

**СООТНОШЕНИЕ ГРАФЕМЫ И ГЛИФА
В ГРАФИЧЕСКОЙ СИСТЕМЕ «ТОДО БИЧИГ»
(«ЯСНОЕ ПИСЬМО») СТАРОКАЛМЫЦКОГО ЯЗЫКА¹**

Проект разработки Национального корпуса калмыцкого языка приобретает особую актуальность в последние годы, поскольку, с одной стороны, реальное функционирование калмыцкого языка ограничено только разговорной сферой коммуникации в среде носителей, принадлежащих к старшему поколению и проживающих в Республике Калмыкия [Баранова 2009; Биткеева 2006], и, с другой стороны, развитие компьютерных технологий уже позволяет создавать подобные электронные ресурсы, аккумулирующие массивы текстов с возможностью расширенного поиска по грамматическим и семантическим пометам (см. подробно: [Куканова 2012]).

Разработка корпуса для любого языка является фундаментальным проектом, и его реализация зависит от выполнения многих задач — теоретических (создание формализованного описания грамматики — морфологии и синтаксиса, семантической разметки и т. д.) и технических (создание сайта, парсера, поискового менеджера, сканирование текстов и др.). К тому же сам корпус может состоять из ряда подкорпусов: основного, газетного, поэтического, параллельного, устного и т.д. Каждая из задач и каждый из подкорпусов могут стать предметом отдельного исследования.

В создании основного подкорпуса калмыцкого языка мы опи-

¹ Статья подготовлена при финансовой поддержке РГНФ в рамках проекта «Национальный корпус калмыцкого языка: создание и разработка» № 12-04-12047/в (2012–2014).

раем на результаты работы по разработке корпусов как в отечественной, так и зарубежной лингвистике [Куканова 2013]. Более новым и совершенно не разработанным направлением в корпусной лингвистике является создание подкорпуса «ранних» текстов, написанных на старокалмыцком языке (на «тодо бичиг» ‘ясном письме’) (см. также: [Бембеев 2012]).

В старописьменном языке почти все (за исключением некоторых) графемы в различных позициях (в начале, середине и в конце) пишутся по-разному. Все графические начертания одной буквы являются аллографами (глифами) и, следовательно, обозначениями аллофонами одной и той же фонемы. В целом, можно констатировать, что на сегодняшний день «Тодо бичиг» — вертикальное письмо — имеет ограниченную поддержку в компьютерной среде: за некоторыми глифами до сих пор не закреплена кодировка Unicode.

Слова разделены пробелом, служащим делиметром между ними, пробелом также разъединяется и написание некоторых словоизменяемых аффиксов. Этот аспект очень важен для нас, поскольку, в отличие от древнерусских текстов, где, переписывая множество документов, писцы не оставляли пробелов между словами и часто пропускали гласные в словах, т. е. писали скорописью² в целях экономии и увеличения скорости. Написание букв различается только по позиции внутри слова. Ср., например, с действием семантического и синтаксического критериев в современном калмыцком языке: с прописной буквы пишутся имена собственные, прописные буквы маркируют начало предложения.

Таким образом, графическая система «тодо бичиг» обладает рядом особенностей, среди которых можно указать следующие:

- 1) вертикальное написание сверху вниз, строки располагаются слева направо;

² При транслитерации аффиксы, которые оформлены через пробел, обозначаются написанием через дефис. См. подробно правила транслитерации текстов, которые были разработаны в рамках реализации одной из задач проекта «Национальный корпус калмыцкого языка: создание и разработка» (№ 12-04-12047/в) [Музраева 2012; Куканова, Бембеев, Музраева 2013].

- 2) сингармонизм («мужские» и «женские» согласные в зависимости от последующей гласной);
- 3) отсутствие прописных букв;
- 4) аллофонный характер письма: три варианта графем (не у всех букв) в зависимости от позиции в слове (начало, середина или конец) и от грамматических правил;
- 5) действие некоторых графических правил (например, необязательное оформление выносного элемента в графеме **ѣ**, если в структуре слова имеется согласная **ѣ**).

