

КВАНТИТАТИВНАЯ ЛИНГВИСТИКА

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 81'322+811.512.145

doi: 10.26907/2541-7738.2021.1.180-189

ЗАКОН МЕНЗЕРАТА – АЛЬТМАННА: ЭКСПЕРИМЕНТЫ С ТЕКСТАМИ НА ТАТАРСКОМ ЯЗЫКЕ

А.М. Галиева

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

Закон Мензерата – Альтманна о связи между длиной языковой единицы и длиной его компонентов считается одним из важных законов квантитативной лингвистики. Эмпирическая проверка действия этого закона на основе данных татарского языка, имеющего агглютинативный морфологический строй, – задача, еще не реализованная в лингвистике. Поэтому основная цель статьи – анализ текстовых данных на татарском языке в аспекте раскрытия количественных связей между длиной словоформ, измеренной числом содержащихся в них слогов, и средней длиной этих слогов. Анализ, проведенный на материале художественных текстов, свидетельствует о том, что в целом наблюдаемые значения средней длины слогов падают с увеличением длины словоформ; при этом в ряде случаев у относительно низкочастотных длинных слов, содержащих четыре и более слогов, средняя длина может немного расти. Делается вывод о том, что формула Г. Альтманна позволяет достаточно хорошо моделировать среднюю длину слогов в зависимости от длины словоформ: для проанализированных текстов коэффициент детерминации модели R^2 составляет от 0.676 до 0.999.

Ключевые слова: длина слова, длина слога, квантитативная лингвистика, закон Мензерата – Альтманна, моделирование длины языковых единиц, татарский язык

Введение

Вычислительные модели имеют большое значение для развития лингвистической теории: они дают сведения об ограничениях или внутренней несогласованности теоретических установок и предъявляют к ним довольно строгие требования. Длина лингвистических единиц – одна из важных формальных характеристик языков, которая находит применение при обработке текста, в алгоритмах проверки орфографии, при обучении языку и т. п. Длина слов изучается специалистами в области лингвистики и анализа текста, а также математиками и статистиками, работающими над смежными вопросами. Основные подходы к изучению длины слова представлены в работах [1–3].

Мерой при определении длины слова может служить количество фонем (графем), морфем, мор или слогов в нем, в зависимости от целей исследования. Языки могут существенно отличаться распределением словоформ по длине,

что требует разных подходов к моделированию. Относительная легкость получения данных о длине языковых единиц предоставляет достаточно широкие возможности для экспериментирования при выборе модели и ее параметров: используются различные модификации закона Ципфа, закон Мензерата – Альтманна, данные аппроксимируются при помощи различных распределений и т. п. Исследователи отмечают, что за кажущейся простотой концептуализации длины слова может скрываться ряд сюрпризов, «так что здесь ничего не помогает, кроме непрерывного тестирования, моделирования, варьирования способов представления данных, модификации гипотез, сбора данных по новым языкам и т. д. Каждый “новый” язык может опровергнуть привычную теорию или вынудить нас изменить ее¹» [2, р. 224].

Целью данного исследования является эмпирическая проверка закона Мензерата – Альтманна на материале татарских художественных текстов. Тексты как производные от языка состоят из большого количества элементов разной природы, связи между которыми имеют сложный характер и в которых немалую роль играют случайные факторы, и последние нельзя не учитывать. Тем не менее языковые феномены, отраженные в текстах, несмотря на отклонения и разнообразие в своем поведении, характеризуются вполне определенной регулярностью, устойчивой относительной частотой. Чтобы можно было увидеть общие тенденции и варьирование в поведении языковых единиц, для исследования были отобраны образцы прозы и стихотворной речи, написанные разными авторами и отличающиеся размерами текста.

Имеется относительно небольшое число специальных исследований, посвященных длине слов в тюркских языках [4; 5], некоторые данные представлены в обзорных статьях и работах, содержащих данные по многим языкам [3; 6]. Специальные работы по эмпирической проверке закона Мензерата – Альтманна на данных тюркских языков нам неизвестны.

