

1. СОЦИОЛИНГВИСТИКА, ЭТНОЛИНГВИСТИКА, ПСИХОЛИНГВИСТИКА КАК СОСТАВЛЯЮЩИЕ ЛИНГВОКУЛЬТУРНЫХ ЗНАНИЙ

УДК 811.512.37 (Калм)

КВАНТИТАТИВНЫЙ АНАЛИЗ ЯЗЫКА КАЛМЫЦКОГО ГЕРОИЧЕСКОГО ЭПОСА «ДЖАНГАР»

Бембеев Е. В.

Федеральное государственное бюджетное учреждения науки «Калмыцкого научного центра
Российской академии наук», г. Элиста, Россия
e-mail: galdma@yandex.ru

В статье предпринята попытка квантитативного анализа языка текстов калмыцкого героического эпоса «Джангар», который занимает особое место в духовной жизни калмыцкого и ойратского этносов. С помощью компьютерных технологий создан частотный список, в котором эксплицированы 28 текстов эпоса «Джангар». Общий список словоформ корпуса текстов эпоса «Джангар» представлен 14366 единицами, включая имена собственные. Приведены «верхушки» (50 единиц) наиболее частотных словоформ, где указаны их следующие статистические показатели: ранг (RANK), абсолютная частота (Частота), общая частота употребления на миллион слов условного корпуса (ipm), показатель для определения количества текстов, в которых встретилось слово (R (%)). Квантитативный подход позволяет классифицировать тексты эпоса «Джангар» в соответствии с авторами и школами, их стилями исполнения и временными данными, в рамках которых эти тексты были записаны. Так как различия между ними могут иметь в основном статистический характер, то можно описать и классифицировать все тексты эпоса «Джангар» на строго объективной базе. Частотные списки позволяют определить границы лексической системы языка эпоса, которая имеет свое ядро и периферийные поля (т. е. частотные и нечастотные элементы).

Ключевые слова: эпос «Джангар», квантитативный анализ, частотный словарь, словоформы, калмыцкий язык

ВВЕДЕНИЕ

В настоящее время с развитием информационных технологий большой интерес приобретают исследования и экспликация языковых явлений с помощью методов количественной математики, так называемой квантитативной лингвистики. Одним из реальных оснований применения статистических методов в изучении языка и речи (текста) следует признать объективную присущность языку количественных признаков. Новые компьютерные программы позволяют создавать частотные списки, в которых эксплицирована структурно-вероятностная модель того или иного языка,

определенного текста и т.д., что позволяет количественно описать поведение различных языковых единиц в письменном тексте, частоту употребления данных единиц, их распределение в текстах разного жанра, сочетаемость с др. единицами и т. п. «Одновременно накапливается обобщенная количественная информация о классах единиц, о языковых конструкциях... Такая информация углубляет описание единиц языка» [1, с. 231]. Одной из важнейших задач квантитативной лингвистики является составление частотных словарей, необходимость использования которых для решения прикладных и исследовательских задач несомненна. Такие словари представляют собой особым образом преобразованный текст. Как отмечает в своей работе В. А. Долинский, словарь подобного рода «...включает в себя упорядоченный список слов или других языковых единиц (словоформы, словосочетания), которые зарегистрированы составителем в обследованном им тексте, фрагменте текста или корпусе текстов и снабжены данными о частоте их употребления в тексте (речи). С его помощью 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.» [2, с. 285].

ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА

Создание частотных словарей на материале монгольских языков связаны с разработкой и созданием Генерального корпуса современного монгольского языка и Национального корпуса калмыцкого языка [3; 4; 5; 6]. Как отмечает С. А. Крылов «...на материале ГКМЯ впервые в монголистике монгольский язык был подвергнут изучению в его количественном аспекте. Это было сделано путём создания частотных словарей (ЧС) монгольского языка» [3, с. 48]. Актуальность создания частотных словарей заключается в том, что они передают (в том или ином виде) информацию о частотности употребления инвентаризуемых в них единицах, где наиболее частотная единица является наиболее значимой в лексической системе данного языка (текста), а нечастотные элементы свидетельствует о ее малоупотребительности. Между тем эти лексические единицы могут содержать

следы исторического развития, что является важным фактором при проведении сравнительно-исторических и типологических исследований.

