

Д.Ш. Сулейманов

К ПРОБЛЕМЕ ВОССТАНОВЛЕНИЯ ТАТАРСКОГО АЛФАВИТА НА ОСНОВЕ ЛАТИНСКОЙ ГРАФИКИ

Введение

Очевидно, что смена алфавита, его упорядочение и приведение в соответствие с базовыми закономерностями языка – не гарантия дальнейшего развития языка и его широкого применения. Привлекательность и живучесть языка – в его активности и прежде всего в том, насколько язык является языком наук, информационных технологий, языком общения на официальном государственном уровне, т.е. межправительственным и внутриведомственным языком, языком передачи и усвоения знаний, языком межнационального и международного общения. Вместе с тем адекватный алфавит, представляющий собой систему базовых элементов языка на стыке слышимого и видимого – важный фактор, упорядочивающий вербальное проявление языка. Для укрепления фонетического иммунитета языка, для обеспечения его устойчивости важно, чтобы графемы соответствовали фонемам языка и в алфавите не было “мертвых” графем, не имеющих соответствующих фонем в языке.

Построение адекватного алфавита воспринимается нами как упорядочение основ орфоэпии и орфографии языка, приведение в соответствие фонетического и графемного рядов, очистка от “неработающих” букв.

Технологический аспект. Рассмотрим ряд аргументов в пользу татарского алфавита на основе латинской графики.

Во-первых, и это, возможно, самое важное, особенно в аспекте обучения языкам: переход на латиницу приводит к четкому графическому разграничению фонетических систем татарского и русского языков.

Очевидно, графическое разграничение фонетических систем татарского и русского языков полезно как для татарского, так и для русского языков. Непосредственное и постоянное соседство двух сильно различающихся фонетических систем под крышей одного алфавита приводит к диффузии, размыванию фонем, соответственно, к разрушению обоих языков. Чуть ли не все буквы кириллического алфавита, имеющие одинаковые начертания, по-разному звучат в русском и татарском языках. Например, как в таких словах, как *акын, Казан, чеки, бала, тын, кол, кот, товар, колхоз, целлофан, чир, карга, кит, тир*, имеющих определенные значения на русском и на татарском языках. В зависимости от того, к какому языку они относятся, – татарскому или русскому – эти слова читаются совершенно по-разному.

Необходимо отметить следующее: в данном аспекте важен даже не столько переход на латиницу, сколько разграничение с кириллицей. То есть проблема разграничения татарского и русского алфавитов стоит даже независимо от того, к каким графемам переходить. Если бы татарский и английский были в такой же близости, как русский и татарский, а русский был бы языком географически далекого народа, как народы с английским языком, то с теми же обоснованиями в интересах татарского языка пришлось бы переходить на кириллицу или на любой другой, отличный от латиницы, алфавит.

Но реальная ситуация такова, что стоит проблема разграничения именно татарского и русского алфавитов, базирующихся на кириллической графике. Латиница в данном случае устраивает по следующим причинам: 1) ее используют практически

Джавдет Шевкетович Сулейманов – член-корреспондент АН Республики Татарстан, доктор технических наук, профессор, заведующий Совместной научно-исследовательской лабораторией АН РТ и Казанского государственного университета “Проблемы искусственного интеллекта” (г. Казань). E-mail: dvdt@telecet.ru.

все народы, формирующие сегодня мировую науку, 2) на ней базируются информационные технологии, 3) нет активного коммуникативного пересечения и взаимовлияния татарского и английского языков.

Дифференциация алфавитов по принципу “разные графемы – разные фонемы” удобна и при изучении языков. Переход на латиницу является выигрышным при изучении татарского языка как для татар, для которых русский и, соответственно, русский алфавит служат первым языком, первым алфавитом, так и для русских, желающих изучать татарский язык, – легче будут усваиваться новые фонемы.

Во-вторых, для того чтобы выжить и развиваться в настоящее время (а в скором будущем тем более), татарский язык должен войти в компьютерные технологии как язык накопления, обработки и передачи информации.

