

КВАНТИТАТИВНАЯ ЛИНГВИСТИКА

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 81'32+811.512.145

doi: 10.26907/2541-7738.2022.5.130-140

МАРКИРОВАНИЕ НАЧАЛА И КОНЦА СЛОВА В ТАТАРСКОМ ЯЗЫКЕ: СИСТЕМА ГЛАСНЫХ

А.М. Галиева¹, Ж.Е. Вавилова²

¹ООО «НекстГИС», г. Москва, 107078, Россия

²Казанский государственный энергетический университет, г. Казань, 420066, Россия

Аннотация

В естественных языках сегментация речевого потока является важнейшим аспектом декодирования речи: значение лингвистического целого можно понять, лишь разложив его на более мелкие компоненты и комбинируя их сообразно грамматическим правилам языка. Следовательно, маркирование сегментов речи должно определяться внутренним устройством языка; необходимым условием распознавания границ сегментов выступает наличие когнитивных механизмов, участвующих в процессах декодирования знаковых систем. В работе исследован вопрос о том, насколько в татарском языке начало и конец слова маркируются на уровне гласных фонем. Использование методов математической статистики при анализе художественных текстов позволило продемонстрировать, что распределение восьми гласных в первых и последних слогах имеет статистически значимые различия. При этом только фонема «а», которая примерно одинаково представлена в начальных и финальных слогах, имеет два позиционно маркированных варианта – для начальных и всех остальных слогов. Такое распределение гласных фонем позволяет, наряду с ударением и гармонией гласных, правильно определять границы словоформ. В перспективе разработанная методология может быть использована для проведения аналогичных исследований на материале других тюркских языков.

Ключевые слова: квантитативная лингвистика, татарский язык, границы словоформы, вокалическая структура словоформы, тест хи-квадрат

Введение

Выявление статистически значимых связей между разными уровнями языка и релевантных закономерностей – задача, успешно решаемая средствами компьютерной лингвистики. Обнаружение таких зависимостей может дать исследователям новую информацию о внутренней структуре языка и законах, ею управляющих. Языковые явления, несмотря на большое разнообразие и некоторую вариативность в своем статистическом поведении, характеризуются четко выраженными регулярностями, относительно устойчивой повторяемостью, определяемой структурой конкретного языка. Тексты как результаты функционирования языка состоят из большого количества элементов разной природы, связи между

которыми сложны и подвержены влиянию множества случайных факторов. Поэтому текстовые данные предоставляют нам информацию о языковых закономерностях в пределах определенных вариаций.

В естественных языках сегментация речевого потока является важнейшим аспектом декодирования речи: значение лингвистического целого можно понять, разложив его на более мелкие сегменты и комбинируя эти сегменты в соответствии с грамматическими правилами языка. Следовательно, маркирование таких сегментов должно определяться внутренним устройством языка и, соответственно, должны существовать когнитивные механизмы распознавания границ сегментов, которые необходимо изучить.

В качестве примера можно сослаться на исследования об обработке слов со сложной морфологической структурой, которые по-разному интерпретируют этот процесс. Сторонники обращения к целому слову (*whole word access*) считают, что мультиморфные слова представлены в лексиконе целиком и дискретно; приверженцы моделей, основанных на декомпозиции, утверждают, что сложные слова распознаются и интерпретируются путем их декомпозиции (разбора на значимые составляющие – морфемы). Третий подход является гибридным и предполагает, что имеются механизмы как обращения к целому слову, так и его декомпозиции, но первый тип преобладает [1].

