



Published in the Russian Federation
 Oriental Studies (Previous Name: Bulletin of the Kalmyk Institute
 for Humanities of the Russian Academy of Sciences)
 Has been issued as a journal since 2008
 ISSN: 2619-0990; E-ISSN: 2619-1008
 Vol. 13, Is. 5, pp. 1438–1445, 2020
 DOI: 10.22162/2619-0990-2020-51-5-1438-1445
 Journal homepage: <https://kigiran.elpub.ru>



УДК 811.512.3

DOI: 10.22162/2619-0990-2020-51-5-1438-1445

Структурно-вероятностная модель монгольской грамматики и измерение употребительности обобщенных грамматических единиц

Сергей Александрович Крылов¹

¹ Институт востоковедения РАН (д. 12, ул. Рождественка, 107031 Москва, Российская Федерация)
 доктор филологических наук, ведущий научный сотрудник

 0000-0001-5207-0949. E-mail: krylov-58@mail.ru

© КалмНЦ РАН, 2020

© Крылов С. А., 2020

Аннотация. *Введение.* Синтаксическая молекула (СМ) — минимальная функционально-синтаксически автономная единица, способная служить ответом на вопрос. *Цель статьи* — описать принципы выделения синтаксических молекул в Генеральном корпусе монгольского языка, принципы и особенности разметки, а также дать общую характеристику количественной модели позиционно-морфологической грамматики монгольского языка. *Результаты.* Сформулированы принципы выделения СМ. Перечислены основные особенности грамматической разметки в Генеральном корпусе монгольского языка ГКМЯ-3б (по сравнению с ранними версиями ГКМЯ): выделение базового лексикона, принципы моносемизации, принципы категориально-семантической нотации, основные отличительные черты грамматической нотации (частеречное разбиение неизменяемых лексем и расщепление номинатива). Изложены принципы упорядочения морфологических структур в количественной модели монгольского языка, а также трактовка супrasegmentных маркеров письменного текста. В заключение вкратце охарактеризовано общее строение количественной модели позиционно-морфологической грамматики монгольского языка.

Ключевые слова: синтаксическая молекула, корпусная лингвистика, монгольский язык, грамматическая разметка, количественная лингвистика, письменный текст, пунктуация, супrasegmentные маркеры, аналитические формы слова, морфологическая структура словоформы

Благодарность. Настоящее исследование выполнено при поддержке комплексной программы фундаментальных исследований ОИФН РАН «Евразийское наследие и его современные смыслы» (направление «Создание версии «3» генерального корпуса современного монгольского языка»), а также РФФИ (грант № 18-012-00611 «Структурно-вероятностная морфология монгольских языков» и № 19-512-44017 «Структурно-вероятностная интегральная модель монгольских языков»).

Для цитирования: Крылов С. А. Структурно-вероятностная модель монгольской грамматики и измерение употребительности обобщенных грамматических единиц // *Oriental Studies*. 2020. Т. 13. № 5. С. 1438–1445. DOI: 10.22162/2619-0990-2020-51-5-1438-1445

UDC 811.512.3

DOI: 10.22162/2619-0990-2020-51-5-1438-1445

Structural-Probabilistic Model for Mongolian Grammar and Measurement of Frequency Use of Generalized Grammatical Units

*Sergej A. Krylov*¹

¹ Institute of Oriental Studies of the RAS (12, Rozhdestvenka St., Moscow 107031, Russian Federation)
Dr. Sc. (Philology), Leading Research Associate
 0000-0001-5207-0949. E-mail: krylov-58@mail.ru

© KalmSC RAS, 2020

© Krylov S. A., 2020

Abstract. Syntactic molecule (SM) is the minimum functional and syntactically autonomous unit able to serve as an answer to the question. *Goals.* The paper formulates the principles of identification of the SM. *Results.* The main features of grammatical tagging in the General Corpus of the Mongolian Language GCML-3b (compared to earlier versions GCML) are delineated: introducing of the basic vocabulary, principles of monosemization, principles of categorial-semantic notation, main distinctive features of grammar notation (part-of-speech splitting of the indeclinable words and splitting of the nominative). The principles of ordering of morphological structures in a quantitative model of the Mongolian language, as well as the interpretation of suprasegmental markers in written text are discussed. In conclusion, a brief survey of the general structure of the quantitative model of the positional-morphological grammar of the Mongolian language is discussed.

