

«РАННИЕ» СЛОВАРИ КАЛМЫЦКОГО ЯЗЫКА И СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ*

Н. Ч. Очирова

В XXI веке разработка информационных технологий считается одним из перспективных задач современного общества. Основные приоритеты данного процесса заключаются в фиксировании, продуцировании, обработке, хранении, передаче информационных данных и т. п., осуществляемые на основе современных средств вычислительной техники, а также на базе разнообразных средств информационного обмена.

Постепенно традиционные методы в лингвистике также заменяются компьютерной обработкой данных, в частности в лексикографии. Современная лексикография представляет собой интегрированную информационно-методологическую науку, которая «лежит на стыке различных информационных технологий, компьютерного языкознания, практических научных экспериментов, общей и педагогической лингвистики и др.» [Дубичинский 2001: 6]. Только компьютер может предоставить немедленный доступ к словнику, цитатному материалу и т. п., обнаружить несогласованность различных текстов словарных статей, левой и правой частей словаря, дефиниций, переводных эквивалентов, а также отслеживать семантические изменения лексем, сферу ее употребления, фиксировать изменения в деривационном, синтаксическом, грамматическом аспектах лексической единицы [Дубичинский 2008: 350].

Компьютеризация лексикографической деятельности заключается прежде всего в создании специализированных машинных фондов и корпусов и в разработке методов их формирования, представления информации и ее использования. Современная лексикография все шире пользуется подобными базами, в частности, большими корпусами текстов, в которых компьютер по запросу осуществляет поиск нужных слов. На этой основе формируется целое новое направление лингвистики и лексикографии — корпусная лингвистика и лексикография [Quirk 1992; Svartvik 1992].

Ю. Д. Апресян отмечает, что во второй половине XX века началась глубокая реформа лексикографии. Она была вызвана несколькими факторами: 1) новым типом лингвистического описания, сложившимся в работах по формальным моделям языка; 2) новыми тенденциями в развитии теоретической лексикографии; 3) поворотом самой лексикографии к назревшим нуждам практического преподавания языков [Апресян 1990: 123 — 124]. Корпусы первого поколения насчитывали до миллиона словоупотреблений и были репрезентативными для некоторых исследований: при создании учитывались разные жанры устной и письменной речи, каждый из которых был представлен определенной выборкой. Корпусы второго поколения создаются в результате сплошного обследования текстов и по объему превышают корпусы первого поколения в несколько раз [Апресян 1995: 19].

* Статья подготовлена в рамках проекта РГНФ «Национальный корпус калмыцкого языка: создание и разработка» (12-04-12047, тип «в»).

Одним из таких проектов является создание Национального корпуса калмыцкого языка (НККЯ), электронного ресурса, который будет доступен любому пользователю сети. В перспективе этот ресурс будет содержать мультимедийный материал, помимо размеченных текстов. НККЯ состоит из двух модулей: текстов в электронном формате и базы данных, в которой дана метаразметка материала. База данных метаописания была сконструирована в базе данных MS Office Access. Набор параметров, обозначенных в узлах классификации, соответствует практике кодирования текстов известных корпусов и следуют рекомендациям TEI [Sperberg-McQueen, Burnard 2001] и EAGLES [Sinclair 1996]. Однако все они не могут учитывать национальную специфику функционирования калмыцких текстов, калмыцкой художественной литературы [Куканова 2011: 140].

В рамках НККЯ создается словарный модуль калмыцкого языка, где особый интерес представляют старописьменные словари калмыцкого языка XVIII — начала XX вв., которые связаны с именами классиков отечественной монголистики Г. А. Цвика, Я. И. Шмидта, К. Ф. Голстунского, О. М. Ковалевского, А. В. Попова, А. М. Позднеева, М. Львовского, Г. И. Рамседта и др. Сохранившиеся до настоящего времени словари на «тодо бичиг» являются свидетельством бытования в языке целого пласта лексики, утраченного в силу разных причин в современном калмыцком языке. В ряде случаев примеры из подобных лексикографических источников, наряду с другими сохранившимися старописьменными памятниками могут оказаться единственными.

