

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУАЗИЯ
ҰЛТТЫҚ УНИВЕРСИТЕТІ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН
ЕВРАЗИЙСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ Л.Н. ГУМИЛЕВА

MINISTRY OF EDUCATION AND SCIENCE
OF THE REPUBLIC OF KAZAKHSTAN
L.N. GUMILYOV EURASIAN NATIONAL UNIVERSITY



16-18 маусым
Нұр-Сұлтан, 2022

«TURKLANG 2022»

«Түркі тілдерін компьютерлік өңдеу»
атты X халықаралық конференция
ЕҢБЕКТЕРІ

ТРУДЫ

X Международной конференции
«Компьютерная обработка тюркских языков»

«TURKLANG 2022»

PROCEEDINGS

of the X International Conference
on Computer processing of Turkic Languages

«TURKLANG 2022»

**ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ**

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН
ЕВРАЗИЙСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ Л.Н. ГУМИЛЕВА**

**MINISTRY OF EDUCATION AND SCIENCE OF
THE REPUBLIC OF KAZAKHSTAN
L.N. GUMILYOV EURASIAN NATIONAL UNIVERSITY**

**«TURKLANG 2022»
«Түркі тілдерін компьютерлік өңдеу»
атты X халықаралық конференция
ЕҢБЕКТЕРІ
16-18 маусым 2022 ж.**

**ТРУДЫ
X Международной конференции
«Компьютерная обработка тюркских языков»
«TURKLANG 2022»
16-18 июня 2022 г.**

**PROCEEDINGS
of the X International Conference
on Computer processing of Turkic Languages
«TURKLANG 2022»
16-18 June 2022**

Нұр-Сұлтан, 2022

УДК 80/81:004
ББК 81.2:32-973
Т 90

Техникалық редакция:

Ергеш Б.Ж.
Елибаева Г.К.
Турсынова Н.А.

Т 90 ТҮРКІ ТІЛДЕРІН КОМПЬЮТЕРЛІК ӨНДЕУ. X халықаралық конференция: Еңбектері = КОМПЬЮТЕРНАЯ ОБРАБОТКА ТЮРКСКИХ ЯЗЫКОВ. X международная конференция: Труды. / - Нұр-Сұлтан: «Булатов А.Ж.» ЖК, 2022.= Нур-Султан: ИП «Булатов А.Ж.»

ISBN 978-601-326-645-9

Жинақта «Түркі тілдерін компьютерлік өңдеу» атты X халықаралық конференция қатысушыларының баяндамалары енген.

Компьютерлік лингвистика бағыты бойынша оқитын студенттерге, магистранттарға, докторанттарға және мамандарға арналған.

Жинақ «BR11765535» Қазақ тілі мәдениетін арттыру және функцияларды кеңейту бойынша ғылыми-лингвистикалық негіздер мен IT-ресурстарды әзірлеу» бағдарламасы есебінен жарияланды.

В сборнике представлены доклады участников X международной конференции «Компьютерная обработка тюркских языков».

Предназначен для студентов, магистрантов, докторантов и специалистов специализирующихся в областях компьютерной лингвистика.

Сборник издан за счет средств программы BR11765535 «Разработка научно-лингвистических основ и IT-ресурсов по расширению функций и повышению культуры казахского языка».

УДК 80/81:004
ББК 81.2:32-973

ISBN 978-601-326-645-9

© Л.Н.Гумилев атындағы Еуразия ұлттық университеті, 2022

© Евразийский национальный университет им. Л.Н. Гумилева, 2022

Шарипбай Алтынбек Амирович
Евразийский национальный университет им. Л. Н. Гумилева
д.т.н, профессор кафедры
Технологии искусственного интеллекта
Нур-Султан, Казахстан
sharalt@mail.ru

ПРОБЛЕМЫ СОЗДАНИЯ АЛФАВИТА КАЗАХСКОГО ЯЗЫКА НА ОСНОВЕ ЛАТИНСКОЙ ГРАФИКИ

Аннотация: В статье даются сведения о письменности казахского языка и её роли, доказывається некорректность действующего алфавита на основе кириллицы и обосновывается необходимость проведения реформы казахской письменности, показывается ошибки и противоречия предложенного на утверждение шестого варианта алфавита на основе латиницы, уточняется звуковая система казахского языка и предлагаются два варианта алфавита на основе латинской графике, первый на основе алфавита турецкого языка с диакритическими знаками, а второй на основе алфавита английского языка с диграфами.

Ключевые слова: аллофон, диакритика, диграф, орфография, фонема, которые определяются:

Аллофон – смыслонеразличительная и обозначаемая знаком, который обозначает в алфавите созвучную с ней фонему, минимальная фонетическая единица языка.

Диакритика – знак, обозначающий созвучные фонемы, передающий фонетические свойства звука, ударение и т. п. и помещаемый над, под или рядом с буквой.

Фонема – смыслоразличительная и обозначаемая отдельным знаком в алфавите минимальная фонетическая единица языка.

Орфография – раздел лингвистики, состоящий систему правил для единообразной передачи слов и грамматических форм речи на письме.

Sharipbay Altynbek Amiruly
L.N.Gumilyov Eurasian National University
Doctor of Technical Sciences
Professor of the Artificial Intelligence Technologies Department
Nur-Sultan, Kazakhstan
sharalt@mail.ru

PROBLEMS OF CREATING THE ALPHABET OF THE KAZAKH LANGUAGE BASED ON THE LATIN GRAPHICS

Abstracts: In the article provides information about the writing of the Kazakh language and its role, proves the incorrectness of the current alphabet based on the Cyrillic alphabet and substantiates the need for a reform of the Kazakh writing system, shows errors and contradictions of the sixth version of the alphabet based on the Latin alphabet proposed for approval. Are clarified the sound system of the Kazakh language and suggests two versions of the alphabet based on the Latin script: the first based on the Turkish alphabet with diacritics, the second based on the English alphabet with digraphs.

Keywords: Phoneme, allophone, diacritics, digraph, spelling.

1. Введение

Казахский язык является одним из тюркских языков, имеющий древнюю историю и формировавшийся на протяжении многих веков как и языки родственных тюркских племён, проживавших на территории современного Казахстана [1, 2]. Поэтому он, как и любой древний язык, нуждается в исследовании его истории, естественных лингвистических особенностей и возможностей.

Казахский язык является национальным языком казахов, проживающих в Республике Казахстан и во многих странах мира [3]. Поэтому его, как и любой национальный язык, служащий основным средством общения носителей языка, надо изучать и развивать для использования его в ежедневной жизни казахского народа.

Казахский язык является государственным языком Республики Казахстан. Поэтому он, как любой государственный язык, должен служить средством общения народов, населяющих территорию Казахстана, поддерживаться государством и использоваться во всех сферах его деятельности, в том числе и во внешних отношениях с другими странами.