Закон Мензерата – Альтманна

В квантитативной лингвистике важное значение имеет закон Мензерата (Menzerath's law), или закон Мензерата – Альтманна (Menzerath–Altmann law), согласно которому с увеличением размера языковой единицы уменьшаются размеры его компонентов и наоборот. Первый вариант закона был сформулирован П. Мензератом на основе данных слогов слов немецкого языка: чем длиннее слова, чем короче оказались составляющие их слоги. Г. Альтманн придал этому закону строгую математическую форму и показал, что он работает на материале не только слов и слогов, но и других структур языка (например, предложений и клауз, клауз и составляющих их слов). При этом обязательной монотонности не требуется, средний размер составляющих является функцией от размера конструкта. Формула, предложенная Г. Альтманном, выглядит следующим образом:

$$y(x) = a \cdot x^b \cdot e^{cx},$$

где y – средняя длина конститuentов, x – длина конструкта, a , b , c – параметры модели [7].

¹ Перевод наш. – А.Г.

Справедливость закона Мензерата – Альтманна была подтверждена в ряде работ по языкам с разной морфологической структурой [8–13]. Вопрос о том, имеется ли связь между информацией, непосредственно передаваемой в коммуникативном акте планом содержания языковых единиц, и структурной информацией, является объектом дискуссий [14].

Результаты, представленные в эмпирических исследованиях о реализации закона Мензерата – Альтманна в языках мира, в свою очередь, становятся объектом дальнейших исследований; в частности, ставятся вопросы о том, имеют ли параметры, формализованные Г. Альтманном, характерные значения в зависимости от уровня лингвистического анализа, коррелируют ли они с типом лингвистических данных [13]. Имеются данные о том, что эти параметры в ряде случаев (в частности, на материале о строении клауз) могут существенно различаться для письменной и устной формы языка и типа текстов [10; 12]. В литературе представлены также некоторые модификации закона Мензерата – Альтманна, например, предложен вариант формулы для дискретных данных [15].

Действие рассматриваемого закона обусловлено глубинной организацией языковых структур и имеет большое значение для современной лингвистической теории, нацеленной на раскрытие связей между количественными параметрами и качественными характеристиками языка. При этом следует отметить, что закон Мензерата – Альтманна считается достаточно универсальным статистическим законом и тестируется на материале многоуровневых систем за пределами квантитативной лингвистики. Так, имеются сведения о том, что голосовое общение самцов обезьян гелад подчиняется этому закону [16]. Получены данные о том, что структура человеческого генома также определяется законом Мензерата – Альтманна [17].

Длина словоформ и длина слогов

В качестве языкового материала были использованы тексты трех сказок Г. Тукая в стихотворной форме, а также фрагменты из прозаических произведений А. Еники, Г. Ибрагимова, Г. Гильмана (см. Источники). В процессе предварительной обработки тексты были приведены к фонологической форме: 1 символ – 1 фонема, а слова (словоформы) были поделены на слоги. Все этапы исследования, включая обработку текстов, визуализацию данных и вычисления, выполнены при помощи языка R [18].

Графики (рис. 1–4) представляют некоторые важные для настоящего исследования количественные аспекты татарского текста на примере фрагмента повести А. Еники «Невысказанное завещание» (анализируемый фрагмент текста после токенизации содержит 2183 словоформы).

Рис. 1 представляет распределение слов разной длины, когда длина словоформ измерена в фонемах. Анализируемый фрагмент текста А. Еники состоит из словоформ с длиной от 1 до 14 фонем, самыми частотными являются единицы, состоящие из 5 фонем.

Словоформы в тексте А. Еники были поделены на слоги. Суммарное количество слогов в тексте – 4962, количество слогов в словоформах лежит в интервале от 1 до 6, самыми частотными являются слова, образованные двумя слогами (рис. 2).

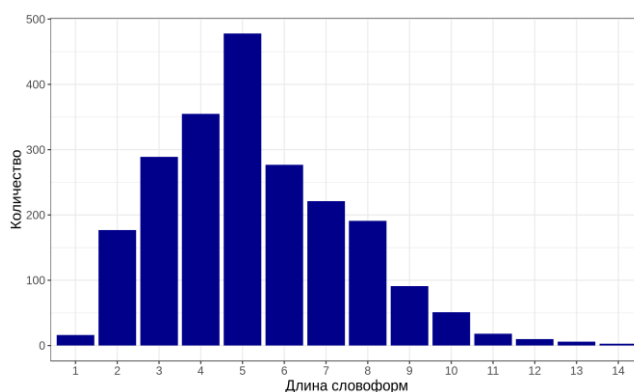


Рис. 1. Распределение слов в татарском тексте по длине (длина словоформ измерена в фонемах)

