***Л.Д. Бадмаева***

**ПРОБЛЕМЫ МОРФОЛОГИЧЕСКОГО ОПИСАНИЯ ЛЕКСЕМ СТАРОМОНГОЛЬСКИХ ТЕКСТОВ[[1]](#footnote-1)**

**Аннотация.** Бурятскиелетописи, составленные на старомонгольской письменности более ста лет назад, получают свое второе рождение в виде публикаций в романизированном виде, что позволяет формировать на их текстовой базе корпусный ресурс. В работе показаны предварительные разработки для формализованного описания словоизменения старомонгольского языка для нужд морфологического аннотирования, созданных на нем летописных текстов бурят.

**Ключевые слова.** Бурятский язык, старописьменные летописи, корпус, морфологическое описание.

1. **Введение**

Труды по проблемам разметки и грамматического описания корпусных материалов освещают различные их аспекты. В российской корпусной лингвистике можно отметить работы А.А. Пичхадзе, С.О. Савчук, С.А. Крылова, А.Е. Полякова, Д.В. Сичинавы, М.А. Даниеля, Т.А. Архангельского, В.В. Кукановой и др. Коллектив разработчиков диахронического корпуса бурятского языка полагается на опыт проведения разметки текстов, излагающийся вышеперечисленными авторами. В начальной стадии формирования корпуса по тому или иному языку в первую очередь проводятся работы по сбору, составлению различных баз данных как текстовых, так и сугубо лингвистических. Последние бывают тесно связаны с теоретическими описаниями самого языка, создаваемого ресурса, далее на основе которых, строятся программные средства [Орехов 2014: 135]. В работе по идентичным проблемам, напр., башкирских текстов справедливо указывается, что «В основе любого автоматического анализа морфологии лежат две составляющие: грамматический словарь и формализованное описание словоизменительной системы языка» [Там же: 135].

Целью данной работы является освещение проблем, связанных с подготовительными работами для проведения разметки при составлении лингвистического диахронического корпуса на материале старописьменных монгольских текстов. Задачами проведенного первичного этапа являлись выбор и подготовка текстовых источников корпуса, формирование частотного словаря на их лексическом материале (выполнено Ринчиновым О.С.), разработка грамматических помет (тегов) для описания словоизменительной системы старомонгольского языка, морфологическое описание лексем указанного словаря, выявление особенностей плана выражения элементов словника частотного словаря.

Для организации названного корпуса выбраны пять транслитерированных электронных версий летописных бурятских текстов, опубликованных на CD-ROM: Летопись баргузинских бурят; Хроника Вандана Юмсунова 1875 г., Летописи хоринских бурят; Хроника Тугултур Тобоева 1863 г., Летописи хоринских бурят; Летописи селенгинских бурят, Хроника Убаши Дамби Джалцан Ломбо Цэрэнова, 1868 г. [Бадмаева 2009] и «Летопись хоринских бурят. Хроника Шираб-Нимбо Хобитуева» в [Бадмаева и др. 2018: 139-188]. Оригиналы данных летописей создавались задолго до начала кириллического периода функционирования письменности у бурят, в пору бытования у них соответственно старописьменного монгольского языка. Письменная графика названного языка использовалась бурятами официально до ее отмены в 1931 г. В научный оборот данные тексты введены в 1 половине ХХ века А.И. Востриковым, Н.Н. Поппе и В.А. Казакевичем.

1. **Основы и некоторые особенности морфологического описания старомонгольских текстов**

В основе грамматического описания старомонгольского языка лежит, в первую очередь, работа Н.Н. Поппе [Поппе 1935]. Автором использованы почти все основные труды классиков монголоведения XIX в., начала XX в., представляющие собой не только грамматические, но художественные сочинения, а также прессу его времени. Сама работа Н.Н. Поппе состоит из Введения и трех частей таких, как Фонетика и письмо, Морфология и Синтаксис. Для составления аннотации, представляющей собой таблицу помет / тегов с дальнейшим их толкованием в общепринятой лингвистической терминологии, использованы материалы второй части, т.е. Морфологии, при этом – это материалы сугубо старомонгольского словоизменения. Кроме названных материалов, использованы в небольшом объеме материалы из третьей части книги, т.е. Синтаксиса, посвященные послелогам, которые сам автор называет словами, несущими служебную функцию [Поппе 1935: 164].