Keywords: syntactic molecules, corpus linguistics, Mongolian language, grammatical tagging, quantitative linguistics, written text, punctuation, suprasegmental markers, analytical word forms, morphological structure of word forms

Acknowledgements. The reported study was funded by Comprehensive Fundamental Research Program of the RAS (Department of Historical and Philological Sciences) ‘Eurasian Heritage and Its Contemporary Meanings’ and RFBR (grants no. 18-012-00611 ‘Structural-Probabilistic Morphology of Mongolic Languages’ and no. 19-512-44017 ‘Structural-Probabilistic Integral Model of Mongolic Languages’).

For citation: Krylov S. A. Structural-Probabilistic Model for Mongolian Grammar and Measurement of Frequency Use of Generalized Grammatical Units. *Oriental Studies*. 2020. Vol. 13(5): 1438–1445. (In Russ.). DOI: 10.22162/2619-0990-2020-51-5-1438-1445



Вводные замечания
Необходимое предварительное пояснение

Синтаксическая молекула (Ш. Балли) (СМ) [= «член предложения» (В. Г. Гак, И. Б. Долинина), «синтаксема» (А. М. Мухин, И. Ф. Вардуль, Г. А. Золотова)] — это

минимальная автономная единица, способная служить ответом на вопрос.

1. Принятые условные соглашения при работе над корпусом с неснятой омонимией

Для разработки методов выделения измеряемых единиц в таких корпусах

разумно принять такие условные соглашения:

(а) любая омонимичная (грамматическая либо лексическая) единица трактуется как имеющая дизъюнктивно организованное означаемое, то есть (а1) любая лексически омонимичная графическая словоформа трактуется как соотнесенная с особой «квазилексемой» (дизъюнктивной), и (а2) любая грамматически омонимичная графическая словоформа трактуется как соотнесенная с особой «квазиграмматемой» (дизъюнктивной);

(б) условно принимается «рабочая» презумпция моносемичности любых словоформ, позволяющая работать с их внешними обликами так, как если бы они были двусторонними (знаковыми) единицами. То есть мы будем исходить из эвристики (нередкой для компьютерной лингвистики), известной под названием «формального» (// «формалистического») подхода¹.

1. Принципы выделения СМ

1) знаменательная словоформа морфологически размечена (tagged) как сочетание лексем (леммы) с грамматемой (по Т. В. Булыгиной) (грамматической характеристикой, «ярлыком», «этикеткой»); «морфологическая грамматема» соответствует синтетически выражаемой грамматической форме.

2) служебные и вспомогательные компоненты аналитических СМ, а также основы местоимений и других «строєвых» («полуслужебных», «функциональных», «грамматизированных», «десемантизованных») слов трактуются как носители грамем («Г-маркеры»). Для их разметки строится полуйскусственный метаязык (практически на основе латыни).

3) пунктуационное (далее Пу-) оформление СМ тоже трактуется как Г-маркер.

4) позиция СМ, т. е. «начало» («In»), «середина» («M») и «конец» («Fi») по от-

ношению к Пу-единицам, т. е. к Пу-аналогам синтагмы («Пу-синтагма»), клаузы («Пу-клауза»), предложения и абзаца) также трактуется как Г-маркер.

5) дистанцированный Г-маркер (см. 2 и 3 выше) трактуется как элемент одной из тех двух СМ, с которыми он линейно контактирует. Таким образом, принимается такая аксиоматика синтаксической теории, где само существование разрывных аналитических форм не допускается в силу сугубо формальных принципов, так как все реально разрывные аналитические формы почти автоматически сводятся к контактными формам.

6) все Г-маркеры подразделяются на препозитивные («проклитикоиды») и постпозитивные («энклитикоиды»).

