

Хертек Аржаана Борисовна, Ооржак Байлак Чаш-ооловна

## **О MORFOЛОГИЧЕСКОЙ РАЗМЕТКЕ ЭЛЕКТРОННОГО КОРПУСА ТЕКСТОВ ТУВИНСКОГО ЯЗЫКА**

В статье освещается ход работы по снабжению научным аппаратом корпуса текстов тувинского языка. При составлении морфологической разметки авторами учитывались особенности изменения тувинской словоформы, были созданы модели изменяемых именных и глагольных словоформ и списки их словоизменительных аффиксов.

Адрес статьи: [www.gramota.net/materials/2/2012/7-2/57.html](http://www.gramota.net/materials/2/2012/7-2/57.html)

Источник

### **Филологические науки. Вопросы теории и практики**

Тамбов: Грамота, 2012. № 7 (18): в 2-х ч. Ч. II. С. 214-218. ISSN 1997-2911.

Адрес журнала: [www.gramota.net/editions/2.html](http://www.gramota.net/editions/2.html)

Содержание данного номера журнала: [www.gramota.net/materials/2/2012/7-2/](http://www.gramota.net/materials/2/2012/7-2/)

### **© Издательство "Грамота"**

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: [www.gramota.net](http://www.gramota.net)

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: [voprosy\\_phil@gramota.net](mailto:voprosy_phil@gramota.net)

## GRAMMATICAL STEM OF MODERN RUSSIAN SENTENCE IN ORTHOLOGICAL ASPECT

**Elena Mukhamedovna Khakimova**, Ph. D. in Philology, Associate Professor  
*Department of Mass Communication  
South-Ural State University  
khakimova-elena@yandex.ru*

By the material of modern mass media texts the author considers the types of syntactic errors made while composing the grammatical stem of a sentence, analyzes the deviations in the use of subject, predicate, the violations of the correlation between the main parts of a sentence, the contradictions between the form of a grammatical stem and its content, as well as the non-normative expansion of the grammatical core of a sentence, and mentions that these deviations from syntactic standard destroy formal and semantic structures of a Russian sentence reducing the overall level of speech culture.

*Key words and phrases:* grammatical stem; orthology; syntactic norm; syntax error; speech culture.

УДК 8

**Филологические науки**

*В статье освещается ход работы по снабжению научным аппаратом корпуса текстов тувинского языка. При составлении морфологической разметки авторами учитывались особенности изменения тувинской словоформы, были созданы модели изменяемых именных и глагольных словоформ и списки их словоизменяемых аффиксов.*

*Ключевые слова и фразы:* морфологическое аннотирование; морфологическая разметка; морфологический анализатор; тексты; грамматические категории; имя и глагол; тувинский язык.

**Аржаана Борисовна Хертек**, к. филол. н.  
**Байлак Чаш-ооловна Ооржак**, к. филол. н.  
*Научно-образовательный центр «Тюркология»  
Тувинский государственный университет  
arzhaanakhertek@yandex.ru; oorzhakb@mail.ru*

**О МОРФОЛОГИЧЕСКОЙ РАЗМЕТКЕ ЭЛЕКТРОННОГО КОРПУСА  
ТЕКСТОВ ТУВИНСКОГО ЯЗЫКА<sup>©</sup>**

*Работа выполнена при поддержке РГНФ, проект № 11-04-12073в  
«Электронный корпус текстов тувинского языка».*

Работа над созданием Электронного корпуса текстов тувинского языка (ЭКТТЯ) была начата тувинскими лингвистами, сотрудниками Научно-образовательного центра «Тюркология» Тувинского госуниверситета при поддержке Российского гуманитарного научного фонда в 2011 г. В 2012 г. к работе по переводу в электронный формат текстов художественной литературы и публицистики подключились учителя тувинского языка и литературы ряда общеобразовательных школ республики. На сегодняшний день в электронном корпусе тувинского языка содержатся в электронном формате тексты тувинской художественной литературы (проза, поэзия, драматургия), фольклорные тексты, тексты официально-деловых документов, диалектные тексты; морфемно-орфографический словарь тувинского языка (М. В. Бавуу-Сюрюн, С. М. Далаа).

