

---

## ЯЗЫКОЗНАНИЕ

УДК 811.512.37 (Калм)

### **КВАНТИТАТИВНЫЙ АНАЛИЗ КАЛМЫЦКОГО ГЕРОИЧЕСКОГО ЭПОСА «ДЖАНГАР»: ЧАСТОТНОСТЬ ЛЕММ**

© **Бембеев Евгений Владимирович**

кандидат филологических наук, старший научный сотрудник,  
Калмыцкий научный центр Российской академии наук  
Россия, г. Элиста, Республика Калмыкия  
E-mail: galdma@yandex.ru

В статье предпринята попытка квантитативного анализа языка текстов калмыцкого героического эпоса «Джангар». С помощью специальной лингвистической программы «TextAnalyzer» обработан и создан частотный список, в котором эксплицированы 28 текстов эпоса «Джангар». Общий список лемм корпуса текстов эпоса «Джангар» представлен 4445 единицами. Приведены «верхушки» (50 единиц) наиболее частотных лемм, где указаны их следующие статистические показатели: ранг (RANK), абсолютная частота (Частота). Частотные списки позволяют определить границы лексической системы языка эпоса, которая имеет свое ядро и периферийные поля (т. е. частотные и нечастотные элементы) как в целом в тексте эпоса «Джангар», так и на уровне подтекстов (авторских стилей), их распределение по текстам, а также во временном континууме (записи XIX и XX вв.).

**Ключевые слова:** эпос «Джангар»; квантитативный анализ; частотный словарь; леммы; калмыцкий язык; частотные элементы.

С развитием информационных технологий большой интерес приобретают квантитативные исследования языковых процессов. За счет автоматизации обработки текстов значительно расширяется база лексикостатистики, а одним из реальных оснований применения статистических методов в изучении языка и речи (текста) следует признать объективную присущность языку количественных признаков. Компьютерные программы позволяют создавать частотные списки определенного текста или корпуса текстов, а на этой основе уже можно количественно описать поведение различных языковых единиц.

В настоящее время остается открытой одна из чрезвычайно важных задач калмыцкой лексикологии — выяснение границ живой лексической системы современного калмыцкого языка. Мы не имеем отчетливых представлений об общей системе лексического состава языка, хотя предположение о наличии в ней общеупотребительного и периферического слоя не вызывает сомнений. Таким образом, использование квантитативных методов и составление частотных словарей, представляющих собой особым образом преобразованный текст, является крайне актуальным вопросом. Более того, частотные словари имеют важное значение для решения разных прикладных и исследовательских задач. С помощью такого рода словарей «можно попытаться ответить на вопросы: как много слов в языке (тексте), с какой интенсивностью они используются в речи, какие из них

предпочтительнее в той или иной сфере коммуникации у того или иного автора и т. д.» [1, с. 285]. Не меньшую ценность имеют и словники по отдельным жанрам, которые дают интересные сведения о качественном расслоении лексики.

В настоящей статье нами предпринята попытка количественного анализа языка текстов калмыцкого героического эпоса «Джангар», который веками слагался в устах рапсодов-сказителей (джангарчи) и прошел сложный путь развития от архаических форм к классическому героическому эпосу, подвергаясь различным трансформациям, напластованиям и переработкам. Главная идея создания частотного списка эпоса «Джангар» заключается в том, чтобы выявить употребительность лексических единиц в тексте на уровне лемм, которые могут дать ценные сведения историко-сравнительного характера по лексике калмыцкого языка, по определению границы лексической системы языка эпоса. Эти данные могут быть использованы для создания структурно-вероятностной модели языка эпоса, а также для решения проблем полисемии и омонимии лексики калмыцкого языка.

