

МОНГОЛОВЕДЕНИЕ

УДК 81'322=512.31

О НЕКОТОРЫХ ОСОБЕННОСТЯХ БУРЯТСКОГО СУФФИКСАЛЬНОГО ФОРМО- И СЛОВООБРАЗОВАНИЯ: ЛИНГВОСТАТИСТИЧЕСКОЕ ОПИСАНИЕ НА ПРИМЕРЕ ХУДОЖЕСТВЕННОГО ТЕКСТА

© *Дырхеева Галина Александровна*

доктор филологических наук, профессор, главный научный сотрудник,
Институт монголоведения, буддологии и тибетологии СО РАН
Россия, 670047, г. Улан-Удэ, ул. Сахьяновой, 6
E-mail: an5dag1@mail.ru

Статья посвящена особенностям структурной организации слова в бурятском языке. В основе анализа лежит обратный алфавитно-частотный словарь, составленный по произведениям классика бурятской литературы Х. Намсараева. Общий объем обработанного текста составил более 272 тысяч словоупотреблений, объем словаря — 36540 словоформ. Разработаны принципы членения словоформ на морфемы в целях автоматической обработки словаря, рассмотрены основные трудности морфологического членения. Разработан алгоритм автоматизированного морфологического анализа, который опирается на формально выраженные элементы морфемного членения. Всего выделено 5028 корней и 271 суффикс. Выявлены наиболее продуктивные и частотные корни и суффиксы, редкие, а также неописанные в научной литературе. Определена функциональная нагрузка отдельных суффиксов, а также степень разнообразия определенного типа производных слов.

Ключевые слова: морфологическая структура слова; бурятский язык; статистика; корень; суффикс; обратный алфавитно-частотный словарь.

Существуют различные подходы к выявлению типологических особенностей того или иного языка. Одним из наиболее эффективных считается количественный или статистический подход, а именно выявление различных частотных характеристик тех или иных лексико-грамматических классов слов, их структурных особенностей. Такой подход позволяет провести более тонкую и точную оценку их грамматических, функциональных и статистических особенностей, распространенность и продуктивность. Достаточно целесообразным считается использование данного подхода для изучения морфологического строя агглютинативных языков, в числе которых и бурятский язык. Традиционно бурятский язык относится к агглютинативным, хотя, как отмечают многие исследователи, нет чисто флективных, агглютинативных, изолирующих и т.д. языков. Проблема морфологической структуры слова, его состава до сих пор относится к одной из сложнейших в бурятском языкознании [1; 2; 3; 4; 5; и др.] и, возможно, ее решению отчасти может помочь именно лингвостатистический подход, в частности, использование для этого обратного алфавитно-частотного словаря. Обратное алфавитное расположение слов позволяет выявить весь спектр слово- и формообразующих показателей и элементов языка, специфику морфологической структуры слова, словообразовательных моделей, омоформ.

В связи с поставленной задачей — исследование лексико-грамматической системы бурятского языка на базе обратного алфавитно-частотного словаря — необходимо было решить ряд задач, связанных с некоторыми частными, но достаточно сложными, теоретическими вопросами бурятского языкознания:

- 1) определение принципов членения словоформ на морфы,
- 2) установление лексических и грамматических значений морфем,
- 3) выявление возможностей появления и, соответственно, учета при слово- и формообразовании элементов значений, формально не выраженных, а также случаев лексико-грамматической омонимии,
- 4) характеристика типов словообразования, особенностей сложного словообразования.

Очевидно, что в небольшой статье, практически, невозможно подробно рассмотреть все указанные аспекты обычно сложных и спорных вопросов морфологического анализа любого языка, поэтому были выделены основные моменты, а также наиболее характерные трудности, с которыми пришлось столкнуться на этапе занесения подготовленного массива словарного материала в электронную базу данных.

Общеизвестно, что при исследовании агглютинативных языков наиболее целесообразным считается использование последовательного порядкового членения, поскольку принцип построения словоформ считается линейным и каждая вычленимая элементарная единица плана выражения более или менее строго соотносена с планом содержания. При этом, очевидно, что производится сопоставление анализируемого слова с его грамматическими формами (для выделения окончания), производящими словами (выделение основы) и с одномодельными словами (выделение словообразовательных суффиксов), то есть учитывается семантическая общность словообразующих аффиксов. Алгоритм нашего автоматизированного морфологического анализа опирается на формально выраженные элементы морфемного членения, выделенные с помощью знаков /, = (для парных слов), * (для устойчивых словосочетаний), а также на приписанный каждой словоформе индекс, обозначающий часть речи.

