

О создании базы данных аутентичных текстов калмыцкого героического эпоса «Джангар»

On the Creation of a Database of Authentic Texts of the Kalmyk Heroic Epic “Dzhangar”

Б. Б. Манджиева (B. B. Mandzhieva)¹

¹ кандидат филологических наук, заведующий отделом фольклора, Калмыцкий научный центр РАН (г. Элиста). E-mail: mbbairta@yandex.ru.
Ph.D. in Philology (Candidate of Philological Sciences), Head of Department of Folklore Studies, Kalmyk Scientific Center of the RAS (Elista). E-mail: mbbairta@yandex.ru.

Аннотация. Статья посвящена одной из актуальных проблем калмыцкой фольклористики — созданию электронной базы аутентичных текстов, в которую будут включены образцы многожанрового калмыцкого фольклора, в частности тексты калмыцкого героического эпоса «Джангар». Материалом для базы данных аутентичных текстов будут являться письменные и устные первоисточники. Письменные первоисточники будут представлены рукописными текстами калмыцкого героического эпоса «Джангар» в записях XIX–XX вв. Устные первоисточники будут состоять из аудиозаписей эпических песен калмыцких сказителей, записанных в 1960–1980-е гг. учеными КНИИЯЛИ (ныне — КалмНЦ РАН). Реализация проекта по созданию базы аутентичных текстов будет иметь важное значение для сохранения фольклорного наследия калмыков и дальнейшего комплексного исследования такого уникального эпического памятника, как героический эпос «Джангар».

Ключевые слова: калмыцкий фольклор, героический эпос «Джангар», текст, письменные источники, устные источники, модуль, метаразметка.

Abstract: The article is devoted to one of actual problems of the Kalmyk folklore — the creation of an electronic database of authentic texts, which will include samples of multi-genre of Kalmyk folklore, in particular, the texts of the Kalmyk heroic epic “Dzhangar”. The material for the creation of authentic texts of the database will be written and oral source material. Writing primary sources will be presented handwriting Kalmyk heroic epic “Dzhangar” in the records of XIX–XX centuries. Oral source material will consist of recordings of songs Kalmyk epic storytellers recorded in the 1960s–1980s by scientists of KNIYALI (now KSC RAS). Implementation of the project to create an electronic database of authentic texts will be important to preserve the folklore heritage of the Kalmyks and further comprehensive study of this unique monument as a heroic epic “Dzhangar”.

Keywords: Kalmyk folklore, heroic epic “Dzhangar”, text, written sources, oral sources, module, metarazmetka.

В условиях современной глобализации каждый народ стремится сохранить традиционную национальную культуру, не утратить своеобразие. С появлением новых компьютерных технологий появляются новые возможности для сохранения нематериального культурного наследия народов Российской Федерации.

В Калмыцком научном центре РАН на протяжении многих лет ведется большая работа по разработке фундаментального проекта — Свода калмыцкого фольклора, в котором представлено во всей полноте устно-поэтическое и музыкальное творчество калмыков. В рамках данного проекта сотрудниками отдела фольклора КалмНЦ РАН подготовлены к изданию тексты песен, калмыцкого героического эпоса «Джангар». Фольклористами собран богатейший материал, который необходим для формирования локальной электронной базы, а в дальнейшем и единой базы данных калмыцкого фольклора. В настоящее время одной из актуальных задач калмыцкой фольклористики является создание электронной базы, содержащей образцы многожанрового калмыцкого фольклора, в частности базы данных аутентичных текстов героического эпоса «Джангар».

До настоящего времени нет единства в трактовке терминов «аутентичность» и «аутентичный материал». «Аутентичный» происходит от латинского *authenticus*, что в переводе означает ‘подлинный, достоверный’. В нашей работе под *аутентичным фольклорным материалом* понимаются образцы фольклора, созданные носителями языка, т. е. сказителями. Аутентичность материала отражает живую речевую деятельность. Многожанровый калмыцкий фольклор во всем своем многообразии является ярким образцом метафоричности языка, содержащим богатый пласт архаики. Оригинальные тексты эпических песен «Джангара» в ранней записи XIX в. включают в себя архаические мотивы, поэтому в них сохранилась древняя лексика, которая, к сожалению, уже не употребляется в современной калмыцкой речи. Поэтому включение текстов эпических песен «Джангара» в базу данных аутентичных текстов является одной из важных задач современной фольклористики, которая будет направлена, прежде всего, на сохранение фольклорного наследия калмыков.