Целью проекта является создание электронного корпуса текстов тувинского языка, систематизированного собрания лингвистических банков данных, предназначенных для последующей комплексной автоматизации научных исследований и прикладных разработок в области тувинского языкознания, реализуемых на персональном компьютере. Основные задачи, которые предполагается решить, заключаются в создании:

- 1) базы данных тувинских текстов современного и советского периодов в локальной версии (формат базы данных *Microsoft Access*) и системы управления базой данных;
- 2) компьютерных программ для автоматизации сбора данных для лингвистических исследований в области тувинского языка (статистический метод);
- 3) базы данных частотных лексем и первичных именных и глагольных основ и системы управления базой данных (электронный словарь частотных лексем и первичных основ);
- 4) сайта электронного корпуса тувинских текстов [4, с. 72].

Корпус некоторого языка – это собрание текстов на данном языке, представленное в электронной форме и снабженное научным аппаратом. Аппарат, «встроенный» в корпус, обычно называется «разметкой» или «аннотацией» корпуса [3, с. 296]. Хорошая разметка позволяет быстро и эффективно найти в корпусе слова (семантическая разметка), грамматические формы слова (морфологическая разметка), конструкции

(морфологическая и синтаксическая разметки), которые нужны исследователю. Именно разметкой в первую очередь отличается электронный корпус от электронной библиотеки.

Из разработанных подходов к морфологической разметке корпуса в качестве приемлемого для разметки текстов на тувинском языке мы выбираем «формально-морфологический», который предполагает, что каждой встреченной в тексте словоформе, отличающейся по набору буквенных знаков от других словоформ, присваивается некоторый ярлык в зависимости от реально стоящей за ней грамматико-семантической или синтактико-семантической информации [2, с. 2]. Таким образом, морфологическая разметка в корпусе текстов тувинского языка представляет собой выделение каждой словоформы и приписывание к ней информации о ее частеречной принадлежности, совокупности ее грамматических признаков, определение ее грамматической формы в некотором контексте.

Для обозначения грамматических признаков вводится система сокращенных грамматических помет на основе латинского алфавита и латинских названий грамматических категорий, используемых в общем языкознании. Например, если в тексте встретилась словоформа *ажылдадым* ‘я работал’ (глагол *ажылда*-, прош. время на =ды, 1-е л. ед. ч. =м), то она будет помечена следующим образом: *Rv=PastI=Isg*.

Для более широкого охвата количества пользователей ЭКТТЯ предполагается, что поиск грамматических форм может производиться как по сокращенным грамматическим пометам на латинском языке, так и по сокращенным терминам на русском и тувинском языках: *DAT* / Д. п. / Б. п.

Грамматические формы и значения, характерные только для тувинского языка, например, пять форм прошедшего времени в морфологической разметке обозначены ярлыками на латинской основе с нумерацией, а обозначения традиционных названий, отражающих семантику форм, принятых в тувинском языкознании, как, например, недавнопрошедшее время (форма на =ды), в морфологической разметке не предусматриваются. Формы прошедшего времени обозначаются как *Past1*, *Past 2...* / Прош. 1, Прош. 2... / Эрт. 1, Эрт. 2 и т.д.

Морфологическая разметка содержит информацию о словоизменительных, но не словообразовательных признаках лексемы. Основы тувинских лексем (непроизводные и производные основы) входят в «Словарь основ тувинского языка», составленный исполнителями проекта. «Словарь основ тувинского языка» будет включен в базу данных, кроме того, он будет использоваться при морфологической разметке для установления границ основ слов.