Мы считаем, что фонологическая структура словоформы также дает ключ к пониманию того, как дискретные слова распознаются в потоке речи. Фонологический уровень языка, блестяще описанный структуралистами и их последователями, принадлежащими к разным научным школам ([2–5] и др.), является хорошим объектом для количественных исследований: текстовый инвентарь фонем поддается различным классификациям и статистическим оценкам, и неслучайно именно исследователи фонологической системы языка высказали мнение о неравномерном распределении фонем в зависимости от их позиции в словоформе [3, 6–9]. Тем не менее фонологическая система татарского языка остается недостаточно хорошо изученной методами современной количественной лингвистики, а в академических грамматиках татарского языка практически не представлены количественные оценки распределения фонем и их сочетаний [10, 11]. При этом нельзя не отметить ставшие уже классическими работы У.Ш. Байчуры ([12] и др.), в которых звуковой строй татарского языка исследовался методами экспериментальной фонетики, и труды Т.И. Ибрагимова ([13] и др.), где татарские фонемы изучались в аспекте формирования структуры слогов. Ритмико-темпоральные характеристики татарского языка также были исследованы в рамках проекта по созданию синтезатора речи [14–16].

Из последних работ по слоговым структурам можно отметить статьи [17, 18], в которых показано, что распределение слоговых структур в начале и в конце слов татарского языка характеризуется статистически значимыми различиями, обусловленными особенностями строения основ и аффиксальных цепочек. Исследование проводилось с выделением гласных и шумных и сонорных согласных в структуре слога, а для оценки статистической значимости этих различий был использован критерий хи-квадрат.

Цель настоящей работы – исследовать, насколько в татарском языке начало и конец слова маркированы на уровне гласных фонем. Маркирование границ

словоформы в языке должно быть значимо с позиций когниции, а именно для восприятия и различения языковых единиц в потоке речи. В плане просодики для определения границ словоформ важна акцентуация, тем не менее на ударение в случае татарского языка нельзя полагаться по целому ряду причин. Природа ударения в тюркских языках является вопросом дискуссионным ([10, 14, 16, 19] и др.), и в явном виде ударение определяется только для отдельно произносимых татарских слов, но не в потоке речи. Границы словоформ в определенной мере маркируются также гармонией гласных, но этот механизм функционирует только при сочетании словоформ, относящихся к разным сингармонистическим типам, например, когда словоформа с гласными переднего ряда является смежной со словоформой с гласными заднего ряда. Однако это требование в потоке речи не выполняется, и делимитативная функция сингармонизма является существенно ограниченной [20, с. 129–130].

С учетом вышесказанного представляют большой научный интерес статистические исследования фонологической структуры тюркской словоформы с точки зрения потенциальных способов маркирования ее границ. Разработанная нами методология направлена на изучение вокалической структуры словоформы татарского языка с выделением гласных фонем в начальных и финальных слогах.

Методология

В качестве эмпирических данных выступили художественные тексты на татарском языке: были взяты образцы из поэзии и прозы, всего десять текстов (для экономии места в рамках данной статьи приводятся данные, полученные по семи текстам: Алиш, Гыйльм., Еники, Ибр., Такт., Тукай1, Тукай2). В процессе подготовки данных текстовый материал был приведен к фонологически релевантной форме: одна буква – одна фонема (см. табл. 1). Позиционные варианты фонем, а также редукция кратких гласных [15] не учитывались.

Общее количество словоформ в отобранных для анализа текстах находится в диапазоне от 419 («Су анасы» Г. Тукая) до 2183 (фрагмент из текста А. Еники); совсем небольшие по объему тексты нами не рассматривались, поскольку они могут характеризоваться определенными флуктуациями, обусловленными множественностью случайных факторов.

Словоформы не были подвергнуты лемматизации. Поскольку нас интересует распределение гласных в начальных и финальных слогах, далее из текстов были отобраны для анализа только словоформы, содержащие два и более слогов.

В табл. 2 представлено распределение гласных фонем в первых и последних слогах многосложных слов, извлеченных из фрагмента повести классика татарской литературы А. Еники «Эйтелмэгэн васыять» («Невысказанное завещание»). Как видно, в начальных слогах представлен весь набор гласных фонем, в том числе лабиализованных. В финальных слогах набор гласных несколько ограничен (см. табл. 2 и 3).