В настоящей статье нами предпринята попытка количественного анализа языка текстов калмыцкого героического эпоса «Джангар», который занимает особое место в духовной жизни калмыцкого и ойратского этносов, являясь сокровищницей традиций, отражающих самосознание и мироощущение кочевого народа. «Джангар» слагался и воссоздавался веками в устах рапсодов-сказителей (джангарчи) и прошел сложный путь развития от архаических форм к классическому героическому эпосу, подвергаясь различным трансформациям, напластованиям и переработкам. Исследователи калмыцкого героического эпоса «Джангар» насчитывают двадцать восемь 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. 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 языке¹. Тексты включены в базу данных Национального корпуса калмыцкого языка и обработаны в специальной лингвистической программе «TextAnalyzer» [7; 8].

¹ Программа «Textanalyzer for the Kalmyk Language». Свидетельство о государственной регистрации программы ЭВМ №2016617062. Авторы: сотрудники КалмНЦ РАН А.Ю.Каджиев, В.В. Куканова, Е.В. Бембеев.

Прежде чем приступить к обработке данных, необходимо отметить, что анализу подвергаются не лексемы, а словоформы, обладающие реальной частотностью в языке текста. Интересно также отметить, что составители частотных словарей отмечают привычность основной словарной единицы – лексемы, а также тот факт, что при сведении словоформ в лексемы лингвисты могут использовать разные принципы, что приводит к более высокой доле субъективности в количественных показателях по лексемам по сравнению с количественными данными по словоформам. Между тем по оценке Л. Лёнгрена: «...количественные языковые факты, опирающиеся только на уровень словоформ, являются более объективными и надёжными» [9, с. 28–29]. Таким образом, в качестве единицы описания выступает слово «от пробела до пробела» в той грамматической форме, в которой она употреблена. Программа «TextAnalyzer» позволяет составить частотный список, выбрав в качестве единицы счета как словоформу, так и лемму (каноническую форму слова).

За основу создания нашего словаря принят Частотный словарь, основанный на материале Национального корпуса русского языка [Ляшевская, Шаров 2009]. Общий список словоформ корпуса текстов эпоса «Джангар» представлен 14366 единицами, включая имена собственные. Каждой словоформе приписан ранг, а также указана абсолютная частота по всем текстам эпоса в целом.

В таблице №1 приведены «верхушки» (30 единиц) наиболее частотных словоформ, где указаны их следующие статистические показатели:

1. Словоформы (приведены по принципу убывания общей частоты встречаемости словоформ).

2. Ранг (R), который присваивается в зависимости от частоты словоупотребления в текстах. Т.е. какое место занимает данная единица в ряду ей подобных, упорядоченному по принципу убывания числа вхождений данной единицы в состав корпуса.

3. Абсолютная частота словоупотребления. Т.е. общее количество вхождений единицы в состав корпуса.

4. Ipm (instances per million words), общая частота употребления на миллион слов условного корпуса.

5. Показатель R (%) для определения количества текстов, в которых встретилось слово.

6. Переводы словоформ на русский язык (в его основном значении).

Таблица 1. «Верхушка» частотного списка словоформ, отсортированного по абсолютной частоте и рангу

	1	2	3	4	5	6
	Текст	RANK	Частота	IPM	R(%)	Перевод
	уга	1	1200	7650	100	постпозитивная частица отрицания “не, нет”
	хойр	2	1036	6605	100	два
	улан	3	948	6044	100	красный
	гиж	4	881	5616	100	соед. деепр. от гих “говорить” что [изъясн. союз]
	цаһан	5	872	5559	96	белый
	хар	6	720	4590	100	черный
	арвн	7	684	4360	96	десять
	гиһэд	8	659	4201	100	разд. деепр. от гих “говорить” что [изъясн. союз]
	мөңгн	9	555	3538	100	серебро; серебряный
	долан	10	547	3487	100	семь
	авад	11	543	3461	100	разд. деепр. от авх “брат”
	шар	12	530	3379	93	желтый
	алтн	13	495	3155	100	золото
	дөрвн	14	488	3111	100	четыре

КВАНТИТАТИВНЫЙ АНАЛИЗ ЯЗЫКА КАЛМЫЦКОГО ГЕРОИЧЕСКОГО ЭПОСА «ДЖАНГАР»

