

знаний; когнитивное моделирование процессов оперирования знаниями; формальное описание активного и пассивного режимов работы лингвокультурного когнайзера с использованием теоретико-множественных, логических, вероятностных и графовых моделей; конструирование компьютерной системы (лингвокультурного когнайзера русского языка), интегрирующей механизмы работы языкового сознания в активном и пассивном режимах.

В рамках проекта предполагается провести прикладные опытно-конструкторские разработки и исследования: интерактивных информационных технологий проведения и представления результатов полиязычных ассоциативных экспериментов; формальное описание АВС на основе нейросетевой модели; экспериментальное исследование эффективности процессов поиска информации в Интернет и автоматического перевода на основе АВС; средств представления знаний и имитации рассуждений в автоматизированных системах контроля знаний; макетов полиязычных лингвокультурных тезаурусов.

Предполагается реализация образовательной программы «Компьютерная лингвистика и семиотика» для подготовки магистров в МГТУ им. Н.Э.Баумана и МГУП им. Ивана Федорова, обучающихся по направлениям 230100 – «Информатика и вычислительная техника» и 230400 – «Информационные технологии и системы».

ПАРАМЕТРИЗАЦИЯ НОВОГО РУССКОГО АССОЦИАТИВНОГО ЭКСПЕРИМЕНТА РАС-2¹⁹

Г. А. Черкасова (Москва)
Институт языкознания РАН
gacherk@mail.ru

Доклад посвящен рассмотрению первых результатов новой серии начатых в 2008 г. ассоциативных опросов, по материалам которых планируется создать *Русский региональный ассоциативный тезаурус* и подготовить *Русский ассоциативный словарь 2 (РАС-2)*.

1. Основные параметры 1-го этапа: а) 1000 слов-стимулов (903 из РАС-1); б) ассоциации записывались в бумажных анкетах, содержащих по 100 слов, одновременно в европейской части России, на Урале и в Сибири; в) на стимулы в каждом регионе получено ≈ 500 реакций, т. е. опрошено более 15 тыс. студентов.

¹⁹ Работа выполнена при финансовой поддержке грантов РГНФ № 12-04-12039в и 12-04-12059в и гранта Президента РФ № НШ-3661.2012.6.

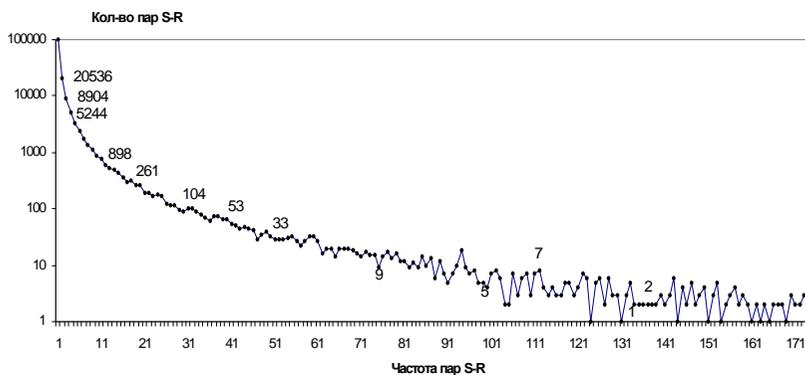
В докладе анализируются данные, полученные при опросе в европейской части России. Из 5000 напечатанных анкет было не заполнено ≈ 250 , и почти 500 номеров были опрошены дважды. Количество анкет, заполненных мужчинами – 2579, женщинами – 2624.

2. После ввода информации из анкет и удаления «сомнительных» реакций в исходной базе данных сформировано 520890 пар «стимул–реакция» (S–R). Объединением одинаковых пар сформирована БД словаря РАС-2, содержащая 170321 запись «стимул–реакция–количество пар» (S–R–K), т. е. исходная база уменьшилась более чем втрое, а повторяющиеся пары S–R составили $\approx 70\%$.

