

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АКТИВНОСТИ КЛЮЧЕВЫХ СЛОВ: АВТОМАТИЗИРОВАННЫЙ ПОДХОД

А. А. Фаустов, И. Е. Воронина, М. Я. Розенфельд, Н. Ю. Свиридова

Воронежский государственный университет

Поступила в редакцию 21.12.2016 г.

Аннотация. В статье обосновывается логика построения ассоциативного поля слов и предлагается программная реализация этого алгоритма. Материалом для построения поля выступают авторские ключевые слова (маркемы) и «обычные» слова, а также лексические реакции автора на них, полученные в ходе психолингвистических экспериментов с его участием. Благодаря автоматизированному построению ассоциативного поля для каждого из экспериментальных слов становится возможным сравнить семантическую активность маркем и «обычных» слов в спонтанной речи автора.

Ключевые слова: (маркема), автор, ассоциативное поле, семантика, психолингвистический эксперимент.

Annotation. The article substantiates the logics of word associated field construction and offers the software implementation of the said algorithm. The material for field construction is provided by author's keywords (markemes) and «ordinary» words as well as author's lexical reactions to them, obtained during psycholinguistic experiments with his personal involvement. Computer-aided construction of the associated field for each of the experimental words creates the possibility of comparison between semantic activity of markemes and «ordinary» words in spontaneous author's speech.

Keywords: (markeme), author, associated field, semantics, psycholinguistic experiment.

ВВЕДЕНИЕ

Цель статьи – выяснить, в какой мере ключевые слова (маркемы), выявляемые в результате автоматической обработки авторского корпуса текстов [1–2], отличаются по своему поведению в обыденной (спонтанной) речи автора от «обычных» слов. Главный критерий, который используется при таком сопоставлении и который как раз предполагает программную реализацию, – сравнение семантической активности слов, а именно – их ассоциативного потенциала. Вспомогательный инструмент – психолингвистические эксперименты с участием автора.

Заметим, что выделение ключевых слов и словосочетаний – одна их актуальных задач современной науки, касается ли это моделирования лингвистической среды, информационного поиска, работы с текстовыми ба-

зами данных, построения высокоуровневого описания содержания документа, решения задач автоматического реферирования и т. д. Это первый из этапов на пути определения семантической близости ключевых слов и формализованного построения поля знаний, а в конечном счете – компьютерных исследований семантики текста [1].

АЛГОРИТМ ПОСТРОЕНИЯ АССОЦИАТИВНЫХ ПОЛЕЙ ЭКСПЕРИМЕНТАЛЬНЫХ СЛОВ

В качестве исследуемых слов при проведении психолингвистических экспериментов выступают – пятьдесят на пятьдесят – авторские маркемы вместе с «фоновыми» («обычными») лексемами. В серию экспериментов входят следующие способы работы с этими словами [4–6]:

- 1) свободный ассоциативный эксперимент;
- 2) эксперимент на свободные дефиниции;

3) направленный ассоциативный эксперимент (вопросы-ограничители – «какой?» и «что делает?»);

4) ряд экспериментов, предполагающих разбиение всех исходных слов на пары по признаку близости значения (симиляры), а также составление пар противоположных по значению лексем (оппозитов).

Обратимся к алгоритму построения ассоциативных полей экспериментальных слов [5, 6]. При создании такого поля используются ответы реципиента, полученные в ходе указанных психолингвистических экспериментов. При этом необходимо учитывать, что для построения поля актуальны только имена существительные, поэтому будут рассмотрены те реакции, в которых есть существительные.

На начальном этапе построения поле удобно представить в виде таблицы, в столбцах которой слева направо располагаются слова-стимулы и слова-реакции. Среди слов-реакций могут быть такие, которые, в свою очередь, оказываются стимулами, вызывающими появление других реакций. Особо стоит оговорить, что при переходе от слова-стимула к его симиляру (оппозиту), а затем от этой ставшей уже стимулом реакции к ее симиляру (оппозиту) мы возвращаемся к начальному слову-стимулу, и процесс построения ассоциативной цепи должен начинаться снова. Поэтому такой – зеркальный, бесконечный – вариант развертывания ассоциативного поля исключается.

