

Г.А.Черкасова.

Квантитативные исследования АВС¹

Составной частью «Информационной системы когнитивных экспериментов — ИСКЭ» является ассоциативный тезаурус (русский ассоциативный тезаурус) представляющий ассоциативно-вербальную сеть АВС, построенную на основе русского ассоциативного словаря (РАС). Для разработки формальных моделей когнайзера весьма важны статистические характеристики распределений слов-реакций на слова-стимулы, а также другие параметры сети. Исходная версия тезауруса ИСКЭ была получена путем «проецирования» АВС на множество слов-знаков когнем, т.е. в ней был выделен фрагмент, в состав которого были включены все слова-стимулы и слова-реакции эквивалентные словам-знакам (словоформам-знакам) базы данных когнем. Изначально тезаурус составлял проекцию 6000 слов-знаков. По мере увеличения базы когнем и списка разных слов-знаков, возникает необходимость расширения экспериментального фрагмента АВС. В итоге экспериментальный фрагмент достигнет объема всей АВС. Настоящий материал (опубликованная ранее статья) содержит практически все статистические характеристики, необходимые для моделирования когнайзера.

Одной из эффективных форм представления результатов ассоциативных вербальных экспериментов в современной психолингвистике являются ассоциативные словари-тезаурусы. Традиционно в ассоциативных словарях полно фиксируются данные о вербальной номенклатуре эксперимента в виде двух основных отношений: “прямого” — стимул-реакция (S–R) и “обратного” — реакция-стимул (R–S). Количественные характеристики в них представлены частично — это абсолютные значения индивидуальных частот реакций и их количество для конкретных стимулов. Применение только этих данных существенно ограничивает продуктивность исследований и обоснованность выводов. Для более полного и корректного использования результатов экспериментов необходимы дополнительные исследования материалов ассоциативных словарей, в числе которых так называемые квантитативные исследования.

Традиционно предметом квантитативных исследований в лингвистике является речь в различных формах своего существования. Наибольшее распространение получил анализ текстов различных языков, стилей и авторов, целями которого являются:

— построение эмпирических функций распределения языковых единиц, идентификация их теоретическими распределениями и на этой основе определение “законов” построения текстов, т.е. “сжатия” массивов текстовых данных до формулы или нескольких чисел с некоторой лингвистической интерпретацией;

— прогнозирование на основе интерполяции изменений словарного запаса, характеристик текстов и других языковых явлений;

— сравнение между собой различных текстов или текстовых выборок на основе модельного (статистического) описания, выявление “лексической концентрации или словарного богатства” текстов и др.

В статье представлены результаты квантитативных (статистических) исследований данных компьютерных версий Русского [РАС] и Славянского [САС] ассоциативных словарей. Целью исследований было построение эмпирических частотных распределений и нахождение значений основных статистик для множеств стимулов, реакций, пар “стимул-реакция” и анкет.

¹ Повторная, дополненная новыми данными версия статьи: *Черкасова Г.А.* Квантитативные исследования ассоциативных словарей // Общение. Языковое сознание. Межкультурная коммуникация. Сб. статей. / Под ред. Н.В. Уфимцевой, Т.Н. Ушаковой. М.-Калуга, 2005. с. 308-318.

Электронная публикация подготовлена при поддержке РФНФ, грант № 12-04-12039 «Информационная система когнитивных экспериментов — ИСКЭ»

РАС сформирован по результатам трех этапного ассоциативного эксперимента. При его проведении авторы, исходя из имеющихся организационных и финансовых возможностей, стремились получить на каждый стимул не менее 100 реакций испытуемых. Анкеты генерировались программно так, чтобы: во-первых, в каждой из них было сто стимулов, случайно выбранных из заданного списка; во-вторых, во всем множестве полученных анкет каждый стимул встречался не менее ста раз [Черкасова 1994].

Вместе с тем на первом этапе эксперимента была предпринята попытка значительно увеличить количество реакций на предъявляемые стимулы. Для этого по 760 стимулам был проведен дополнительный опрос, который увеличил количество реакций на них со 100 до 500. В итоге результаты первого этапа русского ассоциативного эксперимента можно разделить на следующие части:

- 517 стимулов по 100 реакций;
- 756 стимулов, имеющих по 500 реакций (этап 1-д);
- 4 стимула по 1000 реакций².

В таблице 1 приведены сводные количественные характеристики русского ассоциативного словаря (см. также работы автора, в том числе [Черкасова 2004б]).

Таблица 1. Сводные статистические характеристики РАС

Этапы	РАС	1-й	1-д	2-й	3-й
1. Кол-во стимулов S	6624	1277	756	2685	2935
2. Кол-во анкет	10412	4522		2820	3070
3. Всего пар S–R	1037522	451377	399409	281246	304899
4. Разных пар S–R	462531	170750	139903	152598	145354
5. Разных реакций R	103051	52349	45640	50489	46948
6. Добавлено новых R	103051	52349	45640	30938	19764

В отличие от русского славянский ассоциативный эксперимент проводился на списке стимулов из 112 слов, которые присутствовали в каждой анкете, а количество полученных реакций по каждому стимулу не менее 500. В настоящее время обработаны анкеты и созданы базы данных (БД) для болгарского, белорусского, русского и украинского языков. Данные САС приведены в таблице 2.

Язык	русский	болгарский	белорусский	украинский
1. Кол-во стимулов S	112	113	112	112
2. Всего анкет	594	584	643	478
3. Всего пар S–R	65936	65588	71897	53592
4. Разных пар S–R	20152	24819	25222	20045
5. Разных реакций R	10324	11409	13011	10291

Таблица 2. Сводные статистические характеристики САС

В качестве материалов для исследований, представленных в настоящей статье, использованы данные САС по четырем языкам и данные только первого этапа РАС, при этом для сопоставления словарей было использовано подмножество 756 стимулов (1-д) и соответствующие им множества реакций. Такой выбор материалов позволил обеспечить корректное сопоставление словарей, которые были получены в результате экспериментов, различных по количеству участвовавших в них испытуемых.

² При обработке результатов опроса, выяснилось, что четыре стимула: **нести, спина, серьезный, старый** в исходном списке случайно опрашивались дважды, и на них было получено 1000 реакций.

Проведенные исследования условно можно разделить на две группы: а) анализ динамики роста словников реакций ассоциативных словарей; б) влияние количества респондентов на состав и структуру словарных статей Прямого ассоциативного словаря.

Исследование динамики роста словников

Анализ динамики роста словников ассоциативных словарей проводился по следующим направлениям: построение функций наполнения словников, параметризация и прогноз роста объема словников, сопоставительные исследования разных экспериментов.

Уточним используемое в исследовании понимание “словников” в ассоциативных словарях.

Словник стимулов (Ssl) – это множество всех стимулов (слов и словосочетаний), используемых при опросе и включенных в анкеты.

Словником реакций (Rsl) будем называть множество разных ответов, т.е. реакций, полученных в ассоциативном эксперименте. В таблицах 1 и 2 в п. 5 приведены размерности данных словников.

Ассоциативным словником (Asl) будем называть объединение словника стимулов и словника реакций.

