

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

К ПРОБЛЕМЕ ЭЛЕКТРОННЫХ СЛОВАРЕЙ КАЛМЫЦКОГО ЯЗЫКА

В. В. Куканова, Е. В. Бембеев

Калмыцкий институт гуманитарных исследований РАН, Элиста

The article is devoted to the problems of electronic dictionaries of Kalmyk language which belongs to the group of the endangered languages. The authors consider that it is necessary to create and develop the Kalmyk-Kalmyk dictionary with sound module and elements of thesaurus for the frequent words.

Сегодня пристальное внимание лингвистов всего мира направлено на использование информационных технологий в своих исследованиях. Не исключение в этом ряду и калмыцкое языкознание, для которого теория и практика составления электронных словарей является одной из актуальнейших проблем, поскольку в Республике Калмыкия сложилась ситуация угрозы языка как средства познания и коммуникации. Калмыцкий язык причисляют к группе языков, которые находятся на грани исчезновения, поэтому существует настоятельная необходимость создания всевозможных электронных словарей, в первую очередь калмыцко-калмыцких, а также переводных (калмыцко-русских и русско-калмыцких).

В сети Интернет размещен только один электронный словарь (<http://dictionary.bumbinorn.ru/>), представляющий собой базу данных, которая состоит из заголовочного слова и самой словарной статьи (в объеме около 11 тыс. слов). Поиск можно осуществлять только по заголовочной лексеме, при этом элементы словарной статьи не имеют связи с калмыцкими вариантами, поскольку в его основе лежит только один словарь «Я изучаю калмыцкий. Малый русско-калмыцкий словарь» [Илишкин и др., 2003]. Вместе с тем, учитывая то, что других электронных словарей калмыцкого языка нет вообще, то уже сам факт его существования — это большой шаг в развитии, вернее, в зарождении электронной калмыцкой лексикографии, хотя оговоримся, что оцифровку данного словаря проводил не лингвист, которому, по всей видимости, не известны лингвистические принципы электронных словарей.

Следовательно, отсутствуют выработанные теоретически и закрепленные практически принципы работы со словарным материалом на калмыцком языке, именно это обуславливает настоятельную необходимость систематизации лексикографической работы с целью разработки перспективных направлений в области теории и практики компиляции словарей. Для калмыцкого языка наиболее перспективным является создание словаря со звуковым модулем, поскольку по правилам орфографии на письме не обозначаются «неясные» гласные. Несмотря на многочисленные попытки решить эту проблему, лингвисты до сих пор не нашли способа их графического обозначения (см., например, [Харчевникова, 2002]). Учащиеся в процессе обучения калмыцкому языку (по сути, как иностранному) не могут интуитивно определить правила, где необходимо произносить «неясные», что в свою очередь создает большие трудности в произношении сочетаний согласных в калмыцких словах. Школьники, научившись писать, быстро забывают о звуковых образах, для них более значимым становится письменная форма, нежели устная. Соответственно, доминирует буквенное произношение многих слов как в калмыцком, так и в русском языках. Но если в русском языке редко встречаются такие сочетания, то в калмыцком они носят частотный характер, ряд их может состоять из, например, 10 букв: көдлмшч-нр-мдн [ködëlměščinërdën])

рабочий-NOM.PL.-PART.POSS.1PL / рабочий-CAUS.PL.-PART.POSS.1PL
'наши рабочие'/'наших рабочих'

Однако произносительного словаря также недостаточно в целях ревитализации языка, так как необходимо предоставить как можно более полную информацию о слове: о его грамматических, словообразовательных (деривационных) и дистрибутивных свойствах, поскольку, как известно, слово «ра-ботает» в контексте. Создаваемые тем самым связи и отношения между ними расширяют «сеть» лексикона языкового носителя, в результате чего формируемая языковая система становится самодостаточной и может функционировать уже независимо, т. е. без «помощи» другой языковой структуры, хотя на первых порах обучения, конечно, нужно пользоваться переводными словарями, но более «продвинутыми» — желателен калмыцко-калмыцким словарем. С другой стороны, точную семантику той или иной лексической единицы можно передавать только средствами калмыцкого языка, например:

итк-лц-х

доверять- Soc

'доверять друг другу'

Поэтому наиболее востребованным в калмыцком обществе словарем станет электронный словарь со звуковым модулем и элементами тезауруса, который отвечает и образовательным стандартам, т. е. лексическая единица описывается как элемент системы со всеми парадигматическими и синтагматическими связями с другими элементами лексической системы.