Анализ лексики, зафиксированной в лексикографических и других источниках, созданных несколько столетий назад, важен не только для изучения истории калмыцкого языка, изменений словарного состава национального языка, но и для более точной характеристики изменений, происходивших в общественно-политической и экономической жизни калмыцкого народа. Особенно ценным источником для историко-лингвистических исследований являются двуязычные словари (калмыцко-русские, русско-калмыцкие, калмыцко-немецкие и др). На основании комплексного изучения подобных словарей можно судить о важных тенденциях развития соответствующих языков. Привлечение подобных источников может действовать решению ряда проблем в области языкознания, например: выявлению степени участия живой народно-разговорной речи в формировании национального языка; уточнению процессов образования литературного языка; рассмотрению вопросов, касающихся статуса языка (особенно в условиях реальной угрозы исчезновения калмыцкого языка); изучению диалектной дифференциации калмыцкого языка в «ранний» период его становления и т. д.

Следует отметить, что формирование лексической системы калмыцкого языка того периода происходило в условиях междиалектного и межъязыкового общения, поэтому использование материалов ранних источников применимо для изучения факторов взаимодействия различных языковых систем, построения теории языковых контактов. В русле когнитивной лингвистики лексические материалы источника могут быть использованы для построения фрагментов языковой картины мира носителей языка одного региона в определенный исторический период. Кроме того, богатая в тематическом отношении лексика «ранних» источников, отражающая особенности уклада жизни калмыков в определенный исторический период, может использоваться в аспекте культурологической информативности для изучения региональной истории и культуры.

Старописьменные калмыцкие памятники долгое время оставались совершенно неизученными. Однако в последние годы появляются лингвистические исследования источниковедческого характера: «Письма Аюки-хана и его современников (1714–1724 гг.): опыт лингвосоциологического исследования» [Сусеева 2003], «Письма наместника Калмыцкого ханства Убаши (XVIII в.): факсимиле писем» [Гедеева 2004], «Лингвистическое описание памятника старокалмыцкой (ойратской) письменности: “Сказание о хождении в Тибетскую страну Малодербетовского Бааза-бакши”» [Бембеев 2004] и др.

Малоизученность старописьменных текстов XVIII столетия объясняется тем, что исследователи редко прибегают к рукописным материалам, стремясь использовать лишь опубликованные источники. Хотя архивы, в том числе и архив Калмыцкого института гуманитарных исследований РАН, располагают большим количеством источников, относящихся к раннему периоду развития калмыцкого национального языка.

Так, например, в рукописном отделе Российской национальной библиотеки (Эрмитажное собр. № 22) хранится русско-калмыцкий словарь XVIII века «Словарь языка калмыцкого», представляющий интерес не только как источник лексики калмыцкого языка XVIII века и, шире, как памятник истории отечественного калмыковедения и монголоведения в целом. Этот словарь отражает живую калмыцкую речь того периода, изучение стилистических характеристик некоторых редких слов позволит более объемно представить словарный состав калмыцкого языка XVIII века [Омакаева, Бурыкин 1999: 21].

В связи с этим объединение всех словарей в одну большую лексикографическую базу данных по калмыцкому языку, а также переиздание лексикографических источников XVIII – начала XX вв. обеспечат эффективный доступ к памятникам культурного наследия. Наряду с решением проблемы доступности, электронная форма публикаций позволит повысить надежность хранения и удобство распространения материала. Включение различных программных компонентов и баз данных в состав современных электронных изданий по калмыцкому языку позволит эффективно осуществлять работу с материалом, представленном в издании, проводить научные исследования, поиск и обработку информации. В то же время печатные издания традиционны, а, следовательно, привычны и удобны в использовании для большого количества людей, поэтому эта форма переиздания не может быть исключена. Таким образом, современное переиздание должно содержать и печатную, и электронную формы [Филлипович 2007: 4].

В процессе работы по созданию электронных коллекций старокалмыцких письменных памятников лингвисты сталкиваются с проблемами разного характера, начиная с теоретических и заканчивая техническими (см. подробно [Бембеев 2012]). Одной из больших проблем в обработке «ранних» словарей является отсутствие кодировки некоторых символов в соответствии со стандартами UNICODE. Еще не разработан принцип вертикального письма как для *тодо бичиг*, так и для *худма бичиг*. Другой проблемой является фонетический принцип написания слов, т. е. слова фиксировались в ранних памятниках так, как произносились. С одной стороны, это отражает ту речь, которая бытовала в обществе (в особенности это важно для лингвистов), для программистов же этот аспект является проблемным при разработке на словарной основе распознающей программы, поскольку это создает большое количество вариативного написания того или иного слова [Бембеев 2012: 39].