Очевидно, чем ближе алфавит татарского языка к алфавиту языка информационных технологий, каковым сегодня является латиница, и чем меньше промежуточных конвертаций (программных транслитераций – переводов с одного алфавита на другой), тем он эффективнее, так как это приводит к экономии памяти для хранения и сокращению времени обработки, а значит, и экономически выгодно. Уменьшение промежуточных модулей (конверторы, таблицы перехода) приводит к увеличению надежности системы (меньше деталей – больше надежность). Как известно, в технических системах время, память, надежность – это одни из самых критичных и важных показателей. Для татарского языка на основе латиницы – меньше проблем при использовании программ, обрабатывающих чисто латинические тексты. Меньше проблем с совместимостью и адаптацией данных и программ для обработки языковых блоков, с сортировкой, отображением на экране информации, с конвертацией текстов. Меньше проблем с татарской локализацией новых пакетов программ и операционных систем.

В-третьих, сокращение количества графем позволяет разработать более подходящую раскладку клавиатуры для татарского языка.

В-четвертых, упрощается общение в компьютерных сетях на татарском языке. Тексты на латинице, в отличие от кириллицы, читаются всегда корректно. В принципе эта проблема для кириллического текста разрешима и разрешается, однако требует дополнительных усилий по поддержке корректного отображения кириллических букв.

С переходом на латиницу татарские тексты будут читаться стабильно. Во всяком случае, если даже и не будет соответствующих драйверов и шрифтов, и не все буквы будут отображаться корректно, текст будет понятен, а это при переписке, а также при ознакомлении с каким-либо текстом зачастую самое главное: понять, что же там написано.

В-пятых, в настоящее время подготовлена необходимая документальная и программная база, регламентирующая использование татарской латиницы в компьютерных технологиях. Разработана опытно-эксплуатационная версия пакета драйверов и шрифтового обеспечения для татарского языка на основе латиницы. Академией наук Татарстана подготовлены соответствующие материалы и принято постановление Кабинета министров РТ “О стандартах кодировки символов татарского алфавита на основе латинской графики и базовых программах для компьютерных применений” за № 625 от 27 сентября 2000 г. Разработана и реализована базовая версия конвертера татарского текста с кириллического начертания на латиническую. Разработана специальная страница в Интернете www.tatartele.ru для оказания технической и консультационной помощи пользователям при установке драйверов и татарских шрифтов на основе латиницы.

В-шестых, практически все языки развитых информационных технологий используют латиницу.

Татарский язык, в силу исключительной регулярности грамматики, семантической емкости и плотности морфологии является удобным и весьма перспективным языком для компактного хранения и эффективной обработки языковой информации. Следовательно, переход татарского языка на латиническую основу обеспечит естественное вхождение его в компьютерную среду, упростит использование татарского языка в качестве программного инструментария при создании интеллектуальных систем.

В-седьмых, почему нельзя просто модифицировать кириллический (русский) алфавит, адаптируя его к татарскому фонемному ряду? Одна из причин, почему нельзя обойтись простым изменением кириллицы, “подгонкой” кириллического алфавита, заключается в следующем. Анализ кириллического алфавита с точки зрения соответствия русских букв татарскому фонемному ряду показывает, что для приведения кириллического алфавита в соответствие с татарской фонемной системой необходимо внести в алфавит не менее 20 изменений, т.е. добавить в русский алфавит не менее 20 новых символов, причем удалив оттуда ряд “чужих” для татарского языка (*e, ё, я, ю, ц, щ, э, ь, ъ*). Изменению подлежат даже такие буквы, как “а”, “о”, “в”, “ч”, “ы”. Очевидно, это уже будет не измененный кириллический алфавит, а полная эклектика, т.е. новый алфавит, далекий как от кириллицы, так и от латиницы.

Лингвистический аспект. В данном разделе статьи осуществляется анализ татарского алфавита на основе латинской графики из 34 букв, утвержденного указом президента РТ и признанного в настоящее время экспериментальным вариантом, и предлагается новый вариант алфавита, состоящий из 26 букв.

В ряде писем в Академию наук Республики Татарстан и Комиссию по реализации языковой политики в компьютерных технологиях Комитета по реализации Закона “О языках народов РТ” при Кабинете министров Республики Татарстан их авторами предлагалось использовать для отображения татарских букв только латинские буквы английского алфавита, а специфические татарские фонемы обозначать в виде сочетания латинских букв (например, немецкие ‘Ш’ – ‘Sch’ и ‘Ч’ – ‘Tsch’, английские ‘Ч’ – ‘Ch’, ‘Х’ – ‘Kh’).