	эн	15	487	3104	100	этот
	көк	16	435	2773	96	зеленый
	һарад	17	431	2747	100	разд. деепр. от һарх “идти”
	Хоңһр	18	430	2741	96	имя собст. Хонгор
	көвүн	19	425	2709	96	мальчик
	миңһн	20	420	2677	79	тысяча
	нег	21	419	2671	96	вин. падеж от негн “один”
	болад	22	416	2652	100	разд. деепр. от болх “становиться, быть”
	ирв	23	404	2575	96	пришел
	сээһн	24	397	2531	96	красивый; хороший
	алдр	25	381	2429	89	слава; славный; великий
	һурвн	26	372	2371	89	три
	одв	27	369	2352	100	ушел
	тавн	28	367	2339	96	пять
	Жаңһр	29	364	2320	96	имя собст. Джангар
	богд	30	350	2231	96	святейший

По статистическим характеристикам можно определить с заданной относительной погрешностью ту часть словника, в которую входят слова с высокой частотой появления, что позволяет обозначить границы лексической системы языка эпоса, которая имеет свое ядро и периферийные поля (т. е. частотные и нечастотные элементы). Данные показывают, что количество словоформ, имеющих употребление $>$ или $= 10$, насчитывается 1758 единиц, количество словоформ, имеющих употребление $>$ или $= 100$ — 150 лексических единиц. А количество словоформ, употребление которых в тексте эпоса < 5 раз насчитывает 8986 единиц.

Рассмотрим лексические единицы, входящие в первый десяток наиболее частотных «верхушек». Самой частотной единицей на материале эпоса «Джангар» является постпозиционное слово-отрицание уга (не, без, нет), что не безосновательно. В калмыцком языке отрицание уга, выражая значение отсутствия чего-либо, характеризуется весьма широким кругом употребления [11, с. 293; 12, с. 119; 13]. Слово-отрицание уга употребляется с именами, причастиями (многократным, прошедшего времени, будущего времени), разделительным деепричастием. Сочетаясь с именами существительными в атрибутивной позиции, выступает как аффикс, образующий прилагательные со значением отсутствия какого-либо признака, качества, свойства у предмета. Как субстантивное слово, оно склоняется и в предложении выполняет функцию самостоятельного члена предложения, например сказуемого «при отрицании экзистенциальных предложений» [11, 293–294; 14, с. 696].

Следующей абсолютно частотной словоформой в языке эпоса «Джангар» является числительное 2 “*хойр*” (1036 раз), которое помимо основного значения, выражающее понятие отвлеченного числа или количества предметов, может употребляться в функции обобщающего или союзного слова. Например: «*Ирн келн бээнэ: “Асхндаң хаана күүкн күргн хойр хойр бөк ноолдулн гижэ бээнэ”*» – «Придя молвил: “Вечером ханская дочь **и** зять хотят провести борцовский поединок между **двумя** борцами”» [Хондрин гер авлһна бөлг. Ээлян Овла]. В данном примере в первом случае “күүкн күргн **хойр**” («ханская дочь **и** зять») числительное 2 “хойр” выступает в роли союзного слова, во втором случае — как количественное числительное “**хойр бөк**” (два борца). Помимо числительного 2 в первую десятку высокочастотных словоформ эпоса также входят числительные 10 “арвн” (684 раза) и 7 “долан” (547 раз), связанные с древними мифологическими представлениями монголоязычных народов и обладающие сакральной семантикой. Эти данные отчасти подтверждаются материалами Б. Х. Тодаевой, выявленные ею на материале 12 песен эпоса [15, с. 177-513].

Помимо числовых словоформ в список наиболее частотных единиц текста эпоса входят словоформы, обозначающие цветовую символику. Как справедливо отмечает

профессор Г. Ц. Пюрбеев «принцип цветообозначения играет очень важную роль в фольклорном тексте, особенно в эпическом. В языках монгольских народов цветовой признак положен в основу видовых названий животных, пищи и пищевых продуктов, злаков и растений, различной качественной характеристики предметов и социального противопоставления людей» [16, с. 73]. Так на третьей позиции по частоте употреблений находится словоформа “улан” (красный), на пятой — “цаһан” (белый), на шестой — “хар” (черный). Две другие словоформы, обозначающие цветовую гамму и отличающиеся «высокими коннотациями в палитре этнической картины мира всех монголоязычных народов» также входят в число наиболее употребительных и частотных словоформ эпоса «Джангар»: “шар” (желтый) на 12 позиции, “көк” (синий/зеленый) на 16 месте по абсолютной частоте употребления [17, с. 76].