Таким образом, изучение лексикографических памятников письменности, относящихся к «раннему» периоду развития калмыцкого языка, и современный уро-

вень развития лингвистики и лексикографии в частности, а также социальные изменения требуют создания словарей нового типа, в которых лексикографические методы и приемы будут соответствовать современным научным достижениям и требованиям общества. Современные средства автоматизации лексикографической работы позволяют своевременно проследивать изменения в употреблении тех или иных лексических единиц, составлять тексты и текстовые комплексы для их многократного использования. Информационные технологии существенно меняют стиль работы лексикографов и повышают эффективность их деятельности, что позволяет ставить и решать новые задачи в условиях современной модернизации.

Литература

- Апресян Ю. Д.* Словарная фиксация синтаксических свойств лексем: взаимодействие словаря и грамматики // Wiener Slawistischer Almanach. Band 25–26. 1990. С. 31–52.
- Апресян Ю. Д.* Избранные труды. Т. 2. Интегральное описание языка и системная лексикография. М.: Наука, 1995. 414 с.
- Бембеев Е. В.* Лингвистическое описание памятника старокалмыцкой (ойратской) письменности: «Сказание о хождении в Тибетскую страну Малодербетовского Бааза-бакши. Автореф. дис. ... канд. филол. наук. М., 2004. 26 с.
- Бембеев Е. В.* Коллекции рукописей на старокалмыцком (ойратском) языке XVII–XIX вв. в свете компьютерной обработки: постановка проблемы // Мат-лы IV Междунар. науч. конф. «Информационные технологии и письменное наследие». Петрозаводск: ПГУ, 2012. С. 38–40.
- Гедеева Д. Б.* Письма наместника Калмыцкого ханства Убаши (XVIII в.) / отв. ред. Э. У. Омакаева. Факсимиле писем. Изд. текстов, введ., транслитер., пер. со старокалм. на современ. калм. яз., сл. / Д. Б. Гедеева. Элиста: АПП «Джангар», 2004. 196 с.
- Дубичинский В. В.* Современная лексикографическая наука: основные направления и перспективы // Современная лексикография: достижения, проблемы, перспективы: сб. науч. ст. Краснодар, 2001. С. 5–29.
- Дубичинский В. В.* Лексикография русского языка. М.: Наука: Флинта, 2008. 432 с.
- Куканова В. В.* Архитектура метаописания в Национальном корпусе калмыцкого языка // Вестник Калмыцкого института гуманитарных исследований РАН. 2011. № 1. С. 139–145.
- Омакаева Э. У., Бурькин А. А.* Русско-калмыцкий словарь второй половины XVIII века как памятник отечественного калмыковедения // Научная конф. «Актуальные проблемы алтаистики и монголоведения (язык и литература)». Тез. докл. и сообщ. Элиста: АПП «Джангар», 1999. С. 18–21.
- Сусеева Д. А.* Письма Аюки-хана и его современников (1714–1724 гг.): опыт лингвосociологического исследования. Элиста: АПП «Джангар», 2003. 423 с.
- Филлипович А. Ю.* Автоматизированная технология переиздания лексикографических источников XVIII – нач. XIX вв.: на материале Словаря Академии Российской 1789–1794 гг. Автореф. дис. ... канд. техн. наук. М., 2007. 23 с.
- Sinclair J.* Preliminary recommendations on text typology. EAGLES Document EAG-TCWGTTP/P [электронный ресурс] // <http://www.ilc.cnr.it/EAGLES96/texttyp/texttyp.html>. 1996 (дата обращения: 15.04.2011).
- Svartvik J.* (ed.). Directions in Corpus Linguistics, Berlin, 1992.
- Sperberg-McQueen C. M., Burnard L.* Guidelines for Electronic Text Encoding and Interchang [электронный ресурс] // <http://www.hcu.ox.ac.uk/TEI/P4X/index.html>. 2001 (дата обращения: 15.04.2011).