

© 2013 г. С.А. КРЫЛОВ

ОПЫТ ИЗУЧЕНИЯ СОВРЕМЕННОГО МОНГОЛЬСКОГО ЯЗЫКА В КОЛИЧЕСТВЕННОМ АСПЕКТЕ*

Создан Генеральный корпус современного монгольского языка (ГКМЯ), содержащий 966 текстов общей длиной 1 155 583 слов. Разработан морфологический анализатор, грамматический словарь на 63 071 лексем, таблица морфологических омонимов. Процессор анализирует 95% текстовых словоформ, соответствующих 76% словоформам словника конкорданса к ГКМЯ.

Современный монгольский язык (СМЯ) описывается в квантитативном аспекте, в соответствии с структурно-вероятностной моделью (СВМ), содержащей частотные словари (ЧС) разных типов: словоформ, лексем, грамматем, корневых и аффиксальных морфем и алломорфем, флексием, граммем.

СВМ количественно описывает поведение языковых единиц (ЯЕ) в письменных текстах: частоту употребления, дистрибуции и сочетаемости с другими ЯЕ. Возникает перспектива превращения обычной структурной модели СМЯ в его СВМ, основанную на статистическом анализе текстов (в СВМ ЯЕ описываются как обладающие своим «весом», языковые противопоставления и связи оказываются измеримыми).

Приведены «верхушки» некоторых ЧС: ЧС словоформ («верхние» 32 словоформы, превышающие 2091 ipm), ЧС лексем («верхние» 32 лексемы, превышающие 2627 ipm) и ЧС грамматем («верхние» 32 грамматемы, превышающие 3920 ipm).

Ключевые слова: корпусная лингвистика, современный монгольский язык, частотные словари, количественные методы в лингвистике

The paper describes a General Corpus of the Modern Mongolian language (GCML), which contains 966 texts, 1 155 583 words. We also report a morphological analyzer for the Modern Mongolian language (MML), a grammatical dictionary for 63 071 lexemes, a general table of morphological homonymy. The processor analyzes effectively 95% of textual word forms which correspond to 76% word forms from the inputs of the concordance to the GCML.

MML can be described in its quantitative aspect, according to a structural-probabilistic model (SPM) of MML. SPM contains frequency dictionaries (FDs) of MML of different types: FDs of word forms, lexemes, grammemes, root morphemes and allomorphemes, affixal morphemes and allomorphemes, flexionemes, grammemes.

SPM allows to describe behavior of various language units in the written text from the quantitative point of view: their frequency, distribution in texts, compatibility with other units etc. It is possible to transform the usual structural model into an SPM, which is based on statistical analysis of texts (in this model units of language are considered as possessing «the weight», the language oppositions and relations are being measured).

* Настоящее исследование выполнено при финансовой поддержке Программы Президиума РАН по корпусной лингвистике на 2011 и 2012 гг. (направление № 4 – «Создание и развитие корпусных ресурсов по языкам мира»), Фонда фундаментальных лингвистических исследований (проект 2012/2013 отчетного года № С-13. «Квантитативно-реализационный грамматический словарь современного монгольского языка») и Российского гуманитарного научного фонда (проект № 12-04-00357 «Структурно-вероятностная модель современного монгольского языка»). Корпус создавался при участии Г.Ц. Пюрбеева, Н.С. Яхонтовой и М.П. Петровой, которым автор приносит сердечную благодарность.

The paper reports the top lists of some FDs: i. e. FD of word forms (top-list of the upper 32 word forms having frequencies higher than 2091 ipm), FD of lexemes (top-list of the upper 32 lexemes having frequencies higher than 2627 ipm) and FD of grammemes (top-list of the upper 32 grammemes having frequencies higher than 3920 ipm).

Keywords: corpus linguistics, modern Mongolian language, frequency dictionaries, quantitative approach in linguistics

1. О ГЕНЕРАЛЬНОМ КОРПУСЕ МОНГОЛЬСКОГО ЯЗЫКА (ГКМЯ)

Создана первоначальная версия корпуса современного монгольского языка (СМЯ), включающего разные жанры текстов:

- 1) художественную прозу XX века: романы; повести и рассказы; очерки;
- 2) поэзию XX в.;
- 3) перевод эпоса «Сокровенное сказание» на современный язык;
- 4) подборку газетных статей из газеты «Даяар монгол».

Корпус содержит 966 текстов (длиной 1 155 583 слов).

Первоначальная версия ГКМЯ (версия ГКМЯ-1a) в настоящее время доступна для просмотра. Адрес онлайн-ресурса: http://web-corpora.net/MongolianCorpus/search/?interface_language=ru

Создан морфологический анализатор, словарь на 63 071 лексему, таблица омонимов. Проведена лемматизация и глоссирование корпуса (в духе Лейпцигских правил глоссирования¹).

