

РУССКО-КАЛМЫЦКИЙ СЛОВАРЬ ПАРМЕНА СМИРНОВА КАК ИСТОЧНИК ИЗУЧЕНИЯ ЛЕКСИКИ КАЛМЫЦКОГО ЯЗЫКА*

Н. М. Мулаева

В настоящее время «тодо бичиг» ('ясное письмо') не имеет должной компьютерной обработки, поскольку буквы в слове и слова пишутся сверху вниз (в вертикальном направлении), слова в столбцах разделяются пробелами, столбцы располагаются слева направо. Большинство букв имеет три различных написания в зависимости от позиции в слове (в начале, середине и конце).

Работа по созданию электронных коллекций старокалмыцких письменных памятников затрудняется 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 [Бембеев 2012: 39].

В русле исследований, связанных с компьютерной лингвистикой, в Калмыцком институте гуманитарных исследований РАН ведется работа по созданию Лингвистически аннотированного корпуса калмыцкого языка, к направлениям которого относятся разработка подкорпуса «ранних текстов» и создание «Словарного модуля калмыцкого языка». Последний представляет собой электронную лексикографическую базу данных (на основе лексикографических источников XVIII–XX вв.), которая впоследствии станет основой для создания программы, распознающей символы «тодо бичиг». Начата работа по оцифровке имеющихся словарей, приведению их к единообразию в оформлении словарной статьи: уже распознаны и тщательно выверены «Краткий русско-калмыцкий словарь» П. Смирнова [1857] и «Русско-калмыцкий словарь анонимного автора» (18 в.).

Как известно, словарный состав двуязычных словарей отражает демографическую ситуацию в соответствующих странах, уровень их экономического, политического и культурного развития, международные приоритеты [Юсупова 2007: 169]. Одна из целей создания указанных выше двуязычных словарей заключалась в обучении русских калмыцкому языку.

В данной статье предпринята попытка описания лексико-тематических групп в «раннем словаре», являющемся как лексикографическим памятником, так и источником изучения развития лексической системы калмыцкого языка, а также некоторых результатов количественного анализа структуры словника. В качестве материала выступает «Краткий русско-калмыцкий словарь» Пармена Смирнова, оцифрованный для Словарного модуля в корпусе калмыцкого языка.

Известный миссионер, священник Пармен Смирнов провел в калмыцкой степи около десяти лет. Свою церковную деятельность отец Пармен начал после окончания в 1844 г. полного курса Астраханской духовной семинарии. На калмыцкий язык он перевел первоначальные христианские молитвы, символ веры и десять заповедей Божиих, все евангельские чтения на воскресные, праздничные и высо-

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

которожественные дни, «Краткую священную историю Ветхого и Нового Завета», а также составил «Грамматику калмыцкого языка», «Калмыцкий словарь с прибавлениями», «Заметки о некоторых особенностях быта калмыков Астраханской губернии» и др. [Смирнов 1999: 244].

«Краткий русско-калмыцкий словарь» опубликован Парменом Смирновым в 1857 г. в Казани по настоянию Его преосвященства. Словарь знакомит нас с богатой лексикой калмыцкого языка, где можно встретить как устаревшие слова, так и слова, употребляемые в настоящее время. Этот лексикографический памятник составлен по образцу двуязычных словарей, слова в нем расположены в алфавитном порядке. Структура оцифрованного словаря состоит из нескольких элементов: дозволение цензора на печать, посвящение архимандриту Астраханскому и Енотаевскому, при этом в электронной копии отсутствуют несколько страниц, затем последняя страница документа от имени составителя, видимо предисловия, датированного 1852 г., сам словарь, состоящий из 3 184 единиц, комментариев о календаре калмыков на последней странице (на 127-й странице). Слова на странице расположены в две колонки, нумерация начинается с 10-й страницы сверху по центру. Заголовочное слово дается в исходной форме жирными буквами. В качестве заголовочного слова могут выступать и имена собственные (например: *Енотаевск (город)* šine balyasun ‘шинэ балхасун’; *Волга* İjil ‘Ижил’). Калмыцкие слова на «тодо бичиг» написаны не сверху вниз, а слева направо. Структура оригинальной словарной статьи состоит из:

- слова на русском языке, написанного в соответствии с нормами орфографии русского языка до 1917 г.;
- переводного эквивалента на калмыцкий язык, написанного на «тодо бичиг»;
- дефиниций безэквивалентной лексики*;
- в некоторых случаях этнографических комментариев**.