Так как все стимулы ассоциативных экспериментов встречаются в реакциях испытуемых, то далее будем исследовать только словники реакций. Посмотрим, как изменяется количество разных ответов испытуемых в процессе ассоциативного эксперимента, сравним данные разных опросов и проанализируем, что дает одна анкета, один стимул, массив всех анкет и всех стимулов.

1. Количество разных ответов-реакций в одной анкете

В каждой анкете РАС содержится разный набор стимулов, и нет двух одинаковых. После обработки массива 4522 анкет первого этапа получено эмпирическое распределение количества разных ответов в одной анкете, полигон которого представлен на рисунке 1. Кроме полигона на рисунке приведена сглаживающая кривая (пунктирная линия). Значения основных статистик полученного распределения следующие: минимальное значение количества разных реакций – 65; максимальное значение – 100; среднее значение – ~94 (точнее 93.81); мода – 97; среднее отклонение – 24,42; асимметрия – -1,54; эксцесс – 3,60.

Суммарное количество разных ответов по всем анкетам составляет 424 194. Количество анкет, в которых разных ответов минимальное (менее 79) всего – 77, что составляет 1,7 % от общего количества опрошенных. А максимальное число разных ответов 100, т.е. все реакции разные, содержится в 212 анкетах или 4,69 %. Большинство анкет включают от 94 до 98 разных ответов (см. данные рис. 1).



Кол-во разных реакций	Число анкет	Кол-во разных реакций	Число анкет
≤ 79	77	90	182
80	25	91	223
81	23	92	272
82	27	93	294
83	37	94	420
84	48	95	441
85	65	96	498
86	69	97	515
87	91	98	445
88	103	99	326
89	129	100	212

Рисунок 1.

Распределения количества разных реакций в одной анкете для каждого из четырех языков САС в виде полигонов представлены на рисунке 2. Там же приведена и сглаживающая кривая (пунктирная линия).

Сопоставление показывает, что характер распределений (сглаживающих кривых) и количественные показатели для всех четырех славянских языков близки друг к другу, и это, возможно, объясняется тем, что список стимулов во всех экспериментах был один и тот же, а в каждой анкете содержались все стимулы. Анализируя графики, можно отметить, что наибольшее число анкет содержит 95–106 разных реакций. Во всех языках в конечных координатах (точки 112-113) кривые стремятся к нулю.

Рассматривая русскую часть САС, мы имеем следующие результаты: 594 анкеты дали в сумме 556147 разных ответов. Среднее количество разных реакций в одной анкете получилось ~ 98 , что составляет 87.4% от максимально возможного количества ответов 112.

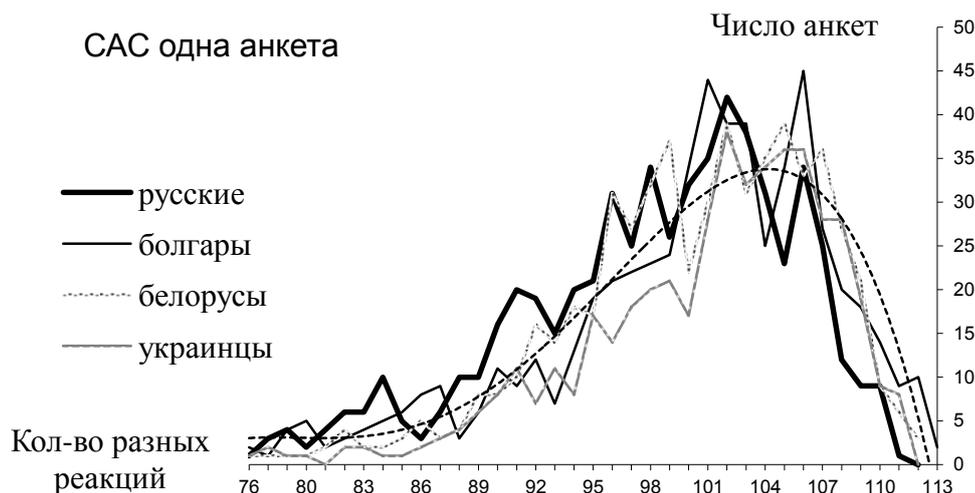


Рисунок 2.

Сопоставляя данные для РАС и САС можно отметить, что распределения количества разных ответов в обоих случаях являются асимметричными и унимодальными. Среднее количество разных реакций для САС — 87.4% от максимально возможного значения, в то время как для РАС этот же процент выше и составил 93.81%, что можно объяснить различием списков слов-стимулов в каждой анкете.

Графики САС в отличие от данных РАС заканчиваются на нулевой отметке, т.е. отсутствуют анкеты, имеющие на все стимулы разные реакции.

2. Динамика роста словника реакций в зависимости от числа испытуемых

Исследование проводилось по следующему алгоритму:

1. Задавался шаг, т.е. количество добавляемых к выборке анкет. В представленных ниже результатах $k_{\text{шаг}} = 10, 50 \text{ и } 500$. Для начальной точки отсчета $k_0 = 0$ принимаем размер словника, равный среднему значению количества разных ответов-реакций в одной анкете.
2. Из всей БД ассоциативного словаря выбирались данные из $k_i = k_0 + k_{\text{шаг}}$ анкет и формировалась очередная выборка реакций.
3. В каждой выборке вычислялось количество встретивших разных реакций Rsl_i . Данные заносились в таблицу парами $k_i - Rsl_i$.
4. Изменялось значение $k_0 = k_i$.
5. Если $k_0 + k_{\text{шаг}} < K$, где K – общее количество обработанных анкет в эксперименте, то осуществлялся переход к п. 2. В случае не- выполнения условия заканчивалось формирование таблицы.
6. По данным, записанным в таблице, вычислялись значения абсолютного прироста словника реакций ($Rsl_i - Rsl_{i-1}$), и процент относительного увеличения объема словника на каждом последующем шаге $(Rsl_i - Rsl_{i-1}) / Rsl_{i-1}$.
7. Результаты отображались в виде графиков.

Для славянского эксперимента задавался шаг в 10 анкет. Результаты расчетов по п.6 алгоритма представлены на рисунках 3–5. Графики на рис. 3 показывают, что количество разных реакций для всех четырех славянских языков возрастает по мере увеличения количества опрошенных. При количестве респондентов 500 объемы словников реакций достигают размеров от 9 до 11 тысяч.

И хотя кривые имеют тенденцию к уменьшению угла наклона, как показывают графики на рис. 4, при выборках более ста респондентов количество реакций, добавляемых на каждом шаге, колеблется в пределах 100–300 слов, т.е. каждая анкета добавляет от 10 до 30 новых реакций на 112 предъявленных стимулов.

