

A.R.Yeshkeyev, O.I.Ulbrikht, R.M.Ospanov

Properties of models of a center and central type of some positive Jonsson theories in an allowed enrichment of signatures

The main purpose of this paper is to describe the structure of the core models of strongly convex Δ - R -perfect theories in a certain enrichment of the signature and consideration by question A.D. Taimanov for positive Jonsson theories. We study the algebraically simple and atomic models of positive Robinson's theories. For them, a criterion is given uncountable categorical was found. Under the terms set out have shown that a positive solution to the issue A.D. Taimanov is the same as for the central type of such a theory, and for the center of this theory.

References

- 1 Itay Ben-Yaacov. *Journal of Mathematical Logic*, 2003, 3, 1, p. 85–118.
- 2 Yeshkeyev A.R. *Jonsson theory*, Karaganda: Publ. KarGU, 2009, p. 250.
- 3 Sikorski R. *Boolean algebra*, Moscow: Nauka, 1969, p. 376.
- 4 Mustafin T.G. *Mathematics and physics research*, Karaganda: Publ. KarGU, 1974, 1, p. 80–84.
- 5 *Handbook of mathematical logic: In 4 parts*, Ed. by Dzh. Barvaysa, Ch. 1, Model Theory / English Transl., Moscow: Nauka, 1982, p. 126.
- 6 Robinson A. *Introduction to Model Theory and to the Metamathematics of Algebra*, Amsterdam, 1963.
- 7 Kueker D.W. *Core structures for theories*, *Fundamenta Mathematicae* LXXXIX, 1975, p. 154–171.
- 8 Mustafin T.G. *Proceedings of the Soviet-French symposium on model theory*, Karaganda, 1990, p. 112–125.
- 9 Palyutin E.A. *Algebra and logic*, 2003, 42, 2, p. 194–210.
- 10 Baldwin J.T., Kueker D.W. *Ann. Math. Logic*, 1981, p. 289–330.

УДК 004.43

С.Ш.Кажикенова, К.В.Мазиева, Е.Г.Шурыгина

Карагандинский государственный университет им. Е.А.Букетова
(E-mail: sauleshka555@mail.ru)

Идеальная лингвоматематическая модель для анализа структуры текста

В статье рассмотрены вопросы изучения текстового материала различных жанров с целью его совершенствования. Показано, что, используя текст в качестве универсальной модели, можно установить те пределы изменчивости, в которых могут осуществляться самоорганизация и развитие лингвистических систем. Путем использования математических расчетов получены значения энтропии буквы с учетом одной, двух, трех, четырех, пяти, шести букв текста на русском и казахском языках.

Ключевые слова: моделирование, самоорганизующаяся система, информация, энтропия, лингвосинергетика, формула Шеннона, иерархическая структура, статистический метод, жанр текста, стиль текста, языковой уровень, модель.

Наши исследования обусловлены необходимостью изучения текстового материала различных жанров с целью его совершенствования. Любой текст должен быть стилистически, грамматически, синтаксически оформлен грамотно, без лингвистических погрешностей. Путем использования математических расчетов нами получены значения энтропии буквы с учетом одной, двух, трех, четырех, пяти, шести букв текста на русском и казахском языках.

При определении количества информации рассматривается языковой текст, который состоит из букв, слов, словосочетаний, предложений и т.д. Появление каждой буквы описывается как последовательная реализация определенной системы. Количество информации, передаваемое указанной буквой, равно по абсолютной величине той энтропии (неопределенности), которая характеризовала систему возможных выборов и которая была снята в результате выбора определенной буквы.

Как известно, для расчета энтропии необходимо иметь полное распределение вероятностей возможных комбинаций. Поэтому для вычисления энтропии той или иной буквы необходимо знать вероятности появления каждой возможной буквы.

Мы предлагаем идеальную лингвоматематическую модель для анализа структуры текста. Она построена на основе фундаментального закона сохранения суммы информации и энтропии с применением формулы Шеннона [1–4].

