

УДК 004.42:159.99

АВТОМАТИЗАЦИЯ ФОНОСЕМАНТИЧЕСКОГО АНАЛИЗА И СИНТЕЗА СЛОВ И ТЕКСТОВ РУССКОГО ЯЗЫКА

С.Н. Верзунов

Рассматривается проблема алгоритмизации фоносемантического анализа и синтеза слов и текстов русского языка, а также разработка соответствующих программных средств. Разработанные алгоритмы протестированы на оригинальном лексическом материале.

Ключевые слова: фоносемантический анализ; синтез слов; шкалы Осгуда; звукоцветовой анализ; реклама.

AUTOMATION ANALYSIS AND SYNTHESIS WORDS AND TEXTS OF THE RUSSIAN LANGUAGE

S.N. Verzunov

The paper is devoted to the problem of algorithmization phonosemantic analysis and synthesis words and texts of the Russian language and development of appropriate software. The developed algorithms are tested on the original lexical material.

Keywords: phonosemantic analysis; word generation; Osgood scales; sound-and-colour analysis; blurb.

Введение. В 1952 г. американский психолог Ч. Осгуд, анализируя публичные выступления политиков, заметил: из двух примерно одинаковых кандидатов (и их столь же идентичных программ) выигрывает тот, кто использует более благозвучную мелодику речи. Осгуд разработал способ управления “музыкой слов”, назвав его методом “семантических дифференциалов”. Фоносемантика – наука, которая рождается и утверждает себя на стыке фонетики (по плану выражения), семантики (по плану содержания), лексикологии (по совокупности этих планов) и психологии (теория восприятия). В основе фоносемантики лежит явление синестезии, состоящее в возникновении ощущения одной модальности под воздействием раздражителя другой модальности. Например, синония, разновидность синестезии, при которой два чувства – зрение и слух – связываются воедино [1].

В последние несколько лет в некоторых работах, например [2], были исследованы языковые и стилистические особенности рекламных текстов. Исследования показали, что наряду с качеством товаров или услуг, определяющим успешность внедрения их на рынок, привлекательность образа фирмы, созданная благозвучной рекламой, также может оказать существенное влияние на этот процесс.

Современный компьютерный анализ позволяет в считанные секунды определять фонетическое содержание слов и даже значительных по объему текстов. Формализованные методы изучения текста воплощены в прикладных компьютерных программах анализа текста на уровне фоносемантики: ВААЛ, DIATON, PSYLINE CD, и др. Однако существующие программные системы либо дорогостоящи, либо не удобны в использовании. Восполнить этот пробел призваны предлагаемые в настоящей работе алгоритмы автоматизации фоносемантического анализа и синтеза и разработанные на их основе программные средства.

Алгоритм фоносемантического анализа слов.

Слово состоит из понятийного ядра и эмоционального ореола. Понятийное ядро включает в себе смысл, ореол – окраску слова. Например, слова папа, папулечка, отец несут один смысл (родитель-мужчина), но несут разную эмоциональную окраску. Но если ореол объективно существует, то его можно измерить. Если взять группу людей, например, из 50 человек, и по шкале от 1 до 5 попросить определить оценку задаваемому слову, то можно измерить частотность оценок. Например, на слово “дом” были получены следующие оценки: 1 (очень хорошо) – 35 человек, 2 (хорошо) – 10 человек,

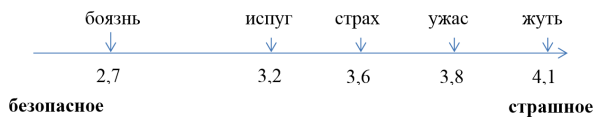


Рисунок 1 – Фоносемантические характеристики некоторых слов

3 (нейтрально) – 3 человека, 4 (плохо) – 2 человека, 5 (очень плохо) – 0 человек. Тогда средняя оценка этого слова – $1*35+2*10+3*3+4*2+5*0 = 72$; $72/50 = 1.44$; таким образом, 1,44 – ореол слова “дом”. Можно измерить ореолы других слов и нанести их на шкалу. Разные слова заняли разное положение на ней, как показано на рисунке 1.