Для определения статистической значимости различий в распределении гласных фонем в зависимости от позиции в словоформе был проведен тест хи-квадрат. Критерий Пирсона (χ^2) – это непараметрический метод, позволяющий оценить значимость различий между наблюдаемым количеством исходов или качественных характеристик выборки, попадающих в каждую категорию, и теоретическим

Табл. 1

Примеры фонологической транскрипции слов

| Словоформа в стандартной графической форме | Фонологическая транскрипция |
|--|-----------------------------|
| <i>урман</i> 'лес' | урман |
| <i>егет</i> 'парень' | јегет |
| <i>елан</i> 'змея' | јылан |
| <i>ямьле</i> 'красивый, приятный' | јәмле |
| <i>якын</i> 'близкий' | јакын |
| <i>кияу</i> 'жених, зять' | кийәw |

Табл. 2

Количество гласных фонем в начальных и финальных словах в тексте А. Еники и результаты теста хи-квадрат

| Фонема | Наблюдаемые значения | | Ожидаемые значения | |
|---------------------------|---|-----------------|--------------------|-----------------|
| | начальные слоги | финальные слоги | начальные слоги | финальные слоги |
| а | 520 | 490 | 505 | 505 |
| ы | 120 | 460 | 290 | 290 |
| у | 204 | 11 | 108 | 107 |
| о | 124 | 8 | 66 | 66 |
| ә | 148 | 338 | 243 | 243 |
| е | 164 | 432 | 298 | 298 |
| ү | 97 | 1 | 49 | 49 |
| ө | 116 | 0 | 58 | 58 |
| и | 300 | 52 | 176 | 176 |
| Результаты теста χ^2 | $\chi^2 = 1054.9, df = 8, p\text{-value} < 2.2e-16$ | | | |

количеством, которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы (см. ожидаемые значения в табл. 2).

Тест хи-квадрат выполняется согласно формуле

$$T = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

где O_{ij} – наблюдаемое значение, E_{ij} – ожидаемое значение, определяемое нулевой гипотезой, r и c – количество строк и столбцов соответственно в таблице сопряженности.

Ожидаемые значения вычисляются по формуле

$$E_{ij} = \frac{S_i S_j}{N},$$

где S_i – сумма в строке i , S_j – сумма в столбце j , N – общее количество наблюдений.

Нулевая гипотеза теста хи-квадрат состоит в том, что между столбцами и строками в таблице сопряженности нет никакой связи: событие «наблюдение в строке i » не зависит от события, когда то же самое наблюдение оказывается в столбце j [21, с. 204–205].

Все этапы работы с текстовым материалом, включая подготовку текстовых образцов, вычисления и визуализацию данных, выполнены на языке R [22].

Результаты и их обсуждение

Как показано в табл. 2, количество гласных фонем в начальных и финальных слогах в тексте А. Еники статистически значимо различается, о чем свидетельствует очень маленькое значение p -value. Обратим внимание на то, что достаточно близки значения только для фонемы «а» (встречается 520 раз в начальных слогах и 490 раз в финальных слогах).

График распределения гласных в начальных и финальных слогах слов в тексте А. Еники (рис. 1) наглядно демонстрирует данные, представленные в табл. 2. В случае отсутствия существенных различий в распределении количества гласных в начальных и финальных слогах эти гласные должны были бы располагаться вдоль диагональной прямой, чего мы не наблюдаем.

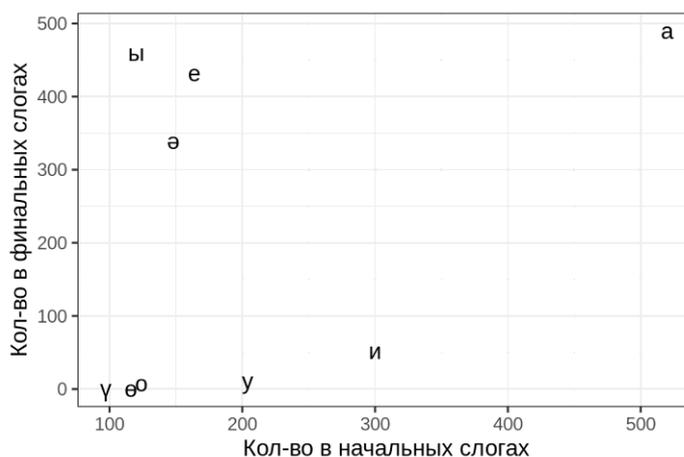


Рис. 1. Количество гласных фонем в первых и последних слогах в тексте А. Еники

В табл. 3 приведены аналогичные результаты для еще шести текстов на татарском языке, а именно наблюдаемые распределения гласных в начальных и финальных слогах и также результаты теста хи-квадрат.