При общей характеристике энтропийно-информационного (*энтропия* — мера беспорядка, а *информация* — мера снятия беспорядка) анализа текстов мы использовали статистическую формулу Шеннона для определения совершенства, гармонии текста:

$$H = - \sum_{i=1}^N p_i \log_2 p_i, \quad (1)$$

где p_i — вероятность обнаружения какого-либо однородного элемента системы в их множестве N ;

$$\sum_{i=1}^N p_i = 1, p_i \geq 0, i = 1, 2, \dots, N.$$

До опубликования созданной Шенноном теории Хартли предложил определять количество максимальной энтропии по формуле

$$H_{\max} = \log_2 N.$$

Большой интерес представляют исследования в области теории информации. Важным для языкознания измерением является энтропия языка. Она является общей мерой вероятностно-лингвистических связей в тексте данного языка. В связи с этим мы проводим сопоставление данных, характеризующих численную оценку этих измерений на казахском и русском языках.

Так как русский алфавит содержит 32 буквы (31 буква, 1 пробел), то согласно этому результату

$$H_0 = \log 32 = 5 \text{ бит.}$$

H_0 — максимальное значение энтропии текста, заключающееся в приеме одной буквы русского текста (информация, содержащаяся в одной букве), при условии, что все буквы считаются одинаково вероятными; *бит* — единица измерения информации.

Казахский алфавит содержит 43 буквы (42 буквы, 1 пробел), согласно этому результату

$$M \frac{\log 43}{\log m} = M \frac{H_0}{\log m}.$$

Здесь $H_0 = \log 43 = 5,4$ бит — энтропия опыта, заключающегося в приеме одной буквы казахского текста (информация, содержащаяся в одной букве), при условии, что все буквы считаются одинаково вероятными.

Здесь нужно отметить, что современный казахский кириллический алфавит используется в Казахстане и Монголии. В принятом в 1940 году алфавите, разработанном С.А.Аманжоловым, 42 буквы, из них 33 буквы русского алфавита и 9 специфических букв казахского языка: *Ә, Ғ, Қ, Ң, Ө, Ұ, Ү, Һ, І*. Первоначально казахские буквы размещались после букв русского алфавита, затем каждую из них поставили после русских букв, сходных по звучанию. Следующие буквы: *в, ё* (с 1957 г.), *ф, х, һ, ц, ч, щ, ь, ь, э* — в исконно казахских словах не употребляются. Буквы *ё, ц, ч, щ, ь, ь, э* используются только в словах, которые заимствованы из русского или через русский язык, и пишутся в соответствии с правилами русской орфографии. Буква *х* в разговорном языке произносится как *қ*. Буква *һ* используется только в арабо-персидских заимствованиях и произносится зачастую как глухая *х*. Буква *е* в абсолютном начале слова произносится как дифтонгоид [*é*]. Буква *э* всегда произносится как *е*. Буква *о* в абсолютном начале слова может произноситься как дифтонгоид [*°о*]. Буквы *і* и *ы* обозначают звуки, подобные старославянским (до падения редуцированных), *ь* и *ь*. Буква *и* обозначает псевдодифтонги *ый, ій*. Буква *у* обозначает неслоговый звук, подобный белорусскому *ў*, и псевдодифтонги *ұу, уу, ыу, іу*.

Следующие буквы (называемые соответственно «мягкие», или «узкие», и «твердые», или «широкие») обозначают пары гласных переднего и заднего ряда: *е – а, ө – о, ү – ұ, і – ы*. В арабо-

персидских заимствованиях также есть противопоставление $\text{ə} - \text{a}$. Поскольку ударение всегда приходится на последний слог, оно на письме не отображается.

В качестве примера был рассмотрен текст, связанный с казахской музыкой. Текст содержит знаков с пробелами — 500, без пробелов — 431.

Ориентировочные значения частот отдельных букв казахского языка представлены в таблицах 1 и 2 (тире здесь обозначает пробел между словами). В таблице 1 буквы расположены в алфавитном порядке, в таблице 2 — по мере убывания относительных частот.

Таблица 1

№ п/п	Буква	Относительная частота	№ п/п	Буква	Относительная частота
1	Пробел	0,138	23	П	0,008
2	А	0,112	24	Р	0,052
3	Ә	0,01	25	С	0,026
4	Б	0,018	26	Т	0,042
5	В	0	27	У	0,022
6	Г	0,004	28	Ұ	0,002
7	Ғ	0,008	29	Ү	0,008
8	Д	0,034	30	Ф	0
9	Е	0,042	31	Х	0,01
10	Ё	0	32	Һ	0
11	Ж	0,014	33	Ц	0
12	З	0,028	34	Ч	0
13	И	0,004	35	Ш	0,006
14	Й	0,018	36	Щ	0
15	К	0,036	37	Ъ	0
16	Қ	0,018	38	Ы	0,124
17	Л	0,036	39	І	0,032
18	М	0,05	40	Ь	0
19	Н	0,044	41	Э	0
20	Ң	0,026	42	Ю	0
21	О	0,014	43	Я	0,004
22	Ө	0,01			