Морфологический анализатор для СМЯ на данный момент работает под управлением информационной среды StarLing². Разработка находится на экспериментальной стадии, эффективно анализируется 95% текстовых словоформ (соответствующих 76% словоформ, являющихся входами в конкорданс словоформ).

Общее количество графических словоформ (лексов) в корпусе ГКМЯ-1a составляет 1 155 583. Общее количество проанализированных (морфологически размеченных как лексически, так и грамматически) графических словоформ (аллолексов) в корпусе ГКМЯ-1a составляет 1 103 233, т.е. 95%.

Таблица 1

	В корпусе:	доля в % (от всего корпуса)	В словнике:	доля в % (от всего слов- ника)
Всего словоформ	1 155 583	100%	89 190	100%
Из них: проанализированных (хоть как-то)	1 123 156	97%	79 137	89%
Из них: проанализированных лексически	1 104 911	96%	68 212	76%
Из них: проанализированных грамматически	1 121 478	97%	78 456	88%
Из них: проанализированных и лексически, и грамматически	1 103 233	95%	67 531	76%

¹ См. [Lehmann 1982; Croft 2003; Касевич 2011: 214–221]; а также www.eva.mpg.de/lingua/resources/glossing-rules.php. Опыт использования подобной нотации в монголистике (применительно к корпусу со снятой омонимией) см., например, в [Баранова, Сай 2009: 10–16, 873].

² Информационная среда StarLing создана С.А. Старостиным (1953–2005) и позже усовершенствована Ф.С. Крыловым.

Всего различных аллолексем в словнике конкорданса к корпусу (и в словнике аллолексем) 89 190. Из них доля распознанных (т. е. проанализированных как лексически, так и грамматически) составляет 67 531, т. е. 76%.

Эффективность работы морфологического анализатора может быть наглядно представлена таблицей 1.

2. МОНГОЛЬСКИЙ ЯЗЫК В КОЛИЧЕСТВЕННОМ ОСВЕЩЕНИИ

На материале ГКМЯ впервые в монголистике монгольский язык был подвергнут изучению в его количественном аспекте. Это было сделано путем создания частотных словарей (ЧС) монгольского языка. При этом само понятие ЧС подверглось некоторому переосмыслению. Дело в том, что традиционно частотные словари понимаются как некоторые инвентари, входами в которые служат лексические единицы. В рамках СВМ понятие ЧС охватывает также грамматические единицы, например, грамемы и грамматемы³. То обстоятельство, что в традиционных ЧС эпизодически включаются некоторые данные о частотности грамматических единиц, не отменяет того несомненного факта, что систематического обследования частотности грамматических единиц в ЧС до сих пор не делалось.

Следует обратить внимание на то, что «ЧС» (в концепции СВМ) понимается не как единый тип словаря (который занимает свое место в типологии словарей и противопоставляется словарю алфавитному и словарю идеографическому). «ЧС» понимается как словарь, обладающий положительным значением одного из дифференциальных признаков, лежащих в основе типологии словарей: это признак «статистичности/нестатистичности». ЧС отличается от «нечастотных словарей» («неЧС») только тем, что он передает (в том или ином виде) информацию о частотности употребления инвентаризуемых в нем единиц, тогда как «неЧС» такой информации не передает.

Из предлагаемого понимания термина «ЧС» с необходимостью вытекает как вывод о проведении различия между разными типами ЧС, так и вывод о проведении различия между понятием ЧС и понятием рангового словаря (РС). РС – это такая разновидность ЧС, в которой инвентаризуемые единицы расположены в порядке убывания употребительности (или, что то же самое, в порядке возрастания рангов). РС противопоставляются неранговым – алфавитным и идеографическим. В отличие от РС (которые по определению являются разновидностью ЧС) алфавитными и идеографическими могут быть и «неЧС»; однако в рамках СВМ (которую интересуют лишь ЧС) по определению разрабатываются лишь ЧС; так что естественными частями СВМ наряду с РС являются алфавитные ЧС и идеографические ЧС. Алфавитные ЧС, в свою очередь, подразделяются на прямые и инверсионные; в прямых упорядочение единиц идет от начала единицы к ее концу, а в инверсионных – от конца к началу.

Следует подчеркнуть, что употребительность слова не представляет собой единого свойства, само собой разумеющегося. Для оценки употребительности слова могут быть использованы разные меры. Простейший и основной способ – это прямой подсчет общего количества вхождений единицы в состав корпуса. Но есть и другие способы оценки употребительности.