2. Особенности корпуса ГКМЯ-36

Изложенный подход применен к корпусу (халха-) монгольского языка (ХМЯ)². Исследование частотности СМ в ХМЯ проводилось на материале Генерального корпуса современного ХМЯ (ГКМЯ-36)³. Относительная частотность, в соответствии с принципами современной корпусной лингвистики, измерялась в ipm, то есть в числе вхождений на 1 млн текстовых словоформ⁴.

3. Основные особенности грамматической разметки в ГКМЯ-36 (по сравнению с ранними версиями ГКМЯ).

3.1. Принципы моносемизации

1) На всем подпространстве словаря словоформ, включающем 300 базовых (по употребительности наиболее частотных) словоформ и 6 026 словоформ, производных от тех же лексем, что и базовые словоформы (в пределах «базового сублексикона»), вручную произведено снятие морфологической

² Таким образом, наше исследование учитывает опыт работ по корпусному изучению монгольского языка, ведущиеся монгольскими (см., напр., [Purev et al. 2003; Purev et al. 2005; Purev, Odbayar 2008]) и китайскими учеными (см., напр., [Dawa et al. 2006]).

³ Задачи корпусной монголистики с точки зрения структурно-вероятностного моделирования были очерчены в работах [Крылов 2012а; Крылов 2012б; Крылов 2013; Крылов 2014; Крылов 2016б; Крылов 2017а; Крылов 2017б; Крылов 2017в; Крылов 2018а; Крылов 2018б; Крылов 2018в].

⁴ См. [Ляшевская, Шаров 2009].

¹ Ср. «дистрибутивно-статистический анализ» (А. Я. Шайкевич), «оптимизационные методы» (Б. В. Сухотин), «диагностику» (И. П. Севбо); умеренно формальный подход в грамматических концепциях А. А. Зализняка и А. К. Поливановой. Основные идеи дистрибутивно-статистического подхода к языку изложены в работах [Шайкевич и др. 2013; Шайкевич и др. 2016].

омонимии. В частности, представляют особый интерес 2 типовые разновидности морфологической омонимии: 1.1) омонимия, вызванная фактором **морфологической конверсии**. В подобных случаях на всем пространстве «базового сублексикона» явочным порядком вручную осуществлена моносемизирующая правка, как правило, в пользу того из членов омографического пучка, который обладает более прозрачной внутренней формой с грамматической точки зрения, если между его внутренней формой и пучком его реальных грамматических свойств в ГКМЯ-36 не обнаруживалось явных противоречий. 1.2) омонимия, резко **неравновесная функционально-статистически**. Тогда на всем пространстве «базового сублексикона» вручную осуществлена моносемизирующая правка (разумеется, в пользу более частотной интерпретации).

3.2. Принципы категориально-семантической нотации

2.) На том же подпространстве (6 326 синтетических словоформ) задана **категориально-семантическая (КС-) нотация**. Она охватывает 6 326 синтетических словоформ, что соответствует 592 613 вхождениям словоформ, принадлежащих «художественному» подкорпусу ГКМЯ-36.

КС-нотация основана на следующих принципах. 2.1) Для тех омонимических пучков словесных сегментов, частотность которых выше пороговой величины D_1 , мы можем интуитивно выделить тот из членов пучка, в значении которого данный сегмент употребляется чаще, чем в остальных значениях; в словарях такие члены обычно подаются под римской цифрой I. 2.2) Для всех словоформ с частотностью выше D_1 мы можем интуитивно выделить то («базовое») из ее значений, которое ощущается как такое (обычно это значение подается в толковом словаре под арабским номером 1). Для всех «базовых» значений слов, входящих в совокупность L_{b1} (= «базовый сублексикон»), мы можем построить их перевод на некоторый искусственный **универсальный семантический метаязык** (УСМЯ), ядро которого составляет известный Лейпцигский стандарт морфологической разметки. В этот УСМЯ входят некоторые элементы латинского языка, а именно, те, на которые удобно переводить (с ХМЯ) основы словоформ, входящих в L_{b1} .

Наиболее систематично КС-нотация на основе УСМЯ представлена в особой таблице. «Означающие» этого УСМЯ приведены в 5-м (ключевом) столбце; а 4-й столбец, где даны толкования, фактически выполняет функцию неформальных пояснений к «означаемым» УСМЯ.