3. На тысячу стимулов получено 48285 разных реакций с частотой в паре от 1 до 378. Распределение количества пар S–R по частоте представлено на графике, который показывает резкое уменьшение количества пар с возрастанием частоты. Единичную частоту в тройках S–R–K имеют 117766 записей в БД РАС-2 или 69,1%. Частоту 2 имеют 20536 пар, 3 – 8904, 4 – 5244 ... 10 – 898, 20 – 261, ... 175 – 2, 176 – 1 и т. д. Наиболее частотные пары следующие:

скорый → поезд 378; **Ленин** → мавзолей 332; **иностранный** → язык 326; **неправда** → ложь 323; **тратить** → деньги 307; **внешний** → вид 281; **русский** → язык 268; **недуг** → болезнь 266; **возмещать** → ущерб 260; **скорлупа** → яйцо 256; **прислать** → письмо 247; **Москва** → столица 245; **зонтик** → дождь 242; **расческа** → волосы 237; **назвать** → имя 235; **образование** → высшее 226; **воскресенье** → выходной 222; **суббота** → выходной 220; **товарищ** → друг 217; **сокол** → птица 216; **сверло** → дрель 209; **сдавать** → экзамен 206; **открывать** → дверь 202; **женский** → пол 201; **кислый** → лимон 199; **прошлый** → год 199; **вернуться** → домой 198; **водить** → машину 198; **трава** → зеленая 197; **смысл** → жизни 195; **ответ** → вопрос 194; **решить** → задачу 193; **страна** → Россия 191; **сдать** → экзамен 189; **уверенный** → в себе 188; **надутый** → шар 187; **молодой** → человек 186; **вранье** → ложь 185; **Достоевский** → писатель 185; **общительный** → человек 185; **добро** → зло 184; **несчастный** → случай 184; **Пушкин** → поэт 184; **остаться** → в живых 183; **секрет** → тайна 183; **проходить** → мимо 182; **Чехов** → писатель 182; **написать** → письмо 181; **покинутость** → одиночество 181; **штурмовать** → крепость 181.

На множестве полученных реакций встретились все стимулы РАС-2 кроме 11 следующих: **буржуазный**, **запирать**, **запутывать**, **нарядный**, **невостребованность**, **предоставлять**, **просеивать**, **ракетный**, **религиозный**, **соучник**, **штурмовать**. В качестве ассоциатов стимулы даны 35867 раз ($\approx 7\%$).



Только 39 стимулов (4%) не имеют отказов от ответа, частота которых колеблется от 1 до 78. Половина – 506 стимулов получили не более пяти отказов, т. е. до $\approx 1\%$, до 10 «не-ответов» (или 2%) еще 197 слов, до 15 – 142, до 20 – 56, до 25 – 26. Более 25 отказов или $>5\%$ имеют 34 стимула:

пересуды и повеса – 78 ($>15\%$); **остальной** – 52 ($\approx 10\%$); **тщеславие** – 50; **вдовый** – 47 (≈ 9); **бюрократия** – 46; **лаконичный** – 45; **отчуждение** – 44; **самобытный** – 43; **надменный** – 42 ($>8\%$); **получаться** – 41; **соученик** – 41; **практичность** – 40; **оказываться** и **приходиться** – 36 ($>7\%$); **карьерист** – 35; **относиться** – 34; **непонимание** – 33; **беспечность** и **гармоничный** – 32 ($>6\%$); **интересовать**, **предприимчивый** и **эффективность** – 31; **воздвигать**, **значить**, **коммунист**, **монополия** и **покинутость** – 29; **начаться** и **показаться** – 28; **безвредный**, **деятельный**, **постоянный** и **стыдливый** – 26 (≈ 5) и т. д.

15151 реакций или 31,4% содержат пробелы, из которых 2495 ассоциатов включает два и более пробелов. Примеры самых длинных реакций: **душа** → *как музей гениальных изделий и редких вещей*; **милый** → *друг не скучай, я вернусь ты так и знай*; **хлеб** → *не ем, идет брожение в теле и в носу* и др.

4. Реакции, ассоциированные количеством стимулов до 99 и определяющие *ядро языковой личности* [Уфимцева], представлены в следующей таблице.