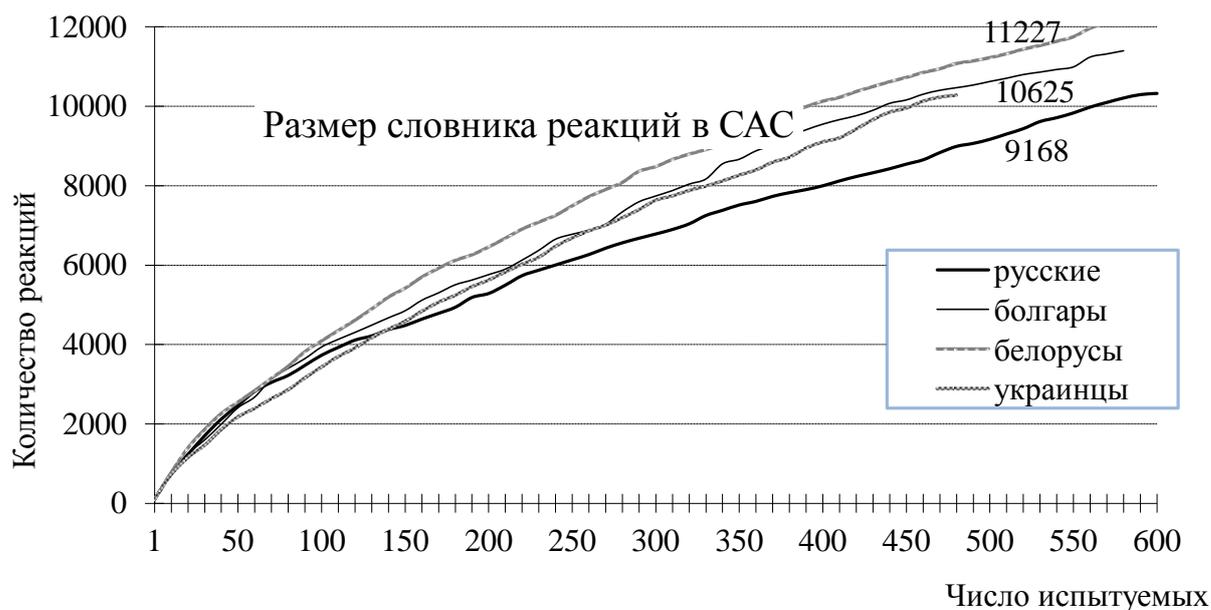


Рисунок 3.

На рисунке 5 “относительный прирост числа реакций (%)” представлено как изменяется процент количества новых реакций на очередном шаге к количеству реакций, полученных на всех предыдущих шагах. Графики сначала резко падают от 100 % до 10 % на первых 50–60 опрошенных, потом медленно уменьшаются, и при достижении выборки 450–500 респондентов относительный прирост новых реакций колеблется в пределах 1 %.



Рисунок 4.

Теперь проанализируем результаты для РАС. На рисунке 6 приведены два графика: 1) рост количества реакций для первого этапа (сплошная линия) и 2) прирост числа новых ответов (пунктирная). Количественные значения на графиках указаны с шагом в 500 анкет.

Рост словника на первом шаге составил ~15 тысяч реакций, добавление на следующих двух шагах приблизительно одинаковое и вдвое меньше: 7952 и 7113. В итоге 1500 анкет породили ~30 тысяч разных ответов. На последующих шести шагах прирост новых реакций уменьшается опять вдвое и колеблется около 3,5–4 тысяч.

Для сравнения данных РАС и русской части САС на рисунке 7 даны графики и конкретные значения объемов словников реакций для первых ста анкет с шагом 10 и для остальных с шагом 50. Анализ показывает, что вид кривых одинаков, но количественные показатели для РАС более чем в 1,5 раза выше, так как число стимулов РАС почти в 10 раз больше, чем число стимулов САС.

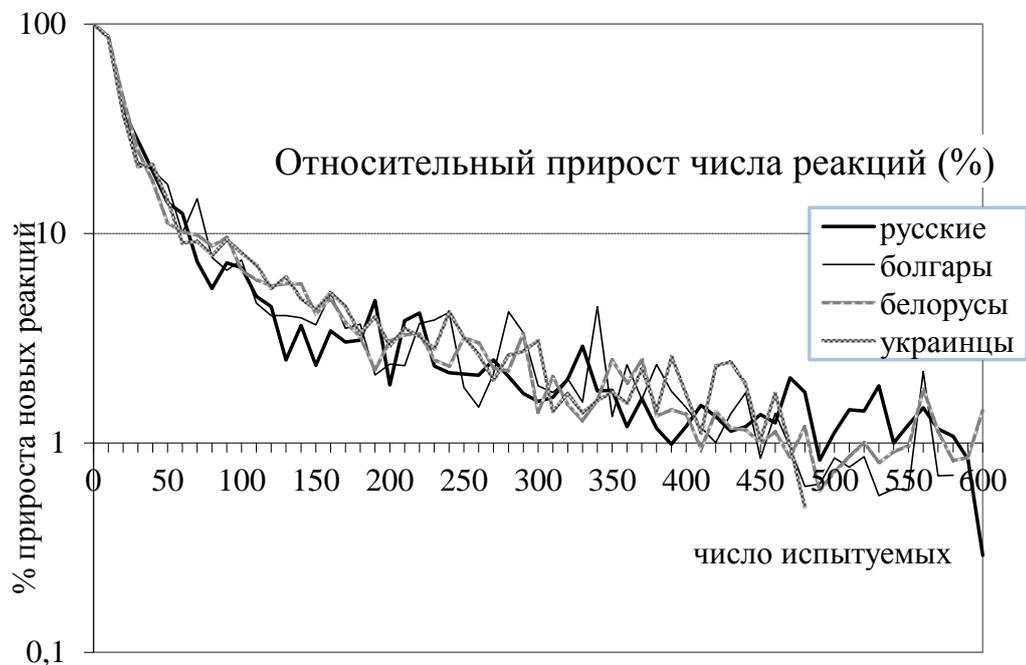
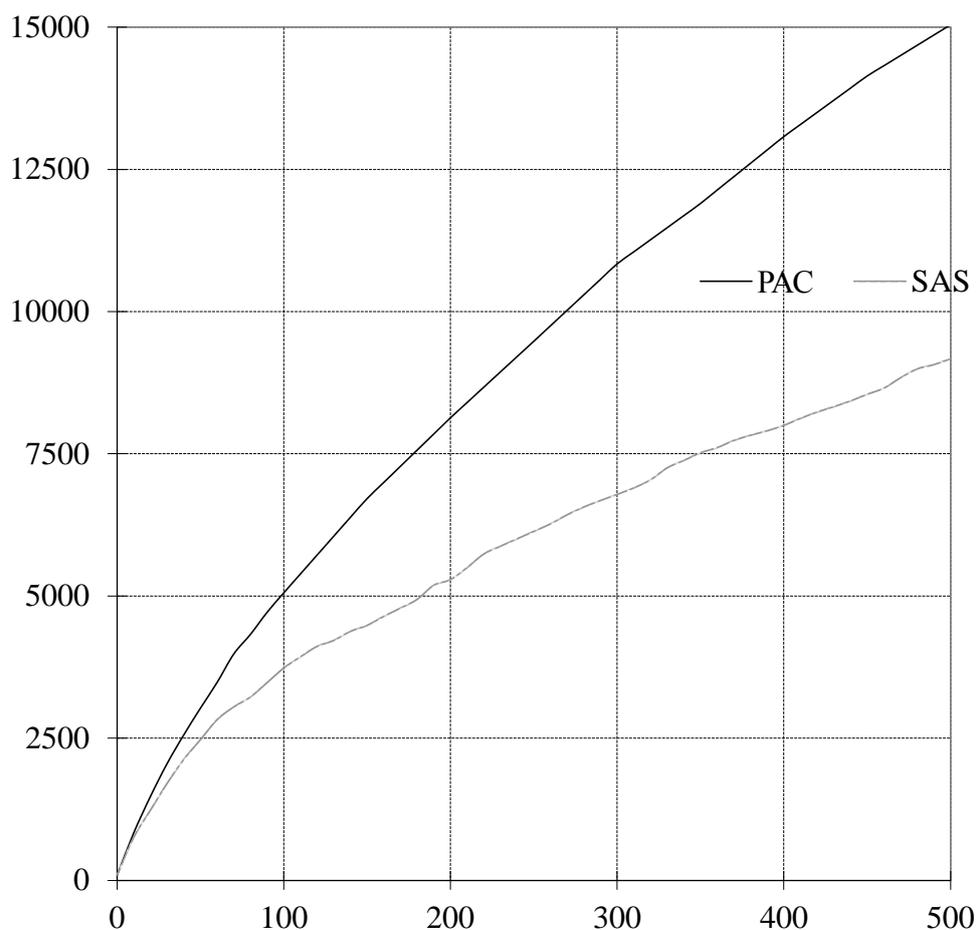