3.3. Черты грамматической нотации (ГН) в ГКМЯ-36

3) Морфологический анализатор, используемый в ГКМЯ-36, основан на ГН, в общих чертах сходной с той, которая использовалась в предыдущих версиях корпуса⁵. Она представляет собой вариант так называемого Лейпцигского стандарта ГН⁶, адаптированный под специфику ХМЯ⁷. Однако эта ГН имеет свою специфику сравнительно с ранними (до 2014 г.) версиями; отметим 2 важнейшие инновации (3.3.1 и 3.3.2).

3.3.1. Частеречное разбиение неизменяемых лексем

(1) Вместо единого символа «0» для ГН графически неизменяемых лексем использовано несколько **разных** символов, условно соотносимых с традиционными обозначениями частей речи (или с их семантико-синтаксическими подклассами): {PART} «частица», {ART} «артикл», {CNJ} «союз», {ADV} «наречие», {POST. REL} «релятор-послелог», {LOC} «локализатор», {TEMP} «темпорализатор», {COMM} «коммуникатив», [Abbr] «неизменяемая аббревиатура», [Alien] «неизменяемое заимствование», [Numer] «число (арабские цифры)», [NumRom] «число (римские цифры)».

3.3.2. Расщепление номинатива

(2) Вместо единого падежа номинатива (NOM) введено два более частных понятия: (а) собственно «**номинатив**» (NOM), или «**номинатив-индефинитив**», функция кото-

⁵ Морфологический анализатор разработан совместно с Ф. С. Крыловым.

⁶ О Лейпцигских правилах глоссирования см. подробнее [Lehmann 1982; Croft 2003]; в начале 2010-х гг. эти правила даже вошли в один из основных отечественных университетских учебников по введению в языкознание, см. [Касевич 2011: 214–221].

⁷ В отечественной монголистике Лейпцигские правила глоссирования применяются к материалу современного калмыцкого языка уже с 2009 г. (см. [Баранова, Сай 2009]).

рого соответствует синтаксической форме «индефинитива» (по И. Ш. Козинскому), или пучку функций так называемого «ректуса» // «прямого падежа» (по В. А. Плунгину); и (б) «**субъектив**» (SBJ), функция которого соответствует синтаксической форме «субъектива» (по И. Ш. Козинскому). Субъектив есть лишь у ничтожно малого количества слов, но зато обладающих столь сверхвысокой употребительностью, что ее достаточно для постулирования особой падежной граммемы. Это **личные местоимения**.

4. Принципы упорядочения морфологических структур в квантитативной модели монгольского языка

В таблицах нашей модели представлены как алфавитный, так и ранговый принцип сортировки данных. Дихотомия («прямой» vs. «инверсионный») имеет ту же функцию для адресата модели, что и так называемая «грамматика порядков» (И. И. Ревзин, Г. Д. Юлдашева). Как известно, порядок следования аффиксов в многоаффиксной словоформе отражает некоторые общие семантико-прагматические закономерности, см. работы Дж. Л. Байби (Хупер). Поэтому порядок грамем в ГН агглютинативных словоформ тоже отражает эти закономерности.

Поэтому максимально наглядным способом презентации цепочек грамем в агглютинативных (а значит, многоаффиксных) словоформах является выбор не однонаправленной подачи, а именно **ободнонаправленной** схемы, предполагающей оба порядка следования: и **прямой** (слева направо), и **инверсионный** (справа налево).

5. Трактовка супraseгментных маркеров письменного текста

Понятие «**супraseгментной граммемы**» охватывает те грамматические значения, которые передаются на письме порядком слов и пунктуацией; она трактуется как комплекс совместно (сопряженно) выраженных элементарных компонентов — «супraseгментных грамем». Как известно, сегментно (в том числе аффиксально) выраженные граммемы определенным образом коррелируют с супraseгментно выраженными грамматическими значениями. Чтобы построить формализованную

количественную модель, отражающую такие (нестрогие) закономерности, целесообразно ввести особое понятие для единицы, соответствующей сочетанию определенной аффиксально выраженной грамматической формы (морфологической граммемы) с определенной супraseгментно выраженной грамматической формой (супraseгментной граммой) — **позиционно-морфологической граммемы**.