Казахстан, интегрируясь в мировое сообщество, должен обеспечивать достаточный (по отношению к мировому уровню) уровень развития государственного языка.

В настоящее время казахи, живущие в разных странах, используют разные алфавиты, основанные на арабице, латинице и кириллице, что мешает их письменному общению и формированию своего единого культурного и информационного пространства.

В эпоху цифровизации во многих странах обсуждается проблема использования единого алфавита в глобальном информационном пространстве, в качестве которого предлагается принять английский алфавит, чтобы сократить затраты на поиск и обработку информационных ресурсов, объем которых растет очень быстро. Кроме того, все производимые в мире компьютеры и другие электронные устройства поддерживают базовый латинский алфавит. Если же алфавит некоторого языка имеет хотя бы одну букву, отличную от буквы базового латинского алфавита, то, чтобы работать на этих аппаратах с алфавитом этого языка, нужно дополнительно создавать шрифты, драйвера, программы сортировки и поиска данных, которые требуют немалых интеллектуальных, временных, трудовых и финансовых затрат.

2. Казахская письменность

Казахская письменность берет свое начало с древнетюркских рунических письменностей VI–X вв. н.э. Один из вариантов древнетюркских алфавитов, знаками которого высечена орхоно-енисейская надпись на надгробном памятнике Кюль-Тегина, впервые был расшифрован (определены значения знаков алфавита) в 1893 году шведским ученым Вильгельмом Томсеном [4].

Близкие к орхоно-енисейским надписям рунические надписи V–III вв. до н.э. были обнаружены в Казахстане в 50 км от г. Алматы, в Иссыкском кургане, в котором был захоронен «Золотой человек». Эти надписи называются «Иссыкское письмо», состоят из 26 знаков, внешне напоминающих орхоно-енисейскую письменность. Поэтому некоторые исследователи признают вероятность того, что потомком Иссыкского письма является более позднее классическое орхоно-енисейское письмо [5].

В связи с распространением ислама в письменном языке тюркского мира с IX по XX век н.э. на протяжении многих веков использовался арабский алфавит, так как на нём была написана священная книга мусульман – Коран. На этом же алфавите многие ученые, педагоги и

поэты более тысячи лет создавали свои бессмертные произведения, вошедшие в мировую сокровищницу науки, образования и культуры.

Однако арабский алфавит, созданный для семитских языков и хорошо приспособленный к требованиям арабского языка, не до конца отражал богатую фонетическую систему тюркских языков: ряд знаков в нем не был нужен тюркским языкам, и, наоборот, немало звуков, имеющиеся в тюркских языках, не находило в нем отражения. В результате появилась необходимость внесения изменений в арабский алфавит, используемый тюркскими языками.

Новый метод тюркского письма на основе арабской графики, называемый «Усуль аль-джадид», был впервые придуман выдающимся крымско-татарским просветителем Исмаилом Гаспралы. Суть его метода заключается в фонетической обработке арабского алфавита, которая ставила в соответствие звуки и буквы, в отличие от старого метода «Усуль аль-кадим», предлагавшего слоговое изучение языка, когда отдельные буквы сливались в слоги, а из слогов потом составлялись слова. Новый метод позволил минимизировать недостатки базового арабского произношения и в 2-3 раза сократить срок обучения грамоте [6].

В XX веке казахский язык поменял свой алфавит 3 раза. В первый раз в 1912 году с помощью метода «Усуль аль-джадид» основоположник теории казахского языка Ахмет Байтурсынов перевел казахскую письменность на новый алфавит на основе арабской графики. Для этого он систематизировал и уточнил звуковую систему казахского языка из 28 звуков, из которых 5 гласные и 19 согласные фонемы (*смыслоразличительная и обозначаемая отдельным знаком в алфавите минимальная фонетическая единица языка*), при этом 4 мягкие гласные считались *аллофонами* (*смыслонеразличительная минимальная фонетическая единица языка, обозначаемая знаком, который обозначает в алфавите созвучную с ней фонему*), на самом деле они являются *фонемами*. Затем он исключил все арабские буквы, обозначающие неказахские звуки и разработал новый алфавит казахского языка, который содержал 24 буквы на основе арабской графики и 1 специальный знак (апостроф). При этом для обозначения мягких гласных использовались диграфы, в которых первыми парами были буквы, обозначающие твердые согласные, а в каждом диграфе второй парой был апостроф. Этим алфавитом в нашей стране пользовались до 1929 года, а казахи, живущие в других странах, в частности, в Китае, Афганистане и Иране, пользуются им и по сей день, добавив еще 3 буквы для обозначения согласных фонем (*в*), (*ф*) и (*х*).

Во второй раз в 1929 году вместо алфавита на основе арабицы был принят алфавит на основе латиницы, который имел 31 букв.

В третий раз в 1940 году вместо алфавита на основе латиницы был принят 42 буквенный алфавит на основе кириллицы: из них 31 обозначают казахских звуков, которые были обозначены буквами алфавита основе латиницы; 9 используются для обозначения не свойственных казахскому языку звуков русского языка; 2 предназначены для обозначения мягкости и твердости согласных звуков. Этот алфавит действует до настоящего времени.

3. Необходимость реформы письменности казахского языка

Как уже сказано в пункте 2 в действующем алфавите казахского языка на основе кириллицы есть 42 букв, из них 15 (Аа, Әә, Ее, Ёё, Ии, Оо,Өө, Уу, Ұұ, Үү, Ыы, Іі, Ээ, Юю, Яя) обозначает гласные фонемы, 25 (Бб, Вв, Гг, Ғғ, Дд, Жж, Зз, Йй, Кк, Ққ, Лл, Мм, Нн, Ңң, Пп, Рр, Сс, Тт, Фф, Хх, Һһ,Цц, Чч, Шш, Щщ) – согласные фонемы, а 2 буквы Ъь – мягкий знак и Ьь – твердый знак.

