

УДК 81'374=161.1:004.415

Чепик Е.Ю.

**ПОЛИТИЧЕСКОЕ СЛОВО В СТРУКТУРЕ ЭЛЕКТРОННОГО СЛОВАРЯ**

*В статье рассматриваются возможности исследования политического слова в компьютерной лексикографии. В качестве фактического материала используется Корпус русского национального языка.*

**Ключевые слова:** компьютерная лексикография, корпусная лингвистика, конкордансы, электронный словарь

*У статті розглядаються можливості дослідження політичного слова в комп'ютерній лексикографії. Як фактичний матеріал використовується Корпус російської національної мови.*

**Ключові слова:** комп'ютерна лексикографія, корпусна лінгвістика, конкорданси, електронний словник

*In article opportunities of research of a political word in a computer lexicography are considered. As an actual material the Case of Russian national language is used.*

**Key words:** a computer lexicography, case linguistics, konkordans, the electronic dictionary

В современной лексикографии в последнее десятилетие отчетливо обозначается перспективное направление – компьютерная лексикография, которая становится важной составляющей прикладной лингвистики. Лингвистами и программистами сегодня обсуждаются такие вопросы, как структура и объем электронных словарей, принципы системной организации лексики, методы лингвистического программного обеспечения, параметры систематизации информационного материала (А.Н. Баранов, А.С. Герд, Б.Ю. Городецкий, Ю.Н. Караулов, А.А. Поликарпов, В.А. Широков, А.В. Зубов, И.И. Зубова). Один из важных аспектов в научных дискуссиях – возможности использования компьютерных технологий при создании и применении словарей. Этим и обусловлена **актуальность** предлагаемого исследования. Специальные программы (базы данных, программы обработки текста) позволяют формировать словарные статьи, хранить информацию и обрабатывать ее в автоматическом режиме. Способы организации корпусного материала с помощью компьютерных методик, возможности отражения семантики слова в электронной структуре словаря определяют **научную новизну** работы.

Как известно, традиционный вид сбора данных для словаря – это каталожная карточка, в которой указываются следующие параметры: описываемое слово, пример употребления, источник примера, автор, сфера функционирования. Этот трудоемкий и длительный процесс может быть упрощен благодаря использованию компьютерных технологий, которые позволяют максимально облегчить сбор и хранение информации, используя для этого расширенную базу данных (база данных – это совокупность определенным образом упорядоченных сведений о некоторых объектах [8, с. 146]). В отличие от обычной бумажной картотеки записи базы данных позволяют легко вносить изменения, автоматически отбирать нужные примеры, сортировать данные по выбранным параметрам. Однако с развитием объема информации возросла сложность его обработки. На данный момент в компьютерной лингвистике не существует специализированных лексикографических баз данных, однако такие базы данных, как ACCESS, FOX-Base, PARADOX, D-Base, могут быть использованы для ведения словарных картотек, а также для их редактирования [8].

Решение задач, связанных с поиском примеров, иллюстрирующих то или иное языковое явление, а также с уточнением употребления отдельных лексем, привело к появлению нового направления в прикладной лингвистике – *корпусной лингвистики* [7], т.е. науки, которая занимается созданием корпусов текстов и их использованием.

За последнее десятилетие корпусная лингвистика быстро переняла тот опыт, который был накоплен в других достаточно развитых областях, объединенных термином АОТ (автоматическая обработка текста), или же АПТ (автоматическое понимание текста – термин Н.Н. Леонтьевой). Как было отмечено на семинаре по проблемам корпусной лингвистики (Санкт-Петербург, 10-15 октября 2005 г.), корпусная лингвистика сегодня имеет две линии развития – лингвистический и информационный анализ текста.

Информационный анализ реализуется в системах типа ИПС (информационно-поисковые системы), а лингвистический анализ стремится интерпретировать, выявить смысл текста, при этом он характеризуется локальностью, т.е. проходит в рамках одного предложения. Информационный анализ, наоборот, «видит» текст как целое. Эти два направления развивались почти параллельно, практически никак не соприкасаясь, вследствие чего не получили достаточного развития. Лингвистический анализ не дал корректного машинного перевода, т.к. одна неправильно интерпретированная фраза в контексте может приобрести совершенно иное значение.

В настоящее время в научном словаре лингвистов появляются очень близкие понятия: «электронные библиотеки», «массив текстов», «коллекция текстов», «электронный архив», «полнотекстовая база данных». Среди них можно выделить лингвистические корпуса, или языковые корпуса. По мнению А.Н. Баранова, корпус текстов – это вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то полные фрагменты макроструктуры текстов данной проблемной области [2, с. 115].

