

Глухов В. А., Глухова Н. Н.

ПАРАМЕТРЫ АУТЕНТИЧНОСТИ МАРИЙСКОГО АЛФАВИТА

Адрес статьи: www.gramota.net/materials/1/2007/3-1/26.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2007. № 3 (3): в 3-х ч. Ч. I. С. 68-71. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2007/3-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

ПАРАМЕТРЫ АУТЕНТИЧНОСТИ МАРИЙСКОГО АЛФАВИТА

Глухов В. А., Глухова Н. Н.
Марийский государственный технический университет
Марийский государственный университет

У любого алфавита и использования его букв/звуков для создания слов существует, как минимум, три количественных показателя его аутентичности: *частота вероятности использования букв тексте, средняя длина слова и длина алфавита.*

Цель статьи - показать аутентичность марийского алфавита, определив его место в системе алфавитов языков стран мира с учетом указанных параметров.

Мари (около 600,000 в 2002) - группа волжских финнов по языковым характеристикам - один из финно-угорских народов, живущих в Европейской части России в Республике Марий Эл, а также в некоторых других регионах Российской Федерации. Марийский язык относится к группе языков с небольшой письменной традицией, так как самые ранние письменные памятники относятся к первой половине XVIII века. Письменность создана на основе русского алфавита с использованием кириллицы. Существующий алфавит, по мнению авторов, не может передать национальной специфики, но частота использования ряда букв в текстах, представленных разными жанрами, может указать на наличие преобладающих букв, передающих звуки, наиболее близкие национальной ментальности.

В разных языках буквы алфавита в текстах используются с различной частотностью. Существует возможность статистически определить главные, часто используемые, и малоиспользуемые буквы в тексте. Результаты такого анализа представлены в работе авторов «Системная реконструкция марийской этнической идентичности» [Глухов, Глухова 2007: 118-128]. Выявленная последовательность распределения букв различных групп частотности характерна только для марийского языка. Сравнение было проведено для русского языка с учетом результатов статистических исследований, полученных Р.Г.Пиотровским [Пиотровский 1977: 120], и использованы уже известные данные для английского языка [Бриллоэн 1960: 24]. Привлечение английского языка для сравнения объясняется тем, что исторически такие статистические исследования были впервые проведены с английским алфавитом в связи с декодированием зашифрованных сообщений во время второй мировой войны и позже [Черчхаус 2005:30-32].

Статистический анализ текста на марийском языке показал распределение букв марийского алфавита на четыре группы. К главным буквам относятся буквы **А, Ы, Е, Н, Т**, так как они составляют 50,2% использования. Вспомогательные буквы - **М, К, О, Л, Ш, Р** - употребляются с частотностью 28,1 %. Дополнительные буквы - **И, Й, Д, У, Ч** составляют 11,6%. На долю оставшихся букв выпадает 10,1 %. Составление такого распределения для марийского языка позволяет получить много неочевидной, но полезной и ценной информации и может быть использовано на практике, в частности теории создания названий, нейминга [Глухов Глухова 2007: 118-128].

Следующим показателем аутентичности алфавита языка является средняя длина слова (в тексте). Существуют два способа определения средней длины слова. Первый - заключается в подсчете длин всех слов в тексте какого-либо определенного объема и последующего вычисления среднего арифметического. Этот способ весьма трудоемок. Второй способ основан на предположении, что в тексте определенной длины количество слов равно количеству пробелов. Размеры средней длины слова в текстах в трех языках - марийском, русском и английском - вычислены по простой формуле и составляют: 7,5 буквы для марийского языка, 5,75 буквы для русского языка, 5 букв для английского языка [Глухов, Глухова, 2007: 134-135].