Следует отметить, что 9 букв Ъь, Ьь, Щщ, Ёё, Ээ, Юю, Яя, Ии, Уу не имеют никакого отношения к казахскому языку. В русском языке буквы Ъь и Ьь используются для обозначения мягкости и твердости согласных звуков. В казахском языке мягкими и твердыми могут быть только гласные звуки. В исконно казахских словах нет звуков, которые могут быть обозначены буквами Щщ, Ёё, Ээ, Юю, Яя. Более того, при использовании в казахской письменности букв Ии, Уу, которые обозначают гласные звуки русского языка, нарушаются:

1) *закон сингармонизма*: в казахских корневых словах гласные звуки должны чередоваться с согласными и при этом в записи должны использоваться только мягкие, либо только твердые гласные, но при использовании указанных букв это условие нарушается. Например, в следующих словах «**қиа** (утес)», «**иә** (да)», «**ауа** (воздух)», «**әуе** (небо)», «**уақыт** (время)», «**уәде** (обещание)», «**кие** (святыня)», «**қиуа** (вдали)», «**қия** (наискось)». «**саяхат** (путешествие)», «**сүю** (любить)», «**сүю** (разжижаться)» буквы гласных звуков встречаются подряд;

2) *орфографические правила*:

- если основа слова (корень или корень + суффикс) заканчивается на твердый (или мягкий) гласный звук, то к нему добавляется притягательное окончание в третьем лице «сы» (или «сі»). Например, «**ана+сы** (мать)», «**әже+сі** (бабушка)»;

- если твердая (или мягкая) основа слова заканчивается на согласный звук, то к нему добавляется притягательное окончание в

третьем лице «ы» (или «і»). Например, «отан+ы (родина)», «ел+і (страна)».

Но при использовании в записи основы слова букв Ии, Уу, которые обозначают гласные звуки русского языка, эти правила нарушаются. Например, реально пишутся «би+і (танец)», «ми+ы (мозг)», «ту+ы (флаг)» и «гу+і (гул)», вместо того, чтобы писать по правилу «би+сі», «ми+сы», «ту+сы» и «гу+сі», которые в казахском языке не имеют никакого смысла.

Указанные примеры доказывают ошибочность действующего алфавита казахского языка на основе кириллицы. Поэтому требуется проведение реформы письменности казахского языка.

4. Анализ предложенного на утверждение алфавита на основе латиницы

С целью реформы письменности казахского языка уже предложены 6 (шесть) версии алфавита казахского языка на основе латинской графики. Все они содержат ошибки и противоречия. Ниже анализируется только шестой вариант.

В таблице 1 представлена шестая версия алфавита казахского языка на основе латиницы, которая размещена на сайте <https://www.qazlatyn.kz/alphabet>.

Таблица 1. Шестая версия алфавита казахского языка на основе латиницы.

№	Baspa túri	Árip ataúy	№	Baspa túri	Árip ataúy
1	A a	[a]	17	Ń ń	[yń]
2	Á á	[á]	18	O o	[o]
3	B b	[by]	19	Ó ó	[ó]
4	D d	[dy]	20	P p	[py]
5	E e	[e]	21	Q q	[qy]
6	F f	[fy]	22	R r	[yr]
7	G g	[gi]	23	S s	[sy]
8	Ĝ ĝ	[ĝy]	24	T t	[ty]
9	H h	[hy]	25	U u	[u]
10	Ĭ ĭ	[i]	26	Ú ú	[ú]
11	I ĭ	[i]	27	V v	[vy]
12	J j	[jy]	28	Y y	[y]
13	K k	[ki]	29	Ý ý	[uý]
14	L l	[yl]	30	Z z	[zy]
15	M m	[my]	31	Sh sh	[Shy]
16	N n	[ny]	32	Ch ch	[chy]

Здесь названия столбцов переводятся так «*Baspa túri* – Печатный вид», «*Árip ataúy* – Название буквы».

Замечание: Далее строчными кириллическими буквами в круглой скобке и буквами международного фонетического алфавита (МФА) в квадратной скобке будем обозначать соответствующие фонемы казахского языка.

Представленный в таблице 1 алфавит не понятен, так как в любом алфавите буквы обозначают не названия букв, а звуков языка, представляя их сначала символами МФА [7, 8] в квадратных скобках, как это принято во всех языках мира, или просто строчными буквами действующего алфавита, заключая их в круглые скобки.

В этом алфавите нет системности: некоторые созвучные твердые и мягкие гласные фонемы обозначены разными латинскими буквами, например, твердая фонема (ы) - [ɯ] обозначена буквой *Yu*, а созвучная с ней мягкая фонема (і) - [ɪ] буквой *Ii*. Звуки (ы) и (і) должны обозначаться одной буквой и для различия для мягкого добавляется диакритический знак, например, *Aa*–(a)–[a], *Áá*–(ə)–[æ], *Oo*–(o)–[ɔ], *Óó*–(ə)–[ø], *Uu*–(ʉ)–[ʊ], *Úú*–(y)–[y]. Использование знака ударения ухудшает читабельность текста.

Указанный алфавит неоднородный, так как для обозначения некоторых фонем используются и диакритические знаки и диграфы. Необоснованно использованы диграфы (ш) – [ʃ] – *Sh sh*, (ч) – [tʃ] – *Ch ch*. При этом в качестве пары диграфа использована латинская буква *Cc*, которая не включена в состав алфавита.

В пункте 3 доказано, что в казахском языке нет гласных звуков (и) и (у). Поэтому их обозначение буквами *Ii* и *Úú* необоснованы, тем более буква *Yu* использована для обозначения несозвучного с ним твердого гласного звука (ы).

Главная ошибка в этом алфавите состоит в том, что в нем обозначены заимствованные из русского языка гласные звуки (и) – [i] и (у) – [u], которые порождают противоречия, указанные в пункте 3.

Учитывая вышесказанные замечания ниже в пункте 5 уточняется звуковая система казахского языка, а в пункте 6 предлагается новый вариант алфавит казахского на латинице, основанный на его уточненную звуковую систему.

5. Уточнение звуковой системы казахского языка

5.1. Гласные звуки

Известно, что в любом естественном языке устная речь первична, а письменная речь вторична. Поэтому для построения алфавита надо сначала уточнить звуковую систему казахского языка, и только потом подбирать буквы (знаки) для обозначения фонем. В некоторых случаях могут быть использованы и аллофоны.

Однако на текущий момент нет общепринятого мнения относительно фонетики казахского языка, до сих пор нет стандарта фонетики, где должны быть уточнены казахские звуки и их классификация. Подобного состояния, как в казахском языке, нет ни в одном другом государственном языке мира. Это связано с тем, что имеющееся учение о фонетике казахского языка унаследовано из русского языка и не отражает точную характеристику казахских звуков, что было показано в работах [9-11]. Поэтому не понятно, что делать с русскими гласными, которые были внедрены в казахский язык без учета его фонетических закономерностей. Один из способов решения этой проблемы - исключение этих звуков из звуковой системы казахского языка, как в азербайджанском языке, и разработка соответствующих орфографических правил.

В казахском языке гласные фонемы, в отличие от согласных фонем, играют важную роль в письменности, так как от их классификационных признаков зависит выполнимость закона сингармонизма и орфографических правил по присоединению аффиксов (суффиксов и окончаний).