Для визуализации категориальных данных удобно использовать мозаичные графики. На диаграмме (рис. 2) представлено распределение гласных фонем в зависимости от их позиции в словоформе в тексте А. Еники; цветом и линией обозначены стандартизованные (пирсоновские) остатки: ячейки, представляющие отрицательные остатки, окрашены в розовый цвет; положительные остатки окрашены синим цветом. Из рисунка видно, что для фонемы «а» позиционные различия статистически не значимы, остальные гласные фонемы недостаточно представлены в одной позиции и избыточно – в другой.

Табл. 3

Распределение гласных в начальных и финальных слогах и результаты теста хи-квадрат для татарских текстов

| Фонема | Г. Тукай «Шүрәле» | | Г. Тукай «Су анасы» | | Г. Ибрагимов «Кызыл чәчәкләр» | |
|---------------------------|---|-----------------|---|-----------------|---|-----------------|
| | начальные слоги | финальные слоги | начальные слоги | финальные слоги | начальные слоги | финальные слоги |
| а | 157 | 176 | 113 | 94 | 110 | 107 |
| ә | 70 | 144 | 25 | 44 | 30 | 77 |
| е | 67 | 171 | 32 | 61 | 29 | 92 |
| и | 120 | 23 | 39 | 10 | 60 | 8 |
| о | 46 | 0 | 18 | 0 | 32 | 0 |
| ө | 35 | 0 | 16 | 0 | 39 | 0 |
| у | 80 | 5 | 43 | 1 | 41 | 6 |
| ү | 55 | 2 | 7 | 1 | 22 | 3 |
| ы | 46 | 155 | 12 | 94 | 19 | 89 |
| Результаты теста χ^2 | $\chi^2 = 393.48, df = 8, p\text{-value} < 2.2e-16$ | | $\chi^2 = 175.21, df = 8, p\text{-value} < 2.2e-16$ | | $\chi^2 = 250.13, df = 8, p\text{-value} < 2.2e-16$ | |
| Фонема | Г. Гыйльман «Очрашу» | | А. Алиш «Койрыклар» | | Х. Такташ «Алсу» | |
| | начальные слоги | финальные слоги | начальные слоги | финальные слоги | начальные слоги | финальные слоги |
| а | 229 | 170 | 186 | 234 | 142 | 94 |
| ә | 101 | 179 | 35 | 197 | 32 | 84 |
| е | 95 | 227 | 89 | 160 | 28 | 129 |
| и | 136 | 33 | 130 | 11 | 57 | 4 |
| о | 37 | 3 | 54 | 0 | 21 | 1 |
| ө | 63 | 0 | 59 | 0 | 27 | 0 |
| у | 80 | 6 | 106 | 6 | 43 | 18 |
| ү | 32 | 1 | 64 | 1 | 66 | 1 |
| ы | 45 | 199 | 33 | 147 | 13 | 98 |
| Результаты теста χ^2 | $\chi^2 = 429.23, df = 8, p\text{-value} < 2.2e-16$ | | $\chi^2 = 574.83, df = 8, p\text{-value} < 2.2e-16$ | | $\chi^2 = 327.67, df = 8, p\text{-value} < 2.2e-16$ | |

Итак, мы показали, что распределение восьми гласных фонем в крайних – первых и последних – слогах татарских словоформ статистически значимо различается. При этом исключением является фонема «а», которая в начальных и финальных слогах характеризуется примерно одинаковой частотностью (см. табл. 2 и 3, рис. 2).