Третьим параметром, анализируемом в статье, является *длина алфавита*. Определению длины алфавита марийского языка способствовал статистический анализ длин алфавитов языков стран мира, который был проведен с использованием двух обобщающих работ. К первой относится справочное издание Р.С. Гиляревского и В.С. Гривнина «*Определитель языков мира по письменности*» [Гиляревский, Гривнин 1961]. Вторая монография «*История письма*» написана И.Фридрихом [Фридрих 1979]. В указанных источниках удалось найти 193 алфавита с длиной от 15 до 72 знаков. Такой большой «разброс» длин свидетельствует о существовавшем разнообразии в условиях формирования письменности и сложной истории этих процессов. С позиции статистики, число 193 незначительно и не может привести к точному результату, но вряд ли можно его значительно увеличить, ведь это почти все существующие алфавиты мира. Данные анализа сведены в Таблицу 1.

Таблица 1. Распределение вероятностей количества букв в алфавитах мира

№	Длина алфавита	Количество алфавитов	Вероятность	№	Длина алфавита	Количество алфавитов	Вероятность
1	15	2	0,01	29	43	2	0,01
2	16	0	0	30	44	1	0,005
3	17	0	0	31	45	2	0,01
4	18	1	0,005	32	46	2	0,01
5	19	1	0,005	33	47	2	0,01
6	20	3	0,016	34	48	0	0
7	21	4	0,021	35	49	1	0,005
8	22	6	0,031	36	50	3	0,016
9	23	6	0,031	37	51	2	0,01
10	24	6	0,031	38	52	0	0
11	25	10	0,052	39	53	0	0
12	26	10	0,052	40	54	1	0,005
13	27	1	0,005	41	55	0	0
14	28	5	0,026	42	56	2	0,01
15	29	11	0,057	43	57	0	0
16	30	8	0,042	44	58	1	0,005
17	31	9	0,047	45	59	1	0,005
18	32	6	0,031	46	60	0	0
19	33	9	0,047	47	61	0	0
20	34	7	0,036	48	62	1	0,005
21	35	9	0,047	49	63	0	0
22	36	14	0,073	50	64	0	0
23	37	7	0,036	51	65	0	0
24	38	8	0,042	52	66	1	0,005
25	39	14	0,073	53	67, 68, 69	0	0
26	40	6	0,031	54	70, 71, 72	0	0
27	41	2	0,01	55	73	1	0,005
28	42	5	0,026	56			

На основании данных таблицы была составлена гистограмма распределения вероятностей количества букв в алфавитах языков стран мира (Рис. 1).

