

*Данилевич Т.А.  
Белорусский государственный университет, Минск  
Науч. рук. – канд. филол. наук, доцент А.О. Долгова*

## **КОМПЬЮТЕРНАЯ ЛЕКСИКОГРАФИЯ КАК ОБЛАСТЬ ПРИКЛАДНОЙ ЛИНГВИСТИКИ**

Компьютерная лексикография как область прикладной лингвистики, занимается составлением электронных словарей. Раньше она преобразовывала существующие словари в электронную форму. Однако, на сегодняшний день компьютерная лексикография – особое направление практической лексикографии со своими собственными подходами к отображению и содержанию словаря. Следовательно, электронный словарь – это особый лексикографический объект, который может реализовать многие продуктивные идеи, не востребованные и не осуществимые в бумажных словарях.

Предлагаем рассмотреть не просто существующие технологии, но и потенциал компьютерной лексикографии. Прежде чем говорить о новых возможностях электронных словарей, стоит обратить внимание на проблемы классической «бумажной» лексикографии. Издания традиционной практической лексикографии имеют три фундаментальных противоречия, которые характерны для этой области человеческой деятельности [1, с.114]:

1. Чем больше объем словаря, полнее и доказательнее описание лексических значений, тем сложнее им пользоваться. Данное противоречие привело к поляризации рынка бумажных словарей: существует большое количество изданий, довольно простых, но при этом удобных в применении. Они противостоят единичным объемным профессиональным изданиям, которые информативны, но не пригодны для быстрого получения информации (например, 20-томный Оксфордский словарь).

2. Чем полнее и точнее описаны лексические значения, тем меньше словарь соответствует текущей языковой и культурной ситуации. Это объясняется тем, что долгий цикл создания и модификации бумажных

словарей приводит к тому, что картина мира, которую они фиксируют, постоянно меняется и уже значительно отличается от действительности.

3. Чем интереснее лексикографическая концепция словаря, тем уже его лексическая база, т.е. универсальные бумажные словари фактически не используют достижения теоретической лексикографии и не применяют их на практике.

Компьютерная лексикография и использование электронных словарей по определению позволяют преодолеть часть указанных проблем. Можно отметить ряд новых возможностей электронного словаря, одной из которых является широкое использование для доступа к содержанию различных лингвистических технологий (морфологический и синтаксический анализ, полнотекстовый поиск, распознавание и синтез звука и т. п.) [4, с. 29].

С позиции пользователя реализация новых технологий означает, что становится возможным быстро получить информацию, которая находится где-то в недрах словаря. При этом ответ на запрос пользователь получает в удобной для него форме. Минимальной единицей доступа при традиционном подходе является лексема (или, другими словами, имя словарной статьи). Для объемных словарей, таких как Оксфордский, это условие представляет серьезную проблему. Например, глагол *set* имеет около 400 основных значений (не говоря о подзначениях и коннотациях). Для пользователя удобно, чтобы словарь максимально точно отображал именно релевантную информацию. При этом мы не говорим об автоматическом выборе переводного эквивалента. Специфика словарного ответа в том, что он выдает информацию о слове или словосочетании, а не просто переводной эквивалент, и предполагает активное взаимодействие с пользователем и выбор из нескольких возможных альтернатив, которые хорошо обоснованы. Однако попытки решить вопрос об адекватной реакции словаря на запрос неизбежно встречают «сопротивление» самого словарного материала, если он перенесен из бумажного словаря.

Таким образом, мы можем видеть новое противоречие между новыми языковыми компьютерными технологиями и старым традиционным словарным содержанием, которое не позволяет воспользоваться этими технологиями в полном объеме. Причина этого противоречия очевидна: словарь является моделью языка, построенной на совершенно иных принципах, чем те формальные модели, на которых основываются эти технологии. При этом стоит отметить, что в области морфологии противоречие еще не очень существенно, но в области синтаксиса и семантики оно почти непреодолимо.

Задача состоит в том, чтобы единицей описания сделать отдельное лексическое значение, а технологии анализа должны устанавливать соответствие между исходным запросом и лексическими значениями, которые релевантны для этого запроса, по синтаксическим и семантическим критериям.