Рисунок 5.



Рисунок 6.



Количество						
анкет	реакций			анкет	реакций	
	САС	РАС			САС	РАС
1	98	94		100	3731	5056
10	756	838		150	4376	6713
20	1231	1477		200	5186	8133
30	1703	2055		250	6134	9479
40	2124	2559		300	6784	10839
50	2471	3029		350	7515	11901
60	2823	3481		400	7998	13077
70	3047	3981		450	8544	14142
80	3223	4326		500	9168	15042
90	3475	4714		600	10324	16797

Рисунок 7.

3. Динамика изменения словника реакций в зависимости от количества ответов на стимулы

Если данные, представленные выше в таблице рисунка 7, разделить на соответствующее количество анкет, то будет вычислено значение “среднего прироста”

словника реакций на одну анкету. Если полученные результаты разделить на количество стимулов в одной анкете, то вычисляется прирост на одно “стимулоупотребление” в анкете. Построенные по рассчитанным данным зависимости даны на рис. 8: для САС — это график (2), а для РАС — график (1). Кривая (2) для САС является одновременно средним приростом количества реакций на одно стимулоупотребление по анкетам и по частоте стимула, поскольку в каждой анкете встречаются все 112 стимулов.

Чтобы определить, как меняется словник реакций в зависимости от количества ответов на каждый стимул для РАС, были сделаны такие выборки из базы данных, в которых каждый стимул в среднем встречался кратное десяти число раз (10, 20, 30 ... 500), и для них подсчитано количество разных реакций. Полученные данные, разделенные на количество стимулов (т.е. 756) и соответствующие частоты встречаемости стимула, представлены на рис. 8 графиком (3). Поскольку в каждой анкете РАС встречается 100 стимулов из 756, то график (3) является графиком (1), сжатым по оси абсцисс почти в семь с половиной раз.

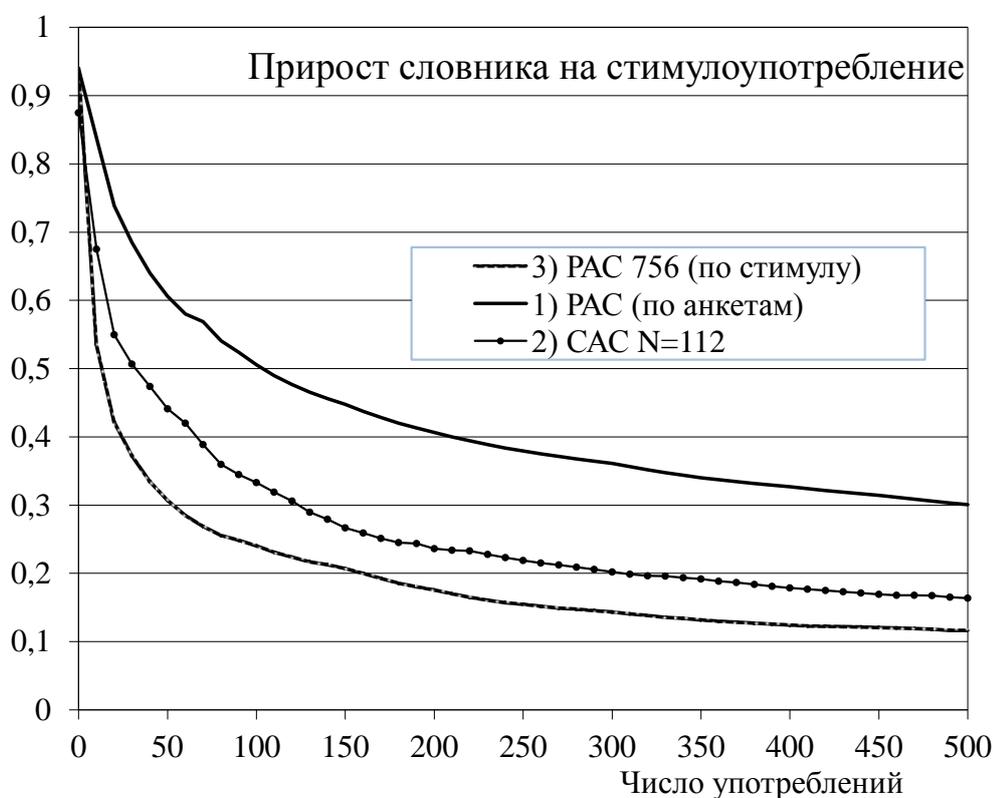


Рисунок 8.

Как показывают графики, прирост словника реакций убывает при увеличении числа опрошенных. График (2) для САС, располагается между графиками (1) и (3) для РАС. Прирост новых реакций на один стимул в САС падает от 0.875 до 0.165, принимая значение 0.3 (т.е. три новые реакции на каждые десять полученных) при 120 ответах и 0.2 при 300 опрошенных.

Для РАС величина прироста только в первой точке больше (0.94), а затем кривая резко опускается, и когда получено по 50 ответов на каждый из 756 стимулов, прирост равен 0.3, а при 150 ответах — 0.2, и опускается он до 0.116 при 500 опрошенных. Получилось, что большее количество стимулов, быстрее насыщает словник реакций R_{sl} .