Таблица 2

Буква отн. частота	—	ы	а	р	м	н	е	т
	0,138	0,124	0,112	0,052	0,05	0,044	0,042	0,042
Буква отн. частота	к	л	д	і	з	ң	с	у
	0,036	0,036	0,034	0,032	0,028	0,026	0,026	0,022
Буква отн. частота	б	й	қ	ж	о	ә	ө	х
	0,018	0,018	0,018	0,014	0,014	0,01	0,01	0,01
Буква отн. частота	ғ	п	ү	ш	г	и	я	ұ
	0,008	0,008	0,008	0,006	0,004	0,004	0,004	0,002

Приравняв эти частоты вероятностям появления соответствующих букв, получим на основании информационной энтропии Шеннона формулу для расчета максимального значения энтропии текста при учете одной буквы казахского текста:

$$H_1 = H(\alpha_1) = b \cdot \log_a b = b \cdot \left(\frac{\ln b}{\ln a} \right),$$

$$H_1 = H(\alpha_1) = -0,138 \cdot \log_2(0,138) - 0,124 \cdot \log_2(0,124) - \dots - 0,002 \cdot \log_2(0,002) \approx 4,3598.$$

Ориентировочные значения частот двухбуквенных сочетаний казахского языка представлены в таблице 3 (тире здесь обозначает пробел между словами). В таблице 3 буквы расположены по мере убывания относительных частот.

Таблица 3

Сочетание отн. частота	ы - 0,032	- м 0,022	ры 0,022	ың 0,020	ң - 0,020	му 0,020	уз 0,020	зы 0,020
Сочетание отн. частота	ык 0,020	ка 0,020	ты 0,018	- т 0,018	та 0,018	н - 0,018	і - 0,016	а - 0,016
Сочетание отн. частота	ыр 0,016	лы 0,016	- б 0,014	ар 0,014	- ж 0,014	мы 0,014	ал 0,012	ық 0,012
Сочетание отн. частота	ас 0,012	сы 0,012	ба 0,012	- к 0,012	ам 0,012	ен 0,012	ер 0,012	- х 0,001
Сочетание отн. частота	ха 0,01	да 0,01	рі 0,01	- о 0,01	ын 0,01	нд 0,01	ан 0,01	де 0,001
Сочетание отн. частота	р - 0,008	қт 0,008	- ә 0,008	ән 0,008	ді 0,008	- д 0,008	п - 0,008	ай 0,008
Сочетание отн. частота	ны 0,008	ла 0,008	ме 0,008	жы 0,008	ні 0,006	із 0,006	жа 0,006	кө 0,006
Сочетание отн. частота	- а 0,006	ды 0,006	кү 0,006	үй 0,006	йл 0,006	ле 0,006	ол 0,006	ыл 0,006
Сочетание отн. частота	- с 0,006	рм 0,006	қ - 0,006	ор 0,004	йт 0,004	ег 0,004	ге 0,004	ім 0,004
Сочетание отн. частота	мі 0,004	ат 0,004	з - 0,004	зд 0,004	ағ 0,004	ға 0,004	л - 0,004	- ө 0,004
Сочетание отн. частота	се 0,004	ед 0,004	аң 0,004	ңа 0,004	ып 0,004	ей 0,004	рл 0,004	аш 0,004
Сочетание отн. частота	- е 0,004	йд 0,004	лм 0,004	ма 0,004	әр 0,002	бі 0,002	ің 0,002	ақ 0,002
Сочетание отн. частота	қс 0,002	өр 0,002	іп 0,002	нд 0,002	өп 0,002	ым 0,002	ыз 0,002	өт 0,002
Сочетание отн. частота	тк 0,002	ке 0,002	са 0,002	йы 0,002	өс 0,002	е- 0,002	тү 0,002	аб 0,002
Сочетание отн. частота	үс 0,002	өб 0,002	бе 0,002	йе 0,002	шт 0,002	си 0,002	ия 0,002	яқ 0,002
Сочетание отн. частота	еш 0,002	шқ 0,002	қа 0,002	ша 0,002	ес 0,002	ск 0,002	кі 0,002	ір 0,002
Сочетание отн. частота	со 0,002	то 0,002	ығ 0,002	ғы 0,002	от 0,002	ра 0,002	ад 0,002	- я 0,002
Сочетание отн. частота	яғ 0,002	ғн 0,002	ни 0,002	и - 0,002	он 0,002	ст 0,002	ау 0,002	у - 0,002
Сочетание отн. частота	бұ 0,002	ұл 0,002						