Некоторые специалисты, учитывая закон сингармонизма гласных в татарском языке, предложили специфические мягкие гласные отображать комбинацией букв – через твердые гласные плюс какие-либо букву или символ, обозначающие смягчение твердой гласной (например, ‘Av’ или ‘Ae’, или ‘Ah’, или ‘A’ вместо татарской буквы ‘ә’). Считая такое предложение достойным внимания, мы провели анализ и пришли к следующему выводу. Специфические согласные буквы из нового алфавита: ‘Ч’, ‘Н’, ‘Ш’, ‘Ж’, действительно, возможно отображать в виде комбинации латинских букв. Причем для их обозначения второй буквой в качестве оператора можно использовать английскую букву ‘h’, по аналогии с принятой нормой в ряде языков на латинической основе (немецкий, английский), достаточно редко встречающуюся в сочетаниях с согласными в татарских словах.

Таким образом, мы получаем следующие обозначения фонем: ‘Ч’ – ‘Ch’, ‘Н’ – ‘Nh’, ‘Ш’ – ‘Sh’, ‘Ж’ – ‘Zh’, являющиеся достаточно естественными и с точки зрения общепринятых норм написания. В этом ряду вне рассмотрения остается буква ‘Ж’, которую можно было бы оставить в том же виде, как и в принятом алфавите – ‘С’. Вместе с тем для фонемы ‘Ж’ возможно ввести следующую комбинацию букв: ‘Jh’. Тогда буквой ‘С’ можно обозначить фонему ‘Ц’, которая в новом алфавите передается сочетанием букв ‘Ts’, или фонему ‘Ч’, которая гораздо более частотная для татарского языка (ср.: коэффициент частотности ‘Ч’ – 0,015, а фонемы ‘Ц’ – 0,0004). В том случае, когда ‘Jh’ передает звук ‘Ж’, буква ‘J’, во изменение нового алфавита, обозначает фонему ‘И’ (‘и’ краткое). Соответственно освобождается буква ‘Y’, которая традиционно обозначает букву ‘Ы’, например, как в слове ‘KRYM’, в результате чего появляется возможность освободиться от целого ряда неприятностей при напи-

сании татарских букв 'Ы' и 'И', вернув для *й̄* обычное написание, принятое в классическом латинском алфавите – 'Y', а 'Ы' обозначив как 'Y'. Эти буквы в новом алфавите отображены 'I' ('I' с точкой или "палка с точкой") и 'I' ('I' без точки или "палка без точки"), соответственно, и порождают массу неприятностей (особенно при разработке шрифтов и их распознавании, путаница с заглавной буквой 'И' в татарском и английском алфавитах, сложности в компьютерных технологиях и др.).

Далее рассмотрим такие одиозные буквы, за которые цепляются некоторые филологи, – это твердые варианты букв 'Г' и 'К'. Они введены в новый алфавит и обозначены буквами 'G' – 'G' с ижицей ("галочкой") и 'Q'. Во-первых, создается прецедент, когда в алфавит вводятся пары согласных по принципу: "мягкий–твердый". Из истории хорошо известно, что у татар уже был такой алфавит – рунический, где все согласные были парные: "твердые–мягкие". Однако в настоящее время такой возврат, тем более в нескольких буквах, не является обоснованным, так как твердость и мягкость татарских согласных в настоящее время однозначно определяется по закону сингармонизма, по мягкости–твердости контекста из гласных букв. Вместо того чтобы установить приоритет закона сингармонизма родного языка, вводя мягкие и твердые пары для букв 'Г' и 'К', мы сознательно идем на его нарушение, т.е. отвергаем правила татарского языка только ради закрепления произношения, пришедшего из другого языка с его фонетическим рядом и ставшего близким нашему слуху (читай: слуху некоторых филологов).