В рамках проекта «Национальный корпус калмыцкого языка: создание и разработка», поддержанного Российским гуманитарным научным фондом, запланировано создание основы, своего рода задела, для подкорпуса «ранних» текстов. В соответствии с этой задачей необходимо выяснить, какие проблемы как теоретического, так и практического характера, существуют в области разработки корпуса рукописных в большей своей массе источников, собрать небольшой текстовый материал, на который можно будет опираться при разработке данного ресурса.

Одной из таких проблем является соотношение графемы и глифа. Два этих понятия тесно связаны, имея отношения абстрактное — конкретное. Действительно, графема, являясь абстрактной единицей, получает свое конкретное воплощение в глифах. В лингвистике принято использование термина «аллограф». Данное понятие под названием «графические альтернативы» было введено в научную практику И. А. Бодуэном де Куртенэ [Бодуэн де Куртенэ 1963]. В типографии с развитием компьютерных технологий большее распространение получил термин «глиф».

К примеру возьмем графему «А», скелет (графическая основа) которой состоит из 3 обязательных элементов: двух линий, образующих угол и соединенных при помощи третьей линии. Глифами этой графемы могут выступать: **A**, **A**, **A**, **A**, **A**, **A** и др. К глифам графемы «а» могут относиться: **a**, **a**, **a**, **a** и др. Скелеты строчной и прописной графем легко узнаваемы носителями того или иного языка, поскольку абстрактный образец — графема — имеет прочную связь со всеми своими аллографами.

Термин «графема» трактуется как единица, которая обладает единством содержания и формы, т. е. означаемого и означающего [Журавлев 1990]. Другими словами, по своему начертанию она должна отличаться от графем того же алфавита и передавать один и тот же звук. Глифы одной и той же графемы должны иметь в своей структуре одинаковую основу, своего рода скелет, но при этом они могут отличаться внутри графемы по своей функции (строчные или прописные: «а» и «А», — курсивные или жирные: «А», «а» и «А», «а») или по своей форме («а» и «а»).

Графемы старокалмыцкого письма, в том числе галики, имеют от одного (например ꚗ, Ꚙ) до пяти (например ꚗ¹) вариантов написания в уставной записи, а в неуставных, т. е. в документах, демонстрирующих отход от графической нормы, закрепленной автором этого письма (Зая-пандитой Намкай Джамцо, 1599–1662), можно встретить большое количество глифов. Что касается графем «тодо бичиг», то внутри одного ряда букв они могут отличаться как по своей форме, так и функционально.

Возьмем к примеру графему «ꚗ¹», которая обозначает фонему <a>. Все глифы данной буквы отличаются функционально: каждое начертание обозначает употребление фонемы <a> в разных позициях слова: в инициальной, медиальной и финальной позициях слова. Иначе говоря, они обозначают варианты фонем — аллофоны. Но при этом все эти глифы, обозначая функциональные особенности графемы, выражают это через свою форму: каждый глиф может отличаться от другого глифа как по употреблению в той или иной позиции в слове, так и внутри одной позиции, если имеется несколько вариантов обозначения того или иного аллофона.

Если почти у всех графем «тодо бичиг» есть кодировка Unicode, которая находится в пределах 1800–18AF [The Unicode Standard 2013], то глифы в «тодо бичиг» реализованы в так называемом FREE VARIATION SELECTOR (см. Таблицу). Это, собственно говоря, не кодировка, а лишь способ решения проблемы создания псевдокодировки у контекстных глифов, т. е. глифов в тех графических системах, выбор которых зависит от контекста. Действительно, это очень неудобно, в некоторых целях необходимо использовать тот

или иной глиф в изолированной позиции, но мы не можем этого сделать, поскольку данный глиф не отображается изолированно, а только в контексте, а происходит это потому, что он не имеет своей собственной кодировки. Конечно, не столь важно создавать кодировку для того или иного стилистического глифа, характерного для уставной или неуставной записи. К примеру, тонкое или жирное написание графемы «а» не влияет на кодировку: она остается неизменной. Однако для палеографов, текстологов и других исследователей важно использовать глифы со своей собственной кодировкой в зависимости от позиции для правильного отражения рукописного текста в том виде, в котором он существует.

