

КОЛЛЕКЦИИ РУКОПИСЕЙ НА СТАРОКАЛМЫЦКОМ (ОЙРАТСКОМ) ЯЗЫКЕ XVII–XIX ВВ. В СВЕТЕ КОМПЬЮТЕРНОЙ ОБРАБОТКИ: ПОСТАНОВКА ПРОБЛЕМЫ¹

Е. В. Бембеев

Калмыцкий институт гуманитарных исследований РАН, Элиста

The article is devoted to the description of problems of creation of digital collection of manuscript monuments, partly problem of absence of type of “todo blchig” and recognized program.

В последние годы в лингвистике для сохранения языка и глубокого его исследования широко используются современные информационные технологии, и, если тексты современного периода уже оцифрованы и созданы на их базе электронные корпуса, то компьютерная обработка ранних текстов только началась. Надо сказать, что некоторые из них (в особенности на европейских языках) стали источниками формирования электронных коллекций текстов — не просто в виде отсканированных изображений, а распознанных в специальных программах.

Что касается ранних текстов на старокалмыцком языке, то работа в этом русле не началась: сформированы лишь небольшие коллекции факсимиле в разных уголках мира, которые еще не получили своего полного библиографического описания. Калмыки использовали «ясное письмо» до 1924 года, а ойраты Монголии — до 1945 года. В настоящее время «тодо бичиг» остается в использовании только у ойратов, проживающих в Синьцзян-Уйгурском автономном районе КНР. Собственно ойрат-калмыцкая литература ведет свое начало с момента создания Зая-пандитой «ясного письма». В настоящее время, по сведениям Х.Лувсанбалдана, найдено 65 переводов, принадлежащих Зая-пандите, и 12 сделанных его учениками [Лувсанбалдан, 1975]. Однако в его «Биографии» говорится о том, что им было переведено 177 сочинений, и 37 переводов сделали его ученики [Radnabhadra, 1959; Бадмаев, 1968; Норбо, 1999]. К большому сожалению исследователей, многие из переводов Зая-пандиты оказались с течением времени утрачены, а те из них, что сохранились до наших дней, хранятся в собраниях рукописных памятников России, Монголии, КНР, Германии, Дании, Франции, США, Японии и т. д.

В России крупнейшими собраниями рукописных памятников на ойратской письменности располагают Институт восточных рукописей РАН и библиотека Восточного факультета Санкт-Петербургского государственного университета [Музраева, 2012]. Небольшая коллекция старописьменных источников на ойратском языке есть и в Калмыцком институте гуманитарных исследований РАН [Чуматов, 1983; Орлова, 2002]. Ценнейшие исторические документы, датированные 1713–1771 гг., находятся в Национальном архиве

¹ Статья подготовлена при финансовой поддержке гранта РГНФ "Национальный корпус калмыцкого языка: создание и разработка" № 12–04–12047в.

Республики Калмыкия [Сусеева, 2003; Гедеева, 2004; Кокшаева, 2011]. Большое количество письменных источников на «тодо бичиг» хранится в фондах Национальной библиотеки, Института языка и литературы АН Монголии, в библиотеке храма Гандантегчинлин в Улан-Баторе, в фондах краеведческих музеев Кобдоского, Убсунурского аймаков, а также в частных коллекциях [Музраева, 2012].

На современном этапе исследователями основное внимание уделяется задачам поиска, создания каталогов и сохранения исторических памятников. Основным методом переноса на новые носители является оцифровка данных, подразумевающая факсимильное копирование источников и сопровождение их библиографическими и археографическими данными. К примеру, такая работа проводится Общественной организацией «Тод номын гэрэл» (Монголия), которая совместно с Американским центром монголоведения (The American Center for Mongolian Studies — ACMS) разместили на сайте 140 рукописных текстов на «ясном письме» (<http://www.dlir.org/archive/orc-exhibit/items/browse/collection/7>).

Важным недостатком такого типа описаний является неполное и ограниченное раскрытие содержания документа, что принципиально не изменяет способа доступа к информации. По-прежнему исследователь должен просматривать значительное количество источников для поиска нужной информации, последовательно «листая» их. В связи с этим важнейшей задачей является решение проблемы поиска информации в созданных электронных хранилищах документов по их содержанию. Хранилища современных документов — это сами документы в текстовом формате, и их неотъемлемая часть — автоматически (автоматизированно) полученные поисковые образы. Такие информационно-поисковые возможности для хранилищ ранних рукописных текстов на старокалмыцком или ойратском письме в настоящее время отсутствуют. Реализация их представляет собой актуальную научно-практическую задачу.

Сама же графическая система, на которой писались эти тексты, — «тодо бичиг» («ясное письмо») — не получила еще должной компьютерной обработки в силу своих особенностей. Ойратский алфавит «тодо бичиг» был создан Зая-пандитой Огторгуйн Далай в 1648 году на основе старомонгольского алфавита «худма бичиг», который был заимствован монголами у уйгуров в XIII веке. Уйгурское письмо восходит в свою очередь к согдийскому, а последний — к арамейскому письму. Ойратское письмо, как и старомонгольское, имеет вертикальное направление, буквы в слове и слова пишутся сверху вниз. Слова в столбцах разделяются пробелами, столбцы располагаются слева направо. Большинство букв имеет три различных написания — в начале, середине и конце. Более того, для обозначения звуков, отсутствующих в ойратском (старокалмыцком) языке, используются дополнительные буквы-«галики». Они встречаются в основном в религиозных текстах для обозначения заимствований из тибетского или русского языка, санскрита. Орфография «тодо бичиг» в основном фонетическая, т. е. каждая буква отражает один звук, что является главным отличием от полифонного старомонгольского письма.