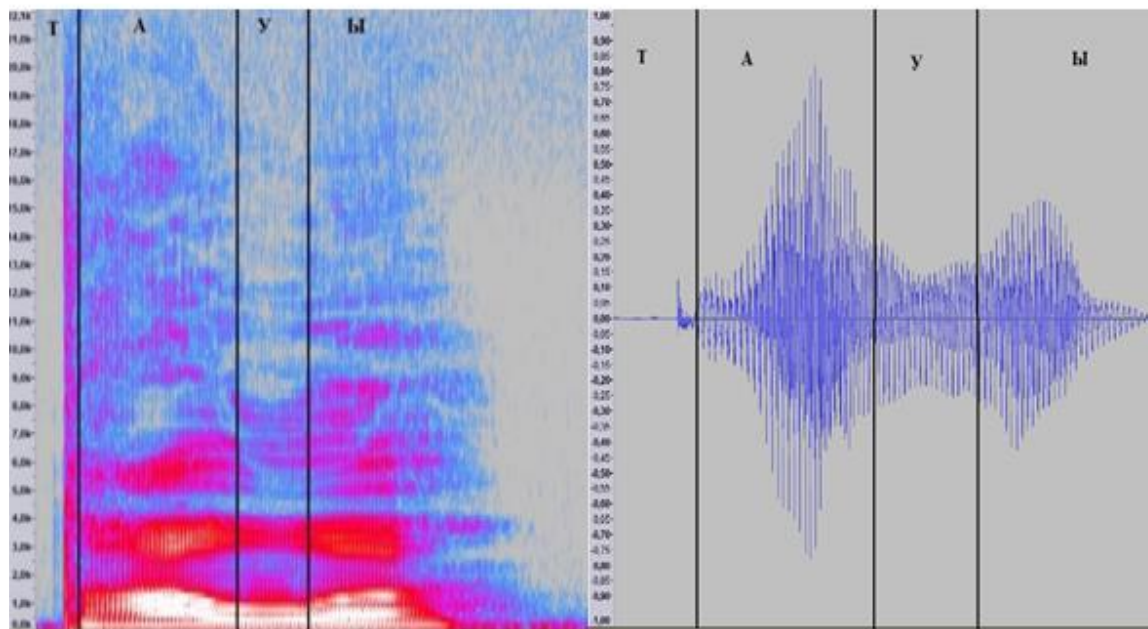
В действующем алфавите казахского языка обозначены 11 гласных звуков (а), (ә), (е), (и), (о), (ө), (ұ), (ү), (у), (ы), (і). Из них звуки (и), (у) заимствованы из русского языка в 1940 году во время перевода казахской письменности на кириллицу. Чтобы проверить обоснованность их внедрения в казахский язык мы провели *перцептивный анализ* казахского языка с целью слуховой идентификации его звуковых единиц. Для этого составили фонетический насыщенный тест (звуки, слоги, слова, фразы), которого продиктовали носители казахского языка из разных регионов, разного пола и возраста.

В результате выяснили, что в исконно казахской речи нет фонемы (и) и (у). Так вместо звука (и) идентифицированы дифтонги (ый) и (ій) в зависимости от твердости и мягкости звучания прочитанного продиктованного теста.

Вместо гласной фонемы (у) идентифицирована поугласная фонема, которая нет в русском языке, но имеется в английском языке и обочена

она латинской буквой Ww, а полугласную фонему будем обозначать как (w).

На рисунке 1 показан результат обнаружения полугласной фонемы (w).



А В

Рисунок 1. Фонетическая сегментация слова «тауы», где А – спектрограмма, В – волна.

На рисунке 1. А можно заметить полугласного звука (w) между гласными звуками (а) и (ы). Звук (w) имеет очень ограниченные препятствия в процессе артикуляции и поэтому он похож на гласные звуки по спектрограмме и звуковой волне, но он отличается от них отсутствием энергии в районе частот 11-15 КГц, изменением формант в районе 2 КГц и 5 КГц и амплитудой, существенно меньшей, чем амплитуды гласных звуков (а) и (ы), что видно на рисунке 1.Б.

На основании проведенного эксперимента звуки (и) и (у) не будут включены в состав казахской звуковой системы как фонема, они могут использоваться как аллофон при произношении некоторых заимствованных из других языков терминов.

Таким образом, в звуковую систему казахского языка будут включены 9 гласных фонем (а) - [a], (ә) - [æ], (е) - [e], (о) - [ɔ], (ө) - [ø], (ұ) - [ʊ], (ү) - [y], (ы) - [ɯ], (і) - [i], которые будут обозначены в алфавите на основе латинской графики отдельными буквами

В некоторых источниках по фонетике казахского языка [7-10] для артикуляционной классификации казахских гласных звуков выделяют 3 бинарных признака.

Для корректного и полного описания характеристик девяти объектов трех бинарных признаков недостаточно, поскольку каждый признак имеет по 2 значения и максимальное количество различных признаков равно $2^3 = 8$. Поэтому в указанных источниках классификация казахских гласных звуков некорректна и не соответствуют классификации звуков в МФА [11, 12], которая имеет 4 бинарных признаков: положение языка, положение челюсти и положение губ с показанными в таблице 1, а также вертикальное положение языка с двумя значениями (*верхние, нижние*). В МФА термин «верхние» называется «close», а «нижние» – «open», подразумевая близость при подъеме языка к небу [13, 14].

В таблице 2 показаны четыре бинарных артикуляционных признака казахских гласных звуков.

Таблица 2. Четыре артикуляционные признаки казахских гласных звуков

Гласные		Артикуляционные признаки							
Обозначение в кириллице	Обозначение в МФА	Положение языка				Положение губ		Положение челюсти	
		Вертикальное		Горизонтальное		Неокругленные	Округленные	Открытые	Закрытые
		Нижние	Верхние	Заднеязычные	Переднеязычные				
(a)	[ɑ]	+	-	+	-	+	-	+	-
(ә)	[æ]	+	-	-	+	+	-	+	-
(ы)	[ɯ]	+	-	+	-	+	-	-	+
(і)	[ɪ]	-	+	+	-	-	+	-	+
(ү)	[ʊ]	+	-	-	+	+	-	-	+
(ү)	[y]	-	+	-	+	-	+	-	+
(o)	[ɔ]	-	+	+	-	-	+	+	-
(ө)	[ø]	-	+	-	+	-	+	+	-
(e)	[e]	-	+	-	+	+	-	+	-

Теперь, используя Булеву алгебру [12], мы построим математическую модель системы гласных звуков [13, 14] в виде алгебраического выражения на основе использования значений 4-х артикуляционных бинарных признака, показанных в таблице 2.