Если сравнить описания послелогов у Н.Н. Поппе с их описаниями в грамматических трудах бурятского языка, то в них они получают описание в морфологическом разделе, называясь служебными словами или служебной частью речи. Поскольку мы озадачены передачей морфологической информации старописьменного монгольского языка и опираемся на конкретную работу Н.Н. Поппе, то в нашей разработанной системе помет для морфологической разметки, послелоги обозначены как элементы синтаксиса и получают соответствующий свой гипероним SYNT (Syntax), т.е. обозначение своей принадлежности к синтаксису. Из синтаксического раздела книги Н.Н. Поппе в нашу аннотационную таблицу включены знаки препинания[[2]](#footnote-2), в старомонгольском обозначении такие, как *čeg* «точка», использующаяся в нем в качестве запятой / обозначения конца предложения, *dӧrbelǰin čeg* «четыре точки» для обозначения конца главы, крупного раздела или строф стиха, *dabqur čeg* «двойные / две точки» – как запятая / конец предложения. Как выше было обозначено, на данном этапе подготовки лингвистических материалов для программных инструментариев дихронического корпуса бурятского языка в них не входят материалы старомонгольского словообразования.

Техническое оформление самой системы помет выполнено в таблице Excel (она доступна в формате PDF на разрабатываемом нами, коллективом разработчиков, сайте диахронического корпуса бурятского языка: <http://annals.imbtarchive.ru/>[[3]](#footnote-3)). В таблице помет отражается в основном идентичная последовательность представления грамматических категорий, наблюдающаяся в книге Н.Н. Поппе. Данная таблица содержит несколько полей (в скобках даны обозначения полей с примерами): A (тег: POS), B (русский термин: Часть речи), C (английский термин: Part of speech), D (гипероним: MORPH ‘Morphology’). В общей сложности сформулированы 122 тега, включая три вида знаков препинания старомонгольского текста. В данной таблице отражены такие грамматические категории старомонгольского языка как части речи, склонение (падеж, двойной падеж), притяжательное склонение, число, время, лицо, наклонение, спряжение.

В таблице выделена отдельная группа существительных, состоящая из семи видов соответственно со своими пометами: имя собственное; личное имя; фамилия; отчество; топоним; организация; существительные, выражающие одушевленность / неодушевленность; аббревиатуры. Следует сказать, что состав видов существительных, входящих в названную выше группу, продиктован соответствующим лексическим материалом летописных текстов, включаемых в разрабатываемый корпус, напр., *ilaγsangdarovisi* «Александрович» – отчество, *erküü* «Иркутск», *eǰen* «владелец, хозяин» – одушевленное имя существительное*, dalai* «море» – неодушевленное имя существительное.

Внутри каждой грамматической категории даны при необходимости дополнительные пометы, представляющие их специфику, напр., помета NG (Noun groups / Группы существительных) и помета DC (Double case / Двойной падеж) имеют соответственно свои дополнительные пометы отдельных подгрупп существительных и конкретных двойных падежей. Примерами могут быть следующие: PROPN – имя собственное – Proper name, PERSN – личное имя – Person name, ORGN – организация – Organization; COM.INSTR – Совместно-орудный падеж – Comitative.Instrumental, COM.ACC – совместно-винительный – Comitative.Accusative.

Т. о., сформулированы пометы по всем частям речи, таким как существительные (N – Noun), прилагательные (A – Adjective), глаголы (V – Verb), наречия (ADV – Adverb), числительные (NUM – Numeral), местоимения (PRON – Pronoun), частицы (PCL – Particle), союзы (CONJ – Conjunction), междометия (INTRJ – Interjection), послелоги (PSTP – Postposition)[[4]](#footnote-4), фиксирующимся в бурятских летописях. К пометам даны их соответствия на русском и английском языках, напр., по числительным, местоимениям, числу, частицам, глагольным формам (финитным, нефинитным), послелогам и т.д. Разработка помет и их соответствий полагалась на общепринятые традиции в соответствии с Лейпцигскими правилами глоссирования, лингвистическим словарем О.С. Ахмановой, а также с учетом уже использованных и представленных в соответствующих трудах. В случае отсутствия в них помет для специфических словоформ старомонгольской письменности формулировались новые в соответствии с названиями / обозначениями, принятыми в монголистике.

Разработанная таблица тегов / помет используется для приписывания грамматических характеристик лексемам старомонгольского языка, которое производится вручную на данном этапе (в перспективе находится решение задачи по разработке инструментария для автоматической разметки старомонгольских транслитерированных / романизированных текстов). Для данной процедуры предварительно составленный частотный словарь словоформ включает 10,3 тыс. языковых единиц по пяти вышеназванным опубликованным летописям бурят, в которых их текст представлен наряду с оригинальной своей старомонгольской версией соответственно в латинской транслитерации. Лексико-грамматическое описание словоформ составленного словаря выполнено в следующей последовательности признаков: словоформа – частотность – лемма – русский перевод – часть речи – одушевленность / неодушевленность – грамматические характеристики. Примерами могут быть следующие: *alba* – 60 – *alba* – дань, повинность; обязанность, служба – N (существительное) – INANIM (неодушевленность) – NOM (именительный падеж)[[5]](#footnote-5); *baraγun* – 53 – *baraγun* – правый – A (прилагательное); *basa* – 52 – *basa* – опять, снова – ADV (наречие); *degürgebe* – 1 – *degürgekü* – наполнить; заканчивать – V (глагол) – IND.PST1 (Изъявительное наклонение. Прошедшее время 1), CAUS (побудительный залог); *dergede* – 53 – *dergede* – возле, около – PSTP (послелог); *ǰaγun* – 150 – *ǰaγun –* сто – NUM (числительное); *ügei* – 271 – *ügei* – нет – PCL (частица); *ba* – 648 – *ba* – и – CONJ (союз); *bide* – 35 – *bide* – мы – PRON(местоимение) – ANIM (одушевленность) – NOM.