Разработанные на подготовительном этапе принципы морфемного членения были дополнены в ходе занесения словарного массива в электронный корпус. А именно, во-первых, рассмотрению подлежало преимущественно синхронное состояние языка, этимология слова учитывалась в случае спорности членения на аффиксы. Во-вторых, соответственно, считается, что в современном бурятском языке имеются полные и краткие типы основ. Одни суффиксы требуют полную основу, другие — краткую; одни суффиксы присоединяются к основам на твердый согласный, другие — на мягкий. В-третьих, соответственно, учитывается орфографический прием передачи твердости-мягкости с помощью йотированного гласного, с помощью которого передается палатализация предыдущего согласного и долгота гласных. При этом йотированный гласный при графическом делении на морфемы включается в состав последующего суффикса. В-четвертых, в целях более корректного морфологического анализа в современных заимствованных словах (преимущественно, из русского языка) выделяются бурятские суффиксы, русская основа морфемному членению не подлежит.

Как и ожидалось, наибольшую трудность при членении на морфемы вызвали случаи, связанные с фузией и с внутренней флексией, появлением вставных

гласных и согласных, выпадением гласных при образовании сложных слов, а также с гаплогогией. Например, в случае если основа заканчивается на гласный, а присоединяемый суффикс имеет в начале долгий гласный или представляет собой долгий гласный, происходит редукция одного гласного на стыке морфем, и, поскольку членение на морфемы происходит справа налево, то образуются как бы усеченные основы или варианты основ: *томо-//том-* ← *томо* «крупный» и *том/оо/рхуу* «важный». В результате наблюдаются случаи, когда, например, корень глагола состоит из одного согласного: *г-//гэ-* ← *г/ээд* и *гэ/хэ* (многозначный глагол, используемый во множестве сочетаний, основное смысловое значение «говорить»). Таким же образом происходит редукция суффиксов, например, в словоформе *хата/ра/лд/уула/н* суффикс *лд-* — вариант суффикса *лда-*, в словоформе *яа/ш/уужа* суффикс *ш-* является вариантом суффикса *ша-*. То есть здесь к тому же наблюдается явление диссимилиации. Данное явление достаточно частое в бурятском языке, особенно при образовании сложных имен существительных, союзов, частиц: *те/эд* (союз «но, однако...») ← *тиихэ гээд* (регрессивная диссимилиация), *аа/бза* (частица модальная и вопросительная) ← *аа бэээ* (прогрессивная диссимилиация).

Выпадение гласного или согласного, а иногда и целых слогов наблюдается в случае образования сложных имен существительных. Например, слово *үхи/бүүн* «ребенок» образовано из *үхин* «дочь, девочка» и *хүбүүн* «сын, мальчик», при этом выпал конечный *н* первого слова и начальный слог второго слова; с помощью выпадения гласного *ү* образовано слово *һайн/дэр* «праздник» (из *һайн* «хороший» и *үдэр* «день»); слова *лам/багай* и *лам/хай* «батюшка-лама» образованы из *лама* «лама» и *абгай, ахай* — (почтительные обращения к старшим) посредством выпадения конечного гласного первого и начального гласного второго слов с дополнительной инверсией в случае *ламбагай*. Явление гаплогогии имеется, например, в слове *зо/мни* «мой народ» ← *зон минии*.

В бурятском языке достаточно часто наблюдаются обратные случаи: использование при слово- и формообразовании вставных гласных (например, в слове *дар/уул/а/н* от *дар/уул/ха* «подавлять») и согласных, чаще это –н, древний маркер именной основы: *би/дэ/н/шиэ* «мест. *мы* в вин.п.». Можно предположить, что вставные согласные и гласные есть результат такой особенности бурятского языка как полногласие и в случае, если основа или суффикс заканчивается согласным, а следующий за ним суффикс также начинается с согласного, между ними по закону сингармонизма будет соответствующий гласный. И, наоборот, между гласными на стыке морфем вставляется согласный, также как существуют варианты словоизменяемых суффиксов для основ, заканчивающихся на гласный и согласный звук.