В калмыцком языке, как и во всех других монгольских языках, речевой или дицендиальный глагол *гих* играет исключительно важную роль. Основная масса придаточных предложений соединяется с главным с помощью разнообразных форм глагола *гих* «говорить, сказывать», а его причастные и деепричастные формы благодаря своей полисемантичности способны выступать не только в функции связи частей сложноподчиненных предложений, но и в качестве служебных слов, выполняющих иные функции. Это положение полностью подтверждается на материале текста эпоса «Джангар». Четвертую и восьмую позиции по частоте употребления занимают словоформы “гиж” (говоря) и “гihэд” (говорив, сказав). В калмыцком языке словоформа *гиж* является грамматическим омонимом и может выступать как глагол прошедшего результативного времени, соединительное деепричастие от глагола *гих* ‘говорить’, как союз (в том числе в составе союзного слова) в сложноподчинительных предложениях, как компонент устойчивых выражений: лексических и синтаксических фразеологизмов, как изменяемое служебное слово при производных глаголах с наречным значением, как утвердительная частица, подчеркивающая объективность повествования и др. Разделительное деепричастие ‘*гihэд*’ также выступает как союз (в том числе в составе союзного слова) в сложноподчинительных предложениях, как компонент

устойчивых выражений: лексических и синтаксических фразеологизмов, как изменяемое служебное слово при производных глаголах с наречным значением.

В десятку наиболее частотных словоформ входит и слово *мөңгн* “серебро; серебряный”, обозначающее названия предметов, сделанных из серебра или покрытых серебром. Как отмечает Г. Ц. Пюрбеев серебро является «самым почитаемым металлом, символизирующим чистоту, святость и благородство. В эпосе насчитывается свыше двадцати названий серебряных вещей, из них не менее десяти связаны с экипировкой всадника и деталями конской сбруи: серебряное стремя (*мөңгн дөрә*), серебряные удила (*мөңгн ууд*), серебряная седельная лука (*мөңгн бүүрг*), серебряная рукоять плети (*мөңгн иш*); украшенное серебром седло (*мөңгн эмэл*), узда (*мөңгн хазар*), поводья (*мөңгн цулвр*) нагрудный ремень (*мөңгн көмлдрһ*), потник (*мөңгн тохм*), подпотник (*мөңгн делтр*), серебряный столик (*мөңгн ширә*), серебряная дверь (*мөңгн үүдн*), серебряная пробка (*мөңгн бөглә*), серебряная чашка (*мөңгн ааһ*) [15, с. 51-52].

Лингвостатистический метод позволяет не только количественно описывать поведение различных языковых единиц (букв, морфем, слов и т. п.) в письменном тексте, но и дает почву для смыслового анализа лексики. Изучение словарных групп и семантических полей показывает, что лексические «пучки» поддерживаются семантическими связями, которые концентрируются вокруг лексем с наиболее общим значением.

2. «Верхушка» частотного списка словоформ, отсортированного по показателю $R(\%)$

Показатель $R(\%)$ указывает на то, какое место занимает та или иная единица в ряду ей подобных, упорядоченному по принципу убывания числа текстов, содержащих данную единицу, т.е. насколько часто данная единица используется в разных текстах эпоса «Джангар».

КВАНТИТАТИВНЫЙ АНАЛИЗ ЯЗЫКА КАЛМЫЦКОГО ГЕРОИЧЕСКОГО ЭПОСА «ДЖАНГАР»

	1	2	3	4	5	6
	Текст	R(%)	Частота	IPM	RANK	перевод
1	авад	100	543	3461	11	<i>разд. деепр. от</i> авх “братъ”
2	барун	100	280	1785	45	правый
3	болад	100	416	2652	22	<i>разд. деепр. от</i> болх “становиться”
4	гиһэд	100	659	4201	8	<i>разд. деепр. от</i> гих “говорить”
5	долан	100	547	3487	10	семь
6	мөңгн	100	555	3538	9	серебро; серебряный
7	хар	100	720	4590	6	черный
8	хойр	100	1036	6605	2	два
9	эс	100	295	1880	42	препозитивная частица отрицания “не, нет”
10	алтн	100	495	3155	13	золото; золотой
11	гиж	100	881	5616	4	<i>соед. деепр. от</i> гих “говорить”
12	деерэ н	100	235	1498	56	послелог
13	дөрвн	100	488	3111	14	четыре
14	ирэд	100	285	1817	44	<i>разд. деепр. от</i> ирх “приходить”