Те лексемы, которые будут давать продолжения, мы предлагаем называть «словами-переключателями» (или «переключателями»). Если на некотором шаге, в одном и том же столбце возникает несколько одинаковых слов, то дальнейшее наращивание поля происходит от любого из них, но при этом фиксируется общее количество таких слов. Если же слово-стимул дает несколько реакций-переключателей, то они размещаются в разных строках одного столбца, и от каждого из них продолжается построения поля. Развертывание отдельных ассоциативных цепей завершается, если на очередном шаге вновь появится лексема, которая уже встречалась ранее, в предыдущих столбцах. Такие лексемы не могут иметь продолжения, и мы будем называть их «словами-замыкателями» (или «замыкателями»). Таким образом, построение ассоциативного поля может быть завершено в одном из случаев:

1) если на некотором шаге возникнет исходная лексема, при этом необходимо «дотянуть» все ассоциативные цепи до того шага, на котором исходная лексема появилась в качестве замыкателя;

2) если среди всех слов-реакций на некотором шаге будут только замыкатели и / или лексемы, не имеющие продолжения.

В качестве примера приведено ассоциативное поле для экспериментального слова «место», представленное в виде табл.

После построения ассоциативного поля для некоторого слова из экспериментально-

Таблица

Ассоциативное поле для лексемы «место»

воспоминание	музыка	гармония	
		Бог, гармония, высота	
		свобода	воздух
		распутница	
	музыка		
	улыбка	ребёнок	
		печаль	воспоминание
			песня
			разлука
			душа

го списка возможно количественно оценить следующие параметры (характеристики) данного поля:

- 1) количество слов, использованных в качестве переключателей (и количество употреблений таких слов);
- 2) количество слов, использованных в качестве замыкателей (и количество употреблений таких слов);
- 3) доля маркем среди слов-переключателей (и доля употреблений маркем);
- 4) доля маркем среди слов-замыкателей (и доля употреблений маркем);
- 5) суммарное количество слов, использованных в качестве переключателей и замыкателей (и количество употреблений таких слов);
- 6) доля маркем в сумме слов-переключателей и слов-замыкателей (и доля употреблений маркем);
- 7) количество столбцов и строк в таблице соответствующего ассоциативного поля (количество шагов и ступеней).

Кроме того, для измерения семантической активности исследуемых слов необходимо учитывать для каждой лексемы из экспериментального списка, сколько раз она выступала в качестве переключателя, в качестве замыкателя и в качестве переключателя и замыкателя в сумме.

Как показывает анализ данных, полученных в ходе психолингвистических экспериментов с двумя воронежскими поэтами – Г. Умывакиной и П. Синёвой, именно этот последний параметр при сравнении ассоциативного потенциала слов обладает наибольшей различительной силой и позволяет количественно оценить соответствующие свойства слов.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Разработанная программа позволяет:

- 1) загружать экспериментальные данные автора (список исследуемых слов – стимулов и реакций);
- 2) рассчитывать описанные выше параметры ассоциативного поля и ассоциативного потенциала слова;

3) определять корреляцию между условными рангами маркем и частотой этих слов в качестве переключателей, замыкателей и переключателей и замыкателей в сумме;

4) сохранять результаты в таблицы Microsoft Excel, причем как по каждому исследуемому слову в отдельности, так и итоговые результаты, средние величины – по всем словам:

- среднее количество переключателей для ассоциативных полей маркем;
- среднее количество замыкателей для ассоциативных полей маркем;
- среднее количество переключателей и замыкателей для ассоциативных полей маркем;
- среднее количество переключателей для ассоциативных полей не-маркем;
- среднее количество замыкателей для ассоциативных полей не-маркем;
- среднее количество переключателей и замыкателей для ассоциативных полей не-маркем;
- среднее количество маркем среди переключателей для ассоциативных полей маркем;
- среднее количество маркем среди замыкателей для ассоциативных полей маркем;
- среднее количество маркем среди переключателей и замыкателей для ассоциативных полей маркем;
- среднее количество маркем среди переключателей для ассоциативных полей не-маркем;
- среднее количество маркем среди замыкателей для ассоциативных полей не-маркем;
- среднее количество маркем среди переключателей и замыкателей для ассоциативных полей не-маркем.