Как известно, фольклорные тексты подразделяются на два вида: письменные и устные. Источником для создания базы данных аутентичных текстов будут являться и письменные, и устные первоисточники.

Письменные первоисточники состоят из рукописных текстов калмыцкого героического эпоса «Джангар»: записи XIX в., включающие тексты трех песен Малодербетовского цикла, две из которых были обнаружены А. Ш. Кичиковым в 1966 г. в архиве К. Ф. Голстунского, хранящиеся в рукописном фонде библиотеки Восточного факультета Санкт-Петербургского государственного университета под шифром 907 Джангар (список Мало-Дербетовского улуса). Calm. C. 4, inventory No. 1834. Old call number Xyl. Q 544. Golstunsky Collection (1862), No. 3 [Джангар. Список Малодербетовского улуса], а также тексты трех песен Багацохуровского цикла, две из которых также находятся в рукописном фонде библиотеки Восточного факультета СПбГУ, одна песнь хранится в архиве РГО; записи XX в. — тексты десяти песен джангарчи Ээлян Овла, записанные Номто Очировым в 1908–1910 гг., тексты шести песен джангарчи Мукебюна Басангова, записанные Басангом Дорджиевым и Михаилом Тюлюмджиевым в 1940 гг., тексты четырех песен джангарчи Давы Шавалиева в записи Алексея Бурдукова в 1939 г., текст одной песни джангарчи Бадмы Обушинова, записанная И. И. Поповым в 1901 г. и одна песнь джангарчи Насанки Балдырова, записанная А. Ш. Кичиковым и Н. Б. Сангаджиевой в 1966 г. [Джангар 1978].

Устные первоисточники состоят из аудиозаписей эпических песен, записанных во второй половине XX в. (1960–1980-е гг.) учеными КНИИЯЛИ (ныне — КалмНЦ РАН) от сказителей: Биткеев Годжга (отрывок), Церенов Эрдни Балдырович (2 песни), Човаев Ара (отрывок), Боожуков Аважн (1 песнь), Гаряев Цюрюм Чонкушевич (1 песнь), Лиджиев Телтя (пролог, 12 песен), Пахитова Ульяна Дорджиевна (пролог), Амбекова Боовуш Пюрвеевна (пролог, 1 песнь), Манджиев Михаил Бадушевич (пролог, 8 песен), Дорджиев Мушка Биджагаевич (пролог, 13 песен), Эрендженев Баснг Бамбаевич (пролог, 3 песни), Сертенов Тапар Серпекович (1 песнь).

Данные аудиозаписи были оцифрованы Б. Х. Борлыковой в лингвофольклорной лаборатории КалмНЦ РАН, которая была созда-

на в целях сохранения записей устной традиции калмыцкого народа конца XX в. и фиксации современного фольклорного материала. В рамках программы Президиума РАН в 2008–2009 гг. решалась задача создания научного архива современного уровня с целью сохранения уникального фонда звуковых записей по фольклору калмыцкого народа. На основании полученных в цифровом формате файлов была создана коллекция презентационных фрагментов, которые будут являться первоисточником базы данных аутентичных текстов калмыцкого героического эпоса «Джангар».

Для создания базы данных аутентичных текстов калмыцкого героического эпоса «Джангар» необходимо выполнить ряд задач. Первоначально необходимо создать два модуля: один модуль будет отвечать за письменный блок, второй — за устный. Для функционирования этих модулей требуется разработать программное обеспечение.

Структура письменного блока будет состоять из четырех частей:

1. Сканированные копии рукописей в формате pdf.
2. Текст в том виде, в каком он представлен в рукописи, что необходимо для осуществления поиска одинаковых элементов в различных фрагментах.
3. Переложение письменного первоисточника на современный калмыцкий язык.
4. Перевод на русский язык.