	1	2	3	4	5	6
15	одв	100	369	2352	27	ушел, отправился
16	орад	100	188	1198	73	<i>разд. деепр. от</i> орх “входить”
17	уга	100	1200	7650	1	постпозитивна я частица отрицания “не, нет”
18	улан	100	948	6044	3	красный
19	эн	100	487	3104	15	Это, этот
20	һарад	100	431	2747	17	<i>разд. деепр. от</i> һарх “выходить”
21	ар	96	212	1351	66	север; северный

«Верхушка» частотного списка словоформ, отсортированного по показателю R(%) показывает, что первые 20 словоформ, используемые во всех 28 текстах эпоса «Джангар». Это говорит о том, что данные единицы носят универсальный характер или присущи текстам определенного жанра. Однако эти единицы не всегда коррелируют с «верхушками» частотного списка словоформ, отсортированного по абсолютной частоте. Например, словоформа бөлг “глава” встретилась 28 раз, ровно столько раз, сколько мы имеем текстов, в виду чего она имеет 100% по показателю R(%). Следует подчеркнуть, что употребительность слова не определяется только лишь показателем R(%). Поэтому для оценки употребительности слова необходимо использовать разные параметры и способы оценки употребительности.

ВЫВОДЫ

Таким образом, приведенные количественные показатели можно использовать для создания структурно-вероятностной модели языка эпоса «Джангар». Во-первых, квантитативный подход позволяет классифицировать сами тексты эпоса «Джангар» в соответствии с авторами и школами, их стилями исполнения и временными данными, в рамках которых эти тексты были записаны. Так как различия между ними могут иметь в основном статистический характер, то можно описать и классифицировать все тексты эпоса «Джангар» на строго объективной базе. Во-вторых, частотные списки позволяют определить границы лексической системы языка эпоса, которая имеет свое ядро и периферийные поля (т. е. частотные и нечастотные элементы). Статистические характеристики дают ценные сведения о семантических полях лексических единиц. В-третьих, структурно-вероятностная модель языка отличается большей реалистичностью, а «верхушки» языковых единиц дают много ценных сведений, которые предоставляют большой интерес для проведения диахронических и сравнительно-сопоставительных исследований с данными других монгольских языков. В-четвертых, в аспекте практической значимости создание частотных списков можно широко использовать в прикладных целях. 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. Статистические данные могут быть широко использованы и при решении других лингвистических задач. Так, обратный статистический словарь даст материалы для анализа словообразования и определения активных средств словообразования калмыцкого языка на материале эпоса «Джангар».

Список литературы

1. Шайкевич, А. Я. Количественные методы в языкознании [Текст] / А. Я. Шайкевич // Лингвистический энциклопедический словарь. – Москва: Советская энциклопедия, 1990. – С. 231.

2. Долинский, В. А. Квантитативная лингвистика в исследовании текста [Текст] / В. А. Долинский // Алфавит: Строение повествовательного текста. Синтагматика. Парадигматика. – Смоленск: СГПУ, 2004. – С. 283–324.
3. Крылов, С. А. Опыт изучения современного монгольского языка в количественном аспекте [Текст] / С. А. Крылов // Вопросы языкознания. – 2013. – № 5. – С. 46–57.
4. 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.
5. Генеральный корпус современного монгольского языка (ГКМЯ) [Электронный ресурс] – Режим доступа: http://web-corpora.net/MongolianCorpus/search/index.php?interface_language=ru. – (дата обращения: 20.10.2017)
6. Национальный корпус калмыцкого языка (НККЯ) [Электронный ресурс] – Режим доступа: <http://kalmcorpora.ru>. – (дата обращения: 20.10.2017)
7. Бембеев, Е. В. Опыт квантитативной обработки текста на старокалмыцком языке: количественные характеристики [Текст] / Е. В. Бембеев // Вестник Калмыцкого института гуманитарных исследований РАН. – 2012. – № 2. – С. 163–168.
8. Бембеев, Е.В., Куканова В.В., Каджиев А.Ю. Частотный словарь современного калмыцкого языка: правила анализа текстового материала [Текст] / Е. В. Бембеев, В. В. Куканова, А. Ю. Каджиев // Вестник Калмыцкого института гуманитарных исследований РАН. – 2014. – № 3. – С. 128–141.
9. Лённгрэн, Л. Частотный словарь современного русского языка [Текст] / The Frequency Dictionary of Modern Russian / Lonngren, Lennart. Acta Univ. Ups., Studia Slavica Upsaliensia Uppsala 32]. – Uppsala, 1993. – 188 с.
10. Ляшевская, О. Н. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка) [Текст] / О. Н. Ляшевская, С. А. Шаров. – Москва: Азбуковник, 2009. – 1112 с.
11. 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.
12. Бембеев, Е.В. К вопросу о частицах калмыцкого языка (историко-сравнительный аспект) [Текст] / Е. В. Бембеев // Вестник Калмыцкого института гуманитарных исследований РАН. – 2011. – №1. – С. 117–123.