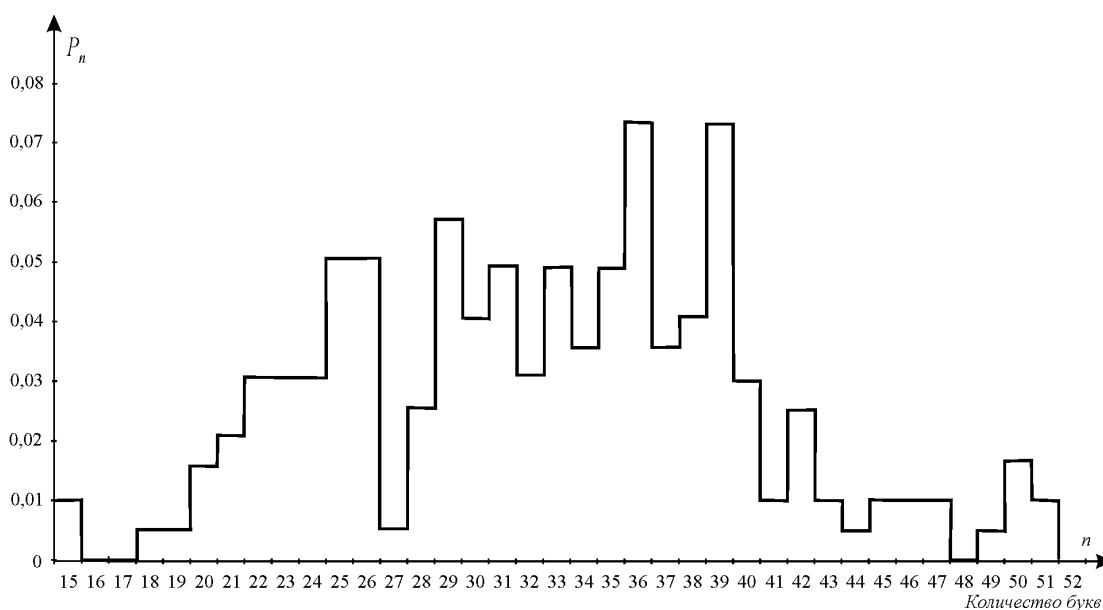


Рис. 1. Гистограмма распределения вероятностей количества букв в алфавитах языков стран мира. Математическое ожидание $M(n) = 33,31$ дисперсия $D(n) = 81,42$.

Полученное распределение имеет сложную форму с «пиками» и «провалами», что следует отнести к недостаточности статистического материала. Количественно математическое ожидание на гистограмме является центром распределения. Математическим ожиданием, или средним значением, $M(X)$ дискретной случайной величины X называется сумма произведений всех ее значений на соответствующие им вероятности:

$$M(X) = \sum_{i=1}^n x_i p_i.$$

Происхождение термина «математическое ожидание» связано с начальным периодом возникновения

теории вероятностей, когда область ее применения ограничивалась азартными играми. Игрока интересовало среднее значение ожидаемого выигрыша или, иначе, математическое ожидание выигрыша [Кремер 2004:98].

В применении к гистограмме распределения вероятностей (Рис. 1) это означает, что суммы площадей столбцов справа и слева от него равны. *Математическое ожидание* этого распределения примерно равно 32. Вполне возможно, что предлагаемый здесь статистический анализ длин алфавитов был уже проделан создателями универсального международного языка - *эсперанто*, поскольку они приняли длину его алфавита, равную 32 буквам.

Марийский луговой литературный язык имеет 36, марийский горный литературный язык - 37 букв, русский - 33 буквы. Все эти значения весьма близки к значению математического ожидания, поэтому их можно считать гармоничными по длине. Английский язык, по мнению авторов, несколько коротковат - 26 букв, поэтому для выражения в нем ряда отдельных звуков используются сочетания нескольких букв.

Вторым количественным параметром любого распределения является дисперсия. Она характеризует степень рассеяния длины алфавитов относительно математического ожидания.

$$D(n) = \sum_{k=1}^n (n_k - M(n))^2 \cdot P_k = 81,42$$

Количественное значение дисперсии получилось в 2,44 раза больше математического ожидания, что говорит о сложной истории создания алфавитов языков стран мира и об их большом разнообразии.

Зная гармоничную длину алфавита в 32 буквы, можно предложить следующие выводы. В марийском и русском алфавитах нет «лишних» букв, редко используемые буквы повышают при необходимости информативность слова: она тем выше, чем больше в нем букв, встречающихся с малой вероятностью. Такие буквы необходимы также для передачи иностранных слов, заимствованных из других языков через письменные каналы. Новые буквы в марийском и русском алфавитах не нужны, а в английском, смеем предположить, - будут полезны для замены сочетаний нескольких букв и упрощения орфографии.

Наглядность гистограммы распределения вероятностей количества букв в алфавитах языков стран мира может быть повышена за счет операции усреднения, т.е. уменьшения числа столбцов с помощью группировки по несколько столбцов и вычисления их среднеарифметического значения вероятности. Такая операция была проделана с данными Таблицы 1, при этом суммирование производилось для каждого трех последующих столбцов, результат помещен в Таблицу 2 и отображен на Рис. 2.