Ранг единицы измерялся двумя способами. Один способ измерения ранга – это измерение того, какое место занимает данная единица в ряду ей подобных, упорядоченному по принципу убывания числа вхождений данной единицы в состав корпуса. Другой способ измерения ранга – это измерение того, какое место занимает данная единица в ряду ей подобных, упорядоченному по принципу убывания числа текстов, содержащих данную единицу. В СВМ МЯ применены оба этих способа.

³ Термин «грамматема», принадлежащий Т. В. Булыгиной (см. [Булыгина 1977: 128–130]), понимается как «единица, являющаяся грамматическим коррелятом лексемы» и получающаяся «в результате абстракции отождествления всех словоформ, имеющих одно и то же грамматическое значение» [Булыгина 1977: 128].

Структурно-вероятностная модель монгольского языка – это комплексное (интегральное) описание, включающее частотные словари (ЧС) словоформ, основ и корневых алломорфем, но также и ЧС флексий, грамматем, флексием, аффиксальных алломорфем, аффиксальных морфем и граммем. Все упомянутые типы ЧС (в СВМ МЯ) организованы не только в ранговом, но и в алфавитном (прямом и инверсионном) порядке, а ЧС граммем, помимо этого, также в идеографическом порядке.

3. СТРУКТУРНО-ВЕРОЯТНОСТНАЯ МОДЕЛЬ МОНГОЛЬСКОГО ЯЗЫКА

Актуальность проблемы изучения количественных характеристик СМЯ обусловлена тем обстоятельством, что большинство этих характеристик до сих пор неизвестны ученым из-за отсутствия представительных и хотя бы относительно сбалансированных корпусов СМЯ, на материале которых могут быть применены дистрибутивно-статистические методы, позволяющие профессионально составлять высококачественные частотные словари и квантитативные грамматики, описывающие частотность единиц морфологии, дериватологии, синтаксиса и лексикологии.

Квантитативный подход позволяет классифицировать сами тексты в соответствии с языковыми стилями и жанрами, в рамках которых эти тексты создавались. Так как различия между этими стилями и жанрами «носят преимущественно статистический характер» [Шайкевич 1990: 231], то таким образом можно основать статистическую стилистику МЯ, описывающую и классифицирующую тексты МЯ на строго объективной базе.

Квантитативный подход к текстам открывает путь к изучению самого МЯ, поскольку сегменты текстов, являющиеся объектами подсчетов, соотнесены с единицами МЯ. Лингвостатистический метод позволяет количественно описывать поведение различных языковых единиц (букв, морфем, слов и т. п.) в письменном тексте: частоту употребления 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 (напр., данные о средней длине слова или предложения, о частоте употребления к.-л. грамматических форм в тех или иных синтаксических функциях и т. п.). Такая информация углубляет описание единиц языка» [Там же]. Например, простая констатация наличия форм мн. ч. существительных в РЯ и МЯ недостаточна для выявления типологических различий, если не учитывать количественные различия в текстовом поведении соответствующих единиц. «Таким образом, создается перспектива превращения обычной структурной модели языка в структурно-вероятностную модель, в которой учитываются результаты статистического анализа текстов (в этой модели единицы языка обладают “весом”, измеряемыми оказываются языковые противопоставления и связи). Структурно-вероятностная модель языка отличается большей реалистичностью, особенно эффективна она в диахронических и типологических исследованиях...» [Там же].

4. ЧАСТОТНЫЕ СЛОВАРИ МОНГОЛЬСКОГО ЯЗЫКА

«Верхушки» некоторых ЧС можно привести уже сейчас [Крылов 2012]. Числовые показатели в этих словарях указывают на относительную частоту (количество вхождений данной единицы на миллион словоформ, ipm – instances per million (words): см. [Sharoff 2002; Ляшевская, Шаров 2009: 9]) и на количество текстов, в которых встретилась данная единица, а также на ранг данной единицы. В интересах компактности изложения ограничимся «верхушками» размером в 32 самые употребительные словоформы, 32 самые употребительные лексемы и 32 самые употребительные грамматические формы (грамматемы).

4.1. Частотность словоформ в монгольском языке

В ЧС, дающих представление о частотности словоформ, входами являются словоформы МЯ с их приблизительными переводами (в его основном значении) на РЯ.