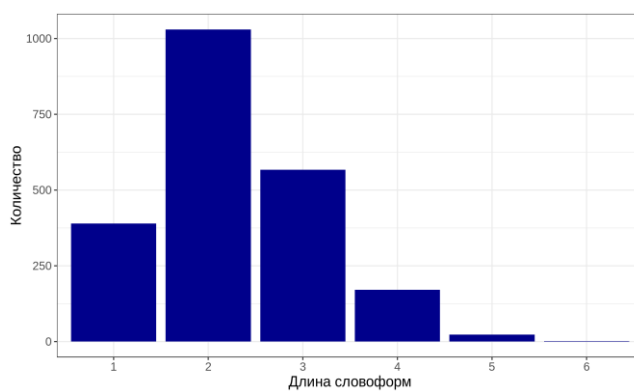


Рис. 2. Распределение слов в татарском тексте по длине (длина словоформ измерена в слогах)

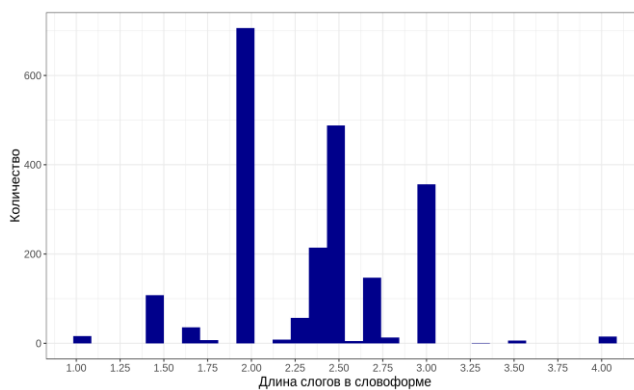


Рис. 3. Распределение слов в тексте в зависимости от средней длины слога

Следует отметить, что анализируемые тексты неоднородны из-за внутренних правил организации, поэтому распределения слов разной длины в отдельных текстах не всегда одинаковы. Во всех изученных текстах наиболее частотными являются словоформы, состоящие из пяти фонем и двух слогов. На рис. 3 представлена гистограмма распределения словоформ с разной средней длиной слогов в тексте А. Еники.

Табл. 1.

Наблюдаемые и ожидаемые значения средней длины слогов в зависимости от длины слов

Кол-во слогов	Г. Тукай «Шүрәле»			Г. Тукай «Кәжә белән сарык хикәясе»			Г. Тукай «Су анасы»		
	Кол-во слов	Наблюдаемые СДС	Ожидаемые СДС	Кол-во слов	Наблюдаемые СДС	Ожидаемые СДС	Кол-во слов	Наблюдаемые СДС	Ожидаемые СДС
1	249	2.67	2.68	93	2.61	2.61	114	2.48	2.49
2	422	2.44	2.44	395	2.38	2.38	198	2.44	2.41
3	196	2.41	2.40	36	2.35	2.36	84	2.33	2.38
4	54	2.44	2.45	55	2.42	2.42	23	2.39	2.37
5	4	2.55	2.54	–	–	–	–	–	–
6	–	–	–	–	–	–	–	–	–
		$a = 2.4223$ $b = -0.2788$ $c = 0.0996$ $R^2 = 0.996$			$a = 2.3382$ $b = -0.2973$ $c = 0.1113$ $R^2 = 0.999$			$a = 2.43968$ $b = -0.07870$ $c = 0.02031$ $R^2 = 0.714$	
Кол-во слогов	Г. Гильман «Очрашу»			А. Еники «Әйтелмәгән васыять»			Г. Ибрагимов «Кызыл чәчәкләр»		
	Кол-во слов	Наблюдаемые СДС	Ожидаемые СДС	Кол-во слов	Наблюдаемые СДС	Ожидаемые СДС	Кол-во слов	Наблюдаемые СДС	Ожидаемые СДС
1	190	2.53	2.52	390	2.50	2.48	64	2.80	2.76
2	445	2.36	2.38	1030	2.30	2.36	189	2.44	2.54
3	252	2.30	2.32	567	2.34	2.31	136	2.40	2.39
4	98	2.26	2.29	171	2.28	2.29	52	2.38	2.27
5	18	2.32	2.28	23	2.35	2.29	4	2.20	2.17
6	3	2.39	2.29	2	2.25	2.29	1	2.00	2.07
7	2	2.21	2.30	–	–	–	–	–	–
		$a = 2.46219$ $b = -0.11559$ $c = 0.02225$ $R^2 = 0.676$			$a = 2.42797$ $b = -0.10871$ $c = 0.02301$ $R^2 = 0.749$			$a = 2.84879$ $b = 0.0717$ $c = -0.03162$ $R^2 = 0.916$	