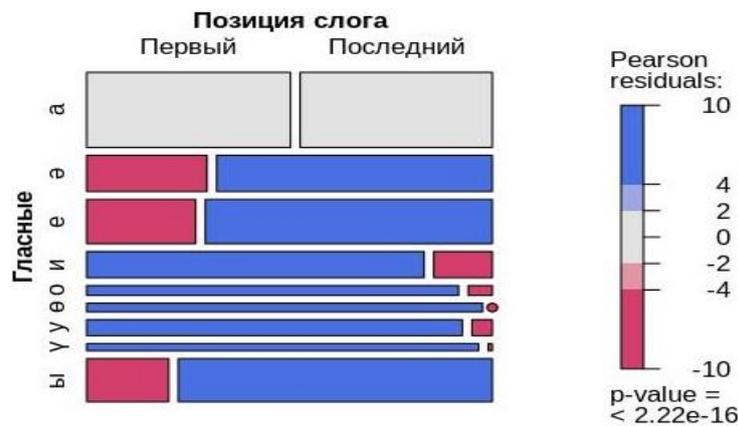


Рис. 2. Гласные фонемы в первых и последних слогах словоформы: мозаичный график

При этом необходимо учитывать, что именно фонема «а» в татарском языке имеет два отличающихся (позиционно маркированных варианта): *a* огубленное («а^о») в первом слоге и обычное «а» во всех прочих [10, с. 90–91], например:

балалар ‘дети’ «ба^олалар»,
авылда ‘в деревне’ «а^овылда»

И это очень важный момент: фонема «а» в начальном слоге татарских слов произносится совсем иначе, чем в других слогах. Можно считать, что с учетом этой особенности частотность распределения различается для всех фонем: гласные «ә», «е», «и», «о», «ө», «у», «ү», «ы» характеризуются разной встречаемостью (количественный фактор) в начальных и финальных слогах (см. эмпирические распределения гласных и результаты теста хи-квадрат в табл. 2 и 3), а фонема «а», у которой частотность в начальных и финальных слогах примерно одинаковая, имеет специфический позиционный вариант для первого слога (качественный фактор). Сочетание количественных и качественных факторов приводит к тому, что вокалическая структура первых и последних слогов татарской словоформы оказывается различной.

Заключение

Таким образом, результаты статистического анализа свидетельствуют о том, что распределение гласных фонем в татарском языке в начальных и финальных слогах статистически значимо различается. Для единственной фонемы «а», частотность которой в начальных и финальных слогах примерно одинакова, в татарском языке имеются позиционные варианты (для других гласных фонем таких явно выраженных вариантов не существует).

Мы можем предположить, что такое распределение гласных фонем обусловлено агглютинативной природой татарской словоформы и позволяет говорящим правильно определять границы словоформ (наряду с ударением и гармонией гласных, которые сами по себе не достаточны для четкого и однозначного определения границ словоформ).

Такие статистически значимые различия в распределении гласных имеют важное когнитивное значение, так как позволяют маркировать начало и конец лексических единиц в потоке речи.

Представляется перспективным проведение исследований на материале других тюркских языков, особенно тех, где не наблюдается лабиализация фонемы <a> в начальных слогах.

Источники

- Алиш – *Алиш А.* Койрыклар. – URL: https://tugan-tel.tatar/adabiyat/proza/Abdulla_Alish/Koiryklar.php, свободный.
- Гыйльм. – *Гыйльман Г.* Очрашу. – URL: <http://baigysh.tatar/ocrasu-gilman/>, свободный.
- Еники – *Еники А.* Әйтелмәгән васыять. – URL: <http://belgech.ru/proizvedeniya/itelmagan-vasuyat>, свободный.
- Ибр. – *Ибраһимов Г.* Кызыл чәчәкләр. – URL: <http://belgech.ru/proizvedeniya/kyzyl-chachaklar>, свободный.
- Такт. – *Такташ Г.* Алсу. – URL: <https://shigriyat.ru/authors/taktash/64>, свободный.
- Тукай1 – *Тукай Г.* Шүрәле. – URL: <http://gabdullatukay.ru/works/poem/1907/shurele/>, свободный.
- Тукай2 – *Тукай Г.* Су анасы. – URL: <https://tatobzor.ru/tatarskie-skazki/su-anasy-kiyat-gabdulla-tukaj.html>, свободный.