Пример работы программы после построения ассоциативных полей для всех слов экспериментального списка показан на рис. 1, 2 и 3.

Примеры отображения результатов в таблицы Microsoft Excel представлены на рис. 4 и 5.

ЗАКЛЮЧЕНИЕ

Построение ассоциативного поля экспериментальных слов на пока еще ограниченном экспериментальном материале позволяет прийти к одному предварительному общему

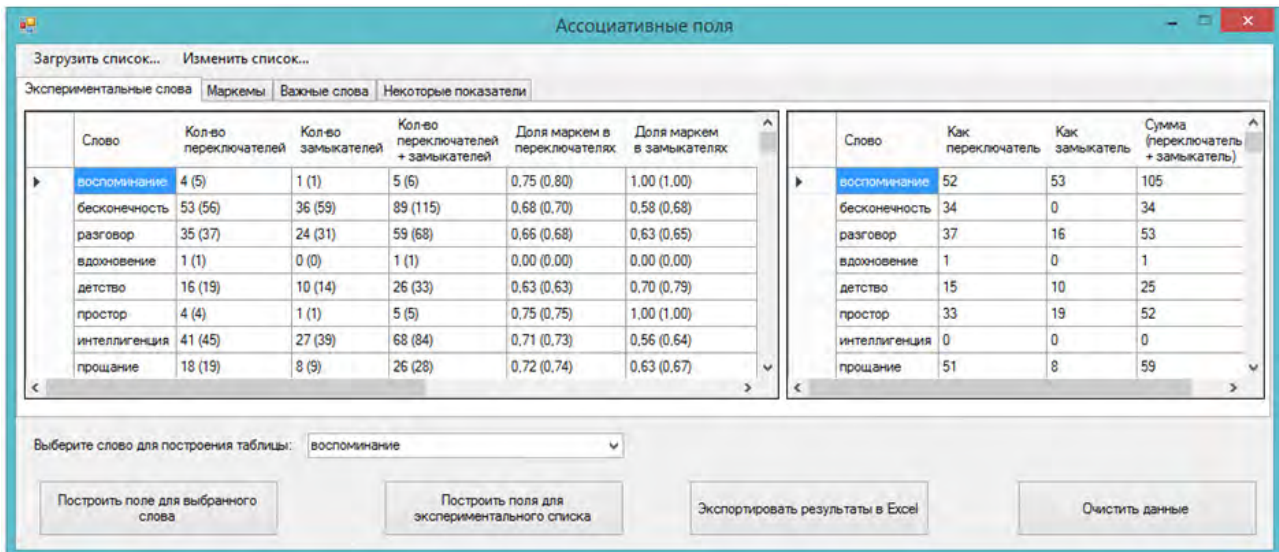


Рис. 1. Окно программы – вкладка «Экспериментальные слова»