В 60-х годах прошлого века Ю. Орлов высказал мысль, что между звучанием и значением есть определенное соответствие. Звуки измеримы по шкале Остуда: О – светлый; Р – устрашающий; К – быстрый; Ш – медленный и т. д. По наличию звуков можно вычислить фоносемантические характеристики слова и определить его местоположение по шкале фоносемантических характеристик.

Вычисление фонетического значения слова производится по формуле:

$$S = \frac{\sum x_i k_i}{\sum k_i},$$

где S – фонетическое значение слова; i – индекс, пробегающий все графоны (звуки) данного слова; x_i – средняя оценка фонетического значения i-го звука (фонетические значения звуков в словах по всем шкалам приведены в работе [3]); k_i – поправочный коэффициент для i-го звука, равный

$$k_i = \frac{P_{max}}{P_i},$$

где P_i – регулярная частотность i-го звука); P_{max} – максимальная частотность звука в данном слове. Для первого графона поправочный коэффициент k_i учетверяется, а для ударного – удваивается. Если первый графон является ударным – коэффициент k_i только учетверяется [3]. Данный алгоритм показан на рисунке 2.

Алгоритм звукоцветового анализа слов. Эксперименты с большим количеством тестируемых людей показали, что в подавляющем большинстве испытуемые ассоциируют, по крайней мере, гласные с цветами вполне определено. Особенно единодушны мнения относительно трех гласных – А, Е, И. Графон А вполне согласованно называют красным, Е – четко зеленым, а И – определенно синим. Графон О все считают светлым и ярким, но хотя многие испытуемые называют его желтым,

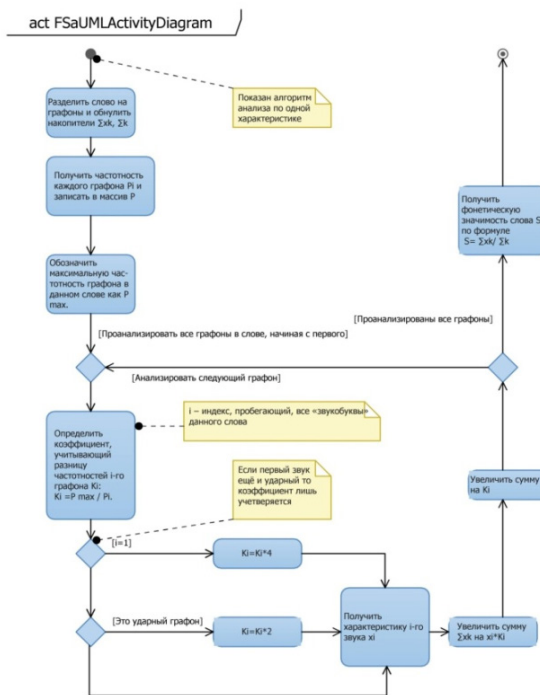


Рисунок 2 – Алгоритм фоносемантического анализа слов

все же довольно часто его определяют как “белый”. Получается, что он солнечный [4].

Алгоритм звукоцветового анализа слов показан на рисунке 3. Алгоритм позволяет визуализировать звукоцветовую оценку слова по каждому символу.

Особенности анализа текстов. При фоносемантическом исследовании текста также анализируются графоны, однако, без деления их на ударные и безударные. Метод, который применялся при расчете фонетической значимости слова, как показывает практика, при анализе текста не подходит. Кроме того, не все признаки, подходящие для оценки слов, подойдут для характеристики значимости звукового тона текста.

Например, вряд ли могут служить для этой цели шкалы “гладкий-шероховатый”, “длинный-короткий” или “округлый-угловатый”; здесь более уместны шкалы экспрессивно-оценочного типа: “нежный-грубый”, “радостный-печальный” и т. п.