Ручное приписывание грамматических характеристик словоформам названного частотного словаря выявляет их особенности, характерные в той иной степени для подобных текстов, созданных до устоявшихся норм литературных языков, напр., русского [Гаврилова и др. 2016: 7]. Можно указать особенности лексики текстовой базы данных – вариативность написания иноязычных слов, личных имен, заимствований в виде русизмов (тибетизмы и санскритизмы встречаются, но редко). Совершенно справедливо указывается в вышеприведенной работе, что «Чем больше вариантов написания будет учтено при разработке ресурсов для аннотации, тем более полно и точно будут проанализированы тексты» [Там же: 10]. Появление русизмов / интернационализмов в бурятских летописях обусловлено соответственно их лингвоспецифичностью, т.е. отсутствием в бурятской культуре в период создания соответствующих текстов тех понятий, которые они выражали.

Примерами вариативности написания иноязычной лексики могут быть следующие: для слова «императрица» в бурятских летописях на старомонгольской письменности встречается десять различных написаний *imperatarsa, imperaturca, impiiratarica, impiratarica, impiratariica, impiratrica, impiraturaca, impiraturica, impiraturiici, imperaturiica*; для слова «министр» встречается три написания – *minister, miniister, miniistar*; Санкт-Петербург встречается в четырех написаниях – *sangpitirburge, sangpitirbürge, sangpitirbüürge, sanktapitirbürge*; слово «адъютант» дано в пяти орфографических формах – *adyudanta, adyüdanta, adyutanta, adyütanta, adyutanti*; для имени *Дамбадугар* встречаются два разных написания – *dambaduүar* и *dambadukar*.

Отдельного описания требует написание падежных форм в рассматриваемых летописях бурят. В письменном монгольском языке данные аффиксы пишутся отдельно от основы слова. В транслитерации указанную отдельность выражают дефисы, напр., *morin-u*, форма родительного падежа,«коня». В материале частотного словаря из семи косвенных падежей (Genitive, Dative-Locative, Accusative, Instrumental, Comitative, Ablative, Indefinite) и четырех двойных падежей (Dative.Locative.Ablative – дательно-местно-исходный, Comitative.Instrumental – совместно-инструментальный, Genitive.Dative.Locative – родительно-дательно-местный, Comitative.Accusative – совместно-винительный) старомонгольского языка выявляются словоформы в пяти, таких как GEN (родительный п.), DAT-LOC (дательно-местный п.), ABL (исходный п.), INSTR (орудный п.), COM (совместный п.). Во всех выявленных падежных формах наблюдается слитное написание соответствующих аффиксов, напр.: *qadaYIN*[[6]](#footnote-6) «горы», GEN, SG; *γajarA* «на территории», DAT-LOC; *angqanAČA* «от / из начала», ABL; *čaʏaǰaBAR* «законом», INSTR; *nökürTEI*«с другом», COM.

1. **Заключение**

Диахронический корпус как онлайн ресурс разрабатывается для его использования при исследованиях этнолингвистического, исторического и географического контекста опубликованных старомонгольских памятников бурят в латинской транслитерации. Представляется, что разрабатываемый корпус важен как для изучения истории языка бурят, так и для поиска возможных путей его дальнейшего развития, поскольку его современное состояние общепризнано как сложное в плане сохранности.

Вышеописанные работы должны послужить формированию грамматического словаря старомонгольского языка, разработке морфологического анализатора текстов на данном языке, а также системы корпусного поиска языковых данных, что представляет основные задачи следующего этапа работ составляемого корпуса. Также важными представляются работы по совершенствованию системы помет (тегов) за счет добавления аффиксов словоизменения (склонение, включая диалектное; спряжение), падежных форм старомонгольских местоимений, частиц в ее таблицу, по наполнению парадигмы вариативности в лексической системе бурятских летописей, в таких группах, включающих заимствования, ономастическую, топонимическую и другую специальную лексику, что в свою очередь в перспективе послужит семантической разметке корпуса.