13. Баранова, В. В. Отрицание в калмыцком языке: изменение системы и распределение форм [Электронный ресурс] // Системные изменения в языках России, СПб., 16-18 октября 2014. – Режим доступа: <http://iling.spb.ru/conf/syschanges2014/materials/baranova.pdf>. – (Дата обращения: 20.10.2017).
14. Сай, С. С. Грамматический очерк калмыцкого языка [Текст] / С. С. Сай // Acta Linguistica Petropolitana. Труды Института лингвистических исследований. Т. V. Ч. 2. Исследования по грамматике калмыцкого языка. – СПб.: Наука, 2009. – С. 622–709.
15. Тодаева, Б. Х. Опыт лингвистического исследования эпоса “Джангар” [Текст] / Б. Х. Тодаева. – Элиста: Калм. кн. изд-во, 1976. – 530 с.
16. 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.
17. Жуковская, Н. Л. Общемонгольские 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 LANGUAGE OF KALMYK HEROIC EPOS DJANGAR

Bembeev E.

**Kalmyk Scientific Center of the Russian Academy of Sciences, Elista,
Kalmyk Republic, Russian Federation
e-mail: galdma@yandex.ru**

Summary. The article attempts to analyze quantitatively the language of the texts of Kalmyk heroic epic Djangar which occupies a special place in the spiritual life of the Kalmyk and Oirat ethnoses. With the help of computer technology, a frequency list has been created, in which 28 texts of the epic Djangar are explicated. The general list of word forms of the body of texts of the epic Djangar is represented by 14366 units, including proper names. The "tops" (50 units) of the most frequent word forms are given, where their statistical indicators are indicated: rank (RANK), absolute frequency (Frequency), total frequency per million words of the conditional case (ipm), an indicator for determining the number of texts in which a particular word is used (R (%)). The quantitative approach allows to classify the texts of the epic Djangar in accordance with the authors and schools, their performance styles and the time data within which these texts were written down. Since the differences between them can have a statistical nature, it is possible to describe and classify all the texts of the epic Djangar on a strictly objective basis. The frequency lists allow us to define the boundaries of epic lexical system of the language which has its core and peripheral fields (i.e. the frequency and nonfrequency elements).
Keywords: epic Djangar, quantitative analysis, frequency dictionary, word forms, Kalmyk language