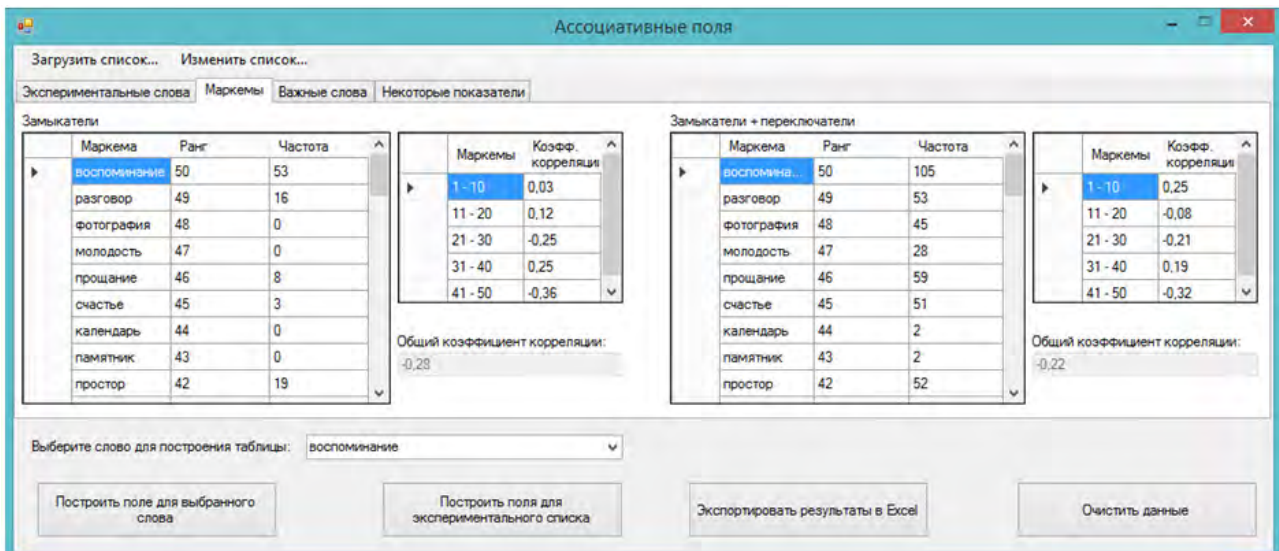


Рис. 2. Окно программы – вкладка «Маркеры»

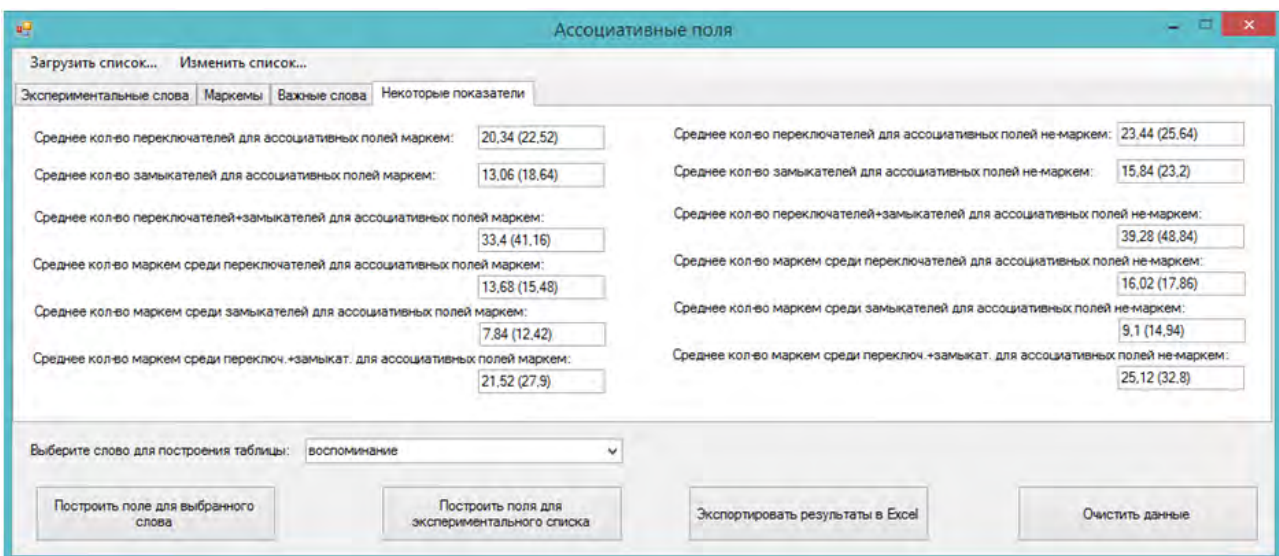


Рис. 3. Окно программы – вкладка «Некоторые показатели»