Ранговый список 32 словоформ с частотами более 2091 ipm

словоформа	перевод	частота	количество текстов	ранг
нь	его, ее, их [притяжат. артикль]	24 463.32	666	1
гэж	что [изъясн. союз, вводящий прямую речь] ⁴	13 884.12	478	2
юм	[показатель ремы]	10 052.32	529	3
ч	же	9882.74	540	4
л	-ка	6798.44	441	5
энэ	этот	6628.87	463	6
тэр	тот, он, она, оно	5801.78	421	7
нэг	один, раз [актантн.]	5658.16	399	8
би	я [подлеж.]	5429.76	446	9
хүн	человек	5260.18	493	10
байна	есть, является, имеется [презенс]	4850.10	437	11
шиг	подобный	4619.10	526	12
хоёр	два, оба, и	4590.55	409	13
минь	мой [притяжат. артикль]	4161.43	496	14
байгаа	есть, является, имеется [прич. дуративного вида]	3666.56	340	15
бол	а, что касается [показатель темы]	3606.86	363	16
дээр	на	3339.53	461	17
чинь	твой [притяжат. артикль]	3109.39	344	18
байсан	был [прич. перф.]	2916.46	337	19
их	большой, очень	2907.81	403	20
дээ	-ка	2858.49	293	21
юу	что? [им. п.]; ли?	2722.66	307	22
гээд	сказав	2625.77	275	23
чи	ты [подлеж.]	2620.58	289	24
уу	ли	2414.67	321	25
бас	тоже, также	2359.30	349	26
билээ	был, находился	2338.53	303	27
байх	быть	2313.44	316	28
сайхан	красивый	2217.41	429	29
байлаа	был	2190.59	265	30
шүү	не так ли?	2188.86	264	31
гэсэн	сказал	2091.10	314	32

⁴ Можно трактовать эту форму как форму конгрессивного деепричастия от глагола гэх «говорить», подвергшуюся грамматикализации и конверсии в другую часть речи – изъяснительный подчинительный союз.

4.2. Частотность лексем в монгольском языке

В ЧС, дающих представление о частотности лексем, входами являются лексемы МЯ (в их исходных формах) с их примерными переводами МЯ (в его основном значении) на РЯ.

Так как работа велась по корпусу с неснятой омонимией, то иногда статус лексем получают не собственно лексемы, а дизъюнктивные пучки (частично) омонимических лексем (ср. в табл. 3 пучки, между членами которых стоит помета дизъюнкции ~). Однако информация о частотности таких единиц в корпусе является не менее ценной, нежели информация о частотности собственно лексем. Во всяком случае, для получения информации о количественных характеристиках лексики и грамматики МЯ не стоит ждать, пока будет создан корпус со снятой омонимией: ждать придется слишком долго, и в любом случае данные, полученные на основе корпуса со снятой омонимией, будут базироваться на слишком малых эмпирических фактах, что обесценит их статистическую значимость. Реально многие из членов пар частично омонимических лексем довольно малоупотребительны (ср., например, *шигэх* в табл. 3); однако для того, чтобы выявить эти маловероятные прочтения (они помечены в таблице символом [?]) и получить «очищенные» данные о речевой статистике таких единиц, следовало бы проделать огромную работу по «ручному» снятию омонимии в корпусе. Трудоемкость этой работы заставляет отложить ее на неопределенное будущее.

Таблица 3

Ранговый список 32 лексем с частотами выше 2627 ipm

лексема	перевод	частота	количество текстов	ранг
нь	его, ее, их [притяжат. артикль]	24 463.32	666	1
гэж	что [изъясн. союз]	13 887.58	478	2
байх	быть	13 638.41	525	3
юм ~ [?] юм(ан) _{им.п.}	[показатель ремы] ⁵ ~ [?] вещь _{им.п.}	10 408.76	531	4
ч	же	9882.74	540	5
болох	стать, становиться	9672.51	528	6
явах	уходить	7327.06	511	7
л	-ка	6845.16	443	8
энэ	этот	6636.66	463	9
тэр	тот, он, она, оно	6024.99	427	10
нэг(эн)	один	5749.00	402	11
хүн	человек	5538.77	497	12
ирэх	приходить	5497.24	439	13
би	я	5434.95	448	14
[?] байн ~ байх _{модификатив}	[?] погода немного ~ быть _{модификативное деепричастие}	5364.87	449	15
хоёр	два; оба; и	5197.03	427	16

⁵ Форма им. п. от существительного *юм(ан)* подверглась грамматикализации и конверсии в другую часть речи – рематическую частицу.

Таблица 3 (окончание)

лексема	перевод	частота	количество текстов	ранг
шиг ~ ?шигэх _{императив}	подобный ~ ?собирайся! //сплачивайся!	4624.29	526	17
дээр	на	4445.20	503	18
байг (= байх _{юссив}) ~ ?бай	пусть будет! ~ ?мишень _{вин.п.} ~ ?расстояние _{вин.п.} ~ ?приз _{вин.п.} ~	4322.35	362	19
минь	мой [<i>притяжат. артикль</i>]	4163.16	497	20
хэлэх	разговаривать	3645.79	348	21
бол	а, что касается [показатель темы]	3609.46	363	22
гарах	выходить	3329.14	392	23
чинь	твой [<i>притяжат. артикль</i>]	3110.26	344	24
орох	входить	3051.43	363	25
гэх	сказать, говорить	3019.42	328	26
их	большой, очень	3011.63	411	27
дээ	-ка	2863.69	294	28
?юу ~ юу(н)	?ли? ~ что?	2725.26	307	29
чи	ты	2661.24	307	30
бодох	думать, вычислять, стремиться	2637.88	303	31
гээд (= гэх _{антecessив}) ~ ?гээ	сказав ~ ?рукоять мотыги _{дат.п.}	2627.50	275	32

Из анализа верхушек ранговых списков можно извлечь немало ценных поучительных сведений и наблюдений типологического характера⁶. Приведем несколько примеров.