В табл. 1 приведены сведения о количестве словоформ в зависимости от числа слогов в словоформе, наблюдаемые (эмпирические) и предсказанные моделью средние значения длины слога (СДС), а также параметры модели для каждого из анализируемых текстов. Для получения наблюдаемых значений средней длины слога словоформы каждого текста были сгруппированы по длине, выраженной количеством слогов, далее для каждого класса была вычислена средняя длина слогов. Ожидаемые СДС были получены на основе формулы, предложенной Г. Альтманном; для подбора оптимальных значений параметров a , b , c была использована функция nls, которая входит в базовый R [18]. Данная функция позволяет получить оценки параметров нелинейной модели методом наименьших квадратов.

Как видно из табл. 1, для высокочастотных структур наблюдаемые значения средней длины слогов падают с увеличением длины словоформ, измеренной количеством слогов; при этом в ряде текстов у относительно низкочастотных

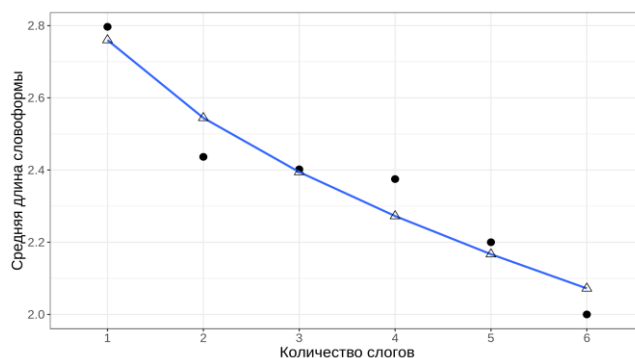


Рис. 4. Наблюдаемые и ожидаемые значения для средней длины слога в тексте Г. Ибрагимова

длинных слов, содержащих четыре и более слогов, средняя длина может немного расти. Исключением является текст А. Еники, где средняя длина трехсложных слов больше, чем у двусложных.

Для оценки качества подгонки модели использован коэффициент детерминации R^2 , который вычисляется по формуле:

$$R^2 = 1 - \frac{\Sigma_{\text{res}}}{\Sigma_{\text{tot}}},$$

где Σ_{res} – сумма квадратов остатков модели, Σ_{tot} – общая сумма квадратов.

Коэффициент детерминации модели R^2 для разных текстов составляет от 0.676 до 0.999, что свидетельствует о том, что формула Г. Альтманна достаточно хорошо описывает данные татарского языка (для анализируемых текстов). При этом модель позволяет отразить не только уменьшение средней длины слогов с увеличением длины слов (монотонность функции), но и ее увеличение (изменение монотонности функции) для ряда текстов.

На рис. 4 представлены наблюдаемые (черные точки) и предсказанные моделью (незакрашенные треугольники) значения средней длины словоформ в зависимости от длины, измеренной количеством слогов, на материале текста Г. Ибрагимова.

Лингвистическая интерпретация статистических моделей часто оставляет место для дискуссий. Хотя действие закона Мензерата – Альтманна можно считать общепризнанным, вопрос об интерпретации его параметров остается открытым [11]. Поэтому в качестве одной из перспектив исследования следует указать раскрытие возможных связей между морфологической структурой татарской словоформы и средней длиной составляющих ее слогов, например, с учетом особенностей строения аффиксальных цепочек.

Заключение

Выявление статистически значимых связей и закономерностей между разными уровнями языка – задача, которая успешно решается средствами современной компьютерной лингвистики. Обнаружение таких зависимостей может дать новые сведения о внутреннем устройстве языка и законах, им управляющих.

Согласно закону Мензерата – Альтманна существует связь между величиной лингвистического конструкта как целого и средней величиной составляющих его частей.

Эмпирическая проверка закона Мензерата – Альтманна показала, что на основе данных текстов на татарском языке закон в целом (как тенденция) соблюдается, но имеются определенные флуктуации, когда средняя длина слогов у достаточно длинных слов оказывается больше, чем у относительно коротких. Моделирование, основанное на формуле, предложенной Г. Альтманном, для татарского языка дало достаточно хорошие результаты: коэффициент детерминации R^2 для разных текстов составляет от 0.676 до 0.999.