При интегрировании размеченного текста в соответствующий корпус появляется возможность быстрого поиска и получения лексического материала при проведении разнообразных лингвистических исследований: анализ контекстов, лексических / статистических / диалектных / архаистичных данных, сочетаемости, синтаксических конструкций и т.д. Посредством разметки совершенствуется грамматический словарь по текстам, базы данных которых включаются в корпус. Создаваемый грамматический словарь по летописным бурятским текстам может послужить формированию исторического словаря бурятского языка.

Работы по диахроническому корпусу бурятского языка требуют своего продолжения, он дополнит такие ресурсы как Бурятский корпус, Параллельный корпус (бурятский). В перспективе при дальнейшем их общем усовершенствовании они составят основные направления / части Национального корпуса бурятского языка.

**Литература**

1. Бадмаева Л.Б. (ред.) (2009), История бурятской книги: справочно-библиографический CD-ROM. Улан-Удэ.
2. Бадмаева Л.Б., Очирова Г.Н. (2018), Летопись Ш.-Н. Хобитуева как памятник письменной культуры бурят. Улан-Удэ.
3. Гаврилова Т.С., Шалганова Т.А., Ляшевская О.Н. (2016), К задаче автоматической лексико-грамматической разметки старорусского корпуса XV-XVII вв. *Вестник ПСТГУ. Серия III: Филология. Вып. 2 (47)*, с. 7-25.
4. Орехов Б.В. (2014), Проблемы морфологической разметки башкирских текстов. *Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2014,* с. 135-140.
5. Поппе Н.Н. (1937), Грамматика письменно-монгольского языка. М.-Л.

**References**

1. Badmaeva L.B. (ed.) (2009), Istoriya burjatskoy knigi: spravochno-bibliograficheskiy CD-ROM. Ulan-Ude.
2. Badmaeva L.B., Ochirova G.N. (2018), Letopis' Sh.-N. Hobitueva kak pamyatnik pis'mennoy kul'tury buryat. Ulan-Ude.
3. Gavrilova T.S., Shalganova T.A., Lyashevskaya O.N. (2016), K zadache avtomaticheskoy leksiko-grammaticheskoy razmetki starorusskogo korpusa XV-XVII vv. *Vestnik PSTGU. Ser. III: Philology. Issue 2 (47), pp. 7-25.*
4. Orekhov B.V. (2014), Problemy morfologicheskoy razmetki bashkirskikh tekstov. *Trudy Kazanskoy shkoly po komp’yuternoy i kognitivnoy lingvistike TEL-2014. Kazan’: «Fen» RT’ Academy of sciences Publ., pp. 135-140.*
5. Poppe N.N. (1937), Grammatika pis'menno-mongol'skogo yazyka. Moscow-Leningrad.

**PROBLEMS OF MORPHOLOGICAL DESCRIPTION OF LEXEMES OF THE OLD-MONGOLIAN TEXTS**

L.D. Badmaeva

**Abstract.** The Buryat chronicles, compiled in the Old-Mongolian Writing more than a hundred years ago, are given their rebirth in the form of publications in a Romanized form from the beginning of the 3rd millennium, which makes it possible to form a corpus resource on their text base, which will multiply the directions of their research by various specialists, primarily linguists. The work will show preliminary developments for a formalized description of the inflection of the Old-Mongolian language for the needs of morphological annotation of the chronicle texts of Buryats created on it. The work performed should serve the formation of a grammatical dictionary of the Old-Mongolian language, the development of a morphological analyzer for texts on this language, as well as a corpus search system for language data.

**Key words.** TheBuryat language, old-written chronicles, corpus, morphological description.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Бадмаева Любовь Дашинимаевна**

ФГБУН Институт монголоведения, буддологии и тибетологии Сибирского Отделения РАН

**Badmaeva Liubov**

FSBSI Institute for Mongolian, Buddhist and Tibetan Sciences, Siberian Branch, RAS

***E-mail:*** [***ldbadm@gmail.com***](mailto:ldbadm@gmail.com)

1. Исследование выполнено в рамках государственного задания (проект «Мир человека в монгольских языках: анализ средств выражения эмотивности»). [↑](#footnote-ref-1)
2. В системе латинской транслитерации, общепринятой в международном монголоведении, знаки препинания вертикальной старомонгольской графики при ее романизации передаются в виде запятой или двоеточий, в зависимости от их количества. [↑](#footnote-ref-2)
3. Разработчик сайта – Ринчинов О.С. [↑](#footnote-ref-3)
4. Выше дано объяснение о разнице описания послелогов в книге Н.Н. Поппе и современных грамматиках бурятского языка. [↑](#footnote-ref-4)
5. Описание грамматических характеристик лексем выполнено Аюшеевой М.В. и автором статьи. [↑](#footnote-ref-5)
6. В данных примерах аффиксы выделены для наглядности прописными буквами. [↑](#footnote-ref-6)