

Информационный анализ текстов различных стилей на казахском языке

Изучение иерархической структуры текста и методов его информационно-энтропийного анализа является одной из тех актуальных проблем современной лингвистики, которые продиктованы не столько его инженерными приложениями, сколько, в первую очередь, необходимостью применения объективной, количественной оценки грамматического строя и семантико-синтаксической организации текста, а также сравнительно-сопоставительного анализа родственных и неродственных языков.

Ключевые слова: энтропия, информация, иерархия, синергетика, информационно-энтропийный анализ.

Нами проведен информационно-энтропийный анализ большого массива современных текстов, принадлежащих к разным жанрам, тематикам и стилям казахского и русского языков.

Информационно-энтропийный анализ структуры текстов осуществлялся на основе информационной энтропии Шеннона с применением формулы классического определения вероятности [1]. Для чистоты эксперимента из текстов были удалены все знаки препинания. В результате исследований нами установлено, что любой языковой текст, от единичного слова до объемного литературного произведения, может быть представлен как система, элементами которой являются отдельные буквы и их сочетания. Кроме того, нами доказано, что энтропия текстов, принадлежащих к разным жанрам и стилям казахского и русского языков, уменьшается при переходе на более высокий уровень организации. Соответственно, информационная емкость текстов возрастает, что полностью соответствует формулировке фундаментального закона сохранения суммы информации и энтропии.

Язык — это средство обработки и передачи информации, получаемой от внешнего мира, следовательно, для более точного вычисления информации, содержащейся в одной букве казахского текста, надо знать вероятности появления различных букв. Определенный интерес в нашем исследовании представляет измерение энтропии и избыточности текста. Количество слов в тексте зависит от среднего числа букв в слове.

В наших измерениях среднее количество букв в слове равно 6. Энтропия Шеннона количественно характеризует достоверность передаваемого сигнала и используется для расчета количества информации, т.е. измерения энтропии и избыточности текста. Казахский алфавит содержит 43 буквы (42 буквы, 1 пробел), тогда по формуле Шеннона: $H_0 = \log 43 = 5,4$ бит. Энтропия опыта, заключается 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 г. 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, Һ, І**. Первоначально казахские буквы размещались после букв русского алфавита, затем каждую из них поставили после русских букв, сходных по звучанию.

Энтропия научного стиля речи. В качестве примера для информационного анализа был рассмотрен казахский текст научного стиля речи. Материалом для эксперимента послужил отрывок из учебного пособия по музыке М.А.Оразалиевой: *Әр халықтың әні, музыкасы бар. Біздің халықтың да жақсы көріп орындайтын, тыңдайтын музыкасы көп. Музыканың тамыры дегеніміз — ата-бабаларымыздың орындаған әндері мен күйлері. Олар жыл өткен сайын өсе түседі, жаңарып көбейеді. Музыка тамыры — халықтың әндері, күйлері, жырлары. Музыка тамыры ағаштың тамыры сияқты ешқашан ескірмейді, солмайды. Жылма-жыл халық музыкасы да жаңа әндермен, күйлермен толығып отырады. Халық музыкасы дегеніміз — барлық музыканың атасы, яғни оның тамыры. Музыка тамыры — бар музыканың бастау бұлағы.*

Текст для подсчета: *Әр халықтың әні музыкасы бар Біздің халықтың да жақсы көріп орындайтын тыңдайтын музыкасы көп Музыканың тамыры дегеніміз ата бабаларымыздың орындаған әндері мен күйлері Олар жыл өткен сайын өсе түседі жаңарып көбейеді Музыка тамыры халықтың әндері күйлері жырлары Музыка тамыры ағаштың тамыры сияқты ешқашан ескірмейді солмайды Жылма жыл халық музыкасы да жаңа әндермен күйлермен толығып отырады Халық музыкасы дегеніміз барлық музыканың атасы яғни оның тамыры Музыка тамыры бар музыканың бастау бұлағы*

Текст содержит знаков с пробелами — 500, без пробелов — 431. Для расчета относительных частот использована формула классического определения вероятности

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

где n — число всех букв; m — число рассматриваемой буквы.