Анализ процессов, происходящих в морфемных швах, был осложнен также в случаях, когда слово является древним заимствованием из тюркских, тунгусо-маньчжурских или из монгольского языков. Так, например, условно-непроизводными самостоятельными основами в ходе анализа были признаны: *са/й/ха* «белеть», *саһан* «снег», *сагаан* «белый», имеющие явное семантическое общее ядро-значение «нечто, связанное с белым»; *дүшэн* «сорок», *дүрбэн* «четыре», *дүнэн* «четырёхлетний», общее значение связано с числом «четыре». Все подобные случаи требуют специального этимологического анализа. Деление на морфемы было затруднено также, например, в случае типа *ами/с/хаа/л/ха* «ды-

шать», где суффикс *с*– является показателем монгольского заимствования. Словообразующий аффикс –*лиг* был заимствован из тюркских языков (*-лык, -лик, -лыг, -лиг*): *тэр/лиг* “национальный халат”, *зар/лиг* “приказ, повеление», *сэсэг/лиг* “цветник». Аффикс –*лиг* и его вариант *-ли(-уули)* образуют не только имена существительные, но и имена прилагательные: *нар/уу/лиг/нар/уули* “солнечный, солнечное место”; *добо/лиг/доб/уули* “холмистый, холм, бугорок”. При этом, можно заметить, что прослеживается неустойчивость конечного согласного «г».

Таким образом, было сделано предположение, что морфологический инвентарь будет достаточно многочисленным. Можно также было ожидать, что вариативность основ и суффиксов будет значительной вследствие многодиалектности языка и многовариантности написания тех или иных слов, например: *хүмбэр* и *сүмбэр* “гора Сумеру”, *үниэн* и *үнэшэн* “сирота” и т. д.

Для лингвостатистической обработки внесенного массива разрабатывался комплекс программного решения, который впервые был применен в данной сфере филологических изысканий, что позволило на качественно новом уровне решать задачи грамматического характера. Проектируемое программное обеспечение позволяет исследователю-филологу работать в комфортной пользовательской среде.

Обратный алфавитно-частотный словарь был составлен на материале прозаических произведений классика бурятской литературы Х.Намсараева, общий объем обработанного текста составил около 272 словоупотреблений (результат морфологического анализа списка словоформ представлен на компакт-диске) [Дырхеева 2005]. Объем словаря — 36540 словоформ. В целом были разработаны следующие принципы морфологического членения словоформ: при исследовании агглютинативных языков наиболее целесообразным считается использование последовательного порядкового членения, поскольку принцип построения словоформ считается линейным и каждая вычленяемая элементарная единица плана выражения более или менее строго соотносена с планом содержания. При этом, очевидно, что производится сопоставление анализируемого слова с его грамматическими формами (для выделения окончания), производящими словами (выделение основы) и с однокоренными словами (выделение словообразовательных суффиксов), то есть учитывается семантическая общность словообразующих аффиксов. Для разграничения морфологических показателей используется косая черта /. Однако, членение на морфемы в бурятском языке затрудняется такими особенностями как фузия, внутренняя флексия, вставные гласные и согласные. Так, в случае, если конечный звук основы гласный и следующий за ней аффикс также начинается с гласного звука, на стыке морфем обязательно происходят изменения основы и аффикса, приводящие к выпадению или, наоборот, появлению вставных фонем. Очевидно, что подобные случаи затрудняют морфологическое членение и требуют тщательного всестороннего анализа.

В итоге обработки всего массива было получено 4 разнообразных списка, в которых все единицы сопровождаются количественной характеристикой их встречаемости в тексте и словаре: прямой алфавитно-частотный список словоформ, разделенных на морфемы; тот же список, упорядоченный в обратном алфавитном порядке; список корней и список суффиксов.

Всего выделено 5028 корней и 271 суффикс. Общее количество выделенных морфем составляет около 5300 единиц. Наиболее частотными и продуктивными,

как и ожидалось, явились глагольные корни *г-* (*гэ-*) 'говорить' (12853 — частота в тексте и 215 — частота в словаре), *бай-* 'быть' (8920/311) и *боло-* 'становиться' (5123/267). Причем из 31 словообразовательной основы (корня), частота встречаемости которых в словаре превышает 100, т. е. они имеют высокий слово- и формообразовательный потенциал, только 2 — *дуун-* (*дуу-*) и *амин-* (*ами-*) — имеют именные признаки, причем они находятся на стыке именного и глагольного словообразования.