Материалом для исследования послужили тексты разновременных записей калмыцкого героического эпоса «Джангар». Исследователи эпоса «Джангар» насчитывают 28 глав (без учета вариантов), записанных в XIX-XX вв. у джангарчи той или иной исполнительской школы и объединенных в репертуарные циклы: Малодербетовский цикл (3 песни), Багацохуровский цикл (3 песни), репертуар Ээлян Овла (10 песен, магтал), Мукебюн Басангов (6 песен), Дава Шавалиев (4 песни, магтал), Насанка Балдыров (1 песня), Бадма Обушинов (1 песня). Все песни уникальны, т. к. записаны в разное время, у сказителей, принадлежащих разным школам исполнительского искусства и проживающих на различных территориях.

История бытования зафиксированных и сохранившихся песен калмыцкого героического эпоса «Джангар» была хронологически реконструирована и подробно изучена А. Ш. Кичиковым, который отмечал, что «Каждая исполнительская школа представлена в «Джангаре» известным количеством песен, объединенных в особый цикл, автономную версию» [2, с. 176]. Исследователь выявил основные ареалы бытования «Джангара» на территории исторического проживания родов калмыцкого этноса и связал их с бытованием основных версий (циклов) и сказительских школ:

1) I Малодербетовская версия (цикл), относимая к Барунской сказительской школе (Нойанакинский аймак Малодербетовского улуса, ныне с. Обильное Кетченеровского района), представителями которой был Поврам, его сын Санджи, сын Санджи Бука (1870–1920-е гг.). История записи и фиксации I Малодербетовской версии (цикла) связана с профессором К. Ф. Голстунским, посетившим калмыцкую степь в 1856 г. Рукопись на ясном письме «тодо бичиг» датируется 1862 г. и не содержит сведений о сказителе, от которого был зафиксирован текст. Известно, что запись организовал зайсанг (владелец) Нойанакинского аймака Джава-Дорджи Кутузов [2, с. 176].

2) Багацохуровская версия (цикл), относимая к Багацохуровской сказительской школе (Багацохуровский улус, ныне Юстинский район). Имена ранних сказителей данной версии неизвестны. По мнению А. Ш. Кичикова «в калмыцкой эпической традиции Багацохуровский цикл представляет раннюю торгутскую версию» [2, с. 167]. Две главы версии были записаны Ш. Санджирхаевым по

просьбе К. Ф. Голстунского в имени хошутовских князей Тюменей, а третья глава была записана с участием сотрудника Русского географического общества Н. И. Михайлова в Багацохуровском улусе [2, с. 167].

3) II Малодербетовский цикл Ээлян Овла (1857–1920 гг.), сказителя Ики-Бухусовской школы, представителями которой в XIX — начале XX вв. были Дэлтэр, Маргаш, Овла, Чюдэд, Окон, Муутал, Бульдин (женщина) [3, с. 122–127]. А. Ш. Кичиков, отмечая сказительскую школу Ээлян Овла, писал: «Она характеризуется устойчивым репертуаром, который передавался, по крайней мере, в семи поколениях с установкой на неизменность текста и точность его усвоения, на высокий профессионализм в исполнении и веру в несомненную достоверность повествований» [2, с. 168]. Репертуар Ээлян Овла был записан летом 1908 г. студентом Восточного факультета Санкт-Петербургского университета Номто Очи-ровым, что стало поистине эпохальным событием в эпосоведении и произвело сенсацию в научных кругах того периода [4, с. 52–54].

С ареалом проживания калмыков на Дону связано имя джангарчи Бадмы Обушинова. В 1901 г. собиратель калмыцкого фольклора И. И. Попов зафиксировал у него одну главу «Об Улан Хонгоре» героического эпоса «Джангар», а также отдельные фрагменты эпического памятника. В 1940 г., в преддверии празднования 500-летия эпоса «Джангар» были открыты имена талантливых джангарчи Мукёбюна Басангова и Давы Шавалиева, а вместе с ними и новые песни эпоса «Джангар» [2, с. 174].

Еще одна песня, которая вошла в свод 28 песен, была записана от джангарчи позднего периода Насанка Балдырова, являющегося прямым последователем творчества Ээлян Овла и его школы.

