

СИНЕРГЕТИЧЕСКИЕ АСПЕКТЫ ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ ИНФОРМАТИКИ

Кажикенова С.Ш., Тажина А.М.

Карагандинский государственный университет им. академика Е.А. Букетова, Караганда,

E-mail: arailim12.86@mail.ru

Подход к языку как иерархической структуре позволяет рассматривать текст как с точки зрения анализа его составляющих, так и с точки зрения синтеза их на высшем языковом уровне.

Мы предлагаем идеальную лингвоматематическую модель для анализа структуры текста. Она построена на основе фундаментального закона сохранения суммы информации и энтропии с применением формулы Шеннона. 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: $H = -\sum_{i=1}^N p_i \log_2 p_i$, где p_i – вероятность обнаружения какой-либо единицы системы в их множестве

N . До опубликования созданной Шенноном теории Хартли предложил определять количество максимальной энтропии по формуле $H_{\max} = \log_2 N$. Важным для языкознания измерением является энтропия языка. Энтропия языка 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 $H_0 = \log_2 32 = 5$ бит. где H_0 - максимальное значение энтропии текста, заключающегося в приеме одной буквы русского текста (информация, содержащаяся в одной букве), при условии, что все буквы считаются одинаково вероятными; бит- единица измерения информации. Для вычисления информации текста нами были подсчитаны вероятности появления каждой буквы в данном отрывке. При подсчете учитывались 31 буква русского алфавита (буквы е и е, ь и ь принимаются как одна буква) и пробел, все остальные знаки (скобки, кавычки, запятые и пр.) не рассматривались. Цифровые данные, содержащиеся в тексте, расписаны прописью. Подсчет вероятности (р) появления различных букв в тексте достигается расчетом относительной частоты отдельных букв. Для определения вероятности появления одной буквы в тексте воспользуемся классической формулой определения вероятности:

$P(\text{одн. буквы}) = \frac{m}{n}$. Нами предложена теоретическая модель текста, учитывающая его иерархическую

структуру, на основе различных характеристик русскоязычных текстов одно, двух, трех, четырех, пяти, шестибуквенных сочетаний. Был проведен информационно-энтропийный анализ отрывка в 500 знаков из лекционного курса по дисциплине "Информатика". Для вычисления информации текста объемом в 500 знаков, без пробелов –429 были подсчитаны *вероятности* появления одной буквы, двухбуквенных, трехбуквенных, четырехбуквенных, пятибуквенных, а также шестибуквенных сочетаний в данном тексте. Приравняв частоты вероятностям появления соответствующих букв, получим для энтропии одной буквы данного текста приближенное значение: $H_1 = H(\alpha_1) \approx 4,3742$

Для дальнейшей характеристики количества информации и энтропии текста аналогично рассчитаем условные энтропии H_2, H_3, H_4, H_5, H_6 . Энтропия текста при учете трехбуквенных сочетаний: $H_2 = H\alpha_1(\alpha_2) \approx 3,0423$ $H_3 = H\alpha_1\alpha_2(\alpha_3) \approx 0,7895$ $H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) \approx 0,5605$

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) \approx 0,0451 \quad H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) \approx 0,0108$$

Результаты исследования заставляют полагать, что любой языковой текст, от единичного слова до объемного литературного произведения, может быть представлен как система, элементами которой являются отдельные буквы, а части представляют собой совокупности одинаковых букв. Соответственно, с помощью синергетической теории информации можно проводить структурный анализ произвольных текстов со стороны их хаотичности и упорядоченности по количеству и числу встречаемости отдельных букв.

Список использованных источников

1. Малышев В.П. Вероятностно-детерминированное отображение. Алматы: Ғылым, 1994. 376 с.
2. Kazhikenova S.Sh., Ospanova B.R. Information-Entropic Analysis of the Text Structure// Germany: LAP LAMBERT Academic Publishing, 2013. – 236 p.