Далее подсчитаем условную энтропию $H_2 = H\alpha_1(\alpha_2)$ опыта α_2 , состоящего в определении одной буквы казахского текста, при условии, что нам известен исход опыта α_1 , состоящего в определении предшествующей буквы того же текста. Согласно сказанному выше H_2 определяется следующей формулой:

$$H_2 = H\alpha_1(\alpha_2) = H(\alpha_1\alpha_2) - H(\alpha_1) = -0,032 \cdot \log_2(0,032) - 0,022 \cdot \log_2(0,022) - \dots - (0,002) \cdot \log_2(0,002) + 0,138 \cdot \log_2(0,138) + 0,124 \cdot \log_2(0,124) + \dots + 0,002 \cdot \log_2(0,002) \approx 2,3444.$$

Аналогично этому можно определить и энтропию H_3 .

Приравняв эти частоты вероятностям появления соответствующих трехбуквенных сочетаний, что находит отражение в разности $H_2 - H_3$, получим для энтропии трех букв казахского текста приближенное значение:

$$H_3 = H\alpha_1\alpha_2(\alpha_3) = H(\alpha_1\alpha_2\alpha_3) - H(\alpha_1\alpha_2) = -0,020 \cdot \log_2(0,020) - 0,020 \cdot \log_2(0,020) - \dots - \dots - 0,002 \cdot \log_2(0,002) + 0,032 \cdot \log_2(0,032) + 0,022 \cdot \log_2(0,022) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,852;$$

для расчета максимального значения энтропии текста при учете четырех букв казахского текста:

$$H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) = H(\alpha_1\alpha_2\alpha_3\alpha_4) - H(\alpha_1\alpha_2\alpha_3) = -0,020 \cdot \log_2(0,020) - 0,020 \cdot \log_2(0,020) - \dots - \dots - 0,002 \cdot \log_2(0,002) + 0,020 \cdot \log_2(0,020) + 0,020 \cdot \log_2(0,020) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,2813;$$

при учете пяти букв казахского текста составляет приближенное значение:

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = -0,020 \cdot \log_2(0,020) - -0,020 \cdot \log_2(0,020) \dots - \dots - 0,002 \cdot \log_2(0,002) + 0,020 \cdot \log_2(0,020) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,1832.$$

Согласно сказанному выше для определения условной энтропии H_6 подсчитали число всех шести-буквенных сочетаний в данном тексте и применили формулу классического определения вероятности:

$$P = \frac{m}{n},$$

где n — число всех 6 буквенных сочетаний; m — число сочетаний, например, музыка.

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = -0,020 \cdot \log_2(0,020) - -0,020 \cdot \log_2(0,020) - 0,012 \cdot \log_2(0,012) \dots - \dots - 0,002 \cdot \log_2(0,002) + 0,020 \cdot \log_2(0,020) + \dots + + 0,002 \cdot \log_2(0,002) \approx 0,1657.$$

В результате были получены следующие значения (в битах):

H_1	H_2	H_3	H_4	H_5	H_6
4,3598	2,3444	0,852	0,2813	0,1882	0,1657.

Обратимся теперь к анализу русскоязычного текста. Нами был проведен информационно-энтропийный анализ отрывка из курса лекций по экономической теории, то есть был изучен текст научного стиля, в котором четко выражены признаки и особенности языка науки.

Для вычисления информации научного текста нами были подсчитаны вероятности появления одной буквы, двух-, трех-, четырех-, пяти-, а также шестибуквенных сочетаний в данном тексте. При подсчете учитывались 31 буква русского алфавита (буквы *е* и *ё*, *ь* и *ы* принимаются как одна буква) и пробел, все остальные знаки (скобки, кавычки, запятые и пр.) не рассматривались.

Кроме того, учитывался и такой аспект в характеристике текста, как мера вероятности прагматической информации, спроецированной на потенциального читателя. Если эту меру вероятности прагматической (новой, полезной, воспринимаемой данным читателем) информации обозначить термином «энтропия», то по отношению к функционально разным текстам эта мера определится следующим образом: для официально-деловых текстов адекватность энтропии принципиальна, для научных — ограничена специальным кругом читателей, потому вполне предсказуема; для публицистических и газетных текстов адекватность энтропии принципиальна, но непредсказуема; для художественных — непринципиальна и непредсказуема.