Словарь был переработан и структура изменена (были добавлены дополнительные графы):

- транслитерация на латинице;
- транслитерация на кириллице;
- современное написание лексических единиц (в соответствии с современной калмыцкой орфографией***).

В «Кратком русско-калмыцком словаре» Пармена Смирнова можно выделить следующие лексико-тематические группы слов:

1. Термины родства: *батьюшка* aba ‘аба’; *дядя по отцу* abaγu ‘абаγу’; *брат старший* аха ‘аха’; *внук* аči ‘ачи’.

Термины родства относятся к одной из наиболее древних групп лексики лютого языка. В них в известной степени отражается и социальное положение носителей языка, ибо термины родства характеризуют систему родственных отношений,

* Например: *вода, подслащенная сахаром и подкрашенная шафраном* aršān ‘аршаан’; *кислое молоко, смешанное с водой* arig ‘ариг’.

** Например: *орел, птица; орел, изображаемый на деньгах, собственно баснословная птица* gekse; *γarudi* ‘гэксэ; харуди’; *памятник надгробный, в роде пирамиды, в них находятся маленькия пирамидки, содержащая прах какого либо вельможи, которая называется* suburγa; *саса* ‘субурγа; паца’.

*** Использовались правила обозначения неясных гласных [Орфографические правила и орфографический словарь калмыцкого языка 2000: 50–52].

разные формы семейно-родственных связей того или иного народа, их исторические взаимодействия с другими народами [Дарбеева 1972: 16].

2. Слова, обозначающие растительный мир (флору): *ир (растение) andis* ‘андис’; *арженик (растение из семейства злаков, похожее на рожь) ariu öbüsün* ‘ариу өбүсүн’; *мята (растение) badaraški* ‘бадарашки’.

По мнению Л. В. Дмитриевой, по наименованиям растений, хорошо знакомым местному населению, по их количеству и разнообразию можно судить о месте жительства того или иного этноса [Дмитриева 1975: 125].

3. В словаре довольно широко представлен животный мир (фауна). Здесь можно выделить следующие подгруппы:

а) слова, обозначающие диких зверей: *лев (зверь) arsalan* ‘арсалан’; *лисица ünegen* ‘үнэгэн’; *барс bars* ‘барс’;

б) домашних животных: *мерин aqta* ‘агта’; *жеребец ajirγa* ‘ажирһа’; *верблюжонок botoxon* ‘ботохон’;

в) названия птиц, насекомых: *куропатка (птица) yotun* ‘йотун’; *воробей boqšurγoy* ‘богшурһой’; *муха bataγana* ‘батаһана’; *комар (насекомое) böküne* ‘бөкүүнэ’;

г) слова, относящиеся к животным, птицам: *горб бүкүн* ‘бүкүн’; *вислюхий delbiñ* ‘дэлбин’; *крыло jigür: jibir* ‘жигүр; жибир’; *гребень птицы zala* ‘зала’.