Сравнительный анализ активности ключевых слов: автоматизированный подход

Слово	Кол-во переключателей	Кол-во замыкателей	Кол-во переключателей+замыкателей	Доля маркем в переключателях	Доля маркем в замыкателях	Доля маркем в переключателях+замыкателях
воспоминание	4 (5)	1 (1)	5 (6)	0,75 (0,80)	1,00 (1,00)	0,80 (0,83)
бесконечность	53 (56)	36 (59)	89 (115)	0,68 (0,70)	0,58 (0,68)	0,64 (0,69)
разговор	35 (37)	24 (31)	59 (68)	0,66 (0,68)	0,63 (0,65)	0,64 (0,66)
вдохновение	1 (1)	0 (0)	1 (1)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
детство	16 (19)	10 (14)	26 (33)	0,63 (0,63)	0,70 (0,79)	0,65 (0,70)
простор	4 (4)	1 (1)	5 (5)	0,75 (0,75)	1,00 (1,00)	0,80 (0,80)
интеллигенция	41 (45)	27 (39)	68 (84)	0,71 (0,73)	0,56 (0,64)	0,65 (0,69)
прощание	18 (19)	8 (9)	26 (28)	0,72 (0,74)	0,63 (0,67)	0,69 (0,71)
счастье	30 (31)	21 (25)	51 (56)	0,67 (0,68)	0,57 (0,60)	0,63 (0,64)
легкомыслие	29 (30)	18 (22)	47 (52)	0,62 (0,63)	0,67 (0,73)	0,64 (0,67)
праздник	50 (58)	36 (57)	86 (115)	0,68 (0,69)	0,56 (0,65)	0,63 (0,67)
фотография	37 (42)	24 (35)	61 (77)	0,65 (0,64)	0,58 (0,66)	0,62 (0,65)
поединок	0 (0)	0 (0)	0 (0)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
календарь	55 (61)	36 (55)	91 (116)	0,65 (0,67)	0,53 (0,64)	0,60 (0,66)
позвоночник	28 (29)	18 (22)	46 (51)	0,64 (0,66)	0,67 (0,73)	0,65 (0,69)
молодость	44 (50)	31 (45)	75 (95)	0,64 (0,66)	0,58 (0,64)	0,61 (0,65)
человек	10 (10)	2 (2)	12 (12)	0,50 (0,50)	0,50 (0,50)	0,50 (0,50)
торжество	46 (50)	34 (51)	80 (101)	0,72 (0,72)	0,53 (0,63)	0,64 (0,67)
улыбка	30 (32)	23 (34)	53 (66)	0,73 (0,75)	0,52 (0,59)	0,64 (0,67)
памятник	29 (35)	15 (25)	44 (60)	0,66 (0,69)	0,80 (0,88)	0,70 (0,77)
монументальность	33 (34)	23 (31)	56 (65)	0,73 (0,74)	0,52 (0,58)	0,64 (0,66)
сердце	26 (29)	19 (23)	45 (52)	0,69 (0,69)	0,68 (0,70)	0,69 (0,69)

Рис. 4. Результат вывода полученных результатов в MS Excel

место	пространство	равнина, горизонт, небо	природа					
		страна	родина					
			место					
		мгновение	капля, безвременье					
			стежок					
	пространство							
	поворот	дорога	мотив, литература					
			пространство					
			распутница					
		судьба	жизнь					
			терпение					
			равновесие					
		прогресс	эволюция					

Рис. 5. Результат вывода полученного ассоциативного поля в MS Excel

выводу (подробному лингво-семантическому анализу полученных данных будет посвящена специальная работа). В спонтанной речи автора наибольшую семантическую активность действительно проявляют маркемы, но не все, а менее половины из них. При этом, с одной стороны, активность «эффективных» маркем не коррелирует с их рангом, а с другой стороны, они порождают ассоциативные

поля, имеющие близкие параметры. Поэтому мы вполне можем говорить о наличии отчетливого кластера таких маркем. И этот результат ставит нас перед очевидной задачей – объяснения того, почему лишь определенная часть ключевых слов наделена повышенным ассоциативным потенциалом, отличающим их от «обычных» лексем. Поиск ответа на этот вопрос ведет нас от психолингвистиче-

ских экспериментов обратно – к исследованию авторского корпуса текстов, к попыткам обнаружить особые семантические приметы тех маркем, которые в спонтанной речи автора оказываются наиболее активными.