Словоформа в тувинском языке строится по принципу тюркской агглютинации, наращивания аффиксов к основе с правой стороны, где каждый аффикс имеет стандартное грамматическое значение и занимает определенное положение в этой цепочке морфем. Структура тувинского имени выглядит следующим образом: *хол=дар=ым=ның*, где *хол*= ‘рука’ > *хол=дар*= ‘руки’ > *хол=дар=ым*= ‘мои руки’ > *хол=дар=ым=ның* ‘моих рук’. При помощи морфологических пометок данная словоформа обозначается: *Rn=Pl=POSS1sg=GEN*. Структура тувинского глагола: *чугаала=ж=ып=ты=выс*, где *чугаала*= ‘говорить’ > *чугаала=ж*= ‘букв. говорить вместе’ > *чугаала=ж=ып*= ‘букв. закончить говорить вместе’ > *чугаала=ж=ып=ты*= ‘букв. закончили говорить вместе’ > *чугаала=ж=ып=ты=выс* ‘букв. мы закончили говорить вместе / мы договорились’. Морфологически помеченная эта словоформа выглядит следующим образом: *Rv=Rec=Perf=PastI=IPl*.

Грамматические формы в тувинском языке в зависимости от качества гласных и согласных звуков в конечном слоге основы имеют несколько фонетических вариантов, которые под одной пометой вместе с основными вариантами морфем вводятся в программу. Фонетические варианты присоединяются к словоформе согласно действующему в тувинском языке закону сингармонизма, определяющего правила сочетаемости гласных и согласных звуков в словоформе. Основной его составляющей частью является гармония гласных, которая определяет последовательность гласных в слове в зависимости от качества гласного начального слога в слове [5, с. 84].

В тувинском языке различают небную (палатальную) и губную (лабиальную) гармонии гласных [Там же, с. 90].

Небная гармония гласных проявляется следующим образом: если предшествующий аффиксу гласный основы является гласным заднего ряда (*a* [a], *ы* [y], *o* [o], *у* [u]), то к основе присоединяются аффиксы с заднерядными гласными (*a* [a], *ы* [y], *у* [u]); если предшествующий аффиксу гласный основы является гласным переднего ряда (*и* [i], *э* [ä], *ө* [ö], *у* [ü]), то к основе присоединяются аффиксы с переднерядными гласными (*и* [i], *э* [ä], *у* [ü]) [Там же, с. 45].

Лабиальная гармония гласных в тувинском языке: если предшествующий аффиксу гласный основы является огубленным гласным (*у* [u], *ү* [ü], *o* [o], *ө* [ö]), то к основе присоединяются аффиксы с огубленными узкими гласными (*у* [u], *ү* [ü]); если предшествующий аффиксу гласный основы является неогубленным гласным (*a* [a], *ы* [y], *и* [i], *э* [ä]), то к основе присоединяются аффиксы с неогубленными гласными (*a* [a], *ы* [y], *и* [i], *э* [ä]). Сфера реализации губной гармонии гласных в большинстве тюркских языков значительно уже, чем палатальной, и исключения многочисленны: *үн=де* ‘в голосе’, *көр=ген* ‘видел’.

Сингармонизм включает в себя и консонантный аспект. Ассимиляция согласных бывает прогрессивной и регрессивной [1, с. 98].

Прогрессивная ассимиляция заключается в том, что к основам, оканчивающимся на глухой согласный, присоединяются лишь те варианты аффиксов, начинающихся с согласных, которые имеют в начале глухие согласные: *аът=ты* ‘лошадь’. У аффиксов с начальным звонким или сонорным согласным, при прибавлении их к основам, оканчивающимся на глухие согласные, начальный согласный звук ассимилируется: если он звонкий шумный, то чередуется с соотносительным ему глухим: *кас + =га* > *кас=ка* ‘гусю’; если же

сонорный и не имеет своей глухой пары, то – с глухим шумным, близким по месту образования: *инек* + =лер > *инек=тер* ‘коровы’. У аффиксов, начинающихся с глухого согласного, при прибавлении к основам, оканчивающимся на сонорные, озвончается начальный согласный: *кел*= + =са > *кел=зе* ‘если придет’.

Регрессивная ассимиляция согласных проявляется при прибавлении аффиксов, начинающихся с гласного, к основам, оканчивающимся на шумные глухие согласные, которые в интервокальном положении озвончаются: *хеп*= + =и > *хеви* ‘его одежда’.