Графики на рис. 8 констатируют, что прирост словника реакций зависит от количества опрошенных респондентов и от числа стимулов в ассоциативном эксперименте, т.е. и размер словника R_{sl} , является функцией двух переменных. При этом отношение размера

словника реакций к величине словника при выборке в сто респондентов не зависит от количества стимулов в опросе, поскольку как показывают графики на рисунке 9 кривые для САС по 112 стимулам и для РАС 1-д по 756 стимулам практически совпадают. Более того, графики на отрезке от 100 до 500 реакций на стимул аппроксимируются прямой линией, уравнение которой дано на рисунке 9 (справа). Получается, что рост словников и для РАС, и для САС при добавлении очередной сотни испытуемых в среднем составляет 36.4%.

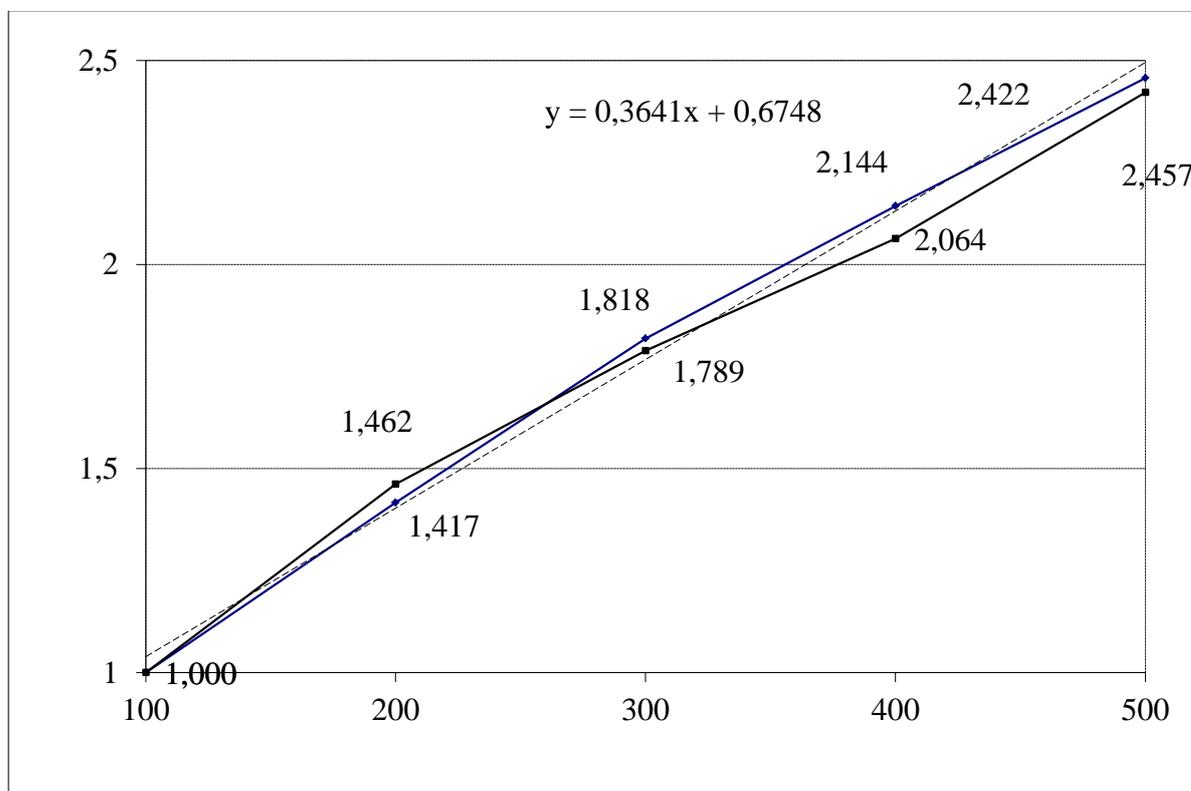
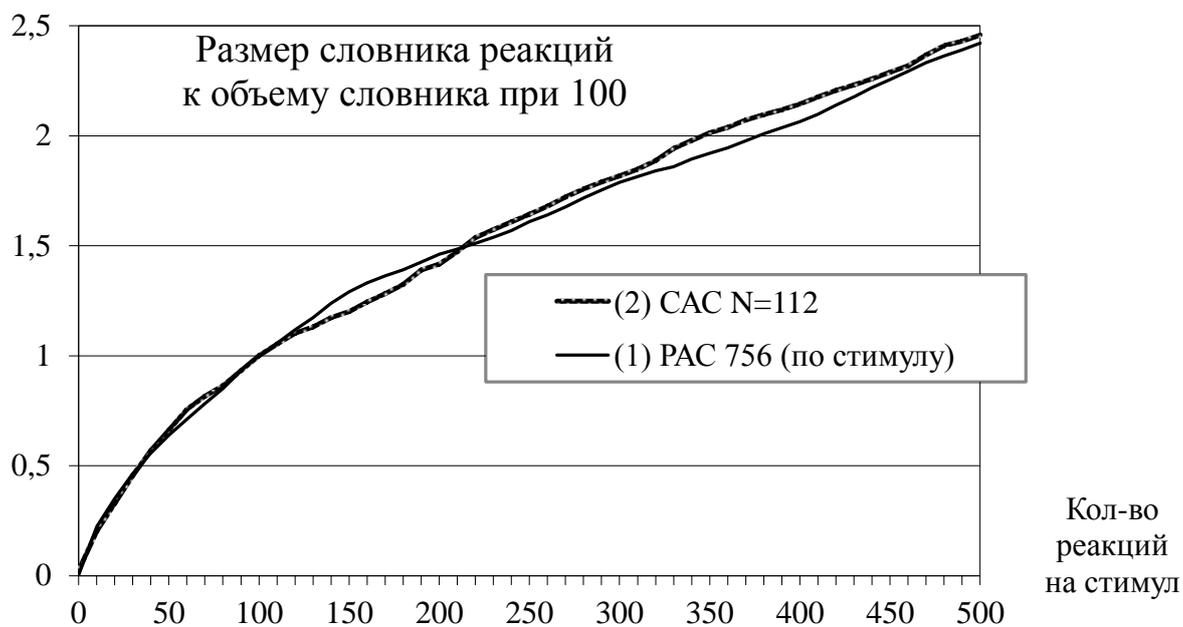


Рисунок 9.

	САС	РАС
100/100	1	1
200/100	1,417	1,462
300/100	1,818	1,789
400/100	2,144	2,064
500/100	2,457	2,422

Частота реакций в ассоциативном поле конкретного стимула при росте объема выборки

Как показано выше, рост словника реакций продолжается при увеличении числа опрашиваемых, а интерполяция графиков, представленных на рисунках 3, 7 и 9, прогнозирует насыщение словников только при увеличении количества респондентов до десятков тысяч человек. В связи с этим встают вопросы: *Какое количество респондентов нужно опросить в ассоциативном эксперименте? Что изменяется в каждой словарной статье при увеличении числа опрошенных? Насколько 100 реакций (выборка, наиболее часто используемая исследователями) достаточно для исследования “ассоциативности” слов-стимулов?*

Для ответа на вопросы, во-первых, проанализируем статьи прямого ассоциативного словаря некоторых стимулов, делая выборки; во-вторых, исследуем как изменяется состав и частоты конкретных реакций при увеличении количества испытуемых от 100 до 600, используя данные по 112 стимулам русского эксперимента в САС.