Таблица 2. Распределение вероятности количества букв в алфавитах мира, усредненное в 3 раза

№	Длина алфавита	Количество алфавитов	Вероятность	№	Длина алфавита	Количество алфавитов	Вероятность
1	15-17	2	0,01	11	45-47	6	0,031
2	18-20	5	0,026	12	48-50	4	0,021
3	21-23	16	0,083	13	51-53	2	0,01
4	24-26	26	0,135	14	54-57	3	0,016
5	27-29	17	0,09	15	58-60	2	0,01
6	30-32	23	0,12	16	61-63	1	0,005
7	33-35	25	0,13	17	64-66	1	0,005
8	36-38	29	0,15	18	67-69	0	0
9	39-41	22	0,111	19	70-73	1	0,005
10	42-44	8	0,042				

На основании данных таблицы была составлена гистограмма усредненного распределения вероятностей количества букв в алфавитах мира (Рис. 2).

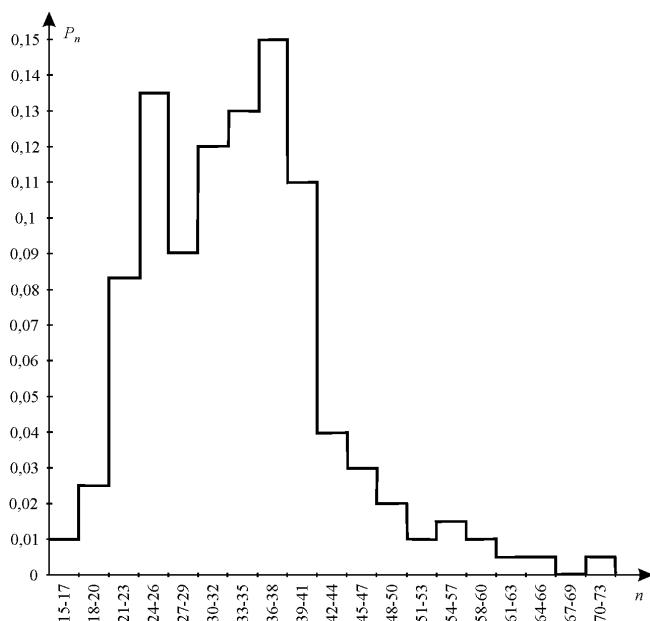


Рис. 2. Усредненная гистограмма распределения вероятностей количества букв в алфавитах мира.

Из таблицы и гистограммы отчетливо видно, что выделяются два максимума для длин алфавитов в 24-26 букв и 36-38 букв. Эти «максимумы» отражают длительный путь развития письменности народов мира на основе латинского алфавита и на основе кириллицы. По-видимому, такие длины алфавитов являются во многих смыслах оптимальными, поскольку именно они успешнее других выдержали трудный путь эволюции письменности. Проведенный статистический анализ показывает, что марийский алфавит с учетом его длины - оптимален, так как находится на одном из максимумов. Как показал известный марийский ученый-фонетист Л.П. Грузов, число букв марийского алфавита превышает число имеющихся фонем. Но при этом существует большое взаимнооднозначное соответствие между буквами алфавита и фонемами [Грузов 1960: 43; 78; 146-150]. Длина марийского алфавита гармонична, т.е. совпадает со вторым максимумом распределения вероятностей длин алфавитов мира, и обеспечивает письменную фиксацию всех необходимых этносу звуков.

Противоположная ситуация сложилась в английском языке с его значительно меньшей длиной алфавита. В нем для передачи одного звука требуется несколько букв. Поэтому часто употребление одной буквы подразумевает вслед за ней появление другой, вполне определенной буквы ($s \rightarrow h$, $t \rightarrow h$, $c \rightarrow h$). Поскольку буквы часто употребляются вместе, их вероятности становятся близки. Если таковых букв много, то у показательной функции на каком-то участке проявляется почти горизонтальное плато, состоящее из ≈ 8 букв для английского языка: **T, O, A, N, I, R, S, H**. В марийском языке плато образуют буквы: **Н, Т, М, К, О, Л, Ш** - 7 букв. Математически, этот участок надо описывать равномерным распределением. Для русского языка это нехарактерно. Русский алфавит, по сравнению с марийским и английским, лучше описывается показательной функцией. При анализе алфавита на нем не видно близкой к равномерному распределению «площадки» в районе номеров от 4 до 10. Это свидетельствует о том, что русская речь представляет собой простейший поток без последствия, т.е. вероятность последующей буквы не зависит от вероятности предыдущей.