В обычной речи звуки встречаются с определенной частотностью, поэтому логично будет предположить, что специальный содержательный звуковой тон текста может быть создан путем увеличения числа звуков с определенными значимостями и уменьшением числа звуков с противоположными значимостями. За норму можно принять

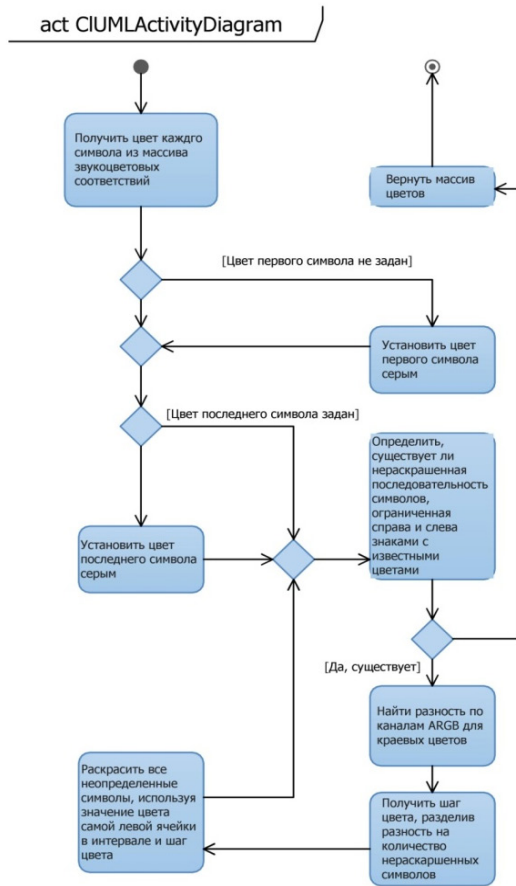


Рисунок 3 – Алгоритм звукоцветовой оценки слов

частотность звуков в обиходной разговорной речи, но уже без деления их на ударные и безударные [3].

Нормальные частотности подвержены колебаниям. Но эти колебания совершаются в некоторых границах. За единицу при измерении размаха колебаний принимают величину $\sigma_{P_i,1}$ равную:

$$\sigma_{P_i} = \sqrt{\frac{P_i(1-P_i)}{n}}$$

где i – индекс, пробегающий все звукобуквы, присутствующие в тексте; n – число всех звукобукв в тексте; P_i – нормальная частотность графона (в тексте разговорного стиля). Звуковой тон создается только теми графонами, частотность которых существенно и неслучайно отклоняется от нормы. Так, для анализа отбираются все отклонения y_i равные:

$$y_i = \frac{F_i - P_i}{\sigma_{P_i}}$$

где F_i – частотность звукобуквы в тексте, если $|y_i| > 1$. Остальные графоны в данном случае роли

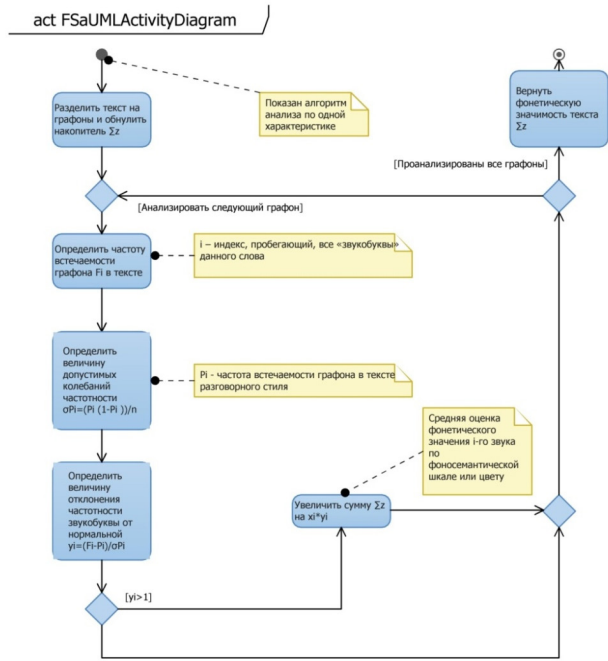


Рисунок 4 – Алгоритм фоносемантического и звукоцветового анализа текста

не играют, поскольку их частотность оказалась в пределах нормальной и, следовательно, они и не привлекают к себе внимание, остаются в тени. Оценка символического значения звуков Z по одной шкале находится по формуле:

$$Z = \sum_i y_i x_i,$$

где x_i – средняя оценка фонетического значения i -го звука по данной шкале. Оценка звукоцветового значения текста по определенному цвету C находится по формуле

$$C = \sum_i y_i w_i,$$

где w_i – средняя оценка цветového значения i -го звука [3]. Алгоритм фоносемантического и звукоцветового анализа текста показан диаграммой на рисунке 4.

Алгоритм синтеза слов с заданными характеристиками. Генерация слов осуществляется по шаблону, причем символы в верхнем регистре и знак “-” означают отдельное вхождение знака, символ “г” – любую гласную, а символ “с” – любую согласную (см. рисунок 5).

Например, шаблон “ДИАссгН” означает, что слово начинается с ДИА, буква Н находится на конце слова, а между ними находится два согласных и один гласный звук. По шаблону “ДИ-АссгН” и характеристикам (например: добрый, яркий, светлый, активный, веселый, радостный,

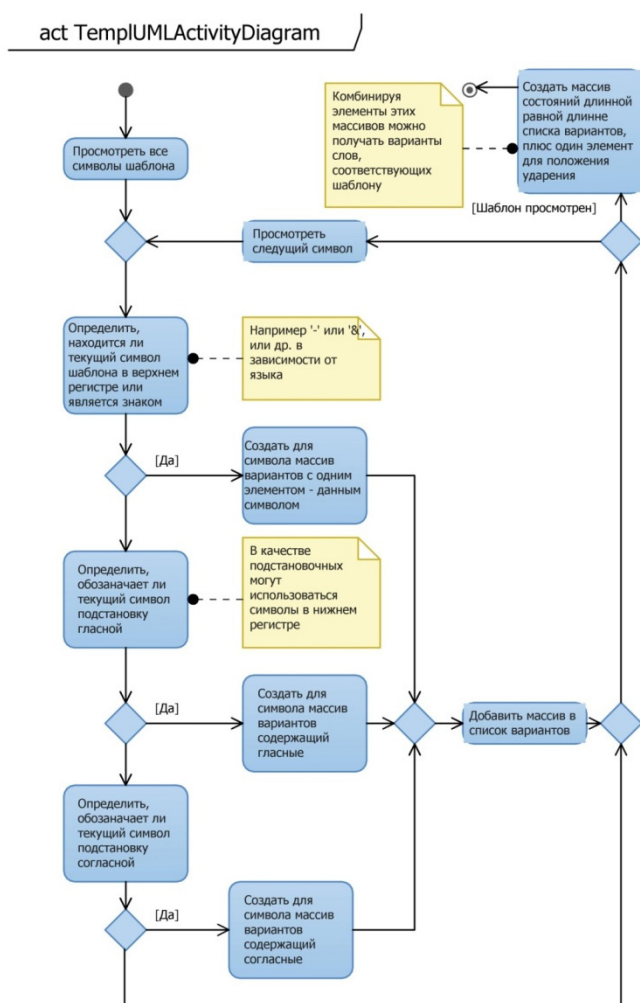


Рисунок 5 – Алгоритм синтеза слов по шаблону

короткий) будет создано 1840 вариантов слов. Необходимо также учитывать, что слова с различным положением ударения имеют разные фоносемантические характеристики.