Обозначим 4 артикуляционных признаков *Вертикальное положение языка* и *Горизонтальное положение, Положение губ, Положение челюсти, языка* казахских гласных звуков через логические переменные x_1, x_2, x_3 и x_4 , соответственно. Эти переменные принимают только значения 1 для истины и 0 для лжи:

- если язык находится в верхнем положении, то $x_1 = 1$, иначе $x_1 = 0$;
- если язык находится в заднем положении, то $x_2 = 1$, иначе $x_2 = 0$;
- если губы находятся в округленном положении, то $x_3 = 1$, иначе $x_3 = 0$;
- если челюсть находится в открытом положении, то $x_4 = 1$, иначе $x_4 = 0$.

На основании установленных значений для переменных можно построить Булеву модель [12] системы казахских гласных звуков, представленной в таблице 3.

Таблица 3. Булева модель системы казахских гласных

Обозначение в кириллице	Обозначение в МФА	x_1	x_2	x_3	x_4
(a)	[ɑ]	0	1	0	1
(ә)	[æ]	0	0	0	1
(ы)	[ɯ]	0	1	0	0
(і)	[ɪ]	0	0	0	0
(ұ)	[ʊ]	1	1	1	0
(ү)	[y]	1	0	1	0
(o)	[ɔ]	1	1	1	1
(ө)	[e]	1	0	1	1
(е)	[e]	1	0	0	1

Далее для каждого из 9 гласных звуков, выписывая признаки в виде их конъюнкции и объединяя эти конъюнкции, получим следующую дизъюнктивную нормальную форму из 9 членов:

$$\left\{ \begin{array}{l} (\bar{x}_1 \wedge x_2 \wedge \bar{x}_3 \wedge x_4) \vee (\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge x_4) \vee (\bar{x}_1 \wedge x_2 \wedge \bar{x}_3 \wedge \bar{x}_4) \vee \\ (\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge \bar{x}_4) \vee (x_1 \wedge x_2 \wedge x_3 \wedge \bar{x}_4) \vee (x_1 \wedge \bar{x}_2 \wedge x_3 \wedge \bar{x}_4) \vee \\ (x_1 \wedge x_2 \wedge x_3 \wedge x_4) \vee (x_1 \wedge \bar{x}_2 \wedge x_3 \wedge x_4) \vee (x_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge x_4) \end{array} \right\} \quad (2.1)$$

Применяя аксиомы Булевой алгебры, получим упрощенное выражение:

$$(x_1 \wedge x_3) \vee (\bar{x}_1 \wedge \bar{x}_3) \vee (x_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge x_4) \quad (2.2)$$

Выражение (2.2) будем называть *функцией принадлежности*, которая характеризует систему казахских гласных в теореме, приведенной ниже.

Теорема принадлежности [13, 14]. Гласный звук λ принадлежит системе казахских гласных тогда и только тогда, когда его артикуляционные признаки x_1 , x_2 , x_3 и x_4 , определенные выше, удовлетворяют дизъюнктивной нормальной форме

$$(x_1 \wedge x_3) \vee (\bar{x}_1 \wedge \bar{x}_3) \vee (x_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge x_4)$$

На основе четырех артикуляционных признаков построена в [16] геометрическая модель гласных звуков казахского языка, показанной на рисунке 1.

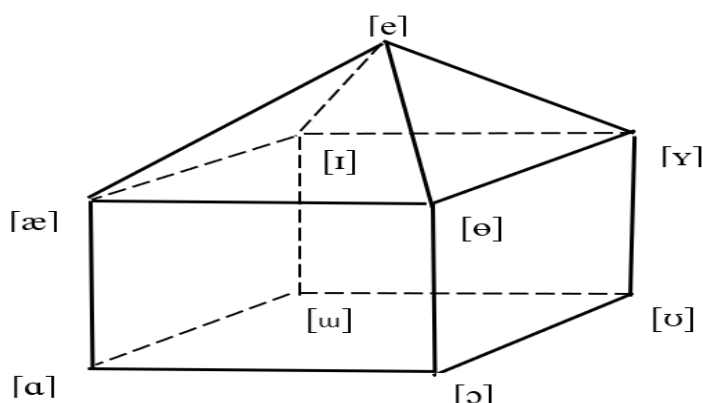


Рисунок 1. Геометрическая модель системы казахских гласных на основе 4-х бинарных артикуляционных признаков.

На рисунке 1 плоскость [a] [ɔ] [ʊ] [ɯ] представляет заднеязычные (твердые) гласные звуки (a) (o) (ʏ) (ы), плоскость [æ] [ə] [ɣ] [ɪ] – переднеязычные (мягкие) гласные звуки (э) (ө) (ү) (і), плоскость [a] [ɔ] [ə] [æ] – открытые гласные звуки (a) (o) (ə) (э), плоскость [ɯ] [ʊ] [ɣ] [ɪ] – закрытые гласные звуки (ы) (ʏ) (ү) (і), плоскость [ɔ] [ʊ] [ɣ] [ə] – округленные гласные звуки (o) (ʏ) (ү) (ө), плоскость [a] [ɯ] [ɪ] [æ] – неокругленные гласные (a) (ы) (і) (э), а вершина [e] представляет особый звук (е).

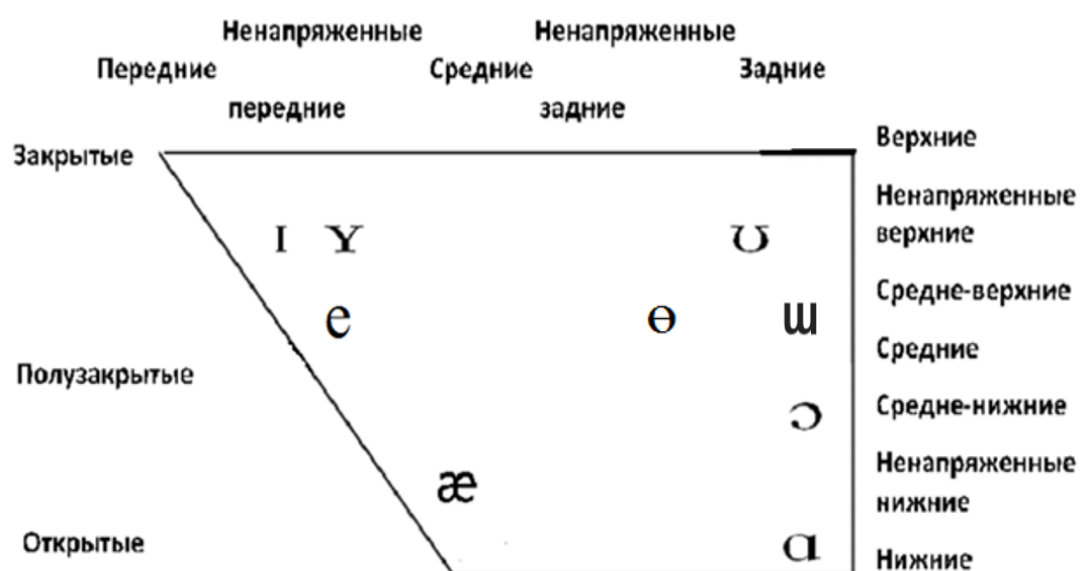
По данной геометрической модели можно установить следующие факты (аксиомы):

1. Никакие вершины (гласные звуки) из нижней и верхней плоскостей, включая [e], не могут оказаться в исконно казахском слове. Этот факт называется небной (палатальной) гармонией гласных или сингармонизмом.