Литература

1. *Badecker W., Allen M.* Morphological parsing and the perception of lexical identity: A masked priming study of stem homographs // *J. Mem. Lang.* – 2002. – V. 47, No 1. – P. 125–144. – doi: 10.1006/jmla.2001.2838.
2. *Jakobson R., Halle M.* Fundamentals of Language. – The Hague: Mouton & Co – S.-Gravenhage, 1956. – 89 p.
3. *Jakobson R., Waugh L.* The Sound Shape of Language. – Brighton: Harvester Press, 1979. – 308 p.
4. *Chomsky N., Halle M.* The Sound Pattern of English. – Cambridge, Mass.: Harper & Row, 1968. – 470 p.
5. *Трубецкой Н.С.* Основы фонологии / Пер. с нем. А.А. Холодовича; под ред. С.Д. Кацнельсона. – М.: Аспект Пресс, 2000. – 352 с.
6. *Байчура У.Ш.* Гласные татарского языка в свете экспериментальных данных: Автореф. дис. ... канд. филол. наук. – М.-Л., 1953. – 18 с.
7. *Щербак А.М.* Сравнительная фонетика тюркских языков. – Л.: Наука, 1970. – 204 с.
8. *Altmann G.* Phonic word structure // *Quantitative Linguistik. Ein internationales Handbuch* / Eds. R. Köhler, G. Altmann, R. Piotrowski. – Berlin; N. Y.: M. de Gruyter, 2005. – P. 191–208.
9. *Баскаков Н.А.* Историко-типологическая фонология тюркских языков / Отв. ред. Э.Р. Тенишев. – М.: Наука, 1988. – 208 с.
10. Татарская грамматика: в 3 т. / Гл. ред. М.З. Закиев. – Казань: Тат. кн. изд-во, 1993. – Т. 1. – 584 с.
11. Татар грамматикасы / Жит. М.З. Зәкиев, ред. Ф.М. Хисамова. – Казан: ТӘҺСИ, 2015. – Т. 1. – 512 б.

12. *Байчура У.Ш.* Звуковой строй татарского языка: Экспериментально-фонетический очерк: в 2 т. – Казань: Изд-во Казан. ун-та, 1959. – Т. 1. – 183 с.
13. *Ибрагимов Т.И.* Изучение образования слогов и структуры их сочетаний в татарском литературном языке: Дис. ... канд. филол. наук. – Казань, 1970. – 167 с.
14. *Сайхунов М.Р.* Ритмо-темпоральные характеристики татарского языка в плане автоматического синтеза речи: Автореф. дис. ... канд. филол. наук. – Казань, 2010. – 26 с.
15. *Ибрагимов Т.И., Сайхунов М.Р.* К проблеме реконструкции звучания письменного текста. Редукция кратких гласных в татарском языке // Учен. зап. Казан. ун-та. Сер. Гуманит. науки. – 2009. – Т. 151, кн. 3. – С. 288–298.
16. *Ибрагимов Т.И., Сайхунов М.Р.* О проблеме словесного ударения в татарском языке (сравнительно с русским языком) // Русская и сопоставительная филология. – Казань: Казан. гос. ун-т, 2010. – С. 56–59.
17. *Галиева А.М.* Структура слога в татарском языке: от данных к модели // Int. J. Open Inf. Technol. – 2020. – V. 8, No 1. – P. 9–16.
18. *Galieva A., Vavilova Z.* Initial and final syllables in Tatar: From phonotactics to morphology // Glottometrics. – 2021. – V. 50. – P. 57–75.
19. Сравнительно-историческая грамматика тюркских языков. Фонетика / Отв. ред. Э.Р. Темишев. – М.: Наука, 1984. – 488 с.
20. *Касевич В.Б.* Морфонология. – Л.: Изд-во Ленингр. ун-та, 1986. – 170 с.
21. *Conover W.J.* Practical Nonparametric Statistics. – N. Y.: Wiley, 1999. – viii+584 p.
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. – Vienna, 2021. – URL: <https://www.R-project.org/>, свободный.