Развитие калмыцкой лингвистики на современном этапе позволяет обобщить результаты многолетних исследований в этой области и сформировать тезаурус наиболее частотных единиц калмыцкого языка¹, а затем уже перейти к составлению словаря «полного» типа. Мы осознаем, что тезаурус всегда будет неполным, так как словари не успевают фиксировать все то, что происходит в языке, настолько стремительно он развивается. К тому же его фрагментарность, неполнота обусловлены особенностями личности, которая принимает участие в составлении словаря. Тем не менее можно выделить ядро тезауруса, который носит, как правило, единый характер и не является вариативной частью системы. Все то, что находится на периферии, носит инвариантный характер. Однако следует не забывать, что четкие границы между ядром и периферией отсутствуют.

При составлении электронного словаря тезаурусного типа из известных лексикографических программ мы остановились на TLex Suite 2010:

Dictionary Production Software (<http://tshwanedje.com>), поддерживающей стандарты XML и UNICODE и состоящей из четырех модулей: TLex Dictionary Compilation Software, tlTerm Professional Termbase Software, tlCorpus Concordance Software, tlReader. Эта программа является профессиональной, т. е. в ней учтены все необходимые потребности лингвиста при компилировании словарей разного типа. Она позволяет значительно уменьшить время и затраты на процесс составления словарей и увеличить качество создаваемого продукта. Если несколько десятилетий назад лингвист тратил время на унифицирование словарных статей и их форматирование, то при помощи этой программы можно с легкостью менять формат представления словарных статей. Более того, эта база данных, которая может хранить весьма разнородный материал, и по желанию лингвиста может фильтровать словарные статьи в зависимости от формата статьи (карманный, средний, полный) или элементы в структуре статьи (например, заголовочное слово и антонимы/синонимы).

Словарная статья тезауруса должна строиться по следующей схеме:

- 1) вход словарной статьи; написание на «тодо бичиг» («ясное письмо»);
- 2) частотность ЛЕ; 3) произношение ЛЕ; звуковой файл; 4) часть речи ЛЕ;
- 5) грамматические пометы ЛЕ; 6) тематические пометы ЛЕ;
- 7) стилистические пометы ЛЕ; 8) этимология ЛЕ; 9) толкования и примеры (с указанием гиперонимов, тематического поля, к которому принадлежит та или иная лекса); 10) сочетания ЛЕ по схемам: Прилаг.+Сущ.; Сущ.+Гл.; Нар.+Прилаг.; Нар.+Гл.1; 11) синонимы каждой отдельной лексы (тезаурус);
- 12) антонимы каждой отдельной лексы; 13) омонимы и их отличия;
- 14) идиомы и фраземы каждой лексы; 15) дериваты; 16) сложные слова.

На данный момент ведутся работы по сведению существующих словарей калмыцкого языка в программу, запись слов, определена частеречная принадлежность, а также уточняются значения некоторых лексических единиц, корпус текстов пополняется новым материалом. Однако в процессе компилирования электронного словаря мы уже стали сталкиваться с определенными проблемами. Например, составные слова — это отдельный лексический вход, или такие единицы нужно размещать в той словарной статье, в которой описывается главное слово? Вообще, наличие большого числа композитов в калмыцком языке является отличительной особенностью монгольских, а также тюркских языков. Скорее всего, в этом вопросе мы будем ориентироваться на опыт составления словарей английского языка, где предпочитают давать сложные слова как отдельное заголовочное слово, но при этом делают ссылку на него в той статье, вокабула которой содержит один из главных компонентов композита.

Бумажные словари уступают электронным словарям европейских языков по своей функциональности и полноте представления словарной статьи, поэтому теорию и практику составления словарей калмыцкого языка в целом необходимо развивать и внедрять информационные технологии в этот процесс. Электронная лексикография калмыцкого языка находится на стадии своего формирования, и именно сочетание разработанных методик в калмыцком языкознании и мирового опыта по составлению, сведению существующих словарей позволяет создать эффективную модель лексической системы калмыцкого языка.

Список литературы

- Биткеев, 2002 — Биткеев П.С. Проблемы неясных гласных в калмыцком языке // Вопросы теоретической грамматики калмыцкого языка: сб. ст. / отв. ред. Э. Р. Тенищев. М.; Элиста, 2002. С. 22–29.
- Харчевникова, — Харчевникова Р.П. Проблемы реформы орфографии калмыцкого языка // Вопросы теоретической грамматики калмыцкого языка: сб. ст. / отв. ред. Э. Р. Тенищев. М.; Элиста, 2002. С. 3–21.
- Илишкин и др., 2003 — Илишкин И.К.; Муниев Б.Д.; Бадмаева В.Д. и др. Я изучаю калмыцкий. Малый русско-калмыцкий словарь. Элиста, 2003.