2. В казахском языке нет исконного слова, содержащего более трех различных гласных звуков, и если оно содержит три гласных звука, то это мягкие гласные, один из которых обязательно [e]. Это можно интерпретировать на рисунке 1 так: никакие 3 гласные нижней или верхней плоскостей не встречаются в казахском слове. Слово, содержащее 3 различные гласные, включает вершины плоскости, проходящей через [e] и любые 2 вершины верхней плоскости.

Теперь можно привести классификацию казахских гласных звуков на основе 4-х бинарных артикуляционных признака (по горизонтальному, по вертикальному положению языка, по положению челюсти, по положению губ), показанную в таблице 4.

Таблица 4. Классификация казахских гласных звуков на основе четырех бинарных артикуляционных признаков.



Здесь сверху показаны признаки по горизонтальному положению языка, слева - признаки по положению челюсти, справа - признаки по вертикальному положению языка, а признаки по положению губ учтены при расположении двух звуков в одной строчке – слева неокругленные, справа – округленные.

Акустический анализ гласных звуков казахского языка основан на на сингармонических тембрах: твердые нижние, мягкие нижние, твердые верхние, мягкие верхние и мягкие низкие. В таблице 5 приведена система сингармонических тембров гласных звуков казахского языка.

Таблица 5. Система сингармонических признаков гласных

Обозначение гласных звуков	Палатальный тембр		Лабialsный тембр	
	твердый	мягкий	низкий	высокий
(а) - [ɑ]	+	-	+	-
(ә) - [æ]	-	+	+	-
(о) - [ɔ]	+	-	-	+
(ө) - [ɵ]	-	+	-	+
(ұ) - [ʊ]	+	-	-	+
(ү) - [y]	-	+	-	+
(ы) - [ɯ]	+	-	+	-
(і) - [i]	-	+	+	-
(е) - [e]	-	+	-	+

Мы провели эксперимент [14], микшируя осциллограммы заднеязычных (твердых) гласных звуков (а) - [ɑ], (о) - [ɔ], (ұ) - [ʊ] и (ы) - [ɯ] с осциллограммой звука [е] - (е) и обнаружили следующие свойства, представленные на рисунке 2:

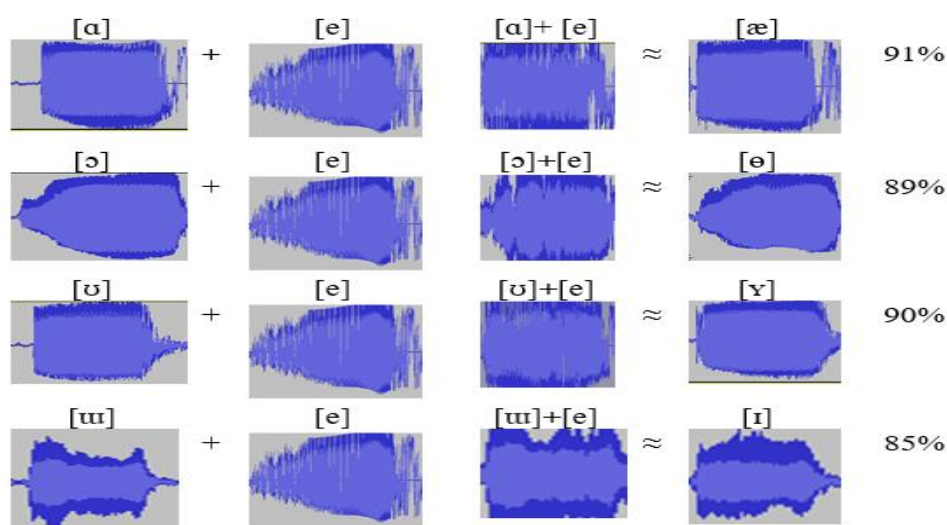


Рисунок 2. Свойства гласных звуков казахского языка

Указанные свойства доказывают, что в корневых словах казахского языка не должны встречаться пары, образованные из подряд стоящих гласных звуков (а) и (е), (о) и (е), (ұ) и (е), (ы) и (е).

Если указанные пары гласных звуков встречаются в составных словах между образующими их корневыми словами, то они должны разделяться пробелами.

В состав звуковой системы казахского не будет включаться только одна согласная фонема, которая обозначена в действующем алфавите буквой Щщ, а все остальные согласные фонемы: (б), (в), (г), (ф), (д), (ж),

(з), (й), (к), (к), (л), (м), (н), (н), (п), (р), (с), (т), (ф), (х), (h), (ц), (ч), (ш) будут включены в неё. Среди фонемы (в), (ф), (х), (ц), (ч) являются заимствованными из русского языка, а остальные являются исконно казахскими фонемами.

Что касается заимствованных фонем (в), (ф), (х), (ц), (ч), то они используются для написания имен людей, наименований стран и местностей, а также международных непереводаемых терминов. (Например, *Вагнер, Вьетнам, вагон, вакуум, вексель, вектор, вето, виза, викторина, вирус, вольт, Форд, Франция, факт, файл, факультет, федерация, физика, филармония, фонетика, формула, функция, Бухарест, Хельсинки, Хинчин, Хомский, Хонекер, Хорватия, Хуавей, химия, хлор, хром, хроника, Цезарь, Циолковский, Цой, Цейлон, Цюрих, цезий, цемент, цензура, циклотрон, цилиндр, циркуль, цунами, Чад, Чайковский, Чаплин, Челюскин, Черногория, Чехия, Чикаго, Чили, Чита, чеченец, чуваш*).

Указанные фонемы так глубоко внедрились в казахский язык, что даже у носителей казахского языка, не знающих русский язык, артикуляционные органы речи приспособились для их произношения. Эти фонемы не входят в состав суффиксов и окончаний, которые применяются при формировании словообразований и слогформ. Поэтому оставление их в составе звуковой системы казахского языка и использовать их при составлении нового алфавита на латинской графике как отдельные фонемы. От них казахский язык никак не пострадает, так как они не будут порождать противоречия в казахской орфографии и орфоэпии, наоборот позволят приблизить написания и звучания международных терминов к их оригиналам.

