

ДЕНОТАТИВНОЕ ПРОСТРАНСТВО ТЕКСТА НАУЧНО-ПОПУЛЯРНОЙ СТАТЬИ

Проблема отражения в тексте мира действительности тесно связана с проблемой изучения денотативного пространства текста. Это является причиной актуальности исследований в данной области.

Материалом нашего исследования послужило денотативное пространство текста научно-популярных статей, взятых из журнала *Discover*.

Денотативная структура текста обозначает структуру, составляющую совокупность денотатов, связанных с предметными отношениями в целостном семантическом комплексе, отображаемых в ситуации языковыми единицами (А. И. Новиков) Денотативная структура текста представляет план содержания текста и может отражать только небольшую часть связи между денотатами, что и является фрагментами действительности. Для того чтобы понять текст, нужно видеть то, как связаны эти фрагменты действительности. Стоит упомянуть, что смысловая нагрузка текста не является равномерной, поэтому важно определить единицы, которые бы являлись наиболее значимыми для понимания текста. По мнению достаточно широкого круга ученых, таковыми единицами являются ключевые слова.

Ключевые слова несут в себе наибольшую информативность, а также могут суммировать смысл текста. Е. М. Кузичева описывает ключевые слова как «веху, указывающую на те или иные компоненты текста, обладающие важным значением для понимания текста как целого».

К сожалению, единой методики по выявлению ключевых слов в тексте не существует. Но в ходе многочисленных исследований были выделены следующие методики: лингвистическая, статистическая, или частотная, машинная и смешанная. Лингвистический метод предполагает морфологический, структурный, синтаксический анализ текста. Лингвистические методы основываются на значениях слов, используют онтологии и семантические данные о слове. К сожалению, эти методы слишком трудоемки на ранних этапах: разработка онтологий, к примеру, весьма трудозатратна (О. Ю. Лазарева).

Статистический метод получил широкое распространение благодаря своей простоте реализации, так как единственным параметром для определения слова к множеству ключевых слов является частота его употребления. Частотный критерий при выявлении ключевых слов является более простым и сподручным. Однако частоту встречаемости нельзя назвать надежным критерием, поскольку в таком случае принимается во внимание не план содержания текста, а план его выражения. На наш взгляд, целесообразно учитывать также и семантический критерий при установлении наиболее частотных смысловых единиц и представить данный метод не как метод отбора ключевых слов, но как метод отбора ключевых денотатов.

Рассмотрим процедуру выявления денотативных областей на примере статьи «Plants may help make your surrounding quieter» by Leslie Nemo. Данный процесс можно разделить на несколько этапов.

Первый этап включает в себе обработку, форматирование текстового документа и приведение его к формату, пригодному в дальнейшем для обработки и распознавания. На этапе подготовки текста к дальнейшему распознаванию производится лексический и морфологический анализ.

Второй этап – это непосредственное выделение ключевых слов. Также на этом этапе проводится сравнение слов, которые могли бы быть емкими лексическими единицами, с эталонным списком ключевых слов. На этом этапе выходит значительный объем единиц:

Noisy, terrible, to sleep, is, noise, pollutant, noise, is, declared, is, banning, muffle, researcher, investigate, studies, greenery, noise, suggested, reduce, foliage, mufflers, studying, interact, pieced, leaves, sound, sound, quiver, waves, making, dissipates, leaves, sound, directions, deflect, bouncing, studies, trees, noise, Labs, researchers, noise, documenting, reduce, simulate, measure, research, study, noise qualities, volume, did, comparing, cut, noise pollution, cacophony of sounds, car, University of Montreal, is, can't, cruising down, noise, directions, means, coming, information, vegetation, noise, Smargiassi, regarding, reduce, says, plants, noise, mask, depends, planted, chosen, dense hedges, noise, better, Kang, blocking off, says, study, sound, layouts, deep, managed, were, branches, foliage, spaces, sound, to extend, to avoid, whistle through, lack, foliage, sounds, ways, can block, combined, landscaped, sound barriers, craft, studies, greenery, architects, Kang, do, see, is, is not designed, says, Kang, designing, greenery, sounds, notes, to keep out loud, isn't taught, not that well-known, Kang, is, says, struggle, sound barrier, noise problems, behave differently, could have, potential, sound buffers, construct, designers, sound, take advantage, perceive, study, Kang, lilac plantings, noise, found, reported, being less annoyed, uncommon, Kang, often, designing, says, could put, peace and quiet, we, crave.

На третьем этапе происходит упорядочивание списка полученных ключевых слов, возможна некоторая визуализация этого списка, а также его сокращение и распределение. На этом этапе нами была исключена всякая лексика, которая повторялась менее трех раз или встречалась исключительно

в одном абзаце. Также была вычеркнута косвенная речь (*he/she said that...*), так как это способ введения в текст статьи чужую речь, а не само ее содержание.

study\ investigate\ research – изучать, исследование

greenery\ foliage\ tree – зелень, листва, деревья

is\to be – есть, являться

reduce\block off – сократить, заблокировать

На этапе упорядочения списка ключевых денотатов была использована следующая формула, которая помогает определить коэффициент важности ключевых слов (В. В. Зубов):

$K = (F \times M / N \times n)$, где

F – абсолютная частота слова в тексте (в нее входит суммарная частота всех синонимов слова и его местоименных замен);

M – число абзацев, в которых встречалось слово;

N – общее число слов в тексте;

n – общее число абзацев в тексте;

K – коэффициент важности слова.

После применения данной формулы были получены следующие данные:

Greenery\ foliage\ tree $K = 0,0234352256186317$

sound $K = 0,0186317322$

noise $K = 0,0174672489082969$

reduce\block off $K = 0,0139737991266376$

is\to be $K = 0,0151382823871907$

researcher $K = 0,0050946142649199$

study\ investigate\ research $K = 0,0034934497816594$

В результате нам удалось выявить следующие денотативные области, представленные посредством полученных нами ключевых денотатов. Рядом указаны количество абзацев и количество повторений в тексте.

Greenery\ foliage\ tree (23, 7) – окружающая среда

sound (16, 7) – физическое явление;

noise (15, 8) – физическое явление;

reduce\block off (16, 6) – человеческая деятельность;

is\to be (8, 5) – тождество;

researcher (7, 5) – вербальное информирование;

study\ investigate\ research (6, 4) – вербальное информирование.

После выявления денотативных областей был высчитан процент частотности указанных денотативных областей: природа (25,49 %), человеческая деятельность (16,99 %), тождество (13,4 %), человек как живое существо (9,08 %), место (8,81 %), универсальные смыслы и отношения (7,39 %), социальная сфера (3,69 %), вербальное информирование (3,54 %), энергия (2,8 %), физическое явление (2,59 %), прием пищи (2,01 %), болезни (1,16 %), время (0,84 %), население (0,79 %), нация (0,69 %), материал (0,42 %), вещество (0,32 %).