Материалом для исследования послужили тексты художественных произведений или фрагменты из них на татарском языке (были взяты образцы как прозы, так и поэзии).

В качестве одной из перспектив исследования можно указать раскрытие возможных связей между морфологической структурой татарской словоформы и средней длиной составляющих ее слогов.

Благодарности. Работа выполнена за счет средств субсидии, выделенной в рамках государственной поддержки Казанского (Приволжского) федерального университета в целях повышения его конкурентоспособности среди ведущих мировых научно-образовательных центров.

Источники

- Гыйльман Г.* Очрашу. – URL: <http://baigysh.tatar/ocrasu-gilman/>, свободный.
- Еники А.* Әйтелмәгән васыять. – URL: <http://belgech.ru/proizvedeniya/itelmagan-vasyyat>, свободный.
- Ибраһимов Г.* Кызыл чәчәкләр. – URL: <http://belgech.ru/proizvedeniya/kyzyl-chachaklar>, свободный.
- Тукай Г.* Шүрәле. – URL: <http://gabdullatukay.ru/works/poem/1907/shurele/>, свободный.
- Тукай Г.* Су анасы. – URL: <https://tatobzor.ru/tatarskie-skazki/su-anasy-kiyat-gabdullatukaj.html>, свободный.
- Тукай Г.* Кәжә белән сарык хикәясе. – URL: <http://gabdullatukay.ru/works/poem/1909/keje-belen-saryk-hikeyase/>, свободный.

Литература

1. *Altmann G.* Aspects of word length // Köhler R., Altmann G. (Eds.). Issues in Quantitative Linguistics. V. 3. – Lüdenscheid: RAM-Verlag, 2013. – P. 23–38.
2. *Popescu I., Naumann S., Kelih E., Rovenchak A., Overbeck A., Sanada H., Smith R., Čech R., Mohanty P., Wilson A., Altmann G.* Word length: Aspects and languages // Altmann G., Köhler R. (Eds.). Issues in Quantitative Linguistics. V. 3. – Lüdenscheid: RAM-Verlag, 2013. – P. 224–281.
3. *Grzybek P.* History and methodology of word length studies // Contributions to the Science of Text and Language. Word Length Studies and Related Issues. – Heidelberg: Springer, 2005. – P. 15–90.

4. *Ризванова Л.М.* Квантитативная характеристика татарского слова (на материале отдельных функциональных стилей татарского языка): Автореф. дис. ... канд. филол. наук. – Казань, 1996. – 24 с.
5. *Galieva A.M.* Word length in Tatar: Selecting relevant parameters for modeling // Elizarov A., Loukachevitch N. (Eds.) Proc. Comput. Models Lang. Speech Workshop (CMLS 2020) Co-Located with 16th Int. Conf. on Comput. Cognit. Linguist. (TEL 2020). Kazan, Russia, Nov. 12–13, 2020. CEUR Workshop Proc. – 2020. – V. 2780. – P. 179–186.
6. *Kromer V.* About word length distribution // Contributions to the Science of Text and Language. Word Length Studies and Related Issues. – Heidelberg: Springer, 2005. – P. 199–210.
7. *Altmann G.* Prolegomena to Menzerath's law // Glottometrika. – 1980. – No 2. – P. 1–10.
8. *Fenk A., Fenk-Oczlon G., Fenk L.* Syllable complexity as a function of word complexity // Proc. VIII Int. Conf. "Cognitive Modeling in Linguistics". – 2006. – V. 1. – P. 324–333.
9. *Fenk A., Fenk-Oczlon G.* Menzerath's law and the constant flow of linguistic information // Contributions to Quantitative Linguistics: Proc. First Int. Conf. on Quantitative Linguistics, QUALICO, Trier, 1991. – Springer, 1993. – P. 11–31.
10. *Xu L., He L.* Is the Menzerath–Altmann law specific to certain languages in certain registers? // J. Quant. Linguist. – 2020. – V. 27, No 3. – P. 187–203. – doi: 10.1080/09296174.2018.1532158.
11. *Mačutek J., Chromý J., Koščová M.* Menzerath–Altmann law and prothetic /v/ in spoken Czech // J. Quant. Linguist. – 2019. – V. 26, No 1. – P. 66–80. – doi: 10.1080/09296174.2018.1424493.
12. *Hou R., Huang C.-R., Ahrens K., Lee Y.-M.* Linguistic characteristics of Chinese register based on the Menzerath – Altmann law and text clustering // Digital Scholarship in the Humanities. – 2020. – V. 35, No 1. – P. 54–66. – doi: 10.1093/llc/fqz005.
13. *Cramer I.* The parameters of the Altmann–Menzerath law // J. Quant. Linguist. – 2005. – V. 12, No 1. – P. 41–52. – doi: 10.1080/09296170500055301.
14. *Milička J.* Menzerath's Law: The whole is greater than the sum of its parts // J. Quant. Linguist. – 2014. – V. 21, No 2. – P. 85–99. – doi: 10.1080/09296174.2014.882187.
15. *Kulacka A., Mačutek J.* A discrete formula for the Menzerath–Altmann law // J. Quant. Linguist. – 2007. – V. 14, No 1. – P. 23–32. – doi: 10.1080/09296170600850585.
16. *Gustison M.L., Semple S., Ferrer-i-Cancho R., Bergman T.J.* Gelada vocal sequences follow Menzerath's linguistic law // Proc. Natl. Acad. Sci. U. S. A. – 2016. – V. 113, No 19. – P. E2750–E2758. – doi: 10.1073/pnas.1522072113.
17. *Li W.* Menzerath's law at the gene-exon level in the human genome // Complexity. – 2012. – V. 17, No 4. – P. 49–53. doi: 10.1002/cplx.20398.
18. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. – Vienna, 2018. – URL: <https://www.R-project.org/>, свободный.