В некоторых случаях наряду с прогрессивной ассимиляцией начальные согласные одних и тех же аффиксов подвергаются также и диссимилиации: *мал* + =лар > *малдар* ‘скот’.

Кроме того, в тувинском языке имеются еще и другие фонетические закономерности, которые надо учитывать при морфологической разметке. В процессе морфологического изменения слов по линии словоизменения наблюдаются образование долгих гласных (*төле*=+эр > *төлээр* ‘заплатить’), выпадение узких гласных (*арын* + ы > *арны* ‘его лицо’), выпадение согласных звуков (*даг* + ы > *даа* ‘его гора’). Данные фонетические закономерности будут рассматриваться нами на втором этапе работы над морфологической разметкой корпуса.

В предлагаемой нами морфологической разметке ЭКТТЯ на данном этапе работы определенными пометами обозначены грамматические признаки имени и глагола, структура словоформ и морфемы с возможными фонетическими вариантами. Образцы морфологической разметки имени и глагола в данной статье представлены в виде таблиц (см. Таблицы 1–4).

Таблица 1.

#### Модель изменяемой именной словоформы и набор словоизменяющих аффиксов имени

0 Rn	1 Comit	2 Num (Pl)	3 Poss	4 Case	
				Case	
				Simple declention	Possessive declention
	<i>лыг</i>	<i>лар</i>	1sgm	Gen <i>ның</i>	Gen <i>ның</i>
	<i>лиг</i>	<i>лер</i>	1sgym	Gen <i>ниң</i>	Gen <i>ниң</i>
	<i>луг</i>	<i>нар</i>	1sgim	Gen <i>нуң</i>	Gen <i>нуң</i>
	<i>лүг</i>	<i>нер</i>	1sgym	Gen <i>нуң</i>	Gen <i>нуң</i>
	<i>ныг</i>	<i>дар</i>	1sgym	Gen <i>ның</i>	Gen <i>ның</i>
	<i>ниг</i>	<i>дер</i>		Gen <i>ниң</i>	Gen <i>ниң</i>
	<i>нуг</i>	<i>тар</i>		Gen <i>нуң</i>	Gen <i>нуң</i>
	<i>нүг</i>	<i>тер</i>		Gen <i>нуң</i>	Gen <i>нуң</i>
	<i>дыг</i>			Gen <i>дың</i>	Gen <i>дың</i>
	<i>диг</i>			Gen <i>диң</i>	Gen <i>диң</i>
	<i>дүг</i>			Gen <i>дуң</i>	Gen <i>дуң</i>
	<i>дүг</i>			Gen <i>дуң</i>	Gen <i>дуң</i>
	<i>тыг</i>			Gen <i>тың</i>	Gen <i>тың</i>
	<i>тиг</i>			Gen <i>тиң</i>	Gen <i>тиң</i>
	<i>туг</i>			Gen <i>туң</i>	Gen <i>туң</i>
	<i>түг</i>			Gen <i>ың</i>	Gen <i>туң</i>
				Gen <i>иң</i>	

Таблица 2.

#### Модель изменяемой тувинской глагольной словоформы и набор словоизменяющих аффиксов глагола

0 Rv	1 Rec	2 Distr	3 Perf	4 Neg	5 Tense (Pres, Past, Future) Mood	6 Person (1, 2), Poss
	<i>ж</i>	<i>гыла</i>	Perf <i>ыт</i>	<i>ба</i>	Past 1 <i>ды</i>	1sgm
	<i>ш</i>	<i>гиле</i>	Perf <i>ут</i>	<i>бе</i>	Past 1 <i>ду</i>	2sgn
		<i>гула</i>	Perf <i>ут</i>	<i>на</i>	Past 1 <i>ду</i>	1plwyc
		<i>гүле</i>	Perf <i>ут</i>	<i>не</i>	Past 1 <i>ду</i>	1plwic
		<i>кыла</i>	Perf <i>ыт</i>	<i>ма</i>	Past 1 <i>ты</i>	1plwyc
		<i>кыле</i>	Perf <i>ыт</i>	<i>ме</i>	Past 1 <i>ти</i>	1plwyc
		<i>кула</i>	Perf <i>ыт</i>		Past 1 <i>ту</i>	2plnar
		<i>күле</i>	Perf <i>ыт</i>		Past 1 <i>ту</i>	2plner
						3plnar
						3plner

Таблица 3.