В РАС четыре стимула **нести, серьезный, старый, спина** имеют частоту предъявления больше тысячи, поэтому разобьем включающие их анкеты на выборки по 100 реакций, и проанализируем статьи прямого ассоциативного словаря этих слов по отдельным выборкам.

Начнем с глагола **НЕСТИ**. Ниже для этого стимула представлено 10 статей прямого словаря, в которых шрифтами выделены следующие классы ответов: жирным прямым — реакции, присутствующие во всех выборках, жирным курсивом — не менее чем в восьми (т.е. в 80% выборок); светлым курсивом помечены неповторяющиеся в других выборках реакции, количество которых в статистических данных указано последней цифрой³. В итоге невыделенными остались реакции, встречающиеся в двух – семи выборках.

1. **бремя** 8; **ношу** 7; **сумку** 7; **тащить** 7; **ведро, груз, крест, тяжело** 5; **тяжесть** 3; **везти, ответственность**, поклажа, **положить, сумка** 2; авоську, арбуз, **беды, бред, бремя на плечах, в люди**, вести, вещь, грести, далеко, **дать простор, делать, добро, заряд, идти, книгу, колбасу, конфеты, коробка, куда, мести, место, нагрузка, образ**, отдать, поклажу, **ползти**, портфель, продукты, **свой крест**, тяжелое, тянуть, уронить, **чемодан, чемоданы**, чепуху, **чушь**, яйца 1; **100+52+0+38+17**

2. **сумку** 18; **ношу, сумка, тяжело** 5; **крест, тяжесть** 4; **груз, чушь** 3; **ведро**, книгу, **ноша, свой крест**, сумки, **чемодан** 2; **брат, в ладони, в сарай, везти, веник**, вести, вещь, вздор, воду, **вор**, грести, **держать, держать в руках, до конца, добро, живот, знания, кладь, книги, кого, кого-нибудь, кости, курица, лабуду, мусор, обуза, ответственность, пол, пчела, рок-н-ролл, с работы, службу**, счастье, **тащить, товар, тяжел, тяжелое что-то, тяжесть жизни**, чепуху, что-то, яйца 1; **100+55+0+41+20**

3. **сумку** 18; **груз** 7; **ношу, тяжесть** 5; **крест**, сумки, **чушь** 4; **бремя** 3; **везти**, вещи, воду, **сумка, тяжело**, цветы 2; авоську, арбуз, в дом, в руках, **веру**, **весть, добычу, домой, дрова**,

³ Напомним, что означают другие цифры: первая — всего получено ответов, вторая — разных реакций, третья — отказов от ответа, четвертая — число реакций с частотой единица.

знамя, кого, корзина, *кошелки*, мужу, *мужчина*, мяч, нагрузку, наказание, науку в массы, обузу, **ответственность**, *плестись*, портфель, *принц*, продукт, радость, ребенка, рюкзак, с базы, с проходной, **свой крест**, свою ношу, *стреля*, сумка в руке, сундук, торт, **чемодан**, яйца 1; $100+52+0+38+19$

4. сумку 16; сумка 8; груз 6; ношу, **тяжесть** 4; вести, *ноша*, **ответственность** 3; **бремя**, **ведро**, портфель, радость, **чемодан** 2; арбуз, *арбузы*, вещи, внести, воду, воз, выбросить, выносить, *горести*, журнал, кара, караул, картошка, клад, кладь, **крест**, легкость, любимую, нагрузку, наказание, несун, пакет, *пальто*, плакат, поклажу, правду, сверток, свое бремя, **службу**, стипендию, стул, тебе, тетрадь, тяжести, уносить, уронить, флаг, хлеб, чепуху, что-то, ярмо 1; $100+55+2+41+20$

5. сумку 18; ношу 11; груз, сумка 6; ведро 5; **ответственность**, **тяжесть**, **чемодан** 4; **службу** 3; **везти**, портфель, радость, сумки, **тащить**, **тяжело**, тяжелое 2; авоська, бросить, вещи, взять, воду, домой, *ересь*, книги, мало, нагрузку, наказание, *ноша*, покупки, попку, посуду, потерю, продукты домой, свет, свою ношу, сетку, тепло, терпеть, *тяжелый*, цветы, **чушь** 1; $100+41+0+25+12$

6. сумку 12; груз 10; **тяжесть** 7; ношу; **ответственность** 5; **везти**, сумка 4; **бремя**, **крест**, **чушь** 3; воду, на руках, **службу**, сумки, **тяжело**, тяжести 2; авоську, **ведро**, воздух, вынести, выносить, *грузчик*, дежурство, занятость, зло, кладь, корзину, круг, мешок, на рынок, *ноша*, плащ, поклажу, помогать, портфель, принести, работа, рюкзак, **свой крест**, свою ношу, сетку, **тащить**, *топор*, цветы, **чемодан**, чепуху 1; $100+47+2+30+12$

7. сумку 17; груз, сумка 5; **тяжело**, **тяжесть** 4; **везти**, *ноша*, ношу, **ответственность**, **службу** 3; **ведро**, корзина, портфель, радость 2; **бремя**, букет, в руках, вахту, вести, весть, воду, друга, жизнь, за пазухой, караул, кому, корзину, **крест**, курица, мешок, мусор, мячик, нагрузку, несун, осторожно, отдать, переносить, поклажа, поклажу, поставить, принести, продукты, радость людям, резать, с собой, сетку, солнце, столб, счастье, **тащить**, тычинка, что, что-то, что-то в руках, **чушь** 1; $100+56+1+42+16$

8. сумку 16; ношу 11; груз 10; **тяжесть** 8; **свой крест** 5; **крест** 3; **бремя**, на руках, сумка, **чушь** 2; авоська, в дом, в массы, в подарок, вахту, **ведро**, ведро воды, весть, вещь, далеко, долг, домой, знания, клад, компот, корзину, лукошко, на горбу, *ноша*, окоlesiцу, остановка, отдать, пакет, письмо, плюнуть, повинность, поклажу, прекрасное, принести, ребенка, с собой, свою ношу, сетку, **службу**, сумки, **тащить**, **тяжело**, *тяжелый* груз 1; $100+49+0+39+18$