6. Общее строение квантитативной модели позиционно-морфологической грамматики ХМЯ

Эта модель включает 4 части. В I части задан набор базовых словоформ (включающий их имена, соотносительность с лексемами, семантическую и грамматико-семантическую субкатегоризацию). Во II части задан набор базовых морфологических грамем. В III части измерена частотность супraseгментных грамем. В IV части измерена частотность базовых позиционно-морфологических грамем.

Заключение

Структурно-вероятностный подход к текстам углубляет наши знания о морфологии ХМЯ, поскольку вхождения обобщенных словоформ и их фрагментов, являющиеся объектами подсчетов, соотносены с морфологическими единицами монгольского языка. Лингвостатистический метод позволяет количественно описывать поведение различных морфологических единиц (обобщенных словоформ, обобщенных лексем и обобщенных грамем) в письменном тексте⁸. На этой основе накапливается обобщенная количественная информация о классах морфологических единиц, о частоте употребления каких-либо грамматических форм в тех или иных синтаксических функциях и т. п. Такая информация позволяет существенно обогатить описание морфологических единиц (ср., напр., такие простейшие констатации, как употребительность деепричастных словоформ, рефлексивно-притяжательных словоформ, комитативных словоформ и т. п.). Тем самым создается перспектива превращения обычной структурной модели морфологии ХМЯ⁹ в структурно-веро-

⁸ Ср. [Шайкевич 1990; Шайкевич 2009].

⁹ Ср. [Крылов 2004].

ятностную модель, в которой учитываются результаты квантитативного анализа текстов (в этой модели единицы ХМЯ обла-

дают «весом», измеряемыми оказываются языковые противопоставления и связи)¹⁰.

Литература

- Баранова, Сай 2009 — *Баранова В.В., Сай С. С.* От составителей // Исследования по грамматике калмыцкого языка (= Acta linguistica petropolitana = Труды Института лингвистических исследований). Т. V. Ч. 2. СПб.: Наука, 2009. С. 7–21.
- Касевич 2011 — *Касевич В. Б.* Введение в языкознание. М.; СПб.: Академия, 2011. 230 с.
- Крылов 2004 — *Крылов С. А.* Теоретическая грамматика современного монгольского языка и смежные проблемы общей лингвистики. Ч. 1. Морфемика, морфонология, элементы фонологической трансформаторики (в аспекте общей теории морфологических и морфонологических моделей). М.: Вост. лит., 2004. 479 с.
- Крылов 2012а — *Крылов С. А.* Структурно-вероятностная модель монгольского языка на базе Генерального корпуса современного монгольского языка // Урало-алтайские исследования. 2012. № 1(6). С. 78–105.