1. Союз *хоёр* 'и' (производный путем конверсии числительного 'два') занимает всего лишь 13-е место в ЧС словоформ (4590.55 ipm) и 16-е место в ЧС лексем (5197.03 ipm). К тому же не надо забывать, что довольно значительная доля вхождений этой словоформы приходится на собственно «количественное» употребление, т. е. на значение 'два', так что собственно союзное (сочинительное соединительное) употребление (ср. *чоно үнэг хоёр* «волк и лиса», букв. «волк, лиса двое» или «волк, лиса вдвоем») несколько менее, чем 5197.03 ipm (если прикинуть «на глазок», вычтя из этого количества собственно «количественное» употребление, т. е. значение 'два', то, по-видимому, оно составляет примерно около 3700 ipm, а значит, соответствует примерно 15-му месту в ЧС словоформ и примерно 21-му месту в ЧС лексем). Союз *бөгөөд* 'и' (производный путем конверсии деепричастия) (1523.55 ipm) занимает всего лишь 52-е место в ЧС словоформ и 67-е место в ЧС лексем. Союз *болон* 'и' (производный путем конверсии деепричастия) (535.54 ipm) занимает всего лишь 215-е место в ЧС словоформ и 279-е место в ЧС лексем.

⁶ Принципы квантитативной типологии разработаны еще в «докорпусную» эпоху, см. [Гринберг 1963; Маслов 1975: 293–298; Касевич 1977: 132–134; Маслов 1987: 233; Касевич, Яхонтов 1982; Сусов 2006: 314; Касевич 2011: 170–172].

Просуммировав данные по союзам *хоёр* 'и' (примерно 3700 ipm), *бөгөөд* 'и' (1523.55 ipm) и *болон* 'и' (535.54 ipm), мы получим число примерно 5760 ipm (соответствующее примерно 7–8 месту в ЧС словоформ и примерно 10–11 месту в ЧС лексем).

Допустим даже, что мы «завысили» оцененную «на глазок» частоту собственно «количественного» употребления, т. е. значения 'два', и не будем вычитать ее вовсе. В этом случае просуммировав данные по союзам *хоёр* 'и' (5197.03 ipm), *бөгөөд* 'и' (1523.55 ipm) и *болон* 'и' (535.54 ipm), мы получим число $5197.03 + 1523.55 + 535.54 = 7256.12$ ipm (соответствующее примерно 4–5 месту в ЧС словоформ и примерно 7–8 месту в ЧС лексем).

Ср.: англ. союз *and* в ЧС (лексем!) занимает 3-е (или 4-е, или 5-е) место; русск. союз *и* в ЧС (как лексем, так и словоформ) уверенно занимает 1-е место. В обоих случаях (и в русском, и в английском) налицо превышение (для русского – даже значительное превышение!) частотности соединительного союза над частотностью его близких аналогов в МЯ.

Откуда столь большая разница? Она объясняется следующими факторами. МЯ предпочитает бессоюзные конструкции. Частота бессоюзия обычно компенсируется явлением так называемого алтайского типа сочинения, предполагающего групповую флексию при сочинении существительных и богатство системы деепричастного такси-са в сфере глагола.

2. Энклитизованные притяжательные местоимения имеют очень высокую частоту.

Например, *нь* (букв. 'его', 'ее', 'их') занимает 1-е место в ЧС лексем и 1-е место в ЧС словоформ!

Ср.: в английских ЧС *his* 'его' занимает 25-е (или 23-е, или 12-е) место, *her* 'ее' занимает 42-е (или 29-е, или 13-е) место, *its* 'его' занимает 78-е (или 77-е, или 142-е) место, *their* 'их' занимает 36-е (или 39-е, или 61-е) место. В русских ЧС слово *его* занимает 41-е (или 50-е) место, *ее* – 72-е (или 121-е) место, *их* – 86-е (или 134-е) место. По частоте эти слова скорее сближаются с суммой частот таких их монгольских эквивалентов, как *үүний* (87.38 ipm) (1565-е место в ЧС словоформ⁷) + *тууний* (1218.15 ipm) (77-е место в ЧС словоформ).