Ориентировочные значения частот отдельных букв казахского языка представлены в таблице 1 (тире здесь обозначает пробел между словами). В таблице буквы расположены по мере убывания относительных частот.

Т а б л и ц а 1

Значения частот отдельных букв

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

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

Ориентировочные значения частот двухбуквенных сочетаний казахского языка представлены в таблице 2 (тире здесь обозначает пробел между словами), буквы расположены по мере убывания относительных частот.

Т а б л и ц а 2

Значения частот двухбуквенных сочетаний

Сочетание	ы -	- м	ры	ың	ң -	му	уз	зы
Относ. частота	0,032	0,022	0,022	0,020	0,020	0,020	0,020	0,020
Сочетание	ык	ка	ты	- т	та	н -	і -	а -
Относ. частота	0,020	0,020	0,018	0,018	0,018	0,018	0,016	0,016
Сочетание	ыр	лы	- б	ар	- ж	мы	ал	ық
Относ. частота	0,016	0,016	0,014	0,014	0,014	0,014	0,012	0,012
Сочетание	ас	сы	ба	- к	ам	ен	ер	- х
Относ. частота	0,012	0,012	0,012	0,012	0,012	0,012	0,012	0,001
Сочетание	ха	да	рі	- о	ын	нд	ан	де
Относ. частота	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,001
Сочетание	р -	қт	- ә	ән	ді	- д	п -	ай
Относ. частота	0,008	0,008	0,008	0,008	0,008	0,008	0,008	0,008
Сочетание	ны	ла	ме	жы	ні	із	жа	кө
Относ. частота	0,008	0,008	0,008	0,008	0,006	0,006	0,006	0,006
Сочетание	- а	ды	кү	үй	йл	ле	ол	ыл
Относ. частота	0,006	0,006	0,006	0,006	0,006	0,006	0,006	0,006
Сочетание	- с	рм	к -	ор	йт	ег	ге	ім
Относ. частота	0,006	0,006	0,006	0,004	0,004	0,004	0,004	0,004
Сочетание	мі	ат	з -	зд	ағ	ға	л -	- ө
Относ. частота	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
Сочетание	се	ед	аң	на	ып	ей	рл	аш
Относ. частота	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
Сочетание	- е	йд	лм	ма	әр	бі	ің	ақ
Относ. частота	0,004	0,004	0,004	0,004	0,002	0,002	0,002	0,002
Сочетание	кс	өр	іп	ңд	өп	ым	ыз	өт
Относ. частота	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002

Продолжение таблицы 2

Сочетание	тк	ке	са	йы	өс	е-	тү	аб
Относ. частота	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
Сочетание	үс	өб	бе	йе	шт	си	ия	як
Относ. частота	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
Сочетание	еш	шк	ка	ша	ес	ск	кі	ір
Относ. частота	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
Сочетание	со	то	ығ	ғы	от	ра	ад	-я
Относ. частота	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
Сочетание	яғ	ғн	ни	и -	он	ст	ау	у -
Относ. частота	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
Сочетание	бұ	ұл						
Относ. частота	0,002	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:

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

Аналогично этому можно определить и энтропию H_3 текста при учете трех букв казахского текста:

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

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

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = \\ = -0,020 \cdot \log_2(0,020) - 0,020 \cdot \log_2(0,020) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,020 \cdot \log_2(0,0020) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,1832;$$

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = \\ = -0,020 \cdot \log_2(0,020) - 0,020 \cdot \log_2(0,020) - 0,012 \cdot \log_2(0,012) - \dots - \\ - 0,002 \cdot \log_2(0,002) + 0,020 \cdot \log_2(0,0020) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,1657.$$

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.

Отсюда можно заключить, что для казахского языка энтропия текста уменьшается при переходе на более высокий уровень организации, при этом увеличивается информационная емкость текста, что подтверждает развитие языка по закону сохранения суммы информации и энтропии.