Финно-угорские алфавиты - преимущественно «длинные», т.е. относятся ко второму максимуму гистограммы. Среди них можно построить следующий рейтинг по степени близости к гармоничной длине алфавита: мордовский - 33, манси - 34, ханты шурышкарские - 34, коми - 35, марийский луговой - 36, марийский горный - 37, удмуртский - 38, венгерский - 38, ханты казымские - 40, ханты ваховские - 44 буквы. Только два финно-угорских народа имеют «короткие» алфавиты, основанные на латинском алфавите: эстонский - 25 и финский - 21 букву. Для компенсации этой особенности в речи и на письме в них существуют удвоенные гласные и согласные.

Путем анализа частотности использования букв во всех финно-угорских языках можно количественно установить степень их родства с помощью вычисления коэффициентов корреляции между распределениями вероятности использования различных букв.

Предложенный алгоритм анализа может быть с большой степенью достоверности применен для определения аутентичности алфавитов других этносов.

Список использованной литературы

1. Бриллюэн Л. Наука и теория информации. - М.: Гос. изд-во физико-математической литературы, 1960. - 392 с.
2. Гиляревский Р.С., Гривнин В.С. Определитель языков мира по письменности. - М.: Изд-во вост. литературы. - 1961. - 303 с.
3. Глухов В.А., Глухова Н.Н. Системная реконструкция марийской этнической идентичности. - Йошкар-Ола, 2007. - 184 с.
4. Грузов Л.П. Современный марийский язык. Фонетика. - Йошкар-Ола: Марийское книжное изд-во, 1960. - 162 с.
5. Кремер Н.Ш. Теория вероятностей и математическая статистика. - М.: Юнити. - 574 с.
6. Пиотровский Р.Г. и др. Математическая лингвистика. - Уч. пособие для пед. ин-тов. - М.: «Высшая школа», 1977. - 383 с.
7. Фридрих И. История письма /пер. с нем. - М.: «Наука», 1979. - 463 с.
8. Черчаус Роберт. Коды и шифры: Юлий Цезарь, «Энигма» и Интернет. - М.: Весь мир, 2005. - 302 с.

СПОСОБЫ ЭКСТАРИАРИЗАЦИИ ВНУТРЕННЕЙ РЕЧИ В ТЕКСТЕ ХУДОЖЕСТВЕННОГО ПРОИЗВЕДЕНИЯ

Глухов Г. В., Кириллова Т. В.

Самарский государственный экономический университет

Человек - единственное живое существом, наделенное даром речи и языковым сознанием. В связи с этим, проблемы языка и языкового сознания, мышления и речи, их взаимосвязанности и взаимообусловленности представляют большой интерес в современном научном сообществе.

В данной статье мы затронем аспект взаимодействия таких сложных процессов, как мышление и речь. Причем нас интересует речь не внешняя, легко поддающаяся регистрации и анализу, а речь внутренняя, непроизносимая, неуловимая, которая неразрывно сопровождает процесс мышления. Мы также рассмотрим различные модели экстариаризации внутренней речи (ВР) в тексте художественного произведения.

В течение многих лет изучением внутренней речи активно занимались представители различных областей знания: нейрофизиологи, психологи, психолингвисты, но, к сожалению, гораздо реже лингвисты (И.Р. Гальперин, В.В. Виноградов и некоторые другие). В силу этого обстоятельства лингвистические ис-