Чтобы вычислить относительную частоту каждой буквы, необходимо количество каждой буквы разделить на общее количество всех знаков (500) (см. табл. 4).

Т а б л и ц а 4

Буква	Число появления буквы: количество всех букв	Отн. частота	Буква	Число появления буквы: количество всех букв	Отн. частота
А	26:500	0,052	Р	27:500	0,054
Б	4:500	0,008	С	24:500	0,048
В	25:500	0,05	Т	29:500	0,058
Г	4:500	0,008	У	11:500	0,022
Д	10:500	0,02	Ф	3:500	0,006
Е	30:500	0,06	Х	2:500	0,004
Ж	5:500	0,01	Ц	1:500	0,002
З	10:500	0,02	Ч	2:500	0,004
И	45:500	0,09	Ш	3:500	0,006
Й	6:500	0,012	Щ	2:500	0,004
К	14:500	0,028	Ы	6:500	0,012
Л	18:500	0,036	Ъ,ъ	2:500	0,004
М	9:500	0,018	Э	5:500	0,01
Н	34:500	0,068	Ю	3:500	0,006
О	55:500	0,11	Я	13:500	0,026
П	14:500	0,028	пробел	58:500	0,116

Расположим относительную частоту знаков последовательно, в порядке убывания (см. табл. 5):

Т а б л и ц а 5

Буква	пробел	о	и	н	е
Частота	0,116	0,11	0,09	0,068	0,06
Буква	т	р	а	в	с
Частота	0,058	0,054	0,052	0,05	0,048
Буква	л	к	п	я	у
Частота	0,036	0,028	0,028	0,026	0,022
Буква	д	з	м	й	ы
Частота	0,02	0,02	0,018	0,012	0,012
Буква	ж	э	г	б	ю
Частота	0,01	0,01	0,008	0,008	0,006
Буква	ф	ш	ъ, ь	х	ч
Частота	0,006	0,006	0,004	0,004	0,004
Буква	щ	ц			
Частота	0,004	0,002			

Вычислив энтропию текста при учете 1, 2, 3, 4, 5, 6 букв, мы пришли к следующим показателям:

$$H_1 = 4,364 \text{ бит};$$

$$H_2 = H_{\alpha_1}(\alpha_2) = H(\alpha_1 \alpha_2) - H(\alpha_1) = 7,3406 - 4,364 = 2,9766;$$

$$H_3 = H_{\alpha_1 \alpha_2}(\alpha_3) = H(\alpha_1 \alpha_2 \alpha_3) - H(\alpha_1 \alpha_2) = 8,123 - 7,3406 = 0,7824;$$

$$H_4 = H_{\alpha_1 \alpha_2 \alpha_3}(\alpha_4) = H(\alpha_1 \alpha_2 \alpha_3 \alpha_4) - H(\alpha_1 \alpha_2 \alpha_3) = 8,4656 - 8,123 = 0,3426;$$

$$H_5 = H_{\alpha_1 \alpha_2 \alpha_3 \alpha_4}(\alpha_5) = H(\alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5) - H(\alpha_1 \alpha_2 \alpha_3 \alpha_4) = 8,5271 - 8,4656 = 0,0615;$$

$$H_6 = H_{\alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5}(\alpha_6) = H(\alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5 \alpha_6) - H(\alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5) = 8,5808 - 8,5271 = 0,0537.$$

Таким образом, дальнейшие подсчеты текстов от одного до шестибуквенных сочетаний для казахского и русского языков неодинаковы.

В казахском языке:

$$\begin{array}{cccccc} H_1 & H_2 & H_3 & H_4 & H_5 & H_6 \\ 4,3598 & 2,3444 & 0,852 & 0,2813 & 0,1882 & 0,1657. \end{array}$$

В русском языке:

$$\begin{array}{cccccc} H_1 & H_2 & H_3 & H_4 & H_5 & H_6 \\ 4,364 & 2,9766 & 0,7824 & 0,3426 & 0,0615 & 0,0537 \end{array}$$

Расчеты показывают, что величина H_{32} в русском языке, так как алфавит содержит 32 буквы (буквы *е* и *ё*, *ь* и *ъ*, которые передаются одной и той же комбинацией, и пробел (-) пустой промежуток между словами), практически не отличается от H_{43} содержания казахского алфавита (42 буквы и 1 пробел):

$$H_0 = \log 32 = 5 \text{ бит};$$

$$H_0 = \log 43 = 5,4 \text{ бит}.$$

Таким образом, полный анализ показывает, что план построения сложной информационной системы может формироваться только на верхних иерархических уровнях и оттуда спускаться на ниже лежащие уровни, задавая на них тот или иной порядок чередования элементов.