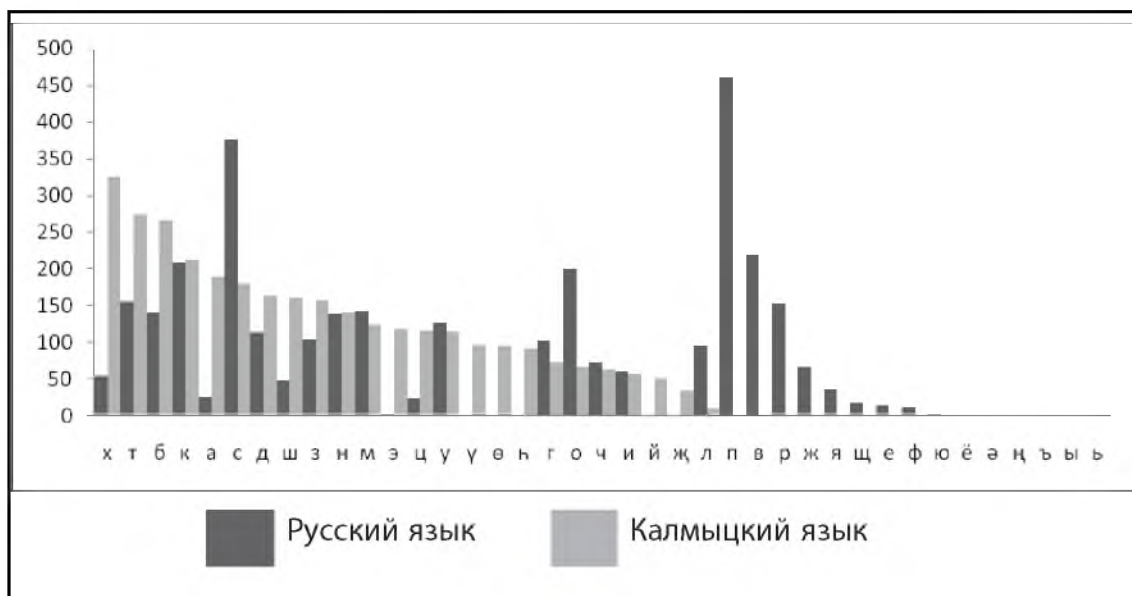
4. Слова, обозначающие географические термины: *остров заселенный dobon* ‘добон; *берег köbē: zaха* ‘көбээ; заха’; *долина ködö* ‘көдөө’.
5. Слова, обозначающие явления природы: *морозец zökün* ‘зөкүүн’; *гроза yike salkitay xura* ‘йикэ салкитай хура’; *мороз, стужа kiyitün* ‘кийитүн’; *иней kirayun: kirü* ‘кираһун; кирүү’.
6. Термины, обозначающие виды пищи, а также напитки: *лепешка пресная, которую пекут в золе под котлом būrsaq* ‘буурсаг’; *если в будан добавляют масло или жир, то называют его булмаг (лакомое кушанье) būlmaq* ‘буулмаг’; *вино arigi* ‘ариги’.
7. Термины, относящиеся к медицине: *лекарь, врач emči* ‘эмчи’; *жаба (болезнь) baqba gem* ‘багба гэм’; *горячка (болезнь) халун ebečin* ‘халуун эбэчин’;
8. Военная, рыболовная и охотничья лексика: *враг, неприятель dayisun* ‘дайисун’; *крюк, снасть рыболовная dügē* ‘дүгээ’; *лук, дуга, перетянутая тетивой numun* ‘нумун’; *дробь свинцовая (горох) burcuq: burčig* ‘бурцуг; бурчиг’.
9. Термины, относящиеся к жилищу человека: *жилище nutuq: süri yazar* ‘нутуг; суури һазар’; *крыша дома debür* ‘дээбүүр’; *кошма işige* ‘ишигэ’.
10. Имена числительные: *десять arban* ‘арбан’; *девятьсто yerin* ‘йэрин’; *девять yesün* ‘йэсүн’.
11. Соматическая лексика: *ладонь alixi: alaxi* ‘алихи; алахи’; *веко anisxay, zobuki* ‘анисхай, зобуки’; *кисть baγalcaq* ‘баһалцаг’; *лоб mañnay: mañlay* ‘маңнай; маңлай’.
12. Названия металлов, их сплавов, предметов и украшений, изготавливаемых из них, названия драгоценных камней: *монета (серебр.) arsalan* ‘арсалан’; *золото (металл) altan* ‘алтан’; *медь (металл) zes* ‘зес’; *алмаз oçir: almas çilün* ‘очир; алмас чилуун’.
13. Музыкальные термины: *дудка, флейта biškiür* ‘бишкиур’; *балалайка dombor* ‘домбор’; *бубен dülbас* ‘дуулбас’.

14. Названия инструментов: *серп* xadūr ‘хадуур’; *молоток* aluxa ‘алуха’; *ножницы* xāji ‘хаажи’; *плуг, инстр. земледельч.* andisun ‘андисун’.
15. Названия посуды: *мешок из коровьей кожи, в который вливают жидкость* arxad ‘архад’; *кувшин глиняный* bumba ‘бумба’; *сковорода* tabā ‘тава (таава)’.
16. Термины, обозначающие сословия, титулы, профессии: *государь, сударь* aldar ‘алдар’; *князь* noyon ‘нойон’; *архиерей* šajin lama. arkirey ‘шажин лама, аркирей’; *купец* xudaldāči ‘худалдаачи’; *кузнец* darxači ‘дархачи’; *наемник* yalči ‘йалчи’; *башмачник* sököči ‘сөөкөчи’.
17. Религиозная лексика: *бог, божество* burxan ‘бурхан’; *ангел* angel ‘ангел’; *освятить* šijitey boluxu ‘шижитэй болхуху’; *благословение* adistid: adis ‘адистид, адис’.