Примеры использования программных средств. Если проанализировать, например, фоносемантические характеристики слова “Родина” (см. рисунок 6), то можно определить, что слово “Родина” производит впечатление чего-то хорошего, большого, грубого, мужественного, активного, сильного, холодного, опасного, величественного, яркого, громкого, храброго, могучего, подвижного. Конечно слово “Родина” женского пола, но, тем не менее, она ассоциируется с чем-то мужественным, суровым, как на известном плакате, показанном на рисунке 7. Не менее показательны фоносемантические и звукоцветовые характеристики



Рисунок 6 – Фоносемантический анализ слова “Россия”

слова “Россия” (см. рисунок 8). Звукоцветовые ассоциации к слову “Россия” – белый, синий, красный, совпадают с порядком цветов в государственном флаге Российской Федерации (см. рисунок 9). Интересно и то, что звукоцвет слова “Родина” абсолютно совпадает со звукоцветом слова “Россия”.

Кроме того, фоносемантические характеристики слова “Россия” говорят сами за себя и в очередной раз показывают глубокую связь между языком, культурой, психологией отдельного человека и общества в целом.

Таким образом, предложенные автоматизированные методы открывают возможность использовать фоносемантический анализ текстов и синтез в различных областях деятельности, таких как поиск наиболее удачных названий, торговых марок, и имен сайтов составлять рекламные тексты



Рисунок 7 – Плакат “Родина-мать зовет!”

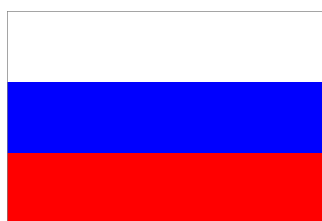


Рисунок 9 – Государственный флаг Российской Федерации

с заранее заданными характеристиками воздействия на потенциальную аудиторию. Новые возможности открываются и в исследовании стихов и прозы.

Литература

1. Богданова Т.В. Психология (включая основы социальной психологии): учебно-методическое пособие для студентов / Т.В. Богданова. Смоленск: Изд-во СмолГУ, 2013. 110 с.

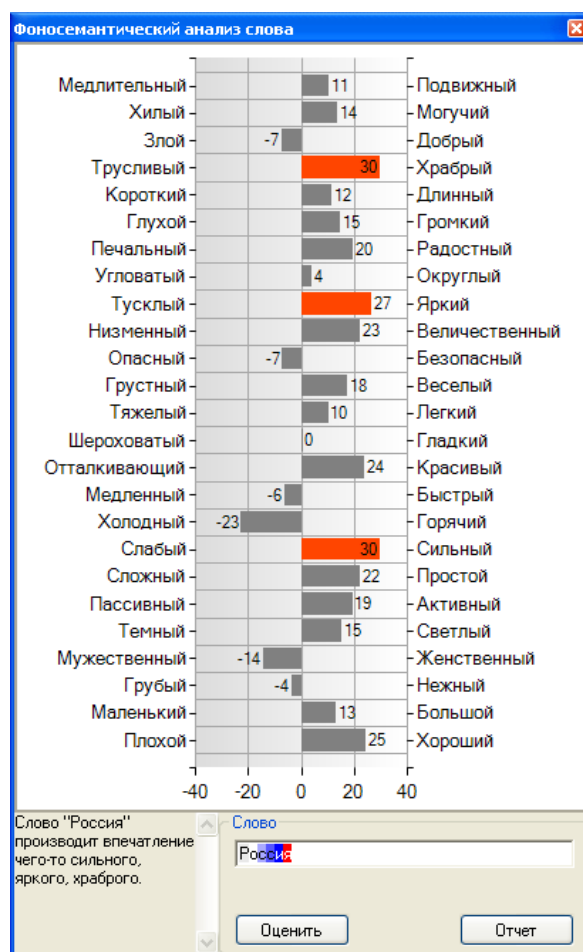


Рисунок 8 – Фоносемантический анализ слова “Россия”

2. Шепелева Г.П. Реклама в Кыргызстане: особенности языка и стиля / Г.П. Шепелева // Вестник КРСУ. 2009. Т. 9. № 3. С. 117–121.
3. Журавлев А.П. Фонетическое значение / А.П. Журавлев. Л.: Изд-во Ленингр. ун-та, 1974. 159 с.
4. Журавлев А.П. Звук и смысл: книга для внеклассного чтения учащихся старших классов / А.П. Журавлев. М.: Просвещение, 1991. 160 с.