Слово *минь* (букв. 'мой') занимает 14-е место в ЧС словоформ и 20-е место в ЧС лексем!

Ср.: в английских ЧС *my* 'мой' занимает 44-е (или 34-е, или 24-е) место. В русских ЧС *мой* занимает 60-е (или 69-е) место. По частоте эти слова скорее сближаются с таким их монгольским эквивалентом, как *миний* (1876.54 ipm) (39-е место в ЧС словоформ).

Слово *чинь* (букв. 'твой') занимает 18-е место в ЧС словоформ и 24-е место в ЧС лексем!

Ср.: в английских ЧС *your* 'твой' занимает 69-е (или 64-е, или 62-е) место. В русских ЧС *твой* занимает 266-е (или 579-е) место. По употребительности эти слова скорее сближаются с таким их монгольским эквивалентом, как *чиний* (423.93 ipm) (292-е место в ЧС словоформ).

Откуда такая существенная разница? Здесь напрашивается такое объяснение. МЯ имеет грамматическую категорию притяжания. Она может быть выражена синтетически или аналитически. Аналитические средства выражения этой категории – энклитизованные притяжательные местоимения. Фактически энклитизованные притяжательные местоимения играют в МЯ роль, во многом сходную с той ролью, которая во многих артиклевых европейских (например, романских и германских) языках выполняется артиклями. Если сравнить западноевропейские притяжательные местоимения

⁷ Для словоформ *тууний*, *үүний*, *миний*, *чиний* не даются их места в ЧС лексем, поскольку это не самостоятельные лексемы, а формы род. п. от соответствующих местоимений, тогда как их русские и английские эквиваленты принято считать отдельными лексемами (притяжательными местоимениями).

с русскими, то можно убедиться, что западноевропейские притяжательные местоимения используются более часто, чем русские (например, при переводе с русского языка на английский и обратно такие факты проявляются особенно ярко: ср. такие переводные эквиваленты, как *жена* ⇔ *my wife (your wife, his wife)*, *мать* ⇔ *my mother (your mother, his mother, her mother)*, *муж* ⇔ *my husband (your husband, her husband)*, *отец* ⇔ *my father (your father, his father,)*, *нос* ⇔ *my nose*, *голова* ⇔ *my head* и т.д.). Процесс грамматикализации притяжательных местоимений в западноевропейских артиклевых языках зашел так далеко, что эти притяжательные местоимения включаются в разряд «детерминативов» (т. е. грамматических эквивалентов артиклей, или «заместителей артиклей») и получают наименование «притяжательных детерминативов», а порой даже «притяжательных артиклей» (в отличие от русских притяжательных местоимений, которые, разумеется, никто не называет «артиклями», хотя бы и «притяжательными»).

Между тем сравнение ЧС показывает, что МЯ продвинулся по шкале грамматикализации притяжательных местоимений еще больше, чем западноевропейские языки. В результате этого продвижения сами притяжательные местоимения подверглись мощному процессу генерализации, транспозиции и десемантизации. Эти местоимения играют уже не только и не столько роль показателей притяжательности в собственном смысле слова, сколько роль показателей определенности, топикальности и субстантивации. Но такой процесс свойствен всем грамматическим категориям, так что сам факт наличия у притяжательных показателей вторичных (непритяжательных) функций на самом деле парадоксальным образом демонстрирует их грамматикализированный характер.

Именно серьезной коммуникативной нагрузкой, состоящей в выражении определенности и топикальности, объясняется столь высокая употребительность энклитизованных форм притяжательных местоимений (а также заодно и постфиксального показателя возвратной притяжательности) в монгольских языках. Этим объясняется и выбираемая в настоящей работе нетрадиционная для монголистики терминология, согласно которой энклитизованные притяжательные формы местоимений могут быть названы «притяжательными артиклями»; впрочем, она нетрадиционна только для МЯ, поскольку в грамматиках германских (английского, немецкого), романских (французского, румынского), армянского или коптского языков этот термин нередко встречается (наряду с такими синонимичными ему терминами, как «притяжательный детерминатив» и т.п.).

4.3. Частотность грамматем в монгольском языке

В ЧС, дающих представление о частотности грамматем, входами являются грамматемы с разъяснением используемой нотации.

Так как работа велась по корпусу с неснятой омонимией, то иногда статус грамматем получают не собственно грамматемы, а дизъюнктивные пучки (частично) омонимических грамматем. Тем не менее информация о частотности таких единиц в корпусе представляет не меньшую ценность, нежели информация о частотности собственно грамматем. Приведенные выше соображения о целесообразности использования корпуса с неснятой омонимией *mutatis mutandis* относятся и к грамматике.