**Модель тувинского причастия и  
набор тувинских словоизменительных аффиксов причастия на =ган**

0 Rv	1 Rec	2 Distr	3 Perf	4 Neg	5 Part	6 Person (1, 2), Poss	7 Case	
							Simple declension	Possessive declension
	ж	гыла	Perfвыт	ба	Part 1 ган	1sgым	Gen ныҢ	Gen ныҢ
	ш	гиле	Perfвут	бе	Part 1 ген	1sgим	Gen ниҢ	Gen ниҢ
		гула	Perfvут	па	Part 1 кан	2sgың	Gen тыҢ	Gen тыҢ
		гуле	Perfvут	пе	Part 1 кен	2sgиң	Gen тиҢ	Gen тиҢ
		кыла	Perfывыт	ма		1plывыс		
		киле	Perfивут	ме		1plивис		
						2plыңар		
						2plиңер		

Таблица 4.

**Модель тувинского деепричастия  
и набор тувинских словоизменительных аффиксов деепричастия на =п**

0 Rv	1 Rec	2 Distr	3 Conv
	ж	гыла	Conv1 п
	ш	гиле	Conv1 ып
		гула	Conv1 ип
		гуле	Conv1 уп
		кыла	Conv1 ҫп
		киле	
		кула	
		күле	

**Необходимые пояснения к таблицам**

В таблицах последовательность следующих за корнем (основой) аффиксов в словоформе представлена пронумерованным обозначением позиций по горизонтали.

Аффиксы-показатели грамматических категорий приводятся со всеми возможными алломорфами.

Одну позицию по горизонтали могут занимать аффиксы одной или нескольких грам. категорий, каждая из которых (кроме *R*) может быть не выражена, например: единственное число, именительный падеж имени.

В глаголе позиции аффиксов совместно-взаимного залога (1-я позиция *Rec*) и многократного вида (2-я позиция *Distr*) могут меняться местами.

**Условные обозначения:**

*Case* – падеж; *Conv* – деепричастие; *Distr* – многократный вид; *Future* – будущее время; *Mood* – наклонение; *Neg* – отрицание; *Num (Pl)* – число (мн.); *Part* – причастие; *Past* – прошедшее время; *Perf* – законченный вид; *Pers* – лицо; *Pres* – настоящее время; *Poss* – принадлежность; *Possessivedeclention* – набор падежных аффиксов притяжательного склонения (после показателя принадлежности); *Rn* – основа имени; *Rv* – основа глагола; *Rec* – совместно-взаимный залог; *Simpledeclention* – набор падежных аффиксов простого склонения; *Tense* – время.

*Список литературы*

- Исхаков Ф. Г., Пальмбах А. А. Грамматика тувинского языка. Фонетика и морфология. М., 1961. 472 с.
- Ляшевская О. Н., Плунгян В. А., Сичинава Д. В. О морфологическом стандарте Корпуса современного русского языка // Сборник научно-технических инноваций. Серия 2. Информационные процессы и системы. 2005. № 6. С. 2–9.
- Плунгян В. А. Зачем мы делаем национальный корпус текстов русского языка? // Отечественные записки. 2005. № 2. С. 296–308.
- Салчак А. Я. Электронный корпус текстов тувинского языка // Тюрко-монгольские народы Центральной Азии: язык, этническая история и фольклор (к 100-летию со дня рождения В. М. Надеяева): материалы международной научной конференции. Кызыл, 2012. С. 72–74.
- Селютина И. Я. Фонологические системы языков народов Сибири. Новосибирск, 2004. 101 с.

