значения и его структуре [10, с. 79]. Это справедливое замечание должно учитываться при систематизации и построении словаря электронного словаря: автократия, амбиция, бюрократия, вертикаль, вето, власть, глобализм, двоевластие, дежа вю, демарш, демократия, европеизация, импичмент, интеграция, клан, коалициант, консенсус, лобби, неофит, олигарх, оппозиция, позиционирование, популизм, портфель, терроризм, толерантность, фигурант, электорат и т.д.

Вопрос о прагматическом компоненте в словарной практике определен и в работах Л.П. Крысина, который предлагает систему прагматических помет: «Одни из этих помет сходны со стилистическими, другие указывают на определенные компоненты в семантике, третьи – на положительные или отрицательные коннотации слова, сопровождающие те или иные его значения» [9, с. 302].

Изложенные предварительные соображения дают основания для вывода о том, что семантическая информация при комплексном описании пересекает границы словарной статьи, представляя политическое слово в понятийно-концептуальном аспекте. Лексикографические стратегии изложенного исследовательского проекта направлены на построение коммуникативно-дискурсивной модели словаря политической лексики с использованием современных технологий. Так, компьютерное обеспечение словаря складывается из двух основных составляющих: 1) базы данных, определяющей системой управления базами данных Access; 2) программы-оболочки, структура которой дает возможность управлять этой базой, а также вносить в нее изменения, не открывая ее непосредственно в Access. Все это позволяет структурировать последовательность получения данных в зависимости от потребностей пользователя. Программная оболочка написана на языке программирования Delphi, что позволяет осуществлять задачи автоматического поиска слов, добавления слов в базу данных, озвучивания заголовочного слова, экспортирования в HTML, обработку информации в пользовательском интерфейсе, а также удобной навигации внутри словаря.

Представленная концепция словарного проекта лежит в основе научного исследования, построенного на материале корпуса политических текстов, включающего факты русского языка конца XX – начала XXI вв. При лексикографическом моделировании такого лексического пространства необходимым условием является организация набора структур знаний, формирующая концептуальную модель мира.

### Список использованных источников

1. Апресян Ю.Д. Прагматическая информация для толкового словаря // Апресян Ю.Д. Избранные труды. – Т. 2. – Интегральное описание языка и системная лексикография. – М., 1995. – С. 135–155.
2. Апресян Ю.Д. Два принципа и два понятия системной лексикографии // Язык. Личность. Текст: Сб. ст. к 70-летию Т.М. Николаевой / Ин-т славяноведения РАН. – М.: Языки славянских культур, 2005. – С. 267–303.

3. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. Изд. 2-е, исправленное. – М.: Едиториал УРСС, 2003. – 360 с.
4. Библиотека Максима Мошкова // <http://lib.ru>. Русская виртуальная библиотека // <http://www.rvb.ru>. Фундаментальная электронная библиотека // <http://feb-web.ru>.
5. Богуславский И.М. и др. Аннотированный корпус русских текстов: концепция, инструменты разметки, типы информации // Труды Международного семинара по компьютерной лингвистике и ее приложениям «Диалог-2000». – Протвино, 2000.
6. Герд А.С., Богданов В.В., Азарова И.В., Аверина С.А., Зубова Л.В. Автоматизация в лексикографии и словари-конкордансы // Филологические науки. – 1981. – № 1 (121). – С. 72–78.
7. Компьютерный корпус газетных текстов русского языка конца 20-го века <http://www.philol.msu.ru/~lex/korpus.html>.
8. Михайлов М.Н. Компьютерное обеспечение корпуса текстов (взгляд пользователя) // Русистика сегодня. – 1998. – № 1–2. – С. 192–201.
9. Крысин Л.П. О типах лексикографической информации в русской части русско-иноязычных словарей // Язык. Личность. Текст: Сб. ст. к 70-летию Т.М. Николаевой / Ин-т славяноведения РАН. – М.: Языки славянских культур, 2005. – С. 285–303.
10. Складневская Г.Н. Реальный и ирреальный мир в толковом словаре (К вопросу о прагматическом компоненте слова) // Семантика и коммуникация. – СПб.: Изд-во СПб ун-та, 1996. – С. 68–80.