Корпус текстов характеризуется четырьмя основными параметрами: во-первых, он должен быть достаточно большого объема; во-вторых, корпус должен быть структурированным или размеченным; в-третьих, тексты, составляющие определенный корпус, должны быть в электронном варианте; в-четвертых, в понятие «электронный корпус» входит, как правило, специальное программное обеспечение для работы с этим корпусом.

Ценность корпуса нам видится в следующем:

- 1) однажды сделанный корпус может многократно использоваться;
- 2) корпус показывает языковые данные в их реальном окружении, что позволяет исследовать лексическую и грамматическую структуру языка, а также непрерывные процессы языковых изменений, происходящие в языке на протяжении определенного отрезка времени;
- 3) корпус характеризуется представительностью, или сбалансированным составом текстов, что позволяет использовать его для тестирования поисковых машин, машинных морфологий, систем перевода, а также использовать его в различных лингвистических исследованиях;
- 4) корпус имеет важное значение для преподавания русского языка, так как с помощью корпуса можно быстро и эффективно проверить особенности употребления незнакомого слова или грамматической формы.

Работа с корпусами, то есть с массивами текстов, представленными в электронном виде, давно уже стала одним из основных методов лингвистических исследований. Так, еще в 1960-е годы создавался Брауновский корпус (США), который включает 1 млн. слов; в 1970-е – LOB корпус (Великобритания, Норвегия), который также включает 1 млн. слов. В 1980-е годы начали создаваться такие корпуса, как Машинный Фонд русского языка, Уппсальский корпус русского языка (Швеция), 1 млн. слов, The Bank of English, Birmingham, 20 млн. слов. В 1990-е создан British National Corpus, который включает 100 млн. слов, а также другие национальные корпуса (венгерский, итальянский, хорватский, чешский, японский) объемом 100 млн. слов. В это же время создавался The Bank of English, Birmingham, включающий 600 млн. слов. В начале XXI в. создавались такие корпуса, как American National Corpus, 100 млн. слов, и Gigaword corpora (английский, арабский, китайский), включающий 1 млрд. слов. Особое внимание следует обратить на русский национальный корпус [10], над которым работают лингвисты Москвы и Санкт-Петербурга.

Существование корпусов текстов дает возможность значительно расширить и автоматизировать анализ языкового материала, который является важнейшей базой любого лингвистического исследования. Чем больше материалов анализируется, тем выше значимость выводов и уровень их достоверности.

Современные компьютерные программы позволяют находить нужные примеры из корпусов текстов, которые хранятся в электронном виде на компьютере. Это экономит значительное количество времени по сравнению с традиционной технологией сбора примеров вручную.

Отметим, что именно аннотация, или разметка, – главная характеристика корпуса, которая и отличает его от электронных коллекций, библиотек, энциклопедий, широко представленных в современном Интернете [4]. Разметка текста – это приписывание тексту определенной информации для более удобного анализа.

Существуют различные типы разметки:

- 1) метатекстовая разметка (автор, название, дата создания, объем, тематика текста и т.д.), которая характеризует текст в целом;
- 2) структурная разметка представляет собой информацию о структуре текста, которая позволяет отделить одно слово от другого, выделить границы словосочетания, предложения, текста;
- 3) лингвистическая разметка заключается в приписывании единицам текста определенной лингвистической информации (отрицательное предложение или вопросительное, управление или примыкание и т.д.).

Как известно, чем богаче и разнообразнее разметка, тем выше научная и учебная ценность корпуса.

Пространство электронных текстовых корпусов дало возможность результативного использования *электронных конкордансов*, которые открывают перспективы моделирования языковой картины мира. Конкорданс является специализированной лингвистической прикладной программой, с помощью которой осуществляется автоматическая выборка заданных языковых единиц из электронных текстов [5, с. 77]. Функцию конкорданса можно сравнить с функцией поиска в текстовом редакторе, однако возможности конкорданса шире – он анализирует не один, а сразу несколько текстов или корпусов электронных текстов, при этом конкорданс выводит на экран информацию о контексте использования заданных языковых единиц. В зависимости от технических возможностей конкорданс может предоставлять информацию о частотности употребления и сочетаемости той или иной языковой единицы, а также дает возможность обращаться к конкретному тексту, в котором был найден пример.