Наличие в составе дизъюнктивных пучков грамматем таких пучков, в которых левый член дизъюнкции тождествен правому, объясняется тем, что в МЯ немало частично-омонимических пар, (а) внутри которых налицо отношение субкатегориальной конверсии (а именно, один из членов пары принадлежит к тематическому склонению, а другой – к атематическому), а также (б) члены которых различаются тем, что в исходе одного из членов пары стоит устойчивое «н», а в исходе другого – неустойчивое «н».

Ранговый список грамматем с частотами выше 761 ipm

грамматема (в виде символической пометы)	пояснение (расшифровка пометы)	показатель (основная форма)	частота	количество текстов	ранг
NOM	им. п.	-0	281 693.26	885	1
0	единств. форма неизм. слова	-0	93 456.52	865	2
CVB.CNGR	деепр. конгрессивное	-ж/-ч	50 195.83	806	3
NOM ~ GEN-NOM	им. п. ~ род. п. в им. п.	-Ы(н)	34 055.37	857	4
PC.PROSP-NOM	прич. проспективного вида, им. п.	-х	30 514.26	794	5
CVB.MOD	деепр. образа действия	-н	23 820.50	733	6
NOM ~ CVB.MOD	им. п. ~ деепр. образа действия	-н	23 777.25	777	7
PC.PRF-NOM	прич. перфективного вида, им. п.	-сАн	23 658.72	767	8
NOM ~ ABS-NOM	им. п. ~ абсолютив в им. п.	-н	23 602.48	753	9
GEN-NOM	генитивный атрибут в им. п.	-ы(н)	16 387.03	779	10
NOM ~ VF.OPT.IMP	им. п. ~ предикативная роль, повелит. наклонение опта- тивной репрезентации	-0	12 322.50	675	11
DAT	дат. п.	-д(т)	11 860.50	753	12
VF.IND.AOR	предикативная роль, аорист индикативной репрезентации	-в	11 502.33	489	13
POSS.REFL	возвр.-притяж.	-АА	10 403.57	664	14
NOM ~ DAT	им. п. или дат. п.	-д(т)	9257.23	639	15
REL-NOM	общий атрибут, им. п.	-н	9122.27	662	16
NOM ~ REL-NOM	им. п. ~ общий атрибутив в им. п.	-н	8674.98	677	17
NOM ~ COM-NOM	им. п. ~ комитативный атри- бутив в им. п.	-тАй	8229.42	646	18
NOM ~ VF.OPT.JUSS	им. п. ~ предикативная роль, юссивное наклонение опта- тивной репрезентации	-г	8070.23	666	19
CVB.ANT	деепр. антецессивное	-ААд	8031.30	548	20
PC.PROSP-DAT	прич. проспективного вида, дат. п.	-хАд	7071.83	490	21
VF.IND.PRS1	предикативная роль, презенс № 1 индикативной репрезен- тации	-нА	6936.00	590	22
GEN/ACC	усеченная форма род.-вин. п.	-ы	6495.64	573	23
ACC	вин. п.	-(ы)г	6383.16	647	24

Таблица 4 (окончание)

грамматема (в виде символической пометы)	пояснение (расшифровка пометы)	показатель (основная форма)	частота	количество текстов	ранг
ABL	аблат. п.	-ААс	4746.28	532	25
PC.US-NOM	причастие узуального вида, им. п.	-дАг	4708.21	477	26
0 ~ VF.OPT.IMP	единств. форма неизменяемого слова ~ предикативная роль, повелит. наклонение оптативной репрезентации	-0	4624.29	526	27
A.COM-NOM ~ DAT	комитативное прилагательное в им.п. ~ дат. п.	-т	4610.45	552	28
NOM ~ ACC ~ VF.OPT.JUSS	им. п. ~ вин. п. ~ предикативная роль, юссивное наклонение оптативной репрезентации	-г	4435.69	377	29
VF.OPT.IMP	предикативная роль, повелительное наклонение оптативной репрезентации	-0	4057.61	457	30
NOM ~ DAT ~ CVB.ANT	им. п. ~ дат. п. ~ антецессивное деепричастие	-ААд/-д	3978.88	329	31
NOM ~ CVB.CNGR	им. п. ~ конгрессивное деепричастие	-ж/ч	3920.05	401	32

Немало ценных сведений можно извлечь из анализа верхушек ранговых списков грамматем. Такие сведения дают импульс для поучительных размышлений типологического характера. Остановимся для примера на следующих примечательных закономерностях.

1. Деепричастные формы глаголов (так называемые конвербы) занимают весьма высокие позиции в ЧС грамматем: например, конгрессивные деепричастия – 50195.83 ipm, модификативные деепричастия – 23820.50 ipm, антецессивные деепричастия – 8031.30 ipm.

