

УДК 801.8

Адживапов Айдер Муратович, аспирант кафедры восточной филологии, ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского», Института филологии, Симферополь

e-mail: ajder.adjivapov@gmail.com

Меметов Искандер Айдерович, кандидат филологических наук, доцент кафедры восточной филологии, ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского», Института филологии

e-mail: isken25@mail.ru

ПРИНЦИПЫ СОСТАВЛЕНИЯ ЭЛЕКТРОННЫХ ПЕРЕВОДНЫХ СЛОВАРЕЙ: НА МАТЕРИАЛЕ КРЫМСКОТАТАРСКОГО ЯЗЫКА

Аннотация. В данной статье авторами рассмотрена проблема современного понимания принципов составления электронных переводных словарей на фоне актуальных проблем крымскотатарской электронной лексикографии. Прежде всего рассмотрены ключевые закономерности в развитии крымскотатарской электронной лексикографии. На фоне актуальной типологической проблематики авторы проводят сравнительный анализ электронных и печатных словарей. Далее авторами раскрыта собственно методология составления электронных словарей на примере крымскотатарско-русской языковой пары. В качестве вывода авторы обозначают ряд приоритетных задач, решение которых, по его мнению, существенным образом поспособствует развитию крымскотатарской электронной лексикографии в частности.

Ключевые слова: электронный словарь, электронная лексикография, корпусная лингвистика, крымскотатарское языкознание.

1. Введение

Формирование крымскотатарской электронной лексикографии обусловлено прежде всего необходимостью сохранения, поддержания и развития крымскотатарского языка. В отличие от бумажной лексикографии, электронная лексикография гораздо менее подвержена внешним деструктивным факторам, что обеспечивает стабильность и безопасность информации. Кроме того, существует ряд других преимуществ электронной формы лексикографии, которые открывают перед исследователями крымскотатарского языка множество возможностей по дальнейшему развитию лексикографической мысли.

Говоря о прошедших тенденциях развития этой мысли, на стадии зарождения электронные словари составлялись наследуя традиции печатного формата. Скорее всего, это было вызвано тем обстоятельством, что и теоретические, и практические возможности не были достаточно осознаны авторами и, следовательно, потенциал электронной версии поначалу не был раскрыт полностью. Анализируя крымскотатарские лексикографические тенденции последних 20 лет, можно констатировать, что электронная лексикография проходит через все стадии развития, свойственные бумажной лексикографии. Как и в случаях с бумажными изданиями, самые первые словари были составлены вовсе не лексикографами. Это, вероятно, обусловлено тем, что понимание практических принципов составления электронных словарей опережало понимание теории. Ведь в случаях, когда авторы опирались на лексикографическую теорию, им приходилось руководствоваться методологией, ориентированной на составление печатных словарей. Такая ситуация непременно провоцирует возникновение с одной стороны – творческой стагнации в практическом аспекте, но вместе с тем – осознание необходимости формирования более современного и адекватного рассмотрения теоретических моделей разработки электронных словарей. Фокус при этом смещается в сторону таких проблемных вопросов, как отбор входящих данных,

формирование структур данных, а также специфику отражения потребностей своих целевых пользователей в своих макро- и микроструктурах.

Вместе с тем необходимо понимать, что в современных реалиях электронный словарь – это уже не просто отсканированная и оцифрованная версия существующего бумажного словаря. Напротив, это полностью самобытная и независимая единица, где форма логично продолжает и дополняет содержание. Это ведёт нас, как исследователей, к осмыслению особенностей общелексикографической типологии и классификации.

2. Проблема типологии

По утверждению Я. Перванова, говоря о противопоставлении бумажных и электронных словарей важно прежде всего то, что текст содержимое и форма не являются взаимно сопоставимыми сущностями. Таким образом, по мнению исследователя, принципиальная разница между электронным и бумажным словарями в контексте разметки и гипертекста, словарных карточек и бумажных страниц со словарными статьями. Напротив, формальные особенности электронных словарей он определяет как способ интерпретации языковой семантики, которому свойственно своеобразное погружение пользователя в виртуальную лексикографию. [5]