Количество словообразовательных суффиксов: С (существительные) — 68 (самые частотные *-хан*, *-л*, *-аан*), П (прилагательные) — 49 (*-ма*, *-та*, *-хи*, *-уу*, *-гар*), Г (глагол) — 26 (*-ла*, *-ра*, *-на*), Н (наречие) — 11 (*-ша*, *-ра*, *-на*), И (изобразительное слово) — 6 (*-га*, *-д*), Ж (междометие) — 3 (*-гшаа*). Среди суффиксальных частиц самыми распространенными являются притяжательная частица *-нь/-ань/-инь/-ынь*, отрицательная — *гуй/-дуй*, уступительная *-шье*. А среди выделенных частей сложных слов *-гара/-гар*, которая встречается только в одном слове *дуу/гара-* (114/1056) и *гэ-/г-* — в двух словах *тши/гэ-* и *ши/гэ-* (65/704).

При этом можно отметить, что обычно высокочастотные слова в тексте имеют высокую частоту и в словаре, т. е. они отличаются разнообразием производных форм. Такими являются словообразовательные суффиксы имен существительных *-л*, *-хан*, имен прилагательных *-гар* и *-та*, у глаголов *-ла* и *-ра*, наречий — *-са*, *-ша*. Отношение частоты в тексте к частоте в словаре выражает функциональную нагрузку слова, словоформы или любого лексического компонента. Среди суффиксов особой функциональной нагрузкой выделяются словообразовательные суффиксы наречия *-доо*, *-ша*, *-ра*, *-уур*. Причем суффикс *-доо*, например, встречается только в одном слове *ходо/доо* (F=72), соответственно, его функциональная нагрузка (Ft/Fc) оказалась самой высокой — 72 (Ft — частота в тексте, Fc — частота в словаре).

То есть при рассмотрении данного показателя очень важным является характеристика множества производных или частота в словаре и, соответственно, функциональная нагрузка может быть низкой из-за множества производных. Таким образом, обратное соотношение — Fc/Ft — выражает степень разнообразия данного типа производных слов. Например, среди суффиксов прилагательных *-гар* при относительной высокой частоте в тексте (794) имеет небольшую функциональную нагрузку — 2.7, поскольку у него самая высокая среди данного группы суффиксов частота в словаре, т. е. он участвует в производстве максимального числа разнообразных слов и форм среди прилагательных.

Как и ожидалось, самыми высокочастотными оказались суффиксы образования деепричастий *-жа/-жо/-жэ*, *-н*, причастий (*-ха/-хо/-хэ*, *-хан/-хон/-хэн*) и глагола настоящего времени *-на/-но/-нэ*. При этом можно отметить, что деепричастные и причастные суффиксы в целом отличаются высокой частотностью использования в тексте. Легче назвать формы, частота которых незначительна: суффикс попутного деепричастия *-нгаа*, заменительного — *-нхаар*, степени действия — *хысо*.

Что касается именных форм, то самыми частотными являются нулевое падежное окончание (8829/79256), суффиксы родительного падежа *-ай/-ой/-эй* (1823/10833), дательного *-да/-до/-дэ* (1656/8417), совместного *-тай/-той/-тэй* (1720/7609), а также винительный падеж возвратного притяжания *-аа/-оо/-ээ/-ёо/-еэ/-яа* (11723/1956).

Что касается слово- и формообразовательной структуры слова, то в словаре примерно 10% — корневые слова, 32% состоят из двух морфем, 36% — трех, 16% — четырех, более 5% — из пяти и более морфем.

В ходе проверок словарных массивов уточнялись принципы морфологического членения, касающиеся словообразования, отнесения той или иной словоформы к заглавной корневой морфеме.

Проведенный анализ позволил выявить не только самые частые, но и самые редкие случаи использования тех или иных суффиксов для словообразования и формообразования. Например, *-ги* в *уран/ги*, *-тар* в *бүхэ/тэр* и др. Хотя, возможно, что некоторые из них являются плодом творчества писателя или диалектными особенностями.