В некоторых разработанных шрифтах, как, например, Mongolian Baiti, реализован формат OpenType, который может включать широкий спектр глифов и лигатур. В зависимости от правого контекста происходит автоматическая замена глифов. Например, если добавляется следующий символ (не знак пробела) к уже напечатанному, меняется глиф в изолированной позиции на глиф в инициальной позиции, если печатается еще один символ, то происходит еще одна замена глифа в инициальной позиции на глиф в медиальной позиции. Такая постоянная смена символов помимо того, что раздражает глаза, не позволяет установить, правильно ли напечатано то или иное слово, еще до того, как оно не будет отражаться на экране полностью, включая пробел.

Изучение соотношения графемы и глифа принципиально важно для создания распознающей графические начертания программы, которая увеличила бы объем будущего подкорпуса «ранних» текстов. Однако для распознающей программы необходимо, чтобы графемы имели поддержку Unicode. Использование шрифтов, поддерживаемых системой Unicode, принципиально важно для филологов: это лучший способ избежать проблем при их публикации как в бумажном, так и электронном виде [Rabus 2009: 91]. В настоящее время ведется работа в архивах Республики Калмыкии по оцифровке материалов писем калмыцких ханов XVIII в., которая заключается в транслитерировании текстов в специализированной программе.

Литература

Rabus A. Unicode and OpenType — a practical approach to producing Church Slavonic scientific editions // Письменное наследие и современные информационные технологии: сб. ст. лекторов международной научной школы для молодежи (Ижевск, 12–15 октября 2009 г.) / отв. ред. В. А. Баранов. Ижевск: Удмуртия, 2011. С. 89–105.

The Unicode Standard 6.3. 1991–2013 [электронный ресурс] // <http://www.unicode.org/Public/6.3.0/charts/CodeCharts.pdf> (дата обращения: 25.06.2013). 2100 p.

Баранова В. В. Языковая ситуация в Калмыкии: социолингвистический очерк // Исследования по грамматике калмыцкого языка / Ред. С. С. Сай, В. В. Баранова, Н. В. Сердобольская. (ACTA LINGUISTICA PETROPOLITANA. Труды Института лингвистических исследований РАН, 2009. Том V, часть 2). С. 22–42.

Бембеев Е.В. Опыт квантитативной обработки текста на старокалмыцком языке: количественные характеристики // Вестник Калмыцкого института гуманитарных исследований РАН. 2012. № 2. С. 163–168.

Биткеева А. Н. Социальные функции национального языка в современном мире: ойрат-калмыцкий язык / дисс. М., 2006. 435 с.

Бодуэн де Куртэнэ И. А. Об отношении русского письма к русскому языку // Избранные труды по общему языкознанию: в 2 тт. Т. 2. М.: Изд-во Акад. наук СССР, 1963. С. 208–235.

Журавлев В. К. Графема // Лингвистический энциклопедический словарь / под ред. В. Н. Ярцевой; Ин-т языкознания АН СССР. М.: Сов. энцикл., 1990. С. 117.

Куканова В.В., Бембеев Е.В., Мулаева Н.М., Очирова Н.Ч. Национальный корпус калмыцкого языка: архитектура и возможности использования // Вестник Калмыцкого института гуманитарных исследований РАН. 2012. № 3. С. 138–150.