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:

- позволяют хранить большой объем информации за счет использования гиперссылок;
- электронный словарь имеет эффективную систему поиска (полнотекстовый поиск, одновременный поиск в нескольких словарях, скорость поиска);
- в электронных словарях могут применяться средства мультимедиа: озвучивание заголовочных слов, введение иллюстративного материала с фотографиями, анимацией, видеофрагментами;
- возможность использования словарей в локальной и глобальной сетях;
- значительная экономия времени и материальных ресурсов при создании компьютерных словарей.

Значительный опыт исследований в компьютерной лексикографии позволяет расширить возможности изучения смысловой структуры политического слова. Так, сегодня в научных проектах активно разрабатывается идея, связанная с организацией концептуального пространства языка политики с позиции гипертекстового единства [3].

Говоря о специфике содержания такого словаря, отметим, что слово в электронном варианте реализуется в семантическом макрополе, состоящем из нескольких полей, выстроенных в определенную систему:

- 1) *поле индексации* служит для удобного поиска в базе данных;
- 2) *поле заголовочного слова*;
- 3) *поле грамматической информации*, прагматическая нагрузка которой связана с конкретным грамматическим статусом в текстах;
- 4) *поле этимологических справок*;
- 5) *поле авторского толкования* строится не по всей семантической парадигме слова, а лишь по ее основной части, тематически связанной с политической сферой употребления;
- 6) *поле иллюстраций* работает для верификации дефиниции, тем самым выявляя семантику политического слова в определенном контекстуальном окружении. Важно отметить, что иллюстрации из текстов различных жанров используются в словаре 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;
- 7) *поле фразеологии*;
- 8) *поле энциклопедических сведений* отражает основные этапы становления данной дефиниции, эволюцию ее доминантных признаков.

Компьютерное обеспечение словаря толково-справочного типа складывается из двух составляющих:

- 1) база данных словаря определяется системой управления базами данных Access;
- 2) интерфейс разрабатывается специально для данного словаря. Базой для его реализации послужил язык программирования Delphi, который позволяет осуществлять задачи автоматического поиска слов, а также удобной навигации внутри словаря.

На рис. 1. приведен лексикографический виртуальный портрет слова *эскалация*.

Слово в электронном словаре представлено в 4-х полях:

- поле 1 состоит из блока заголовочного слова, блока грамматических помет, этимона. В этом же поле обязательным элементом является блок авторского толкования;
- поле 2 представлено иллюстративным материалом Корпуса русского национального языка. Эти иллюстрации помогают описать семантическую структуру слова и определить коммуникативный статус слова;
- поле 3 – микроструктура электронного словаря, включающая дефиниции современных толковых и толково-справочных изданий. Так, например, слово «эскалация» в «Кратком словаре политического языка» В.В. Бакеркиной описывается как постепенное увеличение, усиление, расширение чего-н.;
- поле 4 является, на наш взгляд, необходимой структурной частью электронного справочного издания и содержит энциклопедические сведения о данном понятии. В Большом Энциклопедическом словаре, например, рассматриваемое слово определяется как расширение, наращивание (вооружений и т.п.), постепенное усиление, распространение (конфликта и т.п.), обострение (положения и т.п.).

Именно 3-е и 4-е поля могут описать динамику развития семантики слова. Например, слово *эскалация* в современных словарях политического языка представлено с имплицитным негативным прагматическим компонентом, обусловленным категориальной семой *действие, направленное на отрицательный результат*, что подтверждается «Поле иллюстраций».

Таким образом, словарь представляет собой систематизированный алфавитный массив ядерной политической лексики. Предлагаемое лексикографическое исследование дает возможность выявить основные тенденции в развитии ядерной политической лексики, обеспечить доступность словаря за счет эффективной системы поиска, а также хранить большой объем информации за счет гиперссылок.

**База данных**

Слово:

Заголовочное слово: Эскалация, и, ж. [англ. escalation букв. восхождение с помощью лестницы]. Постепенное усиление, увеличение, расширение, распространение чего-нибудь. Эскалация войны.