- Крылов 2012б — *Крылов С. А.* The general corpus of the modern Mongolian language and its structural-probabilistic model // Компьютерная лингвистика и интеллектуальные технологии: по мат-лам ежегод. междунар. конф. «Диалог» (Бекасово, 30 мая – 3 июня 2012 г.). Вып. 11 (18). М.: Изд-во РГГУ, 2012. С. 331–341.
- Крылов 2013 — *Крылов С. А.* Опыт изучения современного монгольского языка в количественном освещении // Вопросы языкознания. 2013. № 5. С. 46–57.
- Крылов 2014 — *Крылов С. А.* Теоретическая грамматика монгольского языка и смежные проблемы общей лингвистики.: в 6 ч. Ч. 2. Структурно-вероятностная модель современного монгольского языка (на базе Генерального корпуса современного монгольского языка). М.: Наука, Вост. лит., 2014. 637 с.
- Крылов 2016б — *Крылов С. А.* О принципах синтаксического анализа в сводном корпусе монгольских языков // Гуманитарная наука Юга России: международное и региональное взаимодействие. Мат-лы II Междунар. науч. конф., посвящ. 75-летию Калмыцкого института гуманитарных исследований РАН (г. Элиста, 14–15 сент. 2016 г.). Элиста: КИГИ РАН, 2016. С. 198–199.
- Крылов 2017а — *Крылов С. А.* Сочетаемость монгольских синтетических словоформ: количественный аспект // Вестник Калмыцкого института гуманитарных исследований РАН. 2017. № 4. С. 108–133. DOI: 10.22162/2075-7794-2017-32-4-108-133
- Крылов 2017б — *Крылов С. А.* Монгольские аналитические конструкции в количественном аспекте // Вестник Калмыцкого института гуманитарных исследований РАН. 2017. № 5. С. 155–179. DOI: 10.22162/2075-7794-2017-33-5-155-179¹⁰
- Крылов 2017в — *Крылов С. А.* Монгольские аналитические словоформы: опыт лингвостатистического исследования // Вестник Калмыцкого института гуманитарных исследований РАН. 2017. № 6. С. 79–93. DOI: 10.22162/2075-7794-2017-34-6-79-93
- Крылов 2018а — *Крылов С. А.* Опыт дистрибутивно-статистической классификации монгольских аналитических словоформ // Oriental studies. 2018. №. 2. С. 88–101. DOI: 10.22162/2075-7794-2018-36-2-88-101
- Крылов 2018б — *Крылов С. А.* О гибридных жанрах словарей (на материале монгольского языка) // Востоковедные чтения 2018. Лексикология и лексикография. К 200-летию ИВ РАН. Тезисы докладов межинститутской науч. конф. (Москва, 4–6 апреля 2018 г.) / сост. А. С. Панина. М.: ИВ РАН, 2018. С. 33–34.
- Крылов 2018в — *Крылов С. А.* О гибридных жанрах словарей (на материале монгольского языка) // Труды Ин-та востоковедения РАН. Вып. 19. Проблемы общей и востоковедной лингвистики: Лексикология и лексикография / отв. ред. З. М. Шаляпина, ред.-сост. А. С. Панина. М.: ИВ РАН, 2018. С. 156–165.
- Ляшевская, Шаров 2009 — *Ляшевская О. Н., Шаров С. А.* Частотный словарь современного русского языка (на материалах Нацио-