References

1. Shaikovich A. Ya. *Kolichestvennye Metody v Yazykoznanii* [Quantitative Methods in Linguistics]. Linguistic Encyclopaedic Dictionary. M.: Sovetskaya Entsiklopediya Publ., 1990. 231 p.
2. Dolinskii V. A. *Kvantitativnaya Lingvistika v Issledovanii Teksta* [Quantitative Linguistics in Text Research]. *Alfavit: Stroenie povestvovatel'nogo teksta. Sintagmatika. Paradigmatika*. Smolensk: SGPU Publ., 2004. Pp. 283–324.
3. Krylov S. A. *Opyt Izucheniya Sovremennogo Mongol'skogo Yazyka v Kolichestvennom Aspekte* [A Quantitative Study of Mongolian Language]. *Voprosy Yazykoznanii*, 2013, no 5, pp. 46–57.
4. Kukanova V. V., Bembeev E. V., Mulaeva N. M., Ochirova N. Ch. *Natsional'nyi Korpus Kalmytskogo Yazyka: Arkhitektura i Vozmozhnosti Ispol'zovaniya* [National Corpora of the Kalmyk Language: Architecture and Prospects of Usage]. *Vestnik Kalmytskogo Instituta Gumanitarnykh Issledovanii RAN*. 2012, no 3, pp. 138–150.
5. *General'nyi Korpus Sovremennogo Mongol'skogo Yazyka* [General Corpora of the Mongolian Language]. Available at: http://web-corpora.net/MongolianCorpus/search/index.php?interface_language=ru. (accessed 20 October 2017).
6. *Natsional'nyi Korpus Kalmytskogo Yazyka* [National Corpora of the Kalmyk Language]. Available at: <http://kalmcorporu.ru>. (accessed 20 October 2017).
7. Bembeev E. V. *Opyt Kvantitativnoi Obrabotki Teksta na Starokalmytskom Yazyke: Kolichestvennye Kharakteristiki* [Experience of Quantitative Approach to Text Processing in Old-Kalmyk Language: Quantitative Characteristics]. *Vestnik Kalmytskogo Instituta Gumanitarnykh Issledovanii RAN*, 2012, no 2, pp. 163–168.
8. Bembeev E.V., Kukanova V.V., Kadzhiev A.Yu. *Chastotnyi Slovar' Sovremennogo Kalmytskogo Yazyka: Pravila Analiza Tekstovogo Materiala* [Frequency Dictionary of the Modern Kalmyk Language: Rules of Analysis of Text Material]. *Vestnik Kalmytskogo Instituta Gumanitarnykh Issledovanii RAN*, 2014, no 3, pp. 128–141.
9. Lenngren L. *Chastotnyi Slovar' Sovremennogo Russkogo Yazyka* [The Frequency Dictionary of Modern Russian]. Uppsala: Acta Univ. Ups. Publ., 1993. 188 p.

10. Lyashevskaya O. N., Sharov S. A. *Chastotnyi Slovar' Sovremennogo Russkogo Yazyka* [The Frequency Dictionary of Modern Russian]. Moskva: Azbukovnik Publ., 2009. 1112 p.
11. *Grammatika Kalmytskogo Yazyka: Fonetika i Morfologiya* [Grammar of the Kalmyk Language: Phonetics and Morphology]. Elista: Kalmytskoe Knizhnoe Izdatel'stvo Publ., 1983. 335 p.
12. Bembeev E.V. *K Voprosu o Chastitsakh Kalmytskogo Yazyka (Istoriko-Sravnitel'nyj Aspekt)* [On the Particles of Kalmyk Language (Historical and Comparative Aspect)]. *Vestnik Kalmyckogo instituta gumanitarnyh issledovanij RAN*, 2011, no 1, pp. 117–123.
13. Baranova V. V. *Otritsanie v Kalmytskom Yazyke: Izmenenie Sistemy i Raspredelenie Form* [Negation in the Kalmyk Language: Change of System and Distribution of Forms]. *Sistemnye Izmeneniya v Yazykakh Rossii*, 2014. Available at: <http://iling.spb.ru/confs/syschanges2014/materials/baranova.pdf>. (accessed 20 October 2017).
14. Saj S. S. *Grammaticheskii Oчерk Kalmytskogo Yazyka* [Grammar of the Kalmyk Language]. *Acta Linguistica Petropolitana*. T. V. Part 2. *Issledovaniya po Grammatike Kalmytskogo Yazyka*. Saint-Petersburg: Nauka Publ., 2009, pp. 622–709.
15. Todaeva B. H. *Opyt Lingvisticheskogo Issledovaniya Eposa "Djanganr"* [Experience in Linguistic Research of the Epic *Djanganr*]. Elista: Kalmytskoe Knizhnoe Izdatel'stvo Publ., 1976. 530 p.
16. Pyurbeev G. C. *Epos "Djanganr": Kul'tura i Yazyk* [Epic *Djanganr*: Culture and Language]. Elista: Dzhangar Publ., 2015. 280 p.
17. Zhukovskaya N. L. *Obshchemongol'skie Kul'turnye Traditsii v Epose "Djanganr"* [Mongolian Cultural Traditions in the Epic *Djanganr*]. *Dzhangar i Problemy Epicheskogo Tvorchestva*. Elista: Dzhangar Publ., 1990, pp. 74–76.