Энтропия публицистического стиля речи. На стилистику публицистической, прежде всего газетной, речи сильное влияние оказывает массовый характер коммуникации. Республиканская газета «Егемен Қазақстан» — официальная ежедневная правительственная газета на казахском языке, которая дает официальную информацию о деятельности государственных органов власти, публикует нормативные правовые акты, а также информацию уполномоченных государственных органов по вопросам внутренней и внешней политики.

Текст [2] содержит знаков с пробелами — 500, без пробелов — 438. В результате подсчета энтропии текста при учете одной буквы:

$$H_1 = H(\alpha_1) = -0,124 \cdot \log_2(0,124) - 0,086 \cdot \log_2(0,086) - \dots - \\ - 0,002 \cdot \log_2(0,002) \approx 4,4253.$$

Энтропия текста при учете двухбуквенных сочетаний:

$$H_2 = H\alpha_1(\alpha_2) = H(\alpha_1\alpha_2) - H(\alpha_1) = -0,028 \cdot \log_2(0,028) - 0,026 \cdot \log_2(0,026) - \dots -$$

$$-(0,002) \cdot \log_2(0,002) + 0,124 \cdot \log_2(0,124) + 0,086 \cdot \log_2(0,086) + \dots + \\ + 0,002 \cdot \log_2(0,002) \approx 2,7267.$$

Энтропия текста при учете трехбуквенных сочетаний:

$$H_3 = H\alpha_1\alpha_2(\alpha_3) = H(\alpha_1\alpha_2\alpha_3) - H(\alpha_1\alpha_2) = -0,01 \cdot \log_2(0,01) - \\ - 0,01 \cdot \log_2(0,01) - \dots - 0,002 \cdot \log_2(0,002) + 0,028 \cdot \log_2(0,028) + \\ + 0,026 \cdot \log_2(0,026) + \dots + 0,002 \cdot \log_2(0,002) \approx 1,0687.$$

Энтропия текста при учете четырех букв:

$$H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) = H(\alpha_1\alpha_2\alpha_3\alpha_4) - H(\alpha_1\alpha_2\alpha_3) = \\ = -0,008 \cdot \log_2(0,008) - 0,008 \cdot \log_2(0,008) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,01 \cdot \log_2(0,01) + 0,01 \cdot \log_2(0,01) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,3301.$$

Энтропия текста при учете пяти букв:

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = \\ = -0,008 \cdot \log_2(0,008) - 0,008 \cdot \log_2(0,008) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,008 \cdot \log_2(0,008) + 0,008 \cdot \log_2(0,008) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,1198.$$

Энтропия текста при учете шести букв:

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = \\ = -0,008 \cdot \log_2(0,008) - 0,006 \cdot \log_2(0,006) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,008 \cdot \log_2(0,008) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,0657.$$

В ходе нашего исследования при подсчете числа повторений различных буквенных комбинаций в публицистическом тексте были получены следующие значения (в битах): $H_1 H_2 H_3 H_4 H_5 H_6$

$$4,4253 \quad 2,7267 \quad 1,0687 \quad 0,3301 \quad 0,1198 \quad 0,0657.$$

Таким образом, подсчеты показали, что с переходом на более высокий уровень организации происходит уменьшение энтропии.

Энтропия официально-делового стиля речи. Проанализировав текст научного стиля речи, обратимся к официально-деловому стилю [3]. Текст содержит знаков с пробелами — 500, без пробелов — 434. Используя классическую формулу определения вероятности, рассчитаем для него относительные частоты отдельных букв и относительные частоты по мере убывания. 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:

$$H_1 = H(\alpha_1) = -0,132 \cdot \log_2(0,132) - 0,130 \cdot \log_2(0,130) - \dots \\ - 0,002 \cdot \log_2(0,002) \approx 4,3443.$$