Введение твердых 'Г' и 'К' многократно усложняет разработку автоматического конвертора татарских текстов с татарской кириллицы на латиницу, требуя различать в тексте заимствованные и родные для татар слова (что, я думаю, будет не просто не только для машины, но и для носителей языка), а в некоторых случаях делает ее практически невозможной без обращения к смыслу предложения. Скажем, как можно было бы, даже прекрасно зная татарский язык, перевести с кириллицы на латиницу следующий текст: *канатка канат сукты?* Возможны четыре варианта: *kanatqa kanat suqti* ('канат ударил по канату'), *qanatqa kanat suqti* ('по крылу ударил канат'), *qanatqa qanat suqti* ('по крылу ударило крыло'), *kanatqa qanat suqti* ('по канату ударило крыло'), и все они верны с точки зрения языка.

Резюмируя сказанное, нами предлагается в новом алфавите оставить только варианты графем *K, G*, и пусть закон сингармонизма решает их морфонологию. Путаницы будет меньше, чем когда мы искусственно пытаемся внести порядок.

Такая путаница ожидает нас и с буквами 'V' и 'W'. Фамилия 'Василов' должна конвертироваться как 'Wasilov', а 'Васильев' – как 'Vasil'ev'. Пытаясь уйти от одной сложности, мы порождаем другую, когда изучающий татарский язык или даже носитель языка всегда должен помнить, не является ли данное слово заимствованным. Оно уже никогда не станет для него родным, за этим будет следить специальное правило. Лучшим выходом из ситуации было бы обогащение татарского языка новым словом через его фонетическую ассимиляцию.

Буква *w̄* одна вполне справится с отображением фонемы 'В'. Даже использование только буквы 'V' – прямого фонетического аналога кириллической буквы 'В' будет менее проблематично (во всяком случае, с точки зрения татарского языка). Во всяком случае, сложностей будет гораздо меньше, чем при одновременном использовании 'V' и 'W'.

Такая же ситуация с буквами 'X' и 'H'. Даже носитель языка, хорошо владеющий татарским языком, зачастую не знает, когда писать ту и другую буквы, настолько они схожи в речи. 'X', как правило, смягчается, а 'H' произносится чуть тверже, чем при отдельном алфавитном произношении. В речи они практически сливаются.

Очевидный выход – положиться на тот же закон сингармонизма, оставить только 'H', исключив 'X'. Тогда, например, получаем следующее написание слов: 'hat', 'hərħaldə'. Именно вариант 'h' и выбран нами для включения в алфавит на основе

латиницы, предлагаемого в данной статье. Вместе с тем имеется очевидный аргумент против использования в латиническом татарском алфавите буквы 'X'. На латинице эта буква используется для обозначения фонемы 'KS', а не 'XA', как в нашем новом алфавите. То есть, включая букву 'X' в алфавит, заранее готовим себе те же грабли – за одной и той же графемой уже в новом алфавите закрепляются две совершенно разные фонемы. Вместе с тем известно, что в латинических алфавитах, как правило, фонема 'X' обозначается комбинацией двух букв: 'KH', скажем, 'Bukharaev', 'Sakhibullin', 'Khayrullof'. Таким образом, получается, что фонема 'X' из татарского кириллического алфавита переносится в новый алфавит на основе латиницы под той же "крышей", закрепив новую путаницу с фонемами. В такой ситуации очевидно, что букву 'X' лучше не использовать в новом алфавите, а при необходимости (скажем, если даже и отображать данный аллофон в языке графически) для отображения фонемы 'X' в латиническом написании использовать комбинацию букв: 'Kh'.

Переходя к гласным буквам, можно с уверенностью утверждать, что принцип отображения их комбинацией букв или с использованием оператора мягкости или твердости здесь не проходит. Прежде всего потому, что гласные буквы в татарском языке очень мобильны и сами одновременно играют роль операторов мягкости или твердости. В отличие от английского, имеющего главным образом корневое изменение и очень слабое аффиксальное, татарский является агглютинативным языком, в котором к корню могут присоединяться аффиксальные морфемы (потенциально – без ограничений), в которых очень активны гласные фонемы. Словоформа, образованная присоединением к корню даже двух-трех аффиксов с фонемами (а это практически норма для языка), составленными по формуле "буква + оператор", становится практически нечитательной.

В этом легко убедиться на простом примере: 'kuhbahlahklahrgah'. Это простое слово 'кУбэлэкларгэ – бабочкам' образовано с использованием смягчающего оператора 'h'. Использование других символов тоже не разрешает ситуацию: 'kuvbavlavklavrgav', 'kuebaelaeklaergae', 'ku'ba'la'kla'rga'¹. Очевидно, что с татарскими гласными ни вариант комбинации букв, ни применение оператора не являются приемлемыми. В английском алфавите, хотя и имеются схожие фонемы, они передаются не графемой, а сочетанием букв, причем не всегда одинаковым. Таким образом, использование специфических графем оказывается единственно приемлемым вариантом.