Таким образом, в реформе по переводу казахской письменности на латинскую графику будут участвовать **33** фонем, из них **9** гласные фонемы (а), (ә), (е), (о), (ө), (ұ), (ү), (ы), (і) и **24** согласные фонемы (б), (в), (г), (ғ), (д), (ж), (з), (й), (к), (к), (л), (м), (н), (н), (п), (р), (с), (т), (w), (ц), (ф), (х), (ч), (ш). В реформе будут участвовать также полугласные (и) и (у) в качестве аллафонов гласных фонем (і) и (ұ), соответственно.

6. Два варианта алфавита казахского языка на латинице

На основе уточнения в пункте 4 звуковой системы казахского языка в составлении его нового алфавита будут участвовать **33** фонем, из них **9** гласные фонемы и **24** согласные фонемы, а также полугласные звуки (и), (у) в качестве аллофонов гласных фонем (і) и (ұ), которые в МФА будут обозначаться как [i] и [w].

Заимствованные из русского языка гласные звуки (и) и (у), включенные в звуковую систему казахского языка как аллофоны

гласных фонем (i) и (y), будут обозначаться буквами, используемыми для обозначения фонем (i) и (y), соответственно. Например, если гласные фонемы (i) и (y) были обозначены латинскими буквами I и Y, то не составило бы труда писать и читать такие термины, как *internet, institute, university, supremum*. Если в новом алфавите заимствованных гласных (и) и (у) обозначить отдельными буквами, то неизбежно появятся противоречия, упомянутые в пункте 2.

При определении состава нового казахского алфавита фонемы, для которых имеются адекватные латинские буквы будут обозначаться этими буквами без диакритических знаков и диграфов.

Замечание: Диграфы не входят в состав алфавита, они по существу являются орфографическими правилами того или иного языка, в котором они используются.

Фонемы, для которых не существуют адекватные латинские буквы будут создавать проблемы. К ним относятся мягкие гласные фонемы (ə), (e), (y) и (i), а также согласные фонемы (f), (k), (h), (ch) и (sh). Для решения этих проблем предлагаются два варианта алфавита на основе латинской графики:

1) В первом варианте будут использоваться надбуквенные и подбуквенные диакритические знаки, как в алфавите турецкого языка.

2) Во втором варианте будут использоваться диграфы, как в алфавите английского языка, ранее подобный алфавит был предложен в [14] без включения фонем (ц) – [ts] и (ч) - [tʃ] в звуковую систему казахского языка.

Теперь займемся выбором конкретных способов обозначения этих проблемных фонем. Так мягкие фонемы (ə), (e), (y) в первом варианте будут обозначаться с помощью надбуквенного диакритического знака «две точки»: Ä ä, Ö ö, Ü ü, при этом твердая фонема (ы) будет обозначаться буквой без точки I i, а мягкая фонема (i) - надбуквенным диакритическим знаком «одна точка» I i. Во втором варианте алфавита эти фонемы будут представляться диграфами Ae ae, Oe oe, Ue ue.

Согласные фонемы (h), (f), (sh) и (ch) в первом варианте будут обозначаться с помощью надбуквенного диакритического знака «тильда» Ññ и Ğğ, а во втором варианте – диграфами Ng ng, Gh gh.

В звуковую систему казахского языка включены согласные фонемы (ch) и (sh), которые будут обозначены в первом варианте подбуквенным диакритическим знаком «запятая» Ç ç и Ş ş, соответственно, а во втором варианте – диграфами Ch ch и Sh sh.

Предлагается включить в оба варианта алфавита букву базового латинского алфавита Xx, которая будет обозначать не конкретную фонему, а сочетания согласных (k) и (c). Использование этой буквы и не

будет порождать никаких противоречий в казахской письменности и позволяет писать международные термины одинаково с оригиналом или ближе к нему. Например: *axelerat* - акселерат, *axis* - ось, *box* - бокс, *context* - контекст, *expert* - эксперт, *expor* - экспорт, *maximum* - максимум, *mixer* - миксер, *Oxford* - Оксфорд, *taxi* - такси, *xerox* - ксерокс.

Первый вариант предлагаемого алфавита не порождает никаких противоречий в казахской орфоэпии и орфографии, так как он основан на уточненной звуковой системе казахского языка и на принципе «Один звук и одна буква».

При использовании второго варианта предлагаемого алфавита возникают некоторые проблемы орфоэпии и орфографии казахского языка, связанные с использованием диграфов. Но от них можно избавиться, если учитывать следующие свойства и требования:

1) В исконно казахских словах перед согласным звуком (ң) не встречаются согласные звуки, поэтому в любом слове диграф «ng» должен писаться только в одном слого.

2) При прямой кодировке разных по смыслу и записи на кириллице слов латинскими буквами происходит одинаковое их написания:

$$kuengi = \begin{cases} \text{күнгі (дневной)} \\ \text{күңі (служанка)} \end{cases}$$

Поэтому, чтобы недопускать подобные случаи надо:

- учесть свойства **ассимиляции (уподобления)** звуков и ввести специальное орфографическое правило: «Если в слове за звуком (п) - [п] непосредственно следует звук (г), то при чтении такого слова звук (п) произносится как звук (ң) – [ң], т.е. звук (п) - [п] уподобляется (ассимилируется) звуку (ң) – [ң]. Например, **әңгіме** (беседа), **түңгі** (ночной) при чтении произносится как **әңгіме**, **түңгі**

- вместо одной латинской буквы **n** надо писать диграф **ng**, например, «күнгі = kuengi = ku**enggi**», что будет соответствовать закономерности произношения в казахском языке и не порождать неоднозначности в его письменности.

3) В слове “хана - помещение” фонему (х)- [h] надо заменить на фонему (қ) - [q]. Тогда записи комбинированных с его участием не будут создавать никаких орфографических проблем. Например, «кітапхана → кітапхана = kitap**q**ana, қымызхана → қымыз**қ**ана = kitap**q**ana, кітапхана → кітапхана = kitap**q**ana».

4) В комбинированных словах или человеческих именах согласную фонему (x) - [h], которая непосредственно следует после согласной фонемы (c) - [s] или (з) - [z], надо заменить на фонему (к) - [q]. Например, «асхана → асқана = asqana, Асхат → Асқат = Asqat, Досхан → Досқан = Dosqan, қымызхана → қымызқана = qumyzqana, Оразхан → Оразқан = Orazqan. В противном случае, эти же слова «асхана = ashana = ашана, Асхат = Ashat = Ашат, Досхан = Doshan = Дошан, қымызхана = qumyzhana = қымыжана, Оразхан = Orazhan = Оражан».