¹⁰ См. [Крылов 2012а; Крылов 2012а; Крылов 2012б; Крылов 2013; Крылов 2014; Крылов 2016б; Крылов 2017а; Крылов 2017б; Крылов 2017в; Крылов 2018а; Крылов 2018б; Крылов 2018в].

- нального корпуса русского языка). М.: Азбуковник, 2009. (= <http://dict.ruslang.ru/freq.php>).
- Шайкевич 1990 — Шайкевич А. Я. Количественные методы в языкознании // Лингвистический энциклопедический словарь. М.: Сов. энцикл., 1990. С. 231.
- Шайкевич 2009 — Шайкевич А. Я. Количественные методы в языкознании // Большая Российская Энциклопедия. Т. 14 («Киреев — Конго») / гл. ред. С. Л. Кравец. М.: БРЭ, 2009. С. 478.
- Шайкевич и др. 2013 — Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А. Дистрибутивно-статистический анализ языка русской прозы 1850–1870-х гг. Т. 1. М.: Языки славянской культуры, 2013. 504 с.
- Шайкевич и др. 2016 — Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А. Дистрибутивно-статистический анализ языка русской прозы 1850–1870-х гг. Т. 2. М.: Языки славянской культуры, 2016. 850 с.
- Croft 2003 — Croft B. Typology and universals. Cambridge, 2003. 368 p. (In Eng.)
- Dawa et al. 2006 — Dawa I., Husal, Liu Yue, Yue Yao Ming, Uulang, Bai Shuang Cheng, Batsaihan, Arai Y., Mitsunaga M., Isahara H., Nakamura S. Multilingual Text–Speech Corpus of Mongolian // International Symposium on Chinese Spoken Language Processing (ISCSLP 2006. Kent Ridge, Singapore, December 13–16, 2006). Vol. II. Pp. 759–770. ISCA Archive https://www.isca-speech.org/archive_open/archive_papers/iscslp2006/B74.pdf. (In Eng.)
- Lehmann 1982 — Lehmann Ch. Directions for interlinear morphemic translations // Folia linguistica. 1982. Vol. 16. Pp. 193–224. (In Eng.)
- Purev et al. 2003 — Purev J., Hyun Seok Park, Altangerel Ch. Tree adjoining grammars for Mongolian // Proceedings of 3-rd International Conference on East-Asian Language, Processing and Internet Information Technology, EALPIIT 2003. Ulaanbaatar, 2003. Pp. 321–323. (In Eng.)
- Purev et al. 2005 — Purev J., Tsolmon Z., Altangerel Ch., and Cheol-Young O. PC-KIMMO-based Description of Mongolian Morphology // International Journal of Information Processing Systems. 2005. Vol. 1. No. 1. Pp. 41–48. (In Eng.)
- Purev, Odbayar 2008 — Purev J., Odbayar Ch. Corpus Building for Mongolian Language // Proceedings of the 6th Workshop on Asian Language Resources (11–12 January 2008, India). Hyderabad, 2008. Pp. 97–98. (In Eng.)
- Sharoff 2002 — Sharoff S. Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics // Proceedings of Language Resources and Evaluation Conference (LREC02). (May, 2002, Las Palmas, Spain, 2002) // European Language Resources Association (ELRA). 2002. Pp. 447–452 (In Eng.)
- Krylov S. A. Theoretical Grammar of Modern Mongolian and Related Issues of General Linguistics. Vol. 1: Morphemics, Morphonology, Elements of Phonological Transformatomics. Moscow: Vostochnaya Literatura, 2004. 479 p. (In Russ.)
- Krylov S. A. The General Corpus of the Modern Mongolian Language and its structural-probabilistic model. In: Kibrik A. E. (ed.) Computational Linguistics and Intellectual Technologies. Conference Proceedings (Bekasovo; May 30 – June 3, 2012). Vol. 11 (18). Moscow: Russian State University for the Humanities, 2012. Pp. 331–341. (In Eng.)
- Krylov S. A. A structure-and-frequency model of the Mongolian language on the basis of the General Corpus of Modern Mongolian. *Ural-Altaic Studies*. 2012. No. 1(6). Pp. 78–105. (In Russ.)
- Krylov S. A. Investigating modern Mongolian: a quantitative perspective. *Voprosy Jazykoznanija*. 2013. No. 5. Pp. 46–57. (In Russ.)

References

- Baranova V. V., Say S. S. Compilers' note. In: Studies in Kalmyk Grammar. Ser.: Acta Linguistica Petropolitana. Vol. V. Part 2. St. Petersburg: Nauka, 2009. Pp. 7–21. (In Russ.)
- Croft B. Typology and Universals. Cambridge, 2003. 368 p. (In Eng.)
- Dawa I., Husal, Liu Yue, Yue Yao Ming, Uulang, Bai Shuang Cheng, Batsaihan, Arai Y., Mitsunaga M., Isahara H., Nakamura S. Multilingual text — Speech Corpus of Mongolian. In: International Symposium on Chinese Spoken Language Processing (ISCSLP 2006. Kent Ridge, Singapore, December 13–16, 2006). Vol. II. Pp. 759–770. Available at: https://www.isca-speech.org/archive_open/archive_papers/iscslp2006/B74.pdf. (In Eng.)
- Kasevich V. B. An Introduction to Linguistics. Moscow; St. Petersburg: Academia, 2011. 230 p. (In Russ.)