Переложение части материалов на «тодо бичиг» представляется одной из самых трудновыполнимых задач по ряду причин: в первую очередь отсутствует юникодовая кодировка всех графем, пишущихся в середине и в конце слова; отсутствует поддержка вертикального письма в текстовых редакторах; отсутствует распознающая программа, которая бы облегчила подготовку текстов на «тодо бичиг» [см. подробно: Бембеев 2012а; 2012б]. Тем не менее, на наш взгляд, возможно найти решение данной проблемы. Например, в русле данных исследований в КалмНЦ РАН ведется работа по созданию Национального корпуса калмыцкого языка, одним из направлений которого является разработка подкорпуса «ранних» текстов. «В рамках проекта проводится анализ рукопис-

ных и печатных источников XVII–XIX вв., выявляются их палеографические и лексические характеристики, среди которых можно перечислить особенности почерка переписчика или шрифта текста, формат рукописи, качество бумаги, чернил и т. п. Предпринятая работа в дальнейшем существенно облегчит работу по вводу и обработке калмыцких и ойратских текстов и изображений, систем оптического распознавания, систем информационного поиска и автоматического индексирования документов» [Куканова, Бембеев, Музраева 2013: 70–71]. С оцифрованными и транслитерированными текстами на старокалмыцком языке мы можем ознакомиться на сайте «Старокалмыцкий корпус» (<http://kalmcorp.ru/todosearch>). Данный подкорпус текстов начала XVIII — начала XX в. дает возможность получить обширнейший материал для исследования лексики и грамматики старокалмыцкого языка. В нашем проекте представление рукописных текстов «Джангара» XIX в. на «тодобичиг» откроет новые перспективы для всестороннего изучения калмыцкого героического эпоса, в особенности для лингвистов.

Структура базы аутентичных текстов будет иметь общую метаразметку (см. [Куканова 2011]), и в то же время более детализированную, в которой будут учтены сведения, необходимые для дальнейшей комплексной текстологии.

Метаразметка в базе данных аутентичных текстов будет иметь следующую структуру (см. [Куканова 2012: 196–197]):

- 1) идентификационный код текста;
- 2) название текста;
- 3) перевод названия на русский язык;
- 4) исполнитель/сказитель (если имеются сведения);
- 5) ФИО человека, зафиксировавшего текст (собиратель);
- 6) форма фиксации (цифровая/аналоговая запись или письменный текст);
- 7) время фиксации;
- 8) место фиксации;
- 9) место хранения оригинала или записи;
- 10) шифр;
- 11) количество листов;
- 12) наличие копии;

- 13) внешняя сохранность/качество записи;
- 14) датировочные данные;
- 15) сведения о почерке;
- 16) история публикации;
- 17) графическая система;
- 18) жанровая принадлежность;
- 19) комментарий.

Для того чтобы дать общую информацию необходимо создать метаописание текста, которое будет содержать краткие сведения об археографии и текстологии документа. Затем необходимо выполнить переложение с «тодо бичиг» на современный калмыцкий язык. Работа по переложению выполнялась авторским коллективом, так как тексты были подготовлены в процессе составления томов Свода калмыцкого фольклора. После загрузки текст необходимо проверить несколько раз. «Двойная проверка обоснована тем, что транслитерация и расшифровка текстов достаточно сложный процесс, который требует максимум внимания и сосредоточенности» [Куканова 2012: 197].

Алгоритм работы включает следующие действия:

- сканирование;
- загрузку;
- транслитерацию и одновременную сегментацию на строки и предложения;
- переложение на современный калмыцкий язык;
- перевод текста на русский язык.

Разработка данных модулей дает возможность осуществления поиска нужного фрагмента по текстам эпоса, в результате которого выдается список, где этот фрагмент использовался, его звучание (при наличии аудиозаписи), написание, и тогда исследователь, владея обширным материалом, может провести необходимый ему анализ текста (например, выявить исторические изменения или установить ретроспективу языкового явления и т. д.).

Устный модуль, который предлагается разработать по специализированной программе ELAN (программа для полевой работы), содержит стопроцентную аутентичность, т. е. текст без изменений. Применение данной программы в полевой фольклористике позво-

ляет передать особенности устной традиции, например интонации, эмоции. Фиксация фольклора на видеокамеры передает все особенности сказывания текста сказителем, полную картину, мимику, интонацию, жесты и т. д. Данный модуль позволяет создавать несколько уровней метаописания текста: например, можно сделать разметку тех фрагментов, когда используются руки, определенная жестикуляция и т. д.