5. Приведем одну словарную статью на стимул, отсутствующий в РАС-1, которая хорошо характеризует отношение современного студента:

Ранг	Реакция	K_stim	Ранг	Реакция	K_stim	Ранг	Реакция	K_stim	Ранг	Реакция	K_stim
1	человек	510	25	хороший	161	49	быстро	124	73	большая	110
2	жизнь	355	26	есть	161	50	все	124	74	душа	110
3	дом	351	27	зло	160	51	путь	124	75	девушка	108
4	плохо	279	28	парень	160	52	мой	124	76	урок	108
5	деньги	269	29	много	160	53	разговор	122	77	сон	107
6	хорошо	266	30	человека	157	54	труд	122	78	красный	106
7	мир	252	31	жизни	156	55	дорога	120	79	моя	105
8	друг	244	32	отдых	152	56	ответ	120	80	язык	104
9	нет	242	33	предмет	150	57	враг	120	81	помощь	104
10	я	216	34	всегда	150	58	дома	119	82	секс	104
11	радость	203	35	умный	147	59	стол	119	83	ложь	103
12	работа	195	36	люди	147	60	лес	117	84	фильм	103
13	любовь	187	37	страх	145	61	красивый	116	85	вид	102
14	время	184	38	взгляд	141	62	людей	116	86	мужик	102
15	мужчина	182	39	ум	140	63	добрый	115	87	интерес	102
16	день	180	40	характер	137	64	долг	115	88	слово	101
17	сила	180	41	боль	136	65	вопрос	114	89	книга	101
18	большой	177	42	поступок	132	66	мальчик	114	90	сильный	101
19	ребенок	172	43	дурак	132	67	свобода	114	91	долго	101
20	счастье	172	44	машина	128	68	власть	112	92	успех	101
21	город	167	45	да	127	69	красота	111	93	движение	101
22	смерть	167	46	плохой	126	70	народ	111	94	мозг	100
23	свет	165	47	ужас	126	71	студент	111	95	друга	99
24	дело	162	48	учеба	125	72	вода	110	96	далеко	99

ЧИНОВНИКИ: бюрократия **26**; власть **23**; бюрократы, коррупция **19**; взяточники **15**; взятки **14**; взятка, воры **13**; деньги **10**; люди, политика **9**; злые **8**; богатые, бюрократ, государство, правительство **7**; депутаты, уроды **6**; гады, Дума, козлы **5**; администрация, вор, дураки, жадные, продажные **4**; Гоголь, Госдума, зло, лживые, ложь, начальники, олигархи, плохие, сволочи, суки, толстые **3**; бандиты, беспредел, богачи, в думе, воруют, враги, госслужащие, грязь, депутат, документы, Жириновский, закон, мрази, мужчина, мэрия, начальник, обманщики, подлецы, разные, решили, служащие **2**; ад, аморальны, аппарат власти, банкроты, безала-

берные, бездельники, бездушные, беспредельность, богатое, богатство, богатый, большие, бумага, бумаги, буржуи, буржуй, было, в костюмах, в политике, в шляпах, важность, вверху, вера, верх, взяточничество, виновники, властвуют, власти, воровство, ворье, враг, вредные, врут, все, высокий, высокопоставленный, высота, высшее, глупые, города, господа, государства, государственная, государственный деятель, грубость, губернаторы, гуляют, дармоеды, деятели, дипломат, директор, добрались, дяди, есть, жадность, жалуются, ждать, жида, жрецы, жулики, жулье, за все, заелись, зажиточный, запрещают, заседают, звери, значимые люди, идиоты, инструкция, история, их много, кабинет, казны, карьеристы, клерки, коммунисты, корыстные, костюм, кража, круть, куплены, лгуны, лжецы, лживый, лидеры, листы, математика, мелкие, могучие, молодые, монтажники, Москвы, мужчины, мэр, мясо, наглые, наелись, налог, напевы, начальство, не врут, некомпетентность, несправедливость, несправедливы, несправедливые, обаятельные, обеспеч. люди, обман, объективность, ограниченность, ожиревшие, олигарх, Они везде, отдел, отмывание денег, офицер, партия, переехали, плохие люди, плохо, подлые, подлый, политики, полиция, полный мужчина, потеря, правитель, предатели, председатель, презрение, продолжение, птицы, пузатые, работники, рабочие, рабочий, разбойники, роман, с мигалкой, свинья, сволочь, сидят, скоты, слабые, служащий, солдаты, солидные, сослужащие, спекуляция, справедливость, среднего звена, статус, стена, стражи порядка, странные, тварь, тоже люди, толстый, тупые, тщеславие, ублюдки, ублюдок, ужас, управлять, хамы, хачи, хитрые, человеку, черная икра, черный, чинуши, чинят **1; 517+232+17+172**