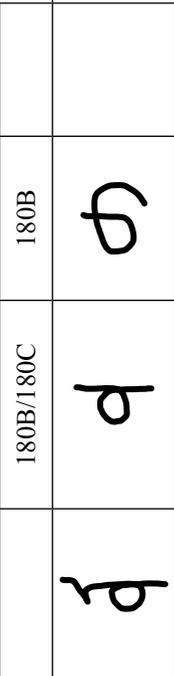
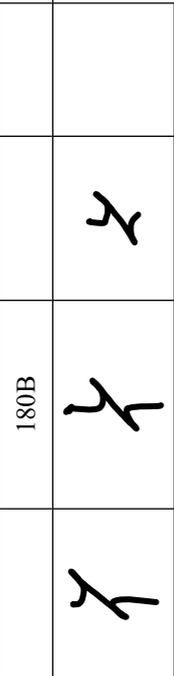
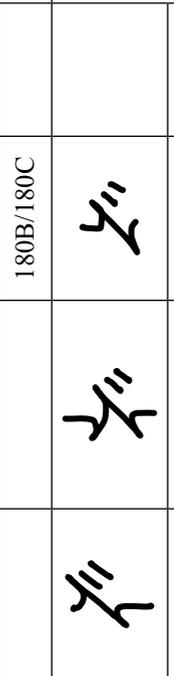
Куканова В. В., Бембеев Е. В., Музраева Д. Н. К вопросу о кодировке UNICODE графической системы «тодо бичиг» и создании базы данных текстов на старокалмыцком языке // Вестник Калмыцкого государственного университета. 2013. № 3. С. 65–75.

Музраева Д. Н. Опыт археографического описания и текстологического анализа рукописного перевода Тугмюд-гавджи (на материале VI главы Oülgun dalai «Моря притч») // Вестник Калмыцкого института гуманитарных исследований РАН. 2012. № 3. С. 167–185.

Таблица 1. Основные типы глифов графем старокалмыцкого письма («тодо бичиг»)

№	Латиница	Название графемы	Глиф/Unicode					Кол-во глифов
			Изолированная позиция	Инициальная позиция	Медальная позиция	Финальная позиция	Другие написания	
1.	a	MONGOLIAN LETTER TODO A	1820					5
				180B/180C	180B			
2.	n		1828					4
				180B	180B/180C			
3.	x	MONGOLIAN LETTER TODO QA	184D					2
					180B			

4.	γ	MONGOLIAN LETTER TODO GA	184E			—		2
					180B			
5.	i	MONGOLIAN LETTER TODO I	1845					4
					180B			
6.	l	MONGOLIAN LETTER TODO LA	182F					3
7.	m	MONGOLIAN LETTER TODO MA	182E					3
8.	o	MONGOLIAN LETTER TODO O	1846					3
					180B			

9.	u	MONGOLIAN LETTER TODOU	1847		3
10.	ü	MONGOLIAN LETTER TODOUE	1849		3
11.	s	MONGOLIAN LETTER TODO SA	1830		3
12.	š	MONGOLIAN LETTER TODO SHA	1831		3
13.	d	MONGOLIAN LETTER TODO DA	1851		4

14.	t	MONGOLIAN LETTER TODO TA	1850		—		2
15.	e	MONGOLIAN LETTER TODO E	1844		180B		2
16.	y	MONGOLIAN LETTER TODO YA	1855		—		2
17.	r	MONGOLIAN LETTER TODO RA	1837				5

18.	ö	MONGOLIAN LETTER TODO OE	1848	  	3
19.	g/q	MONGOLIAN LETTER TODO GAA	1858	  	4
20.	k	MONGOLIAN LETTER TODO KA	1857	  	2
21.	k	MONGOLIAN LETTER TODO	183B	  	2
22.	b	MONGOLIAN LETTER TODO BA	182A	  	4

23.	P	MONGOLIAN LETTER TODO PA	184C	𐎆		𐎆	3
24.	c	MONGOLIAN LETTER TODO TSA	1854	𐎇	—		1
25.	z	MONGOLIAN LETTER TODO JA	1853	𐎈	—		1
26.	č	MONGOLIAN LETTER TODO CHA	1852	𐎉	—		1

27.	j	MONGOLIAN LETTER TODO JIA	185A			—		3
28.	v	MONGOLIAN LETTER TODO WA	1856			—		2
29.	ŋ	MONGOLIAN LETTER TODO ANG	184A	—				2

1829