Поступила в редакцию
28.12.2020

Галиева Альфия Макаримовна, кандидат философских наук, доцент, старший научный сотрудник НИЛ «Квантитативная лингвистика»

Казанский (Приволжский) федеральный университет
ул. Кремлёвская, д. 18, г. Казань, 420008, Россия
E-mail: amgalieva@gmail.com

ORIGINAL ARTICLE

doi: 10.26907/2541-7738.2021.1.180-189

The Menzerath–Altmann Law: Experimenting with Tatar Texts

A.M. Galieva

Kazan Federal University, Kazan, 420008 Russia

E-mail: amgalieva@gmail.com

Received December 28, 2020

Abstract

The Menzerath–Altmann law on the relationship between the length of linguistic units and the length of their components is one of the important laws of quantitative linguistics. This law is a result of an advanced linguistic structures organization and is of great importance for the modern theory of language aimed at revealing the relations between qualitative features and quantitative parameters of the language.

The validity of the Menzerath–Altmann law has been confirmed in a number of works on languages with different morphological structures. The main purpose of this paper is empirical testing of the Menzerath–Altmann law on the Tatar language with the help of various fiction texts (both poetry and prose).

The distribution of word forms in the Tatar language by length, observed values of the average syllable length depending on the word length, average values of the syllable length predicted by the model, as well as the model parameters were investigated for the analyzed texts. To assess the goodness of fitting of the model, the coefficient of determination R^2 , which for different texts ranged from 0.676 to 0.999, was used. It was concluded that G. Altman's formula is in good agreement with the data of the Tatar language. The model predicts not only the decreasing average syllable length with the increasing word length (function monotonicity), but also its subsequent increasing (change in the function monotonicity) for a number of texts.

Keywords: word length, syllable length, quantitative linguistics, Menzerath–Altmann law, modeling of length of linguistic units, Tatar language

Acknowledgments. The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

Figure Captions

Fig. 1. Distribution of words in the Tatar text by length (the length of word forms is measured in phonemes).

Fig. 2. Distribution of words in the Tatar text by length (the length of word forms is measured in syllables).

Fig. 3. Distribution of words in the text depending on the average syllable length.

Fig. 4. Observed and predicted values of the average syllable length in G. Ibragimova's text.

References

1. Altmann G. Aspects of word length. In: Köhler R., Altmann G. (Eds.). *Issues in Quantitative Linguistics*. Vol. 3. Lüdenscheid, RAM-Verlag, 2013, pp. 23–38.
2. Popescu I., Naumann S., Kelih E., Rovenchak A., Overbeck A., Sanada H., Smith R., Čech R., Mohanty P., Wilson A., Altmann G. Word length: Aspects and languages. In: Altmann G., Köhler R. (Eds.). *Issues in Quantitative Linguistics*. Vol. 3. Lüdenscheid, RAM-Verlag, 2013, pp. 224–281.
3. Grzybek P. History and methodology of word length studies. In: *Contributions to the Science of Text and Language. Word Length Studies and Related Issues*. Heidelberg, Springer, 2005, pp. 15–90.