Далее посчитаем условную энтропию текста при учете двух-, трех-, четырех-, пяти- и шестибуквенных сочетаний:

$$H_2 = H\alpha_1(\alpha_2) = H(\alpha_1\alpha_2) - H(\alpha_1) = -0,028 \cdot \log_2(0,028) - 0,028 \cdot \log_2(0,028) - \dots - \\ - (0,002) \cdot \log_2(0,002) + 0,132 \cdot \log_2(0,132) + 0,130 \cdot \log_2(0,130) + \dots + \\ + 0,002 \cdot \log_2(0,002) \approx 2,6006.$$

$$H_3 = H\alpha_1\alpha_2(\alpha_3) = H(\alpha_1\alpha_2\alpha_3) - H(\alpha_1\alpha_2) = \\ = -0,016 \cdot \log_2(0,016) - 0,012 \cdot \log_2(0,012) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,028 \cdot \log_2(0,028) + 0,028 \cdot \log_2(0,028) + \dots + 0,002 \cdot \log_2(0,002) \approx 1,0225.$$

$$H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) = H(\alpha_1\alpha_2\alpha_3\alpha_4) - H(\alpha_1\alpha_2\alpha_3) = \\ = -0,012 \cdot \log_2(0,012) - 0,008 \cdot \log_2(0,008) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,016 \cdot \log_2(0,016) + 0,012 \cdot \log_2(0,012) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,2665.$$

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = \\ = -0,008 \cdot \log_2(0,008) - 0,006 \cdot \log_2(0,006) - \dots - 0,002 \cdot \log_2(0,002) + \\ + 0,012 \cdot \log_2(0,012) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,2012.$$

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = \\ = -0,006 \cdot \log_2(0,006) - 0,006 \cdot \log_2(0,006) - 0,006 \cdot \log_2(0,006) - \dots -$$

$$-0,002 \cdot \log_2(0,002) + 0,008 \cdot \log_2(0,008) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,095.$$

В результате были получены следующие значения (в битах): H_1 H_2 H_3 H_4 H_5 H_6
4,3443 2,60061,0225 0,2665 0,2012 0,095.

Энтропия разговорно-бытового стиля речи. Для информационно-энтропийного анализа разговорно-бытового стиля речи рассмотрен отрывок из повести С.Муратбекова «Жабайы алма» («Дикая яблоня») [4]. Текст содержит знаков с пробелами — 500, без пробелов — 434. В результате исследования энтропия текста при учете одной буквы казахского текста равна

$$H_1 = H(\alpha_1) = -0,162 \cdot \log_2(0,162) - 0,1 \cdot \log_2(0,1) - \dots - 0,002 \cdot \log_2(0,002) \approx 4,3873.$$

Энтропия двухбуквенных сочетаний $H_2 = H\alpha_1(\alpha_2)$ опыта α_2 , состоящего в определении одной буквы текста, при условии, что нам известен исход опыта α_1 , состоящего в определении предшествующей буквы того же текста, равна:

$$H_2 = H\alpha_1(\alpha_2) = H(\alpha_1\alpha_2) - H(\alpha_1) = -0,022 \cdot \log_2(0,022) - 0,020 \cdot \log_2(0,020) - \dots - (0,002) \cdot \log_2(0,002) + 0,162 \cdot \log_2(0,162) + 0,1 \cdot \log_2(0,1) + \dots + 0,002 \cdot \log_2(0,002) \approx 2,7843.$$

Энтропия текста при учете трехбуквенных сочетаний:

$$H_3 = H\alpha_1\alpha_2(\alpha_3) = H(\alpha_1\alpha_2\alpha_3) - H(\alpha_1\alpha_2) = -0,012 \cdot \log_2(0,012) - 0,01 \cdot \log_2(0,01) - \dots - 0,002 \cdot \log_2(0,002) + 0,022 \cdot \log_2(0,022) + 0,02 \cdot \log_2(0,02) + \dots + 0,002 \cdot \log_2(0,002) \approx 1,0557.$$

Энтропия текста при учете четырех букв:

$$H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) = H(\alpha_1\alpha_2\alpha_3\alpha_4) - H(\alpha_1\alpha_2\alpha_3) = -0,006 \cdot \log_2(0,006) - 0,006 \cdot \log_2(0,006) - \dots - 0,002 \cdot \log_2(0,002) + 0,012 \cdot \log_2(0,012) + 0,01 \cdot \log_2(0,01) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,3187.$$