Таким образом, реализация проекта по созданию электронной базы аутентичных текстов будет иметь важное значение для сохранения фольклорного наследия калмыков, а также откроет новые пути для комплексного текстологического изучения фольклора, всестороннего лингвистического анализа языка.

Литература

Бембеев Е. В. Коллекции рукописей на старокалмыцком (ойратском) языке XVII–XIX вв. в свете компьютерной обработки: постановка проблемы // Информационные технологии и письменное наследие. EI' Manuscript-2012: Мат-лы IV Междунар. науч. конф. (Петрозаводск, 3–8 сентября 2012 г.). Петрозаводск, Ижевск, 2012а. С. 31–34.

Бембеев Е. В. Опыт квантитативной обработки текста на старокалмыцком языке: количественные характеристики // Вестник Калмыцкого института гуманитарных исследований РАН. 2012б. № 2. С. 163–168.

Джангар. Калмыцкий героический эпос / сост. А. Ш. Кичиков; ред. Г. И. Михайлов. М.: Наука, Глав. ред. вост. лит., 1978. Т. 1. 441 с. Т. 2. 417 с.

Джангар. Список Малодербетовского улуса. Рук. на старокалм. письм. Оригинал записи поэмы «Догшн Шар Гүргү» // Рукописный фонд науч. библиотеки Восточ. фак-та СПбГУ. Шифр: 325.

Куканова В. В. Архитектура метаописания в Национальном корпусе калмыцкого языка // Вестник Калмыцкого института гуманитарных исследований РАН. 2011. № 1. С. 139–145.

Куканова В. В. Фольклорный подкорпус: проблемы, структура и перспективы использования // Участие калмыков в укреплении российской государственности: Мат-лы регион. науч.-практ. конф., посвящ. 1150-летию российской государственности и Году российской истории. Элиста: КИГИ РАН, 2012. С. 193–198.

Куканова В. В., Бембеев Е. В., Музраева Д. Н. К вопросу о кодировке UNICODE графической системы «тодо бичиг» и создания базы данных текстов на старокалмыцком языке // Вестник Калмыцкого университета. 2013. № 3 (19). С. 65–75.

References

Bembeev E. V. Manuscript Collections in the Old Kalmyk (Oirat) Language of 18th–19th Centuries in the Light of Computer Processing: Problem Statement. In: Information Technologies and Written Heritage. Conf. proc. (Petrozavodsk, 3–8 September 2012). Petrozavodsk, Izhevsk, 2012a. Pp. 31–34. (In Russ.)

Bembeev E. V. Experience of Quantification Text Processing in the Old Kalmyk Language: Quantitative Characteristics. *Bulletin of Kalmyk Institute of Humanitarian Research of the RAS*. 2012b. No. 2. Pp. 163–168. (In Russ.)

Dzhangar. Kalmyk Heroic Epic. A. Sh. Kichikov (comp.). G. I. Mikhailov (ed.). Moscow: Nauka, Vost. lit., 1978. Vol. 1. 441 p. Vol. 2. 417 p. (In Russ.)

Dzhangar. Copy of Maloderbet ulus. Original Record of the Poem “Dogshn Shar Gyrgy”. In: Manuscript Collection of the Scientific Library of Eastern Department of St. Petersburg State University. Cipher: 325. (In Russ.)

Kukanova V. V. Architecture of Meta-description in the National Corpus of the Kalmyk Language. *Bulletin of Kalmyk Institute of Humanitarian Research of the RAS*. 2011. No. 1. Pp. 139–145. (In Russ.)

Kukanova V. V. Folklore Sub-corpus: Problems, Structure and Prospects of Use. In: Participation of Kalmyks in the Strengthening of the Russian Statehood. Conf. proc., devoted to the 1150th anniversary of the Russian statehood and the Year of Russian history. Elista: Kalmyk Institute of Humanitarian Research of the RAS, 2012. Pp. 193–198. (In Russ.)

Kukanova V. V., Bembeev E. V., Muzrayeva D. N. Concerning UNICODE Coding of the Graphic System “todo bichig” and Creation of Database of Texts in the Old Kalmyk Language. *Bulletin of Kalmyk University*. 2013. No. 3 (19). Pp. 65–75. (In Russ.)