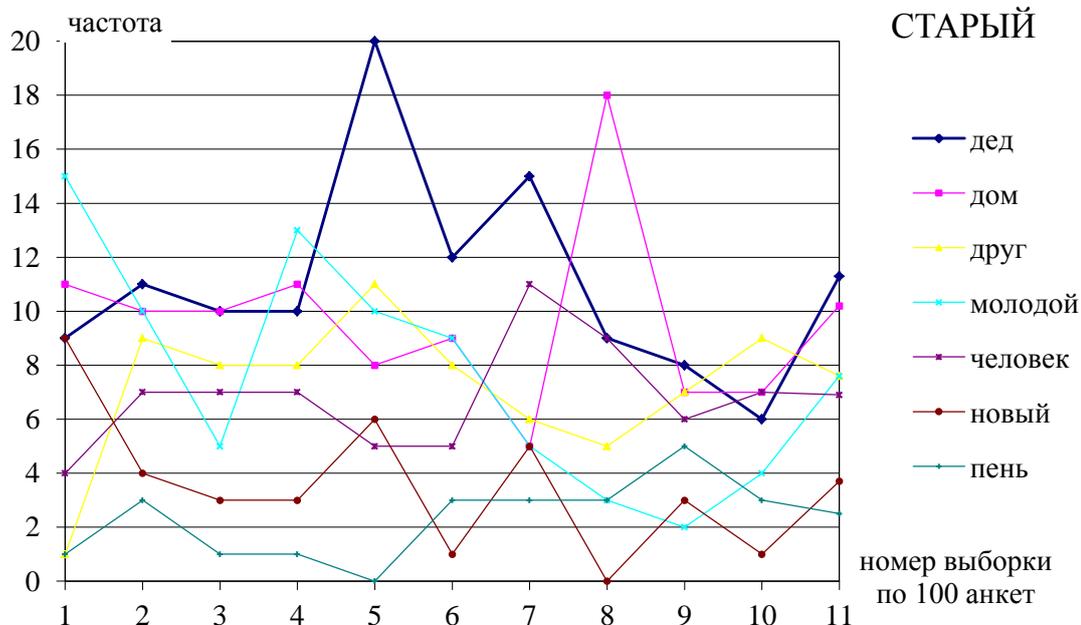
9. сумку 12; груз 9; **крест**, **ответственность** 5; ношу 4; ребенка, **тащить** 3; в массы, *ноша*, рюкзак, **службу**, сумка, *тяжелую сумку*, **тяжесть**, что-то 2; боль, бревно, **бремя**, быстро, в руке, вахту, **ведро**, **везти**, вес, вещь, грех, добро, дом, *жизненный крест*, золото, кого, мешок, наказание, новость, носилки, окоlesiцу, передвигать, печаль, плохо лежит, поднять **тяжесть**, попку, портфель, правду, решето, самому, **свой груз**, **свой крест**, свою ношу, сгорбась, сетку, сумка с продуктами, **тяжело**, тянуть, удовольствие, учить, **чемодан**, шоколад, яйца 1; $100+58+0+43+23$

10. сумку 10; груз 8; ношу 7; **ведро**, вещи, **крест**, **тяжело** 4; **тяжесть**, яйца 3; **бремя**, **везти**, окоlesiцу, **ответственность**, с работы, сумка, сумки, **чемодан**, чепуху 2; автомат, вещь, вздор, далеко, дипломат, долю, домой, ее, идти, камень, корзина, на плечах, на себе, *ноша*, от ответственности, очередь, перетаскивать, поклажу, помидоры, продукты, радость, руки, **свой крест**, сетку, Сидор, скряга, **службу**, стакан, стул, **тащить**, удалять, флаг, через проходную, **чушь**, яйцо 1; $100+53+0+38+18$

Реакция сумку оказалась самой частотной почти во всех выборках и имела частоты от 7 до 18, другая форма сумка и еще три слова груз, ношу, тяжесть встречаются во всех статьях. Кроме указанных пяти ответов во всех выборках присутствуют от 7 до 11 реакций, выделенных жирным курсивом, то есть для стимула нести в приведенных статьях прямого словаря повторяющиеся ответы составляют от 22% до 34% от общего количества реакции. Частоты конкретных реакций в отдельных выборках сильно

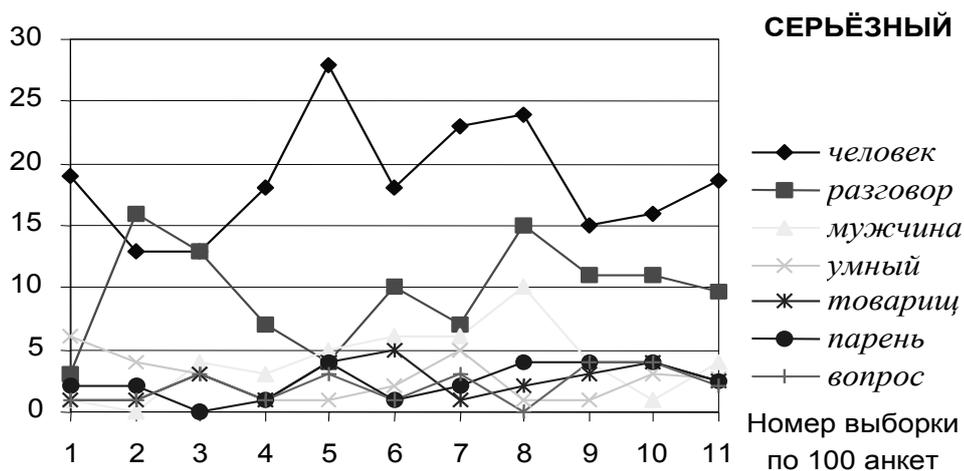
колеблются, кроме того, переменными являются и последовательности упорядоченных по убыванию частоты реакций.

Всего на стимул **нести** получено 254 разных ответа, а в отдельных выборках их встретилось от 41 до 58. Одиночные реакции во всех статьях составляют от 60% до 80%. Неповторяющихся (данных только в одной из выборок) ответов получилось 170 реакций, и почти все они имеют частоту 1, кроме двух: *положить* и *тяжелую сумку*, встретившихся дважды. В конкретных выборках число таких реакций составляет примерно половину “одиночных” (две последние цифры в словарных статьях).



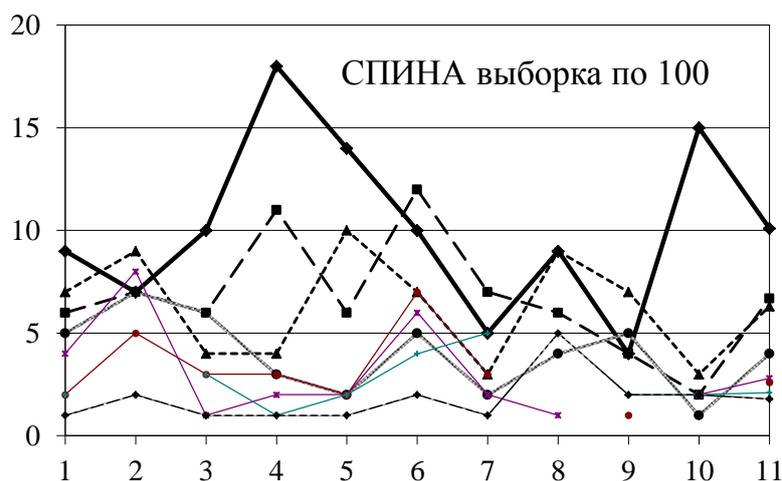
По другим стимулам картина аналогичная, поэтому приведем только первые (частотные) реакции в виде графиков. Последовательность реакций задается величиной “средней частоты” по всем выборкам, которая дана на шкале абсцисс последней 11-ой точкой.