Такой подход разделяют и другие ученые, исследующие эволюцию лексикографии, начиная от бумажного словаря и заканчивая современными цифровыми базами данных. Последние, с развитием технических возможностей, значительно диверсифицируют собственные функциональные и количественные показатели, которые открывают возможности обработки языковых данных в формате, схожей с методами обработки лексикона мозгом человека. Очевидно и то, что нынешние электронные словари больше не являются копиями печатных аналогов, главной формальной методикой которых является демонстрация оцифрованных словарных статей. Вместо этого, фокус сместился в сторону разработки преимущественно интерактивных системы,

представляющих информацию разнообразного характера и разнообразным образом, выходя порой за рамки собственно лингвистической области познания. [3]

По мнению В. Селегея главной особенностью электронного, отличающей его от бумажного является именно его способность реализации совокупности идей, нереализуемых в бумажных словарях. Исследователь приводит следующие антиномии бумажных словарей:

- рост объема словаря усложняет его использование;
- усиление конкретики в описании лексических значений приводит к несоответствию современной лингвокультурной ситуации;
- усиление специализированности словарной концепция ведёт к сокращению лексической базы. [6]

Электронная реализация словаря с переенным успехом даёт возможность преодолеть выше упомянутые проблемы. Прежде всего, за счет более прогрессивных возможностей отображения словарной статьи (например, поиск с фильтрацией критериев), а также с применением лингвистических технологий (морфологический и синтаксический анализ, полнотекстовый и фонетический поиск и т. д.)

3. Принципы составления электронных переводных словарей

Составление электронного словаря связано, прежде всего, с формированием подходов к организации лексической информации. На сегодняшний день нами выделено два таких подхода:

1. Грамматический подход.
2. Статистический подход.

Грамматический способ основан на строгоопределённых параметрах соответствия между лексическими единицами. Предположим, что перед нами стоит задача добавить в словарь перевод слова «эв» на русский язык. В этом случае, мы представляем данные как 2 таблицы, где в одной (Таб. 1) записаны лексемы, а во второй (Таб. 2) – соотношение между ними:

№	Слово	Язык
1	эв	крымскотатарский
2	дом	русский

Таб. 1. Список лексем

Слово 1	Слово 2
эв	дом

Таб. 2. Список соотношения лексем

Следует отметить, что это упрощенное представление, призванное описать принцип применения грамматического способа составления электронных переводных словарей. На практике же организация информационной инфраструктуры состоит из гораздо большего количества звеньев цепи и может охватывать сразу несколько лингвистических аспектов. Приведем пример включения в структуру семантики. Так, нам необходимо определить, что такое «эв» в крымскотатарском языке с точки зрения содержания понятия. Мы можем описать лексему «дом», как «Здание, строение, предназначенное для жилья, для размещения различных учреждений и предприятий». Таким образом нужно добавить в систему новую семантическую единицу:

№	Описание
1	Здание, строение, предназначенное для жилья, для размещения различных учреждений и предприятий

Таб. 3. Список семантических единиц

Следует обратить внимание, что данная таблица не содержит полей «Язык». Это связано с тем, что семантические значения по большей части

являются общими для множества языков, в то время, как средствам выражения (орфографическим, фонетическим) в разных языках для каждого конкретного значения свойственно отличаться. [4]

При включении семантики более нет необходимости задавать прямое соотношение между лексемами (Таб. 2), а вместо этого следует отнести собственно лексемы к семантическим значениям, которые они описывают:

Номер слова	Значение
1	1
2	1

Таб. 4. Список соотношения лексем со значениями

Теперь можно смоделировать получение перевода слова «эв» на русский язык. В контексте такой структуры получение представления данных производится по следующему алгоритму:

Шаг	Описание	Результат
1	Поиск слова «эв» в списке лексем (Таб. 1)	Слово № 1
2	Поиск значений искомого слова по номеру (Слово №1) в таблице соотношений (Таб. 4)	Значение № 1
3	Поиск собственно значения (Значение № 1) в списке значений (Таб. 3)	Здание, строение, предназначенное для жилья ...
4	Поиск других слов, соответствующих найденному значению (Значение №1) в таблице соотношений (Таб. 4)	Слово № 2
5	Поиск слов по целевому языку («русский») номеру, найденному в Шаге 4 (№ 2) в списке слов	Слово «дом»

Таб. 5. Алгоритм получения перевода

Суммируя результаты из данного алгоритма и опуская ненужную для пользователя информацию (номера лексем, значений), по запросу «эв» на «русском языке» мы получаем: «эв» (Здание, строение, предназначенное для жилья ...) – «дом». Разумеется, точность такой системы следует повысить

(например, конкретизируя значения и т.п.), но главную свою задачу (демонстрацию работы грамматического метода) данная система выполняет успешно.