Если символы монгольского письма («худма бичиг») уже получили кодировку UNICODE и уже создан пакет программ с использованием этого письма (Windows, Microsoft Office), то для «тодо бичиг» отсутствует кодировка некоторых символов в соответствии с вышеупомянутыми стандартами. Однако пока еще не разработан принцип вертикального письма как для «тодо бичиг», так и для «худма бичиг» (хотя найдена в Интернете программа, в которой соблюдается этот принцип для двух графических систем (<http://www.dusal.net/downloads/vertNote.rar>), но которая не интегрирована в систему Microsoft Office). 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 порождает большое количество вариативного написания того или иного слова.

Работа по созданию электронных коллекций старокалмыцких письменных памятников сопряжена с рядом таких проблем, как отсутствие до настоящего времени символов «тодо бичиг» в соответствии со стандартами UNICODE, распознающей программы, базы данных лексического ядра или словаря ранних текстов, необходимого для правильного распознавания линейного потока. Кроме того, отдельную проблему составляют интерпретация того или иного плохо сохранившегося графического знака, адекватная передача графических особенностей памятника (например, необычной лигатуры), нечеткое начертание «диакритики», отсутствие унификации и др.

В русле данных исследований в Калмыцком институте гуманитарных исследований РАН ведется работа по созданию «Лингвистически аннотированного корпуса калмыцкого языка», одним из направлений которого является разработка подкорпуса «ранних текстов», а также создание «Словарного модуля калмыцкого языка», который представляет собой электронную лексикографическую базу данных (на основе лексикографических источников XVIII и первой половины XX вв.) и который впоследствии составит основу для создания распознающей программы. Начата работа по оцифровке имеющихся словарей, приведению их к единообразию в оформлении словарной статьи. Параллельно проводится анализ рукописных и печатных источников XVII–XIX вв., выявление их палеографических и лексических характеристик, к которым относятся особенности почерка переписчика или шрифта текста, формат рукописи, качество бумаги, чернил и т. п. следы времени. Все это в дальнейшем облегчит работу по вводу и обработке текстов и изображений, систем оптического распознавания, систем информационного поиска и автоматического индексирования документов.

Список литературы

Бадмаев, 1968 — Бадмаев А.В. Зая-Пандита (списки калмыцкой рукописи «Биографии Зая-Пандиты»). Элиста, 1968.

Кокшаева, 2011 — Кокшаева Н.О. Языковые особенности эпистолярного наследия калмыцкого хана Дондук-Даши (середина XVIII в.). Элиста, 2011.

Лувсанбалдан, 1975 — Лувсанбалдан Х. Тод үсэг, түүний дурсгалууд. Улаанбаатар: Шинжлэх ухааны академийн хэвлэх үйлдвэр, 1975. 356 х.

Музраева, 2012 — Музраева Д.Н. Изучение ранних этапов становления буддийской литературы у ойратов (Зая-пандита и его наследие) в отечественном и зарубежном монголоведении // Вестник КИГИ РАН. 2012. № 1. (в печати).

Норбо, 1999 — Норбо Ш. Зая-Пандита (Материалы к биографии) / пер. со старописьменного монгольского Д. Н. Музраевой, К. В. Орловой, В. П. Санчирова / науч. ред. В. П. Сачиров. Элиста, 1999.

Орловаб 2002 — Орлова К.В. Описание монгольских рукописей и ксилографов, хранящихся в фондах Калмыкии // Бюллетень Общества востоковедов. Вып. 5. М., 2002.

Гедеева, 2004 — Письма наместника Калмыцкого ханства Убаши (XVIII в.). Факсимиле писем. Издание текстов, введение, транслитерация, перевод со старокалмыцкого на современный калмыцкий язык, словарь Д.Б.Гедеевой. Элиста, 2004.

Сусеева, 2003 — Сусеева Д.А. Письма Аюки-хана и его современников (1714–1724 гг.): опыт лингвосоциологического исследования. Элиста, 2003.

Чуматов, 1983 — Чуматов В.О. Старописьменные памятники КНИИ ИФЭ // Монголоведные исследования. Элиста, 1983. С. 116–131.

Яхонтова, 1996 — Яхонтова Н.С. Ойратский литературный язык XVII века. М., 1996.

Radnabhadra, 1959 — Radnabhadra. Rabjamba Zay-a bandida-yin tuyuji saran-u gerel kemekъ ene metъ bolai. Redigit Prof. Dr. Rinchen. Ulanbator, 1959.

СОСТАВЛЕНИЕ ПЕРЕЧНЯ СОКРАЩЕННЫХ НАЗВАНИЙ ЯЗЫКОВ В РАМКАХ ПРОЕКТА ДИГИТАЛИЗАЦИИ «СЛОВАРЯ ПОЛЬСКОГО ЯЗЫКА» С. Б. ЛИНДЕ

Иоанна Билиньска

Варшавский университет, Варшава, Польша

The paper is devoted to the digitalisation of Dictionary of Polish by Samuel Bogumił Linde. The author gives a brief description of a new lexicographic browser available at the Warsaw University and analyses the building process and usage of an index of languages' names abbreviations for digitalisation purposes.

Шеститомный «Словарь польского языка» (1807–1814) авторства Самуила Богумила Линде [Linde, 1807] — выдающееся научное достижение своей эпохи, удостоившееся высокой оценки не только в Польше, но и во всем мире. Нынешняя цифровая версия словаря, размещенная, в частности,