Энтропия текста при учете пяти букв имеет следующее приближенное значение:

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = -0,004 \cdot \log_2(0,004) - 0,004 \cdot \log_2(0,004) - \dots - 0,002 \cdot \log_2(0,002) + 0,006 \cdot \log_2(0,006) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,1265.$$

И, наконец, энтропия при учете шестибуквенных сочетаний данного текста равна

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = -0,004 \cdot \log_2(0,004) - 0,004 \cdot \log_2(0,004) - 0,004 \cdot \log_2(0,004) - \dots - 0,002 \cdot \log_2(0,002) + 0,004 \cdot \log_2(0,004) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,056.$$

В результате были получены следующие значения (в битах): H_1 H_2 H_3 H_4 H_5 H_6
4,3873 2,78431,0557 0,3187 0,1265 0,056.

Нами были проанализированы тексты *научного, публицистического, официально-делового, разговорно-бытового* стилей речи казахского языка. Цифры показывают, что энтропия H_1 , H_2 , H_3 , H_4 , H_5 и H_6 для данных стилей практически равна (в битах). Отличие заключается в следующем:

Т а б л и ц а 3

Сравнительные значения энтропии текста научного, публицистического, официально-делового и разговорно-бытового стилей речи

Энтропия	Стили речи			
	научный	публицистический	официально-деловой	разговорно-бытовой
H_1	4,3598	4,4253	4,3443	4,3873
H_2	2,3444	2,7267	2,6006	2,7843
H_3	0,852	1,0687	1,0225	1,0557
H_4	0,2813	0,3301	0,2665	0,3187
H_5	0,1882	0,1198	0,2012	0,2165
H_6	0,1637	0,0657	0,095	0,056

Таким образом, подсчеты показали, что исследованные стили языка имеют примерно одинаковый уровень избыточности и энтропии. Полный анализ показывает, что план построения сложной информационной системы может формироваться только на верхних иерархических уровнях и оттуда спускаться на лежащие ниже уровни, задавая на них тот или иной порядок чередования элементов [5, 6].

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 заключает 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. В настоящей работе 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., Гладкова М. Информационно-энтропийный анализ сложных иерархических систем // Вестн. Караганд. ун-та. — Сер. математика. — 2011. — № 2 (62). — С. 58–65.
- 2 Жәрімбетова Н. Ауыл шаруашылығына оң өзгерістер қажет // Егемен Қазақстан. — 2012. — 22 маус. — 2-б.
- 3 Дүйсембекова Л. Қазақ тілі: Іс қағаздарын жүргізу. — Алматы: «Мемлекеттік тілде дамыту институты» ЖШС, 2010. — 400 б.
- 4 Мұратбеков С. Жабайы алма: Повесть. — Алматы: Атамұра, 2002. — 400 б.
- 5 Ospanova B., Kazhikenova S.Sh. Information-Entropy Analysis of the Text Structure // Germany: LAP LAMBERT Academic Publ., 2013. — 236 p.
- 6 Ospanova B., Kazhikenova S.Sh., Sadykov M., Hasan M. Determination and stochasticity of hierarchical systems in lingual synergy and their application in evaluation of the writing and speaking styles of professionals // 4th International Conference on Graphic and Image Processing (ICGIP 2012). — Singapore October 6–7, 2012. — Vol. 8. — P. 768.

С.Ш.Қажыкенова, Б.Р.Оспанова

Қазақ тіліндегі әр стильді мәтіннің ақпараттық талдауы

Мәтіннің иерархиялық құрылымын оқып-үйрену және оның ақпаратты-энтропиялық талдауының әдістері қазіргі заманғы лингвистиканың өзекті сұрақтарының бірі болып табылады. Бұл мәселелер оның инженерлік қосымшаларымен айтылмаған. Ең алдыменен, бұл грамматикалық қатардың және мәтіннің семантика-синтаксикалық ұйымдастыруының объективті, сандық бағалануына пайдалану қажеттілігінен туындаған. Сонымен қатар туыстық және туыстық емес тілдердің салыстырмалы талдауын жасауда екені белгілі.