СПИСОК ЛИТЕРАТУРЫ

1. Кретов А. А. Метод формального выделения тематически нейтральной лексики (на примере старославянских текстов) / А. А. Кретов // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2007. – № 1. – С. 81–90.

2. Кретов А. А. Понятие маркемы: методика выявления и практика использования / А. А. Кретов // Универсалии русской литературы 2. / Под ред. А. А. Фаустова. – Воронеж: НАУКА-ЮНИПРЕСС, 2010. – С.138–153.

3. Воронина И. Е. Алгоритмы определения семантической близости ключевых слов по их окружению в тексте / И. В. Попова, И. Е. Воронина, А. А. Кретов // Вестник Воронежского государственного университета. Серия Системный анализ и информационные технологии. – 2010. – № 1. – С. 148–153.

Фаустов Андрей Анатольевич – д-р филол. наук, профессор, зав. каф. русской литературы, филологический факультет ВГУ
E-mail: aafaustov@list.ru

Воронина Ирина Евгеньевна – д-р техн. наук, профессор кафедры программного обеспечения и администрирования информационных систем факультета прикладной математики, информатики и механики
E-mail: irina.voronina@gmail.com

Розенфельд Марьяна Яковлевна – канд. филол. наук, доцент кафедры общего языкознания и стилистики, филологический факультет ВГУ.
E-mail: maryanka.08@mail.ru

Свиридова Наталья Юрьевна – студентка 4 курса кафедры программного обеспечения и администрирования информационных систем факультета прикладной математики, информатики и механики
E-mail: sviridovanatalia.may@gmail.com

4. Фаустов А. А. О субъектно-рефлексивном анализе литературного текста. Эпизод II. Маркемы в индивидуальном лексиконе Г. М. Умывакиной: общий обзор / А. А. Фаустов, М. Я. Розенфельд // Вестник Воронежского государственного университета. Серия Лингвистика и межкультурная коммуникация. – 2013. – № 2. – С. 95–100.

5. Фаустов А. А. Алгоритм построения ассоциативного поля слов (по результатам психолингвистических экспериментов с ключевыми словами автора) / А. А. Фаустов, М. Я. Розенфельд, Н. Ю. Свиридова, И. Е. Воронина // Компьютерная лингвистика: XVI международная конференция «Информатика: проблемы, методология, технологии», (Воронеж, 11–12 февраля 2016 г.). – Воронеж, 2016. – С. 240–245.

6. Фаустов А. А. О субъектно-рефлексивном анализе литературного текста. Эпизод II. 2. Поведение маркем в индивидуальном лексиконе Г. М. Умывакиной / А. А. Фаустов, М. Я. Розенфельд // Универсалии русской литературы. 5 / Под ред. А. А. Фаустова. – Воронеж: ООО ИПЦ «Научная книга», 2013. – С. 77–590.

Faustov Andrey A. – Dr. habil. of Philology, Professor, Head of the Russian Literature Department, Faculty of Philology.
E-mail: aafaustov@list.ru

Voronina Irina E. – Professor of Software Development and Information Systems Administration Department of Applied Mathematics, Informatics and Mechanics Faculty, PhD, Dr. habil. In Technical Sciences.
E-mail: irina.voronina@gmail.com

Rozenfeld Maryanna Ya. – PhD in Philology, Lecturer of the General Linguistics and Stylistics Department, Faculty of Philology.
E-mail: maryanka.08@mail.ru

Sviridova Natalya U. – 4th year student of of Software Development and Information Systems Administration Department of Applied Mathematics, Informatics and Mechanics Faculty.
E-mail: sviridovanatalia.may@gmail.com