Все перечисленные 28 текстов песен эпоса «Джангар» включены в базу данных Национального корпуса калмыцкого языка и обработаны в специальной лингвистической программе «TextAnalyzer». Данная программа представляет собой электронный ресурс научного назначения, предназначенный для обработки текстового материала на калмыцком языке. Он выполнен в виде приложения для операционной системы MS Windows (.exe) и связанной с ней Базы данных (БД), где хранится информация об автоматическом разборе текстов на калмыцком языке<sup>1</sup>.

Как известно, частотные словари составляются с опорой на различные единицы счета: словоформы, лексемы, леммы, словосочетания. «Обычно противопоставляют друг другу две методики составления частотных словарей: за единицу словника принимается либо словоформа, либо лексема» [5]. Программа «TextAnalyzer» позволяет составить частотный список, выбрав в качестве единицы счета как словоформу, так и лемму (каноническую форму слова). В настоящей статье нами предпринята попытка создания частотного списка, где за элемент словника принимается лемма. Переход от словоформ к леммам осуществлен по правилам сведения графических (текстовых) слов к исходным формам слова, разработанным для Национального корпуса калмыцкого языка [6; 7]. Между тем необходимо отметить долю субъективности при создании правил

---

<sup>1</sup> Программа создана сотрудниками КалмНЦ РАН А. Ю. Каджиевым, В. В. Кукановой, Е. В. Бембеевым.

идентификации лемм калмыцкого языка, так как в настоящее время еще до конца не отработаны четкие и надежные алгоритмы преобразования текстовых слов в леммы и не составлена аналитическая грамматика современного калмыцкого языка.

Общий список лемм 28 текстов песен эпоса «Джангар» представлен 4445 единицами, включая имена собственные. Для сравнения, общий список словоформ корпуса текстов эпоса «Джангар» представлен 14366 единицами, т.е. лексем почти более чем в 3 раза меньше количества словоформ. В таблице №1 приведены «верхушки» (50 единиц) наиболее частотных лемм с показателями: каждой лемме приписан ранг (RANK), а также указана абсолютная частота (Частота) по всему тексту в целом, относительная частота (количество вхождений данной единицы на миллион словоформ (ipm) [8, с. 9].

Таблица 1

	Текст	RANK	Частота	IPM	Перевод
1.	гих	1	2416	15403	говорить
2.	болх	2	1914	12202	становиться
3.	бээх	3	1656	10557	быть
4.	һарх	4	1232	7854	выходить
5.	уг <sup>2</sup> (уга)	5	1200	7650	нет, не
6.	ирх	6	1187	7567	приходить
7.	хойр	7	1031	6573	два
8.	улан	8	995	6343	красный
9.	авх	9	901	5744	взять
10.	хан	10	881	5616	хан, правитель
11.	бий	11	868	5534	себя
12.	йовх	12	854	5444	идти
13.	Жаңһр	13	831	5298	имя собств. Джангар
14.	одх	14	762	4858	отправляться
15.	Хонһр	15	741	4724	имя собств. Хонгор
16.	хар	16	723	4609	черный
17.	арвн	17	686	4373	десять
18.	цаһан	18	656	4182	белый
19.	шар	19	637	4061	желтый
20.	көвүн	20	636	4054	мальчик
21.	эн	21	618	3940	этот
22.	би	22	588	3748	я
23.	долан	23	586	3736	семь
24.	мөңгн	24	574	3659	серебро, серебряный
25.	орх	25	570	3634	входить
26.	суух	26	565	3602	сидеть
27.	богд	27	528	3366	святой, император
28.	миңһн	28	523	3334	тысяча
29.	өгх	29	518	3302	давать
30.	негн	30	508	3238	один
31.	дөрвн	31	506	3226	четыре