Чтобы продемонстрировать сказанное на примере можно взять практически любой глагол, принадлежащий ядру языка. Например, глагол «развести» может иметь следующие контексты: разводить руками; разводить мосты; разводить супругов; разводить кроликов; разводить дерущихся; разводить пилу; разводить спирт водой; разводить / разбивать сады (и стоит отметить, что английские эквиваленты разные: bring; conduct; part, separate; mix; dissolve; divorce; breed; plant, etc.).

Таким образом, создание словарного содержания, которое позволило бы использовать в качестве единицы анализа отдельное лексическое значение, а не словоформу, видится нам наиболее перспективным направлением в компьютерной лексикографии. Для его реализации требуется «синхронизация» формальных моделей и словарных описаний, которые используются технологиями анализа. В идеале это должно быть единое интегральное лексико-синтаксико-семантическое описание.

Интегральный подход к лексическим описаниям позволяет также решить проблему «монофункциональности» бумажных словарей [3, с. 208]. К примеру, особенностью большинства бумажных переводных словарей является ориентация описания структуры лексического значения в исходном языке на лексическую систему языка перевода и на реализацию ровно одной функции – собственно перевода с одного языка на другой при учете, что один язык иностранный, а другой – родной. Подобное ограничение делает словарь неудобным при переходе от пользовательской модели к модели создателя. На сегодняшний день такого рода модели реализуются разными типами словарей, что достаточно неудобно. Таким образом, интегральный подход к лексическим описаниям оправдан не только методически (и, что немаловажно, экономически), но и практичен с точки зрения интересов пользователя.

Не стоит забывать о проблеме актуальности словарного содержания. Для электронных словарей, которые являются массовыми программными продуктами, характерны модернизация и совершенствование версий, наличие обратной связи с пользователями и дружественный интерфейс. Следовательно, компьютерная

лексикография – это актуальная и важная часть лексикографии как науки.

Функционирование и работа электронного словаря должна быть схожа с особенностями других программных средств: т. е. с возможностью исправления пользователями замеченных ошибок. Подобный подход лишь фиксирует настоящее положение дел: коллективное авторство на словарное содержание принадлежит всем носителям языка, при этом основная задача лексикографа – фиксация языковых фактов и их методически правильное описание.

Можно сказать, что сегодня разрыв между лексикографической теорией и практикой довольно значителен. Хотелось бы выделить следующие теоретические концепции, реализация которых является полезной и перспективной [2, с. 87]:

1. Необходимо ввести понятие «лексической функции» с целью системного описания несвободных словосочетаний (например, «ответственность несут», «кандидатов выдвигают»).

2. Важно объяснять семантику с учетом практической реализации грамматического словоизменения и словообразования. Каждому языку свойственны свои собственные способы грамматического кодирования смысла, но эти способы не описываются системно в массовых словарях. Например, как передать по-английски смысл «довыпендриваться», даже если знаешь, как передать «выпендриваться»?

3. Ситуация в области учета синтаксических отношений и описаний наиболее сложна, так как в массовых словарях отсутствует даже система понятий, с помощью которой синтаксическая информация могла бы быть доведена до пользователя. Это происходит вследствие того, что принято, что за составление предложений отвечает грамматика, которая изложена в справочнике, а словарь должен обеспечивать именно перевод отдельно взятых слов. Такая идея уже является устаревшей и не выдерживает критики с точки зрения современных представлений о центральной роли слова в синтаксисе.

Можно заключить, что выход из сложившейся ситуации следующий – необходимо создание интегральных словарных описаний, основывающихся на формальных моделях и учитывающих практически все аспекты слова. На этих же моделях должны базироваться технологии доступа к словарному содержанию.

Отрыв лексикографической теории от практики составления словарей – все еще актуальная проблема. В развитии отмеченных идей лежит будущее как лексикографии вообще, так и практики составления электронных словарей.

## ЛИТЕРАТУРА

1. Берков, В.П. Двухязычная лексикография / В.П. Берков. – Москва: Астрель, Транзиткнига, 2004. – 236 с.
2. Введенская, Л.А. Русская лексикография / Л.А. Введенская. – Ростов н/Д.: МарТ, 2007. – 352 с.
3. Марчук, Ю.Р. Компьютерная лингвистика / Ю.Р. Марчук. – Москва: Восток-Запад, 2007. – 220 с.
4. Потапова, Р.К. Новые информационные технологии и лингвистика / Р.К. Потапова. – М.: МГЛУ, 2002. – 575 с.