Начертание мягких гласных можно оставить такими, как они приведены в новом алфавите ('a' – 'ə', 'u' – 'u' ('u' умляут), 'o' – 'o'), однако более прогрессивным и удобным, с точки зрения технологий, представляется унифицированный вариант пар гласных букв: 'a' – 'a' ('a' умляут), 'u' – 'u' ('u' умляут), 'o' – 'ö' ('o' умляут). Вполне приемлемо для татарского языка (как это имеет место и в алфавитах других языков, скажем, буква / в английском алфавите), когда рукописные и наборные начертания мягких гласных различаются, поддерживая в рукописном варианте "прошлый" привычный образ гласных фонем (привычный – без "рожков и двоеточий"). В этом консерватизме, похоже, есть нечто стабилизирующее, связующее и обеспечивающее некое смягчение перехода на новый алфавит, возможно, связанное с тонким телом нации. В таком случае в алфавите в качестве примера рукописного варианта будут фигурировать пары мягких гласных: 'əə', 'Aa'; 'Өө', 'Oö'; 'Yy', 'Uü'.

Ниже представлены таблицы с вариантами татарского алфавита на основе латиницы, состоящего из 26 букв, предлагаемыми автором статьи. Нами разработаны три варианта алфавита, различающиеся только лексико-графической упорядоченностью (ЛГУ) на основе различных принципов.

Вариант № 1: порядок графем скоординирован с порядком соответствующих графем алфавита английского языка (т.е. на основе ЛГУ английского алфавита).

Вариант № 2: тот же алфавит по варианту № 1 упорядочен (уточнен) дополнительно таким образом, что графемы сгруппированы по принципу фонемной близости.

Вариант № 3: порядок графем соответствует фонемному порядку алфавита русского языка. Данный вариант предложен в целях соотнесения татарского и русского алфавитов для удобства составления параллельных словарей, что особенно актуально в условиях функционирования двух государственных языков (т.е. татарский алфавит на основе латиницы упорядочен на базе ЛГУ русского алфавита по фонемному принципу).

Вариант № 4: порядок графем определен на основе статистических данных частотности фонем (на статистическом материале, приведенном Р.Г. Бухараевым, Т.И. Ибрагимовым, А.И. Еникеевым в сборнике трудов [4]).

Вариант № 5: тот же алфавит (вариант № 4) упорядочен по принципу фонемной близости графем.

Кроме того, для каждого из пяти вариантов разработаны три уточненные таблицы (А, В, С), отображающие те же алфавиты для случаев, когда буква С используется для отображения фонем 'Ц', 'Ж' или 'Ч', соответственно.

На первый взгляд, букву 'Ч' целесообразно отображать привычной комбинацией 'Ch', что, как правило, и предлагается сторонниками передачи фонем в виде комбинации букв. Однако, как показывает изучение частотности этих трех фонем и букв, фонема 'Ч' (коэффициент частотности: 0,015) в 3 раза чаще встречается в татарских текстах, нежели фонема 'Ж' (коэффициент частотности: 0,005) (фонема 'Ц', как считается, не присуща для литературного татарского языка), а буква 'Ч' даже еще более частотна, чем буквы 'Ж' и 'Ц'. Коэффициент частотности буквы 'Ч' – 0,017, буквы 'Ж' – 0,004, буквы 'Ц' – 0,0004. То есть буква 'Ч' встречается более чем в 4 раза чаще буквы 'Ж' и более чем в 40 раз чаще, нежели буква 'Ц'. Этот показатель для буквы 'Ц' относительно 'Ч' может варьировать от 10–12 раз в технических текстах до более чем 400 раз в татарской прозе. Например, в произведении Н. Гиматдиновой "Пэри утарында" [5] буква 'Ч' встречается более 1200 раз, тогда как буква 'Ц' – всего в трех случаях.