<sup>2</sup> Автоматический разбор приводит для слова-отрицания *уга* лемму *уг*

32.	күн	32	493	3143	человек
33.	келх	33	491	3130	сказать
34.	баатр	34	488	3111	богатырь
35.	оркх	35	486	3098	ставить
36.	көк	36	473	3015	синий, зеленый
37.	орн	37	470	2996	страна
38.	тавн	38	468	2983	пять
39.	күрх	39	450	2869	достигать, доходить
40.	күлг	40	430	2741	боевой конь
41.	зеерд	41	397	2531	рыжий (масть коня)
42.	хурвн	42	397	2531	три
43.	сээхн	42	397	2531	красивый, красиво
44.	мөрн	42	395	2518	лошадь
45.	деер	45	393	2505	вверху, высоко, на
46.	тер	46	391	2492	тот
47.	һазр	47	384	2448	земля
48.	алдр	48	381	2429	великий
49.	зандн	49	368	2346	сандал, сандаловый
50.	тал	50	360	2295	сторона, грань

Из таблицы №1 мы можем наблюдать, что наиболее частотной леммой в языке эпоса «Джангар» является речевой или дицендиальный глагол *гих* «говорить, сказывать», что вполне закономерно. Благодаря своей полисемантической, разнообразным формам глагола *гих* выступают не только в качестве элементов, соединяющих основную массу сложноподчиненных предложений, но и способны выступать в составе устойчивых выражений (фразеологизмов), в качестве изменяемых служебных слов при непроемных глаголах с наречным значением, как утвердительная частица, подчеркивающая объективность повествования и др.

Следующими по частотности леммами в языке эпоса «Джангар» зафиксированы глаголы *болх* «становиться», *бәәх* «стать, становиться», *авх* «брать». Частотность их употребления объясняется тем, что, помимо собственно функции основного глагола, они часто выступают как вспомогательные глаголы-связки при деепричастных и причастных формах. Отметим, что в список наиболее частотных единиц также попали глаголы движения *һарх* «выходить» и *ирх* «приходить».

Лемма *уга* (постпозиционное слово-отрицание *уга*) «не, без, нет» также является наиболее употребительной единицей. В калмыцком языке отрицание *уга*, выражая значение отсутствия чего-либо, характеризуется весьма широким кругом употребления. Слово-отрицание *уга* употребляется с именами, причастиями (многократным, прошедшего времени, будущего времени), разделительным деепричастием [9, с. 293].

Отметим, что в список частотных лемм входит числительное 2 «хойр», которое помимо основного значения, выражающее понятие отвлеченного числа или количества предметов, может употребляться в функции сочинительно-соединительного союза *хойр* ‘и’, производного путем конверсии числительного “два”.

Среди частотных единиц можно обнаружить цветообозначающую лемму *улан* «красный». Как справедливо отмечает профессор Г. Ц. Пюрбеев «принцип цветообозначения играет очень важную роль в фольклорном тексте, особенно в эпическом» [10, с. 73]. Это подтверждается тем, что все пять основных лемм (*улан* “красный”, *цаһан* “белый”, *хар* “черный”, *шар* “желтый”, *көк* “синий/зеленый”), обозначающие основную цветовую гамму «в палитре этнической картины мира всех монголоязычных народов» входят в «верхушку» 50 наиболее частотных единиц [11, с. 76].

Замыкает десятку наиболее частотных единиц — лексема *хан* «хан, царь; царский», присущая эпическим и сказочным произведениям калмыцкого фольклора.

Таким образом, в данной работе представлены предварительные итоги распределения частотных характеристик «верхушек» (50) лемм в языке героического эпоса «Джангар». Анализ ранговых списков лемм дает сведения по лексической системе эпического памятника, выявляет границы «ядра» и периферии лексических единиц как в целом в тексте эпоса «Джангар», так и на уровне подтекстов (авторских стилей), их распределение по текстам, а также во временном континууме (записи XIX и XX вв.). В аспекте практической значимости создание частотных списков можно широко использовать в прикладных целях. К примеру, при обучении языку отдавать предпочтение наиболее частотным единицам, отмечая при этом значения многозначных и однозначных лексем, их поведение в письменном тексте, сочетаемость с другими единицами и т.п.