В тех европейских языках, где есть деепричастия, они употребляются с гораздо более низкой частотой. Откуда такая разница? Она допускает следующее объяснение. МЯ характеризуется так называемым алтайским типом сочинения, предполагающим (в сфере глагола) богатую систему деепричастно выраженного таксиса. Таким образом, он предпочитает бессоюзное сочинение. Относительная редкость употребления союзов компенсируется богатством системы деепричастий.

2. Возвратно-притяжательные формы употребляются исключительно часто: «простые» возвратно-притяжательные формы – 10403.57 ipm (да еще к этой частоте следовало бы прибавить употребительность возвратно-притяжательных форм перфективных причастий – 2336.80 ipm; возвратно-притяжательных форм дательного падежа перспективных причастий – 2001.12 ipm и т.д.). Наиболее знакомые нам западноевропейские языки вообще не имеют такой грамматической формы. Русский язык, правда, имеет ближайший ее переводной эквивалент – возвратное местоимение *свой*. Но сравнение относительной частотности местоимения *свой* с относительной частотностью возвратно-притяжательных форм показывает, что *свой* употребляется с гораздо меньшей частотой (3825.5 ipm).

Откуда же столь значительное количественное различие? Его можно пояснить следующим образом. Возвратно-притяжательные формы монгольских языков используют-

ся как одно из важнейших средств выражения кореферентности. Западноевропейские артиклевые языки предпочитают иной способ выражения кореферентности (в этой функции используются преимущественно определенные артикли). Русский язык предпочитает так называемые «нулевые» формы выражения кореферентности. Фактически они носят не столько «нулевой», сколько просодический (супрасегментный) характер; но просодические механизмы не находят своего последовательного отражения в письменной форме языка (или, во всяком случае, находят его чрезвычайно редко).

Сказанное выше о степени грамматикализованности «притяжательных артиклей» в еще большей степени относится и к возвратно-притяжательным формам. Вспомним, что теоретики «артиклеведения» признают, что артикли могут быть суффигированными (и охотно говорят о суффигированных артиклях в скандинавских или в балканских языках). Почему бы в таком случае не признать, что «притяжательные артикли» (а они, как мы убедились выше, в монгольских языках есть и являются энклитиками) тоже бывают суффигированными? Именно такой случай представлен примером возвратно-притяжательного показателя в монгольских языках. Он не только энклитизован, но также (в отличие от других монгольских притяжательных артиклей – а именно, от энклитик, требующих отдельного графического написания) входит в сферу действия сингармонизма (и потому вполне закономерно его слитное написание с предшествующей основой). Следовательно, перед нами не просто энклитизованный, но и более того – суффигированный притяжательный артикль.

СПИСОК ЛИТЕРАТУРЫ

- Баранова, Сай 2009 – В.В. Баранова, С.С. Сай. От составителей // Исследования по грамматике калмыцкого языка. Труды Института лингвистических исследований. Т. V. Ч. 2. СПб., 2009.
- Булыгина 1977 – Т.В. Булыгина. Проблемы теории морфологических моделей. М., 1977.
- Гринберг 1963 – Дж. Гринберг. Квантитативный подход к морфологической типологии языков // Новое в лингвистике. Вып. 3. М., 1963.
- Касевич 1977 – В.Б. Касевич. Элементы общей лингвистики. М., 1977.
- Касевич 2011 – В.Б. Касевич. Введение в языкознание. М.; СПб., 2011.
- Касевич, Яхонтов 1982 – В.Б. Касевич, С.Е. Яхонтов (ред.). Квантитативная типология языков Азии и Африки. Л., 1982.
- Крылов 2012 – С.А. Крылов. Структурно-вероятностная модель современного монгольского языка на базе Генерального корпуса монгольского языка // Урало-алтайские исследования. 2012. № 6.
- Ляшевская, Шаров 2009 – О.Н. Ляшевская, С.А. Шаров. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М., 2009. (<http://dict.ruslang.ru/freq.php>).
- Маслов 1975/1987 – Ю.С. Маслов. Введение в языкознание. М., 1975 (2-е изд. М., 1987).
- Сусов 2006 – И.П. Сусов. Введение в языкознание. М., 2006.
- Шайкевич 1990 – А.Я. Шайкевич. Количественные методы в языкознании // Лингвистический энциклопедический словарь. М., 1990.
- Croft 2003 – B. Croft. Typology and universals. Cambridge, 2003.
- Lehmann 1982 – Ch. Lehmann. Directions for interlinear morphemic translations // Folia linguistica. 1982. V. 16.
- Sharoff 2002 – S. Sharoff. Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics // Proceedings of language resources and evaluation conference. May, 2002. Las Palmas (Spain), 2002.

Сведения об авторе:

Сергей Александрович Крылов
Институт востоковедения РАН
krylov-58@mail.ru

Статья поступила в редакцию 03.05.2012.