Таким образом, устранив недостатки приведенного в таблице 1 шестого варианта алфавита казахского языка, предлагается на основе уточненной звуковой системы казахского два варианта алфавита на основе латинской графике: первый вариант на основе алфавита английского языка с использованием диграфов, второй вариант на основе алфавита турецкого языка с использованием диакритических знаков.

Предлагаемые варианты алфавита казахского языка обладают следующими свойствами:

1 при создании и обработке информационных ресурсов на казахском языке не нуждаются в программных средствах, которые требуют дополнительные затраты на изготовление:

- шрифтов и драйверов, так как он состоит только из букв английского алфавита, размещенного на клавиатуре всех типов компьютеров с многочисленными шрифтами и драйверами для всех типов устройств ввода и вывода данных;

- программ сортировок и поиска информации, так как порядок следования букв в нём совпадает с порядком следования букв в английском алфавите;

2 алфавит позволяет сблизить написание непереводаемых на казахский язык международных заимствованных терминов - слов с их написанием в оригинале без изменения звуковой системы казахского языка;

3 алфавит позволяет легко и быстро набрать казахский текст с помощью клавиатуры любого компьютера и смартфона.

В таблице 7 представлен алфавит казахского языка с использованием диакритических знаков на основе турецкого алфавита,

В таблице 8 представлен алфавит казахского языка с использованием диграфов на основе английского алфавита.

В таблице 9 приводятся орфографические правила представления фонем с помощью диграфов.

Таблица 7. Вариант алфавита казахского языка на основе турецкого алфавита

№	Латиница	Кириллица	МФА	№	Латиница	Кириллица	МФА
1	A a	А а	[ɑ]	18	N n	Н н	[n]
2	Ä ä	Ә ә	[æ]	19	Ñ ñ	Ң ң	[ŋ]
3	B b	Б б	[b]	20	O o	О о	[ɔ]
4	C c	Ц ц	[tc]	21	Ö ö	Ө ө	[ɵ]
5	Ç ç	Ч ч	[tʃ]	22	P p	П п	[p]
6	D d	Д д	[d]	23	Q q	Қ қ	[q]
7	E e	Е е	[e]	24	R r	Р р	[r]
8	F f	Ф ф	[f]	25	S s	С с	[s]
9	G g	Г г	[g]	26	Ş ş	Ш ш	[ʃ]
10	Ğ ğ	Ғ ғ	[ɣ]	27	T t	Т т	[t]
11	H h	Х х	[h]	28	U u	Ұ ұ	[ʊ u]
12	I ı	Ы ы	[ɯ]	29	Ü ü	Ү ү	[y]
13	İ i	І і	[ɪ, i]	30	V v	В в	[v]
14	J j	Ж ж	[ʒ]	31	W w		[w]
15	K k	К к	[k]	32	X x		[ks]
16	L l	Л л	[l]	33	Y y	Й й	[y]
17	M m	М м	[m]	34	Z z	З з	[z]

Таблица 8. Алфавит казахского языка на основе английского алфавита

№	Латиница	Кириллица	МФА	№	Латиница	Кириллица	МФА
1	A a	А а	[ɑ]	14	N n	Н н	[n]
2	B b	Б б	[b]	15	O o	О о	[ɔ]
3	C c	Ц ц	[tc]	16	P p	П п	[p]
4	D d	Д д	[d]	17	Q q	Қ қ	[q]
5	E e	Е е	[e]	18	R r	Р р	[r]
6	F f	Ф ф	[f]	19	S s	С с	[s]
7	G g	Г г	[g]	20	T t	Т т	[t]
8	H h	Х х	[h]	21	U u	Ұ ұ	[ʊ u]
9	I i	І і	[ɪ, i]	22	V v	В в	[v]
10	J j	Й й	[j]	23	W w		[w]
11	K k	К к	[k]	24	X x		[ks]
12	L l	Л л	[l]	25	Y y	Ы ы	[ɯ]
13	M m	М м	[m]	26	Z z	З з	[z]

Таблица 9. Орфографические правила для диграфов

№	Латиница	Кириллица	МФА	Пояснение
1	Ae ae	Ә ә	[æ]	
2	Oe oe	Ө ө	[ø]	
3	Ue ue	Ү ү	(y)	
4	Ch ch	Ч ч	[tʃ]	Как в английском
1	Gh gh	Ғ ғ	[ɣ]	
5	Ng ng	Ң ң	[ŋ]	Как в английском
6	Sh sh	Ш ш	[ʃ]	Как в английском
7	Zh zh	Ж ж	[ʒ]	

7. Заключение

В эпоху цифровизации общества многие естественные языки развиваются с помощью компьютерных программ: *созданы электронные словари, мультимедийные вопросно-ответные системы, интеллектуальные системы обучения и оценки знаний, машинные переводчики с одного языка на другой, системы распознавания и синтеза письменной и устной речи и т.д.* Основами этих работ являются математические модели грамматических правил этих языков.

Подобные проблемы можно ставить и успешно решать и для казахского языка. Однако указанные ошибки не дают возможность формализовать морфологических правил казахского языка и автоматизировать морфологический анализ и синтез казахских слов. Предложенный новый алфавит позволяет успешно решать эти проблемы.

Список литературы

1. Казахский язык // Большая советская энциклопедия : в 30 т. / 3-е изд. – М.: Советская энциклопедия, 1969–1978.
2. https://ru.wikipedia.org/wiki/Казахский_язык
3. <https://ru.wikipedia.org/wiki/Казахи>.
4. https://ru.wikipedia.org/wiki/Древнетюркское_руническое_письмо.
5. https://ru.wikipedia.org/wiki/Иссыкское_письмо.
6. <http://www.niac.gov.kz/kz/expertiza/item/138-dzhadidizm-i-reforma-musulmanskogo-obrazovaniya>.
7. International Phonetic Association. Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet. Cambridge, U.K., 2003.
8. https://en.wikipedia.org/wiki/International_Phonetic_Alphabet.
9. Мырзабеков С. Қазіргі қазақ тілінің фонетикасы. Алматы, Дәуір, 2013. – 219 с.
10. Калиева Б.А. Қазақ тілінің фонетикасы. Алматы: Эверо, 2014. – 162 с.
11. http://onstudy.kz/dausty_dybys.

12. https://ru.wikipedia.org/wiki/Булева_алгебра.

13. Z. Yessenbayev, M. Karabalayeva and A. Sharipbayev, "Formant Analysis and Mathematical Model of Kazakh Vowels," 2012 UKSim 14th International Conference on Computer Modelling and Simulation, Cambridge, 2012. - P.427-431.

14. Шарипбай А.А. Проблемы перевода казахской письменности на латинский алфавит (на казахском и русском языках). Монография, Астана, ЕНУ имени Л.Н.Гумилева, 2017, -138 С.