Поступила в редакцию
27.06.2022

Галиева Альфия Макаримовна, кандидат философских наук, доцент, специалист по анализу данных

ООО «НексГИС»

ул. Новая Басманная, д. 23Б, стр. 20, г. Москва, 107078, Россия

E-mail: amgalieva@gmail.com

Вавилова Жанна Евгеньевна, кандидат философских наук, доцент кафедры философии и медиакоммуникаций

Казанский государственный энергетический университет

ул. Красносельская, д. 51, г. Казань, 420066, Россия

E-mail: zhannavavilova@mail.ru

Marking the Beginning and End of the Tatar Word: The System of VowelsA.M. Galieva^{a*}, Zh.E. Vavilova^{b**}^aNextGIS Company, Moscow, 107078 Russia^bKazan State Power Engineering University, Kazan, 420066 Russia

E-mail: *amgalieva@gmail.com, **zhannavavilova@mail.ru

Received June 27, 2022

Abstract

In natural languages, segmentation of the speech flow is the most important aspect of speech decoding: the meaning of a linguistic whole can be understood only if it is broken down into smaller elements and these are combined according to the grammatical rules of the language. Thus, the marking of speech segments must be determined by the internal structure of the language. A necessary condition for defining the boundaries of such segments is the presence of cognitive mechanisms involved in the processes of decoding sign systems. For a number of reasons, certain difficulties arise when identifying the boundaries of word forms in the Tatar speech. Therefore, statistical studies on the phonological structure of the Turkic word form and on potential ways of marking its boundaries are of great scientific interest. Through studying the vocal structure of the word form with the allocation of vowel phonemes in the initial and final syllables, this article investigates how the beginning and end of Tatar words are marked at the level of vowels. Literary texts were used as the empirical material for this study. With the methods of mathematical statistics, it was demonstrated that the distribution of eight vowels in the first and last syllables has statistically significant differences. At the same time, only the phoneme ⟨a⟩, which is roughly equally represented in the initial and final syllables, has two different (positionally marked) variants: ⟨a⁰⟩ (labialized ⟨a⟩) in the first syllable and the usual ⟨a⟩ in all other syllables. This distribution of vowels enables Tatar speakers to correctly define the boundaries of word forms, along with stress and vowel harmony, which alone are not sufficient for a clear and unambiguous definition of the boundaries of word forms. The results obtained suggest that the statistically significant differences revealed in the distribution of vowel phonemes are of great cognitive importance for the perception and discrimination of signs because they mark the beginning and end of lexical units in the speech flow. In the future, the methodology developed in this work can be used for similar studies on other Turkic languages.

Keywords: quantitative linguistics, Tatar language, word form boundaries, vocal structure of word form, chi-square test

Figure Captions

Fig. 1. The number of vowel phonemes in the first and last syllables in A. Eniki's text.

Fig. 2. Vowel phonemes in the first and last syllables of the word form: mosaic plot.

References

1. Badecker W., Allen M. Morphological parsing and the perception of lexical identity: A masked priming study of stem homographs. *Journal of Memory and Language*, 2002, vol. 47, no. 1, pp. 125–144. doi: 10.1006/jmla.2001.2838.
2. Jakobson R., Halle M. *Fundamentals of Language*. The Hague, Mouton & Co – S.-Gravenhage, 1956. 89 p.
3. Jakobson R., Waugh L. *The Sound Shape of Language*. Brighton, Harvester Press, 1979. 308 p.