#### Литература

1. Долинский, В. А. Квантитативная лингвистика в исследовании текста [Текст] / В. А. Долинский // Алфавит: Стрoение повествовательного текста. Синтагматика. Парадигматика. Смоленск: СГПУ, 2004. С. 283–324.
2. Кичиков А. Ш. Героический эпос «Джангар». Сравнительно-типологическое исследование памятника. Изд. 3-е. М.: Вост. лит., 1997. 320 с.
3. Кичиков А. Ш. Исследование героического эпоса “Джангар” (Вопросы исторической поэтики). Элиста: Калм. кн. изд-во, 1976. 156 с
4. Очиров Н.О. О записи оригинала «Джангара» // Ученые записки КНИИЯЛИ. Вып. 5. Сер. филол. Элиста, 1967. С. 52–54. (публ. А. Ш. Кичикова).
5. Частотный словарь русского языка (под редакцией Л. Н. Засориной) [Электронный ресурс] — Режим доступа: <http://project.phil.spbu.ru/lib/data/slovari/zasorina/zasorina.html>. — (дата обращения: 20.10.2017)
6. Куканова В. В. Национальный корпус калмыцкого языка: архитектура и возможности использования [Текст] / В. В. Куканова, Е. В. Бембеев, Н. М. Мулаева, Н. Ч. Очирова // Вестник Калмыцкого института гуманитарных исследований РАН. 2012. № 3. С. 138–150.
7. Национальный корпус калмыцкого языка (НККЯ) [Электронный ресурс] — Режим доступа: <http://kalmcoyropa.ru>. — (дата обращения: 20.10.2017)
8. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009. 1112 с.
9. Грамматика калмыцкого языка [Текст]: фонетика и морфология / П. Ц. Биткеев [и др.] отв. ред. Г. Д. Санжеев. Элиста: Калм. кн. изд-во, 1983. 335 с.
10. Пюрбеев, Г. Ц. Эпос «Джангар» [Текст]: культура и язык (=Жаңһр дуулвр: сойл болн келн) на русском и калмыцком языках / Г. Ц. Пюрбеев. Изд. 2-е, перераб. Элиста: ЗАОр «НПП “Джангар”», 2015. 280 с.

11. Жуковская, Н. Л. Общомонгольские культурные традиции в эпосе «Джангар» [Текст] / Н. Л. Жуковская // «Джангар» и проблемы эпического творчества: тез. докл. и сообщ. междунар. науч. конф., г. Элиста. 22–24 авг. 1990 г. Элиста, 1990. С. 74–76.

QUANTITATIVE ANALYSIS OF THE KALMYK HEROIC EPOS «DJANGAR»:  
FREQUENCY OF LEMMAS

*Evgrbiy V. Bembeev*

Ph.D. of Philology, senior research worker

«Kalmyk Scientific Center of the Russian Academy of Sciences»

Elista, Kalmyk Republic, Russia

E-mail: galdma@yandex.ru

The article attempts to quantitatively analyze the language of the texts of the Kalmyk heroic epic "Djanganar". With the help of the special linguistic program "TextAnalyzer", a frequency list has been processed and created, in which 28 texts of the epic "Djanganar" are explicated. The general list of lemmas of the body of texts of the epic "Djanganar" is represented by 4445 units. The "tops" (50 units) of the most frequent lemmas are given, where their statistical indicators are indicated: rank (RANK), absolute frequency (Frequency). Frequency lists allow us to determine the boundaries of the lexical system of the language of the epic, which has its core and peripheral fields (ie, frequency and non-frequency elements).

*Keyword:* epic «Djanganar»; quantitative analysis; frequency dictionary; lemma; Kalmyk language; frequency elements.