В ходе анализа были выявлены суффиксы, не описанные в научной литературе. В слове *мордо/хуул/ха* «отправлять, провожать» выделение залогового суффикса *-хуула* обусловлено наличием ряда *яб/уула — бай/гуула — мордо/хуула* и т.д. Можно также отметить, что имеется вариант *морд/уула*. То есть данный суффикс, скорее всего является вариантом суффикса *-уула/-гуула*, а его написание, возможно, обусловлено вариативностью передачи старомонгольской графики кириллицей. Возможно, также с этим связано выделение глаголообразующего (от прилагательного) суффикса *-мһара (-мһар)* в слове *һай/мһар/ха* «хвастаться, кичиться» и он является фонетическим вариантом суффикса *-маар*. К разговорному варианту (род.п. *-ай*) можно отнести суффикс *-и* в слове *газар/и* «обычно бранное, в сочетаниях усиливает негативный оттенок», поскольку возможны случаи типа, *ман/ай хүбүүн/и һамган* «жена моего сына» вместо *ман/ай хүбүүн/эй һамган*. Выявлены также варианты суффиксов *-жала/-шала*, *-саа/-шаа*, *-хоор/-оор*, *-гоор*. Отмечены редкие случаи образования имен существительных с помощью суффикса *-дхэл*: *зүү/дхэл*, *үмэ/дхэл*. Возможно, данный суффикс состоит из суффикса *-дхэ*, образующего форму страдательного залога и *-л*, с помощью которого обычно образуются имена существительные.

Для лингвостатистической обработки внесенного массива был разработан комплекс программного решения, который впервые применен в данной сфере филологических изысканий, что позволило на качественно новом уровне решать задачи грамматического характера. Проектируемое программное обеспечение позволяет исследователю-филологу работать в комфортной пользовательской среде. В частности, оно позволяет вывести всю совокупность словоформ с тем или иным суффиксом с соответствующими частотными характеристиками.

Литература

1. Бертагаев Т. А. Морфологическая структура слова в монгольских языках. М.: Наука, 1969. 183 с.
2. Дондуков У.-Ж. Ш. Аффиксальное словообразование частей речи в бурятском языке. Улан-Удэ: Бур.кн.изд-во, 1964. 246 с.
3. Дырхеева Г. А., Ринчинов О. С. Морфологическая структура слова в бурятском языке: лингвостатистическое описание (на материале художественного текста). Улан-Удэ: Изд-во БНЦ СО РАН, 2005. 98 с.
4. Санжеев Г. Д. Несколько замечаний о структуре слова и его границах // Морфологическая структура слова в языках различных типов. М.; Л.: Изд-во АН СССР, 1963. С. 270–272.

Г. А. Дырхеева. О некоторых особенностях бурятского суффиксального формо- и словообразования: лингвостатистическое описание на примере художественного текста

5. Харанутова Д. Ш. Словообразование бурятского языка. Улан-Удэ: Изд-во Бурят. гос. ун-та, 2012. 270 с.

ON SOME FEATURES
OF THE BURYAT SUFFIX FORM- AND WORD FORMATION:
THE LINGVOSTATISTIC DESCRIPTION ON THE EXAMPLE
OF THE ART TEXT

Galina A. Dyrkheeva

doctor of Philology, professor, main research worker of Institute
for Mongolian, Buddhist and Tibetan Studies of the Siberian Branch
of the Russian Academy of Sciences
Sakhjyanova St., 6, Ulan-Ude, 670047, Russia
E-mail: an5dag1@mail.ru

Article is devoted to the features of the structural organization of the word in the Buryat language. The return alphabetic frequency word book made on works of the classic of the Buryat literature H. Namsarayev is the cornerstone of the analysis. The total amount of the processed text made more than 272 thousand word usage, dictionary volume — 36540 word forms. The principles of partitioning of word forms on morphemes for automatic processing of the dictionary are developed, the main difficulties of morphological partitioning are considered. The algorithm of the automated morphological analysis which leans on formally expressed elements of morphemic partitioning is developed. In total it is allocated 5028 roots and 271 suffixes. The most productive and frequency roots and suffixes, rare, and also undescribed in scientific literature are revealed. Functional loading of separate suffixes, and also degree of a variety of a certain type of derivative words is defined.

Keywords: morphological structure of the word; Buryat language; statistics; root; suffix; return alphabetic frequency word book.