В ходе количественного анализа были включены все буквы калмыцкого и русского алфавитов, хотя мы отдаем себе отчет, что некоторых букв, как например *э*, нет в старокалмыцком языке. Количественный анализ свидетельствует о дисперсии, т. е. распределении, слов по алфавиту. Самой большой группой слов является слова, которые начинаются на букву *х*-, самой маленькой – группа слов на *л*-.

В связи с необходимостью обработки больших объемов текстовой информации и растущими возможностями информационных технологий в настоящее время исследуются закономерности организации текста, в частности его структура. Результаты таких исследований используются при разработке эффективных методов и алгоритмов обработки текстов, например, для его идентификации и классификации по тематическим критериям [Ломакин, Панкратова, Суркова 2011: 196].

Распределение слов в русско-калмыцком словаре П.Смирнова по первой букве



База данных по словарю П. Смирнова, которая структурирована в MS Office Excel, позволяет получать списки в алфавитном порядке по каждой графе, фильтровать данные для быстрого поиска необходимого материала. В дальнейшем планируется перевести табличный вариант базы данных в формат XML, что со-

ответствует всем стандартам ТЕІ, принципам подготовки текстов и материала для публикации в Интернете.

Составление Парменом Смирновым «Краткого русско-калмыцкого словаря» связано с введением в обиход калмыков религиозных терминов, связанных с христианством. Священники-миссионеры составляли двуязычные словари (русско-калмыцкие) с целью облегчить свое общение с представителями калмыцкого народа, и, конечно, это было связано с желанием обратить их в христианство. В современном калмыцком языке некоторые русские заимствования, отмеченные в двуязычном словаре П. Смирнова, уже не употребляются или изменили свое значение.

Данные оцифрованного и переработанного словаря составят основу лексического ядра для распознавания программы. Создание распознающей программы необходимо в целях быстрой обработки большого свода документов, написанных на «тодо бичиг» и хранящихся в Национальном архиве Республике Калмыкия и научном архиве КИГИ РАН, кроме того, слова, обработанные в ходе оцифровки словарей, войдут в словник «Грамматического словаря старокалмыцкого языка», который крайне необходим для автоматической обработки текстов (для создания лемматизатора).

Литература

- Бембеев Е. В.* Коллекции рукописей на старокалмыцком (ойратском) языке XVII и XIX вв. в свете компьютерной обработки: постановка проблемы // Материалы IV Междунар. науч. конф. «Информационные технологии и письменное наследие». Петрозаводск: ПГУ, 2012. С. 38–40.
- Дарбеева А. А.* О соотношении назывных и звательных слов в терминах родства в монгольских языках // Проблемы алтаистики и монголоведения. Вып. 2. М.: Наука, 1972.
- Дмитриева Л. В.* Фитонимы и фитонимическая терминология, общее для тюркских и монгольских языков // Проблемы алтаистики и монголоведения. Вып. 2. М.: Наука, 1975. С. 125–142.
- Ломакин Д. В., Панкратова А. З., Суркова А. С.* Золотая пропорция как инвариант структуры текста // Вестник Нижегород. ун-та им. Н. И. Лобачевского, 2011, №4(1). С. 196–199.
- Орфографические правила и орфографический словарь калмыцкого языка: около 23000 слов /РАН. КИГИ/ отв. ред. Э. У. Омакаева. Элиста: Калм. кн. изд-во. 2000. 480 с.
- Смирнов П. А.* Краткий русско-калмыцкий словарь (в пер.). Казань: Тип. Казан. Ун-та, 1857. 127 с.
- Смирнов Пармен.* Путевые записки по Калмыцким степям Астраханской губернии / под ред. А. Б. Лиджиева. Элиста: Калм. кн. изд-во, 1999. 248 с.
- Юсупова А. Ш.* Двуязычные словари татарского языка XIX века как лексикографические памятники // Вестник Челяб. гос. ун-та. 2007. № 22 (100). С. 169–177.