- Krylov S. A. Theoretical Grammar of the Mongolian Language and Related Issues of General Linguistics. In 6 vol. Vol. 2: A Structure-and-Frequency Model of Modern Mongolian. Moscow: Vostochnaya Literatura, 2014. 637 p. (In Russ.)
- Krylov S. A. A Consolidated Corpus of Mongolic Languages: principles of syntactical analysis revisited. In: Humanities in Southern Russia. International and Regional Cooperation. Conference Proceedings (Elista; September 14–15, 2016). Elista: Kalmyk Humanities Research Institute (RAS), 2016. Pp. 198–199. (In Russ.)
- Krylov S. A. Compatibility of Mongolian synthetic word forms: a quantitative aspect. *Bulletin of the Kalmyk Institute for Humanities of the RAS (Oriental Studies)*. 2017. No. 4. Pp. 108–133. (In Russ.) DOI: 10.22162/2075-7794-2017-32-4-108-133
- Krylov S. A. Mongolian analytical constructions: a quantitative perspective. *Bulletin of the Kalmyk Institute for Humanities of the RAS (Oriental Studies)*. 2017. No. 5. Pp. 155–179. (In Russ.) DOI: 10.22162/2075-7794-2017-33-5-155-179
- Krylov S. A. Mongolian analytical word forms: an effort of quantitative research. *Bulletin of the Kalmyk Institute for Humanities of the RAS (Oriental Studies)*. 2017. No. 6. Pp. 79–93. (In Russ.) DOI: 10.22162/2075-7794-2017-34-6-79-93
- Krylov S. A. Hybrid genres of dictionaries revisited: a case study of the Mongolian language. In: Panina A. S. (comp.) *Oriental Studies Readings 2018. Conference Proceedings* (Moscow; April 4–6, 2018). Moscow: Institute of Oriental Studies (RAS), 2018. Pp. 33–34. (In Russ.)
- Krylov S. A. Mongolian analytical word forms: an effort of distributive and statistical classification. *Oriental Studies*. 2018. Vol. 36. Is. 2. Pp. 88–101. (In Russ.) DOI: 10.22162/2075-7794-2018-36-2-88-101
- Krylov S. A. On hybrid dictionaries, exemplified by Mongolian. In: Shalyapina Z. M. (ed.), Panina A. S. (ed., comp.) *Institute of Oriental Studies of the RAS. Transactions*. Vol. 19: Issues of General and Oriental Linguistics. Lexicology and Lexicography. Moscow: Institute of Oriental Studies (RAS), 2018. Pp. 156–165. (In Russ.)
- Lehmann Ch. Directions for interlinear morphemic translations. *Folia linguistica*. 1982. Vol. 16. Pp. 193–224. (In Eng.)
- Lyashevskaya O. N., Sharov S. A. Frequency Dictionary of Modern Russian: Based on the Russian National Corpus. Moscow: Azbukovnik, 2009. Available at: <http://dict.ruslang.ru/freq.php> (In Russ.)
- Purev J., Hyun Seok Park, Altangerel Ch. Tree adjoining grammars for Mongolian. In: Proceedings of 3rd International Conference on East-Asian Language, Processing and Internet Information Technology, EALPIIT 2003. Ulaanbaatar, 2003. Pp. 321–323. (In Eng.)
- Purev J., Odbayar Ch. Corpus building for Mongolian language. In: Proceedings of the 6th Workshop on Asian Language Resources (11–12 January 2008, India). Hyderabad, 2008. Pp. 97–98. (In Eng.)
- Purev J., Tsolmon Z., Altangerel Ch., Chol-Young O. PC-KIMMO-based description of Mongolian morphology. *International Journal of Information Processing Systems*. 2005. Vol. 1. No.1. Pp. 41–48. (In Eng.)
- Sharoff S. Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics. In: Proceedings of Language Resources and Evaluation Conference (LREC02). (May of 2002, Las Palmas, Spain). European Language Resources Association (ELRA), 2002. Pp. 447–452. (In Eng.)
- Shaykevich A. Ya. Quantitative research methods in linguistics. In: *Encyclopedic Dictionary of Linguistics*. Moscow: Sovetskaya Entsiklopediya, 1990. P. 231. (In Russ.)
- Shaykevich A. Ya. Quantitative research methods in linguistics. In: Kravets S. L. (ed.) *Great Russian Encyclopedia*. Vol. 14: 'Киреев – Конго'. Moscow: Bolshaya Rossiyskaya Entsiklopediya, 2009. P. 478. (In Russ.)
- Shaykevich A. Ya., Andryushchenko V. M., Rebetskaya N. A. Russian Prose, 1850s–1870s: Distributive-Statistical Analysis of Its Language. Vol. 1. Moscow: Yazyki Slavyanskoy Kultury, 2013. 504 p. (In Russ.)
- Shaykevich A. Ya., Andryushchenko V. M., Rebetskaya N. A. Russian Prose, 1850s–1870s: Distributive-Statistical Analysis of Its Language. Vol. 2. Moscow: Yazyki Slavyanskoy Kultury, 2016. 850 p. (In Russ.)