Говоря об оценке эффективности систем на основе грамматического подхода, следует выделить точность и стабильность, которые обусловлены собственно философией данного подхода. Структура данных здесь точно предопределена, что обеспечивает стабильность результатов. Отсюда же просходит и главный недостаток: громоздкость. Чем больше языковых аспектов затрагивает система, тем строже необходим контроль и тем больше сложность её внутренней организации. Всё это ведёт к потере гибкости и адаптивности, так как любое изменение в структуре в связи с динамически меняющимися социокультурными реалиями требует всесторонней структурной работы, что в свою очередь, может быть весьма дорогостоящим решением, прежде всего с точки зрения затрат ресурсов.

Статистический подход заключается в имплементации элемента статистики в контексте предоставления перевода лексических единиц. Здесь в качестве исходных данных используются синхронные тексты рабочих языков. Главной отличительной особенностью статистического метода является его масштабируемость — чем масштабнее система, тем точнее результат. В среде исследователей эту особенность принято называть самообучаемостью. Однако, ввиду того, что строго определённые грамматические характеристики и шаблоны не фигурируют в системе, она вместо этого полагается исключительно на статистику, что обуславливает появление грамматических ошибок и неточностей в результирующих текстах.

Также следует выделить и **смешанные подходы**, которые объединяют в себе подходы, описанные выше. Главной задачей появления таких систем является устранение недостатков и сохранение преимуществ грамматических и статистических подходов к составлению электронных переводных словарей.

Таким образом мы выяснили, что в составлении электронных словарей прежде всего следует отказаться от тезиса «электронный словарь = оцифрованный бумажный словарь». Многочисленные образцы современных достижений в области электронной лексикографии неоднократно подтверждают ошибочность такого утверждения. Однако, говоря о ситуации с развитием крымскотатарской электронной лексикографии, по-прежнему существует нужда в применении целого ряда современных методик составления словарей, таких как использование языкового корпуса, машинного обучения и многих других. В нашем исследовании мы постарались кратко описать самые основные принципы составления электронных словарей, однако искренне надеемся, что работа в этом направлении в самом ближайшем будущем будет продолжена, а её плоды станут поводом для гордости и общим достоянием для всей крымскотатарской лингвистической науки и всего крымскотатарского общества.

Список использованных источников

1. Абрамова Н.Н., Глобус Е.И. Формирование многоязычных словарей и их использование при кросс-языковом поиске информации – URL: http://elar.usu.ru/bitstream/1234.56789/1386/1/IMAT_2005_02.pdf (дата обращения: 09.03.2024).
2. Авдеева М. В. Компьютерная лексикография в системе практической и теоретической лексикографии [Электронный ресурс] // Социосфера. 2012. № 44. С. 26–28. – URL: <http://goo.gl/i9op6> (дата обращения: 12.03.2024).
3. Ваулина Е. Ю. Толковый словарь. Возможности электронной версии [Электронный ресурс] // Структурная и прикладная лингвистика. 2007. С. 187–196. – URL: goo.gl/IrAD5 (дата обращения: 08.03.2024).
4. Герд А.С. Прикладная лингвистика. – СПб.: Издательство Санкт-Петербургского университета, 2005. – 268 с.

5. Перванов Я.А. Заметки по электронной лексикографии. [Электронный ресурс]. – URL: <http://arxiv.org/pdf/1107.1753.pdf> (дата обращения: 09.03.2024).
6. Селегей В. П. Электронные словари и компьютерная лексикография [Электронный ресурс] URL: <http://goo.gl/ujMnG>. (дата обращения: 09.03.2024).
7. Шляхова А. В. Электронный словарь и его специфика [Электронный ресурс] // Вестн. Тюмен. гос. ун-та. 2008. С. 181–185. – URL: goo.gl/voMFG (дата обращения: 08.03.2024).