4. Rizvanova L.M. Quantitative features of the Tatar word (based on the material of certain functional styles in the Tatar language). *Extended Abstract of Cand. Philol. Diss. Kazan*, 1996. 24 p. (In Russian)
5. Galieva A.M. Word length in Tatar: Selecting relevant parameters for modeling. *Computational Models in Language and Speech: Proc. Comput. Models Lang. Speech Workshop (CMLS 2020) Co-Located with 16th Int. Conf. on Comput. Cognit. Linguist. (TEL 2020). Kazan, Russia, Nov. 12–13, 2020. CEUR Workshop Proc.* Elizarov A., Loukachevitch N. (Eds.), 2020, vol. 2780, pp. 179–186.
6. Kromer V. About word length distribution. In: *Contributions to the Science of Text and Language. Word Length Studies and Related Issues*. Heidelberg, Springer, 2005, pp. 199–210.
7. Altmann G. Prolegomena to Menzerath's law. *Glottometrika*, 1980, no. 2, pp. 1–10.
8. Fenk A., Fenk-Oczlon G., Fenk L. Syllable complexity as a function of word complexity. *Proc. VIII Int. Conf. "Cognitive Modeling in Linguistics"*, 2006, vol. 1, pp. 324–333.
9. Fenk A., Fenk-Oczlon G. Menzerath's law and the constant flow of linguistic information. *Contributions to Quantitative Linguistics: Proc. 1st Int. Conf. on Quant. Linguist., QUALICO, Trier, 1991*. Springer, 1993, pp. 11–31.
10. Xu L., He L. Is the Menzerath–Altmann law specific to certain languages in certain registers? *Journal of Quantitative Linguistics*, 2020, vol. 27, no. 3, pp. 187–203. doi: 10.1080/09296174.2018.1532158.
11. Mačutek J., Chromý J., Koščová M. Menzerath–Altmann law and prothetic /v/ in spoken Czech. *Journal of Quantitative Linguistics*, 2019, vol. 26, no. 1, pp. 66–80. doi: 10.1080/09296174.2018.1424493.
12. Hou R., Huang C.-R., Ahrens K., Lee Y.-M. Linguistic characteristics of Chinese register based on the Menzerath–Altmann law and text clustering. *Digital Scholarship in the Humanities*, 2020, vol. 35, no. 1, pp. 54–66. doi: 10.1093/lc/fqz005.
13. Cramer I. The parameters of the Altmann–Menzerath law. *Journal of Quantitative Linguistics*, 2005, vol. 12, no. 1, pp. 41–52. doi: 10.1080/09296170500055301.
14. Milička J. Menzerath's law: The whole is greater than the sum of its parts. *Journal of Quantitative Linguistics*, 2014, vol. 21, no. 2, pp. 85–99. doi: 10.1080/09296174.2014.882187.
15. Kuřacka A., Mačutek J. A discrete formula for the Menzerath–Altmann law. *Journal of Quantitative Linguistics*, 2007, vol. 14, no. 1, pp. 23–32. doi: 10.1080/09296170600850585.
16. Gustison M.L., Semple S., Ferrer-i-Cancho R., Bergman T.J. Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, vol. 113, no. 19, pp. E2750–E2758. doi: 10.1073/pnas.1522072113.
17. Li W. Menzerath's law at the gene-exon level in the human genome. *Complexity*, 2012, vol. 17, no. 4, pp. 49–53. doi: 10.1002/cplx.20398.
18. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, 2018. Available at: <https://www.R-project.org/>.

⟨ **Для цитирования:** Галиева А.М. Закон Мензерата – Альтманна: эксперименты с текстами на татарском языке // Учен. зап. Казан. ун-та. Сер. Гуманит. науки. – 2021. – Т. 163, кн. 1. – С. 180–189. – doi: 10.26907/2541-7738.2021.1.180-189. ⟩

⟨ **For citation:** Galieva A.M. The Menzerath–Altmann law: Experimenting with Tatar texts. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Gumanitarnye Nauki*, 2021, vol. 163, no. 1, pp. 180–189. doi: 10.26907/2541-7738.2021.1.180-189. (In Russian) ⟩