Используемый теорией информации статистический метод учета межбуквенных корреляций в литературных текстах обоих языков зависит от смыслового контекста: и одна, и две, и три буквы и т.д. могут быть в одних случаях самостоятельным словом, а в других — входить в состав других слов.

Очевидно, что рассматриваемые сочетания букв относятся к различным иерархическим уровням текста, однако подобное разграничение уровней может осуществляться только по смыслу, который включает в себе анализируемый текст.

Причины возникновения исследуемого порядка всегда остаются за пределами компетенции статистических методов. Находясь как бы на нижних уровнях некой упорядоченной иерархической структуры, вооруженная статистическими методами наука исследует не само действие порождающих исследуемый порядок причин, а лишь его результат. Вероятностная функция энтропии нами использована для строгого определения количества информации и энтропии текстов на уровне букв, но не на уровне слов, так как из слов можно составить практически неограниченное количество текстов.

Используя текст в качестве универсальной модели, можно установить те пределы изменчивости, в которых могут осуществляться самоорганизация и развитие лингвистических систем.

Список литературы

- 1 Кажикенова С.Ш., Оспанова Б.Р. Информационно-энтропийный анализ структуры текста. — Караганда: Изд-во КарГТУ, 2012. — 251 с.
- 2 Кажикенова С.Ш., Оспанова Б.Р. К вопросу о формировании концептуальной системы целевого языка в структуре коммуникативной компетенции // Язык и культура. — 2012. — № 3. — С. 111–121.
- 3 Кажикенова С.Ш., Оспанова Б.Р. О некоторых аспектах языковой модели в теории информации // Международный журнал экспериментального образования. — 2012. — № 8. — С. 115–120.
- 4 Кажикенова С.Ш., Оспанова Б.Р. Лингвосинергетический подход к исследованию текста как самоорганизующегося объекта // Хаос и структуры в нелинейных системах: Материалы междунар. науч.-практ. конф. (18–20 июня). — Караганда: Изд-во КарГУ, 2012. — С. 546–550.

С.Ш.Қажыкенова, К.В.Мазиева, Е.Г.Шурыгина

Мәтін құрылымын талдауға арналған ізгі лингвоматематикалық үлгі

Мақалада мәтіндік материалдың түрлерін зерттеу сұрақтары жетілдеру мақсатымен қарастырылған. Мәтінді әмбебаптық үлгі ретінде қолдана отырып, өзін-өзі ұйымдастыру және лингвистикалық жүйе дамуы жүзеге асырылатын құбылудың шектерін қалыптастыруға болатыны көрсетілді. Қазақ және орыс тілдеріндегі бір, екі, үш, төрт, бес, алты мәтін әріптерінің математикалық есептеу арқылы әріптің энтропия мағынасы алынған.

S.Sh.Kazhikenova, K.V.Maziyeva, Ye.G.Shurygina

An ideal lingua-mathematics model for analysis of the text`structure

The researches of study of the text material of different genres with the purpose of its improvement are presented in this article. It is shown that using the text like a universal model it can be established these limit of changing in which the development of linguistic systems and self-organization can be done. By means of using mathematics calculation the values of entropy of letter given with one, two, three, four, five, six letters of the text in Russian and Kazakh languages are obtained.

References

- 1 Kazhikenova S.Sh., Ospanova B. R. *Information and entropy analysis of structure of the text*, Karaganda: Publ. House of KarGTU, 2012, p. 251.
- 2 Kazhikenova S.Sh., Ospanova B.R. *Language and culture*, 2012, 3, p. 111–121.
- 3 Kazhikenova S.Sh., Ospanova B.R. *The International magazine of experimental education*, 2012, 8, p. 115–120.
- 4 Kazhikenova S.Sh., Ospanova B.R. *Lingvosinergetichesky approach to text research as self-organizing object*, Chaos and structures in nonlinear systems. Materials international scientific and practical conference (on June 18–20) / KarGU, Karaganda: Publ. House KarGU, 2012, p. 546–550.