Приведенные частотные показатели достаточно убедительно демонстрируют, что татарские тексты будут существенно короче, и потребуются меньше нажатий на клавишу, когда именно фонема 'Ч', а не фонемы 'Ж' и 'Ц', будет отображаться одной буквой. В нашем случае – латинской буквой 'C' (как это было в Yanalif 1). Исходя из этого, наиболее предпочтительными во всех пяти вариантах нам представляются алфавиты, приведенные в таблицах С).

Вариант № 1 (на основе английского алфавита)

А) Ц – С, Ж – Jh, Ч – Ch

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	O	P	R	S	T	U	U	W	Y	Z
a	э	б	ц	д	э	ф	г	ь	И	й	к	л	м	н	о	ө	п	р	с	т	у	У	в	ы	з

В) Ж – С, Ч – Ch, Ц – Ts

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	Ä	B	C	D	E	F	G	H	I	J	K	L	M	N	O	O	P	R	S	T	U	U	W	Y	Z
a	э	б	ж	д	э	ф	г	ь	и	й	к	л	м	н	о	ө	п	р	с	т	у	У	в	ы	з

С) Ч – С, Ц – Ts, Ж – Jh

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	O	P	R	S	T	U	U	W	Y	Z
a	э	б	ч	д	э	ф	г	ь	и	й	к	л	м	н	о	ө	п	р	с	т	у	У	в	ы	з

Вариант № 2 (вариант № 1, в котором графемы сгруппированы справа–налево при признаку фонемной близости)

А) Ц – С, Ж – Jh, Ч – Ch

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	P	C	Z	S	E	Y	I	J	F	W	U	U	K	G	H	L	R	M	N	O	Ö	T	D
a	ə	б	п	ц	з	с	э	ы	и	й	ф	в	у	У	к	г	ь	л	р	м	н	о	ө	т	д

В) Ж – С, Ч – Ch, Ц – Ts

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	P	C	G	K	H	D	T	E	Y	F	W	I	J	L	R	M	N	O	Ö	S	Z	U	U
a	ə	б	п	ж	г	к	ь	д	т	э	ы	ф	в	и	й	л	р	м	н	о	ө	с	з	у	У

С) Ч – С, Ц – Ts, Ж – Jh

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	P	C	K	G	H	D	T	E	Y	F	W	I	J	L	R	M	N	O	Ö	S	Z	U	U
a	ə	б	п	ч	к	г	ь	д	т	э	ы	ф	в	и	й	л	р	м	н	о	ө	с	з	у	У

Вариант № 3 (на основе русского алфавита по фонемному принципу)

А) Ц – С, Ж – Jh, Ч – Ch

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	W	G	D	Z	I	J	K	L	M	N	O	O	P	R	S	T	U	U	F	H	C	Y	E
a	ə	б	в	г	д	з	и	й	к	л	м	н	о	ө	п	р	с	т	у	У	ф	ь	ц	ы	э

В) Ж – С, Ч – Ch, Ц – Ts

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	B	W	G	D	C	Z	I	J	K	L	M	N	O	O	P	R	S	T	U	U	F	H	Y	E
a	ə	б	ж	д	з	ф	г	ь	И	й	к	л	м	н	о	ө	п	р	с	т	у	У	в	ы	э

С) Ч – С, Ц – Ts, Ж – Jh

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	Ä	B	W	G	D	Z	I	J	K	L	M	N	O	O	P	R	S	T	U	Ü	F	H	C	Y	E
a	ə	б	в	г	д	з	и	й	к	л	м	н	о	ө	п	р	с	т	у	У	ф	ь	ч	ы	э

Вариант № 4 (на основе частотности фонем)

А) Ц – С, Ж – Jh, Ч – Ch

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	Ä	N	Y	E	R	L	K	T	I	J	G	M	B	D	S	U	Z	P	U	O	O	W	F	H	C
a	ə	н	ы	э	р	л	к	т	и	й	г	м	б	д	с	у	з	п	У	ө	о	в	ф	ь	ц

В) Ж – С, Ч – Ch, Ц – Ts

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	Ä	N	Y	E	R	L	K	T	I	J	G	M	B	D	S	U	Z	P	U	O	O	W	C	F	H
a	ə	н	ы	э	р	л	к	т	и	й	г	м	б	д	с	у	з	п	У	ө	О	в	ж	ф	ь