## ON TUVINIAN LANGUAGE TEXTS ELECTRONIC CORPUS MORPHOLOGICAL MARKING

Arzhaana Borisovna Khertek, Ph. D. in Philology  
Bailak Chash-oolovna Oorzhak, Ph. D. in Philology  
Scientific-Educational Center “Study of Turkic Languages”  
Tuva State University  
arzhaanakhertek@yandex.ru; oorzhakb@mail.ru

The authors cover the progress of the work on supplying the corpus of the Tuvinian language texts with scientific framework, while making morphological marking they considered the features of the Tuvinian word form change, and created the models of variable nominal and verbal word forms and the lists of their word changing affixes.

*Key words and phrases:* morphological annotation; morphological marking; morphological analyzer; texts; grammatical categories; noun and verb; Tuvinian language.

УДК 8

**Филологические науки**

*В статье изучается проблема метафоры «воинские подразделения», функционирующей в современных китайских и российских спортивных репортажах. В большинстве случаев использование метафорического метода в русских и китайских спортивных статьях совпадает. Эти лексемы применяются как средства речевой выразительности, увеличивая информативную ценность и образность сообщения, многие метафорические образы в русском и китайском лингвокультурном сознании сходны, но существуют некоторые различия в создании и использовании военных метафор в спортивной журналистике. Авторы спортивных текстов стремятся к экспрессии в описании спортивных зрелищ.*

*Ключевые слова и фразы:* метафора войны; воинские подразделения; пресса; армия; спорт.

**Чжэнь Чжэнь**

*Кафедра русского языка*

*Московский педагогический государственный университет*

*zhenzgirl@mail.ru*

### ВОЕННАЯ МЕТАФОРА В РУССКОЙ И КИТАЙСКОЙ СПОРТИВНОЙ ЖУРНАЛИСТИКЕ<sup>©</sup>

Детальная структурированность *воинской иерархии* нередко оказывается востребованной при метафорическом обозначении отношений между субъектами спортивной деятельности (рядовые, командиры, старшины, полководцы и др.). См. развернутую метафору (тренер – полководец, команда – армия, дворовая команда – батальон) в спортивном дискурсе: «*Полководец*, обладающий огромной *армией* и проигрывающий *сражение* одному *батальону*, раньше стрелялся. Тренер, располагающий подбором высококлассных и дорогостоящих футболистов и проигрывающий дворовой команде, должен уйти в отставку» [8].

Под воинскими подразделениями мы понимаем различные войсковые единицы: армия, боевое охранение, взвод, войско, гвардия, главные силы, десант, дивизион, дивизия, дружина, легион, полк, колонна, штаб, эшелон. В текстах спортивных репортажей метафоры этой группы обозначают совокупность людей, прежде всего спортсменов, значительно реже – болельщиков.

Слово *армия* значит: «1. Вооружённые силы государства. 2. Сухопутные вооружённые силы в отличие от морских и воздушных сил. 3. Оперативное войсковое объединение из нескольких корпусов, дивизий. 4. В дореволюционной России: обычные сухопутные войска в отличие от гвардии» [27, с. 29]. В российских спортивных репортажах данное слово используется в значении «5. *перен., кого*. Вообще совокупность большого количества чем-то объединённых людей» [Там же] в сочетании с конкретизатором, выраженным именем существительным в родительном падеже: «*Армия спортсменов* увеличится втрое» [52]; «Смодиш точен с линии штрафных и тут же забивает из-под кольца. “Матъяж...” – кричит ведущий, “Смодиш!” – отвечает пятидесятая *армия болельщиков*» [1]. В китайском языке слово *армия* выступает как синоним слова *команда*: 兰帕德领军的切尔西和小罗领军的巴塞罗那那对决中，切尔西队获胜，但小罗以压倒性的优势获得金球奖。 – «Команда “Челси”, в которой Лэмпард руководит *армией*, играла с командой “Барселона”, и “Челси” выиграла, но Роналдиньо получил кубок “золотой мяч” с подавляющим преимуществом» [9]. 郜林一蹴而就，以5粒进球继续占据卡家军队内头牌射手的位置。 «Гао Линь способствовала удача, и он продолжает занимать первую позицию в своей армии с пятью голами» [37].