4. Chomsky N., Halle M. *The Sound Pattern of English*. Cambridge, Mass., Harper & Row, 1968. 470 p.
5. Trubetskoi N.S. *Osnovy fonologii* [Fundamentals of Phonology]. Katznelson S.D. (Ed.). Moscow, Aspekt Press, 2000. 352 p. (In Russian)
6. Baichura U.Sh. Vowels of the Tatar language in the light of experimental data. *Extended Abstract of Cand. Philol. Diss.* Moscow, Leningrad, 1953. 18 p. (In Russian)
7. Shcherbak A.M. *Sravnitel'naya fonetika tyurkskikh yazykov* [Comparative Phonetics of Turkic Languages]. Leningrad, Nauka, 1970. 204 p. (In Russian)
8. Altmann G. Phonic word structure. In: Köhler R., Altmann G., Piotrowski R. (Eds.) *Quantitative Linguistik. Ein internationales Handbuch*. Berlin, New York, M. de Gruyter, 2005, pp. 191–208.
9. Baskakov N.A. *Istoriko-tipologicheskaya fonologiya tyurkskikh yazykov* [Historical and Typological Phonology of Turkic Languages]. Tenishev E.R. (Ed.). Moscow, Nauka, 1988. 208 p. (In Russian)
10. *Tatarskaya grammatika* [Tatar Grammar]. Vol. 1. Zakiev M.Z. (Ed.). Kazan, Tatar. Kn. Izd., 1993. 584 p. (In Russian)
11. *Tatar Grammar*. Vol. 1. Khisamova F.M. (Ed.). Kazan, TehSI, 2015. 512 p. (In Tatar)
12. Baichura U.Sh. *Zvukovoi stroi tatarskogo yazyka: Eksperimental'no-foneticheskii ocherk* [Tatar Sound Structure: An Essay in Experimental Phonetics]. Vol. 1. Kazan, Izd. Kazan. Univ., 1959. 183 p. (In Russian)
13. Ibragimov T.I. A study of syllable formation and the structure of their combinations in the Tatar literary language. *Cand. Philol. Diss.* Kazan, 1970. 167 p. (In Russian)
14. Saykhunov M.R. Rhythmic and temporal characteristics of the Tatar language in terms of automatic speech synthesis. *Extended Abstract of Cand. Philol. Diss.* Kazan, 2010. 26 p. (In Russian)
15. Ibragimov T.I., Saykhunov M.R. To the problem of sounding reconstruction of written language. Reduction of short vowels in the Tatar language. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Gumanitarnye Nauki*, 2009, vol. 151, no. 3, pp. 288–298. (In Russian)
16. Ibragimov T.I., Saykhunov M.R. On the problem of accentuation in the Tatar language (compared to the Russian language). In: *Russkaya i sopostavitel'naya filologiya* [Russian and Comparative Philology]. Kazan, Kazan. Gos. Univ., 2010, pp. 56–59. (In Russian)
17. Galieva A.M. Syllable structure in Tatar: From data to modeling. *International Journal of Open Information Technologies*, 2020, vol. 8, no. 1, pp. 9–16. (In Russian)
18. Galieva A., Vavilova Z. Initial and final syllables in Tatar: From phonotactics to morphology. *Glottometrics*, 2021, vol. 50, pp. 57–75.
19. *Sravnitel'no-istoricheskaya grammatika tyurkskikh yazykov. Fonetika* [Comparative Historical Grammar of Turkic Languages. Phonetics]. Tenishev E.R. (Ed.). Moscow, Nauka, 1984. 488 p. (In Russian)
20. Kasevich V.B. *Morfonologiya* [Morphonology]. Leningrad, Izd. Leningr. Univ., 1986. 170 p. (In Russian)
21. Conover W.J. *Practical Nonparametric Statistics*. New York, Wiley, 1999. viii+584 p.
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, 2021. Available at: <https://www.R-project.org/>.

⟨ **Для цитирования:** Галиева А.М., Вавилова Ж.Е. Маркирование начала и конца слова в татарском языке: система гласных // Учен. зап. Казан. ун-та. Сер. Гуманит. науки. – 2022. – Т. 164, кн. 5. – С. 130–140. – doi: 10.26907/2541-7738.2022.5.130-140. ⟩

⟨ **For citation:** Galieva A.M., Vavilova Zh.E. Marking the beginning and end of the Tatar word: The system of vowels. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Gumanitarnye Nauki*, 2022, vol. 164, no. 5, pp. 130–140. doi: 10.26907/2541-7738.2022.5.130-140. (In Russian) ⟩