С) Ч – С, Ц – Ts, Ж – Jh

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	Ä	N	Y	E	R	L	K	T	I	J	G	M	B	D	S	U	Z	C	P	U	O	O	W	F	H
a	ə	н	ы	э	р	л	к	т	и	й	г	м	б	д	с	у	з	ч	п	У	ө	о	в	ф	ь

Вариант № 5 (по созвучию фонем на основе алфавита по частотности)

А) Ц – С. Ж – Jh. Ч – Ch

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	N	M	Y	E	R	L	K	G	H	T	D	I	J	B	P	S	Z	C	U	U	O	O	W	F
а	э	м	н	ы	э	р	л	к	г	ь	т	д	и	й	б	п	с	з	ц	у	У	о	ө	в	ф

В) Ж – С. Ч – Ch. Ц – Ts

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	N	M	Y	E	R	L	K	C	G	H	T	D	I	J	B	P	S	Z	U	U	O	Ö	W	F
а	э	м	н	ы	э	р	л	к	ж	г	ь	т	д	и	й	б	п	с	з	у	У	о	ө	в	ф

С) Ч – С. Ц – Ts. Ж – Jh

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	Ä	N	M	Y	E	R	L	K	C	G	H	T	D	I	J	B	P	S	Z	U	U	O	Ö	W	F
а	э	н	м	ы	э	р	л	к	ч	г	ь	т	д	и	й	б	п	с	з	у	У	о	ө	в	ф

Ниже приведена таблица, отображающая два алфавита: Yanalif 2, принятый по указу президента Республики Татарстан и признанный в настоящее время экспериментальным, и Yanalif-SLT (26 букв), предлагаемый в качестве варианта.

Известно, что порядок расположения букв и их численное значение имеют огромную важность в эзотерических толкованиях, что все чаще находит свое научное обоснование. Поэтому при окончательном утверждении алфавита представляется необходимым (целесообразным) проведение анализа не только состава алфавита (количество графем и их начертаний), но и исследование порядка следования графем и расположение их исходя из метафизического смысла их численных значений, проверяя их соответствие на ключевых словах мироздания. Тем более что разница между числовыми значениями по частотности в большинстве своем настолько несущественна, что они вполне могут быть смещены в определенных пределах, возможно даже, что полученные данные могут изменяться (уточняться) в ту или другую сторону при увеличении тестового объема текстов. (Кроме фонемы 'А', имеющей наибольший коэффициент встречаемости: 0,113, и занимающей, безусловно, первую позицию. Другие фонемы по частотности расположены в следующих интервалах: от 'А'(0,072) до 'Т'(0,050); от 'Г'(0,042) до 'У'(0,025); от 'З'(0,016) до 'О'(0,010); от 'W'(0,007) до 'Н'(0,003). Расположение графем по варианту 4 С.)

Очевидно, что такая работа сама по себе очень емкая, ответственная и требует проведения глубоких исследований на основе огромного текстового материала с разработкой лингвистических моделей и применением математического аппарата.

Заключение

На семинарах в лаборатории обсуждались разные варианты латиницы и даже был выпущен специальный сборник, в котором, в частности, анализируется ряд вариантов алфавита на основе латиницы и предлагается некий собственный алфавит, предложенный специалистами лаборатории. Однако, поддерживая переход на латиницу в принципе и беспокоясь, что в мелких спорах (и как оказалось – обоснованно) можно потерять саму возможность перехода на латинскую графику, мы практически отказались от продвижения своего варианта.

Пользуясь тем, что эксперимент по поводу перехода на латиницу продолжается, в данной статье проанализирован алфавит, принятый указом президента РТ, и предложен новый вариант татарского алфавита на основе латиницы, состоящий из 26 букв.