[Поле иллюстраций](#)  
[Поле толковых словарей](#)  
[Поле энциклопедических сведений](#)

Амбиции  
Вето  
Лобби  
Люстрация  
Неофит  
▶ Эскалация

Поиск

**Поле иллюстраций**

В Питере продолжается эскалация цен на осмотр достопримечательностей [<http://rostrizm.ru/news/394>]

Банковская система Украины находится под угрозой дестабилизации в случае продолжения и эскалации политического кризиса, который может негативно повлиять на макроэкономическую стабильность, - говорится в распространенном сегодня сообщении международного рейтингового агентства Standard & Poor's. [<http://www.ufs.com.ua/news/news.php?NewsID=90922>]

Европа интересуется ликвидацией Масхадова и опасается эскалации насилия в Чечне 9.03.2005 17:23 // ЗАВТРА // [<http://www.zavtra.com.ua/news/mir/17157/>]

Эскалация террора в России - то, о чем пишут практически все российские газеты, о чем практически каждый день говорят по телевидению и радио. [<http://www.svoboda.org/programs/pr/2004/pr.090204.asp>]

**Поле толковых словарей**

Бакеркина В.В. Краткий словарь политического языка: Более 2 тыс. терминов и терминолог. сочетаний / В.В. Бакеркина, Л.Л. Шестакова. - М.: ООО "Издательство АСТ"; ООО "Издательство Астрель"; ООО "Русские словари", 2002. - 288с.; с.282.  
Эскалация, и, ж. (Англ. escalation букв. восхождение с помощью лестницы). Постепенное увеличение, усиление, расширение, чего-н. Эскалация войны.  
Большой иллюстрированный словарь иностранных слов: 17 000сл. - М.: ООО "Русские словари"; ООО "Издательство Астрель"; ООО "Издательство АСТ", 2003. - 957, [3]с.; с.449  
Эскалация, -и, ж. (англ. escalation букв. восхождение с помощью лестницы) - постепенное увеличение, усиление, расширение чего-л., напр. Э. вооружений, агрессии.  
Новейший словарь иностранных слов и выражений. - М.: Современный литератор, 2005. - 976 с.  
Эскалация [англ. escalation восхождение при помощи лестницы] постепенное расширение, наращивание ч.-л., напр. вооружений; постепенное усиление, распространение или обострение ч.-л., напр. Э. конфликта.  
Краткий политический словарь / Абаренков В.П., Адр.; Сост. и общ. ред. Л.А. Ониква, Н.В. Шишлина. - Политиздат, 1987. - 509 с.  
Эскалация (англ. escalation – постепенное усиление, наращивание, усиление (напр., Э. конфликта). Терм. употребление в связи с расширением агрессии СС нач. 70-х гг. 20 в.  
Словарь С.И.Ожегова  
Эскалация, -и, ж. (книжн.). Увеличение, рост, расш

**Поле энциклопедических сведений**

Большой энциклопедический словарь  
Эскалация (англ. escalation), расширение, наращивание (вооружений и т. п.), постепенное усиление, распространение (конфликта и т. п.), обострение (положения и т. п.).  
[<http://encycl.accoona.ru/?id=74227>]  
Эскалация (англ. escalation), расширение, наращивание (вооружений и т. п.), постепенное усиление, распространение (конфликта и т. п.), обострение (положения и т. п.).

Рис. 1. Лексикографический виртуальный портрет слова *эскалация*

**Источники и литература**

1. Бакеркина В.В. Краткий словарь политического языка: Более 2 тыс. терминов и терминолог. сочетаний / В.В. Бакеркина, Л.Л. Шестакова. – М.: ООО «Издательство АСТ»: ООО «Издательство Астрель»: ООО «Русские словари», 2002. – 288с. – С.282.
2. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – изд. 2-е, исправленное. – М.: Едиториал УРСС, 2003. – 360 с.
3. Бессонова Л.Е. Лексикографическое моделирование политической лексики: опыт составления электронного словаря (в печати).
4. Библиотека Максима Мошкова <http://lib.ru/>. Русская виртуальная библиотека <http://www.rvb.ru>. Фундаментальная электронная библиотека <http://feb-web.ru>. Библиотека художественной литературы <http://www.lib.com.ua>
5. Бовтенко М.А. Компьютерная лингводидактика: Учебное пособие / М.А.Бовтенко. – М.: Флинта: Наука, 2005. – 216 с.
6. Богуславский И. М. и др. Аннотированный корпус русских текстов: концепция, инструменты разметки, типы информации // Труды Международного семинара по компьютерной лингвистике и приложениям «Диалог-2000». – Протвино, 2000.
7. Британский национальный корпус (BNC) <http://sara.natcorp.ox.ac.uk/lookup.html>. Чешский национальный корпус <http://ucnk.ff.cuni.cz>. Упсальский машинный фонд русского языка <http://www.slaviska.uu.se/korpus.htm>
8. 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.
9. Компьютерный корпус газетных текстов русского языка конца 20-го века. – <http://www.philol.msu.ru/~lex/korpus.html>.
10. Национальный корпус русского языка. – <http://ruscorpora.ru>.

*Поступила в редакцию 14.07.2005 г.*