S.Sh.Kazhikenova, B.R.Ospanova

Information analysis of texts different styles of the kazakh language

Exploring the hierarchical structure of the text and methods of information-entropy analysis is one of the urgent problems of modern linguistics, which dictated not so much its engineering applications, much in the first place, the need to apply an objective, quantitative evaluation of grammatical structure and semantic and syntactic organization of the text, and well as comparative analysis of related and unrelated languages.

References

- 1 Kazhikenova S.Sh., Gladkova M. *Bull. Karaganda University, Ser. Mathematics*, 2011, 2 (62), p. 58–65.
- 2 Zharimbetova N. *Egemen Kazakhstan*, 22, 2012, 2 p.
- 3 Dyisembekova L. *Kazakh language: Is kazazdaryn zhyrgizu*, Almaty: «Memleketik tilde damyту institutions» ZSHS 2010, 400 p.
- 4 Мұратбеков С. *Zhabaiy Alma*, Almaty: Atamura, 2002, 400 p.
- 5 Ospanova B., Kazhikenova S.Sh. *Information-Entropic Analysis of the Text Structure*, Germany: LAP LAMBERT Academic Publishing, 2013, 236 p.
- 6 Ospanova B., Kazhikenova S.Sh., Sadykov M., Hasan M. *4th International Conference on Graphic and Image Processing (ICGIP 2012)*, Singapore October 6–7, 2012, 8, p. 768.

УДК 656.22

А.Ш.Кажикенова, К.М.Турдыбекова

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

Анализ вязкости цезия в зависимости от температуры с учетом ассоциации кластеров из кристаллоподвижных частиц

В статье рассмотрена температурная зависимость вязкости согласно концепции хаотизированных частиц. Проанализированы модели зависимости вязкости от температуры с учетом различного содержания частиц: кристаллоподвижных, жидкоподвижных и пароподвижных. Предложена новая кластерная модель температурной зависимости вязкости, позволяющая выявить поведение вязкости на широком диапазоне температур. Рассчитана энергия активации, соответствующая энергии вандерваальсовского притяжения, которая позволяет связывать между собой кластеры.

Ключевые слова: вязкость, концепция хаотизированных частиц, кристаллоподвижные частицы, реперная точка, ассоциации кластеров.

Эффективность многих металлургических процессов зависит от правильно и качественно выбранных условий течения процесса, а также от строгого учета физико-химических характеристик вещества, среди которых вязкость имеет существенное значение. Вязкость является наиболее структурночувствительной характеристикой расплава, дающей общее представление о внутренних силах взаимодействия частиц, и ее изучение представляет значительный научный и практический интерес. Ограниченность и неудовлетворенность существующих теоретических методов расчета вязкости, основанных на структурных модельных теориях, вызывает необходимость совершенствования существующих и разработки новых моделей, позволяющих адекватно описывать в широком диапазоне температур вязкость металлических и шлаковых расплавов и использовать результаты моделирования в решении практических задач металлургического производства.

Сотрудниками Химико-металлургического института (г. Караганда) была разработана концепция хаотизированных частиц [1], основанная на распределении Больцмана. Согласно этому подходу, все три агрегатных вещества рассматриваются с единой точки зрения по его бесструктурной составляющей, которая численно определяется долей сверхбарьерных и подбарьерных по теплосодержанию в точках плавления $RT_{пл}$ и кипения $RT_{кун}$ частиц согласно распределению Больцмана.

Зависимость вязкости расплавов от температуры с точки зрения концепции хаотизированных частиц может быть выражена следующими уравнениями [2]:

– с учетом тормозящего влияния на текучесть только кристаллоподвижных частиц:

$$\nu = \nu_{pen} T_{pen} / T, \quad (1)$$

где ν_{pen} и T_{pen} — соответственно кинематическая вязкость и абсолютная температура для некоторой реперной точки, выбираемой произвольно в качестве наиболее надежного экспериментального определения;