	Yanalif-2	Произношение	Кириллица	Yanalif-STL	Произношение	Кириллица
1	A, a	a	А, а	A, a	A	А, а
2	Ə, ə	ə	Ә, ә	A, a; Ə, ə	ə	Ә, ә
3	B, b	be	Б, б	B, b	Be	Б, б
4	C, c	ce	Ж, ж	C, c	Che	Ч, ч
5	Ç, ç	che	Ч, ч	–	–	–
6	D, d	de	Д, д	D, d	De	Д, д
7	E, e	e	Е, е(ə)	E, e	E	Е, е(ə)
8	F, f	ef	Ф, ф	F, f	Ef	Ф, ф
9	G, g	ge	Г, г	G, g	Ge	Г, г
10	G, ğ	gı	Гъ, гъ	–	–	–
11	H, h	he	Һ, һ	Ĥ, h	He	Һ, һ
12	I, ı	ı	Ы, ы	–	–	–
13	İ, i	i	И, и	I, i	I	И, и
14	J, j	je	Ж, ж	J, j	Jod	Й, й
15	K, k	ke	К, к	K, k	Ke	К, к
16	Q, q	q	Къ, къ	–	–	–
17	L, l	el	Л, л	L, l	El	Л, л
18	M, m	em	М, м	M, m	Em	М, м
19	N, n	en	Н, н	N, n	En	Н, н
20	Ŋ, ŋ	eng	Ŋ, ŋ	–	–	–
21	O, o	o	О, о	O, o	O	О, о
22	Ө, ө	ө	Ө, ө	O, o; Ө, ө	ө	Ө, ө
23	P, p	pe	П, п	P, p	Pe	П, п
24	R, r	er	Р, р	R, r	Er	Р, р
25	S, s	es	С, с	S, s	Es	С, с
26	Ş, ş	şa	Ш, ш	–	–	–
27	T, t	te	Т, т	T, t	Te	Т, т
28	U, u	u	У, у	U, u	U	У, у
29	Ü, ü	ü	У, у	U, ü; Ü, ü	U	У, у
30	V, v	ve	В, в	–	–	–
31	W, w	we	В, в (y)	W, w	We	В, в (y)
32	X, x	xa	Х, х	–	–	–
33	Y, y	y	Й, й	Y, y	Ы	Ы, ы
34	Z, z	ze	З, з	Z, z	ze	З, з

Ц – Ts, Ж – Jh, Ш – Sh, Щ – Tc/ Tsh, Ж – Zh, Ĥ – Nh

Примечание

¹ Принцип сингармонизма в татарском языке предоставляет интересный вариант применения оператора “смягчения”. В том случае, когда слово является “мягким” (составленным из гласных букв “переднего ряда”), оператор можно вывести и ставить перед словом (при первой букве слова), обозначая тем самым, что все слово произносится “мягко”, т.е. все гласные в слове переднего ряда. Тогда слово *ku'ba'la'kla'rga'* запишется следующим образом: *kubalak-larga*. На наш взгляд, такой подход достаточно привлекателен, однако совершенно нетрадиционен, непривычен для татарского алфавита. Возможно, он затрагивает основы менталитета нации, поэтому требует проведения специальных исследований и в рамках данной статьи далее не развивается.

Литература

- Сулейманов* 1998 – *Сулейманов Д.Ш.* Формальная эlegantность и естественная сложность морфологии татарского языка // Электронная конф.: Информационные технологии в гуманитарных науках (Казань, 5–31 мая, 1998 г.). Казань, 1998. http://www.kcn.ru/_tat_ru/universitet/gum_konf/ot7.htm.
- Сулейманов* 1999 – *Сулейманов Д.Ш. (Кыпчак И.И.)*. Татарский язык – язык компьютерных технологий // Матер. научн.-практ. конф. “Языковая ситуация в республике Татарстан: состояние и перспективы” (17 ноября 1998 г.). В 2-х ч. Ч. 1. Казань, 1999. С. 196–198.
- Suleymanov* 1999 – *Suleymanov D.S.* Natural possibilities of the Tatar morphology as a formal base of the NLP // In Proceedings of the First International Workshop “Computerisation of Natural Languages” (Varna, Sept. 3–7, 1999). Sofia, 1999. P. 113–117.
- Татарский язык... 1995 – Татарский язык и информационные технологии. Вып. 2. Казань, 1995.
- Гыйматдинова* 1996 – *Гыйматдинова Н.* Пәри утарында // “Идел” журналы. 1996. № 10. С. 4–17, 32–39 б.
- Папюсъ* 1910 – *Папюсъ*. Каббала или наука о боге, вселенной и человеке / Пер. с фр. А.В. Трояновского. СПб., 1910.
- Плешанов* 2000 – *Плешанов А.Д.* Русский алфавит как инструмент научного познания вселенной. М., 2000.