Стимул **старый** получил 238 разных реакций, а в каждой выборке их было около 50. В шести выборках первый, т.е. частный ответ — *дед*, в двух — *молодой*, в 10-й — *друг*, а в 3-ей их два — *дед* и *дом*. Первые пять, из указанных на рисунке, получены во всех десяти выборках, а реакции *новый*, *дедушка* и *пень* встретились в девяти статьях, *дряхлый*, *дуб*, *хрыч*, *старик*, *дурак*, *осел*, *хрен* присутствуют в восьми. Реакции с частотой 1 составляют 70-86%, из которых от 35% до 55% неповторяющиеся ответы.

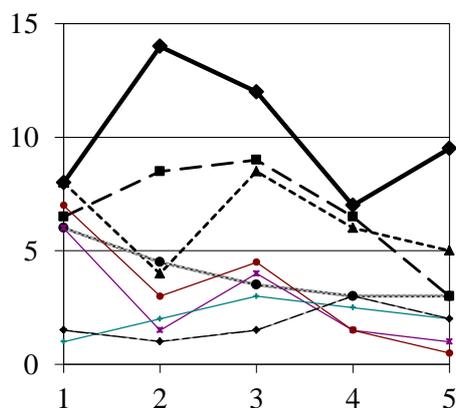


Стимул прилагательное **серьезный**, представлен на рисунке 11. Для него самыми частотными ответами были *человек* и *разговор*, причем первая реакция в восьми выборках явно превосходит по частоте вторую, и только во 2-й выборке *разговор* имеет большую частоту (16 против 13), и в 3-ей оба ответа имеют одинаковую частоту 13. Следующий ответ *мужчина*, встречался от 0 до 10 раз, далее идут реакции *умный*, *товарищ*, *парень*, *вопрос* и др.

Для стимула **спина** результаты представлены двумя вариантами выборок по 100 и 200 опрошенных, и даны значения частот, приведенных. Так как частоты конкретных реакций в отдельных выборках сильно варьируются, ломаные линии отдельных слов на рисунках пересекаются друг с другом. Увеличение количества респондентов вдвое как видно из рисунка, представленного ниже, стабилизирует результаты. Так при выборках 200 испытуемых реакция *широкая* стала всегда первой, хотя в некоторых выборках по 100 ответы *болит* и *прямая* имели большие частоты и линии разных реакций пересекаются значительно меньше.



- ◆— широкая
- прямая
- ▲— болит
- белая
- *— грудь
- живот
- ▲— горбатая
- +— мокрая



Результаты показывают, что только реакции, данные более чем пятью процентами респондентов, встречаются во всех выборках, причем их количество весьма мало, и в представленных статьях составляет всего 3–6 слов (из общего числа более 250). Данный факт показывает, что реакции с относительной частотой менее пяти являются “вероятностными”.

Для подтверждения данного предположения, были проведены исследования по русской части САС. Последовательно выбирались первые 100 анкет, затем 200, 300, 400, 500 и все 594, по которым строились статьи прямого словаря. Затем анализировалось распределение ответов по частотным группам, сколько реакций добавлялось с каждой новой сотней опрошенных и в какие частотные группы. Полученные результаты представлены в таблицах 4 и 5.

В таблице 4 показано, что в 112 словарных статьях, сформированных по первой выборке в 100 опрошенных, было получено 6156 пар стимул–реакция; все стимулы содержали реакции с частотами 1 и 2; только 103 стимула имели ответы с частотой 3, а 85 с частотой 4, и т.д., всего 747 комбинации *стимул–реакция–частота*.

Таблица 4

Кол-во анкет	Число + пар S-R	Частота реакций										
		1	2	3	4	5	6	7	8	9	10	>10
100	6156	112	112	103	85	68	49	42	29	22	21	104
200	3311	112	64	11	1	1						
300	3003	112	93	16	6	2						
400	2631	112	76	9	1	1						
500	2520	112	68	9	1	0						
594	2531	112	52	10	0	0						

Добавление ста анкет породило еще 3311 пар *стимул–реакция*, которые добавили всем 112 стимулам “одиночные” реакции, 64 стимула получили ассоциации с частотой 2, 11 новых реакций встретились трижды, а стимул **родина** четырежды получил реакцию *страна*, и на стимул **думать** пять респондентов ответили *мыслить*.

При дальнейшем пошаговом росте количества опрошенных на 100 (до 594), как показывает таблица, всем стимулам добавлялись новые единичные реакции (количество которых дано в таблице 5). Реакции с частотой 2 добавляло более половины стимулов, а с частотой 3 около десятка. Частоты для новых реакций 4 и 5 встретились довольно редко, а добавления новых реакций с частотой встречаемости более пяти не было. Это позволяет подтвердить вывод о том, что реакции, имеющие относительную частоту встречаемости более пяти, получены уже в первой выборке из 100 респондентов, и констатировать, что рост словника реакций Rsl происходит только за счет добавления “низкочастотных” ответов (с частотами 1, 2 и 3). Следует отметить, что указаны абсолютные частоты, а относительные частоты будут меньше в двое-пятеро раз.

В таблице 5 стимулы приведены не в алфавитном порядке, а упорядочены по убыванию количества разных реакций в словарной статье прямого словаря по выборке 100 опрошенных. Числовые данные показывают, что каждая новая сотня анкет добавляет к уже имеющимся единичным реакциям столь много новых, что их количество обычно превышает 50% от полученных на первой сотне.

Таблица 5.

Стимул	10	20	30	40	50	59	Стимул	10	20	30	40	50	59
	0	0	0	0	0	4		0	0	0	0	0	4
смерть	67	31	32	37	25	28	земля	40	18	14	16	15	19
человек	64	44	41	25	32	22	плохо	40	39	28	27	29	31
помогать	56	29	33	30	17	29	справедливо	40	34	29	17	27	29
							сть						
хорошо	56	36	36	24	16	27	старый	40	16	12	20	13	17
дочь	54	31	25	24	23	24	счастье	40	36	32	21	24	27
народ	54	31	26	26	30	23	вместе	39	37	22	18	11	22
обещать	54	38	32	30	22	30	глупый	39	27	16	23	24	21
душа	53	38	37	27	28	22	девочка	39	37	22	24	20	18
памятник	53	27	15	17	22	19	рот	39	20	23	18	21	21
сила	53	44	27	31	28	25	стыд	39	31	31	26	25	27
Бог	52	39	32	31	20	33	утро	39	24	20	19	13	24
вечность	52	38	38	27	22	7	враг	38	25	32	23	24	25
работа	52	33	28	30	24	21	молодой	38	19	20	23	21	22
радость	52	29	21	27	27	25	руки	38	38	25	24	17	19
вспомина	51	33	34	23	22	24	чистый	38	28	23	20	25	17
							ть						
голос	51	31	32	18	29	23	большой	37	27	33	16	30	25
есть	51	33	22	19	23	21	мальчик	37	35	22	21	18	19
лицо	51	32	26	32	32	24	огонь	37	22	13	19	16	18
ребенок	51	29	25	32	22	24	свет	37	25	28	28	22	27
говорить	50	23	33	30	33	20	слово	37	35	27	28	34	29
хотеть	50	37	24	25	18	24	богатый	36	31	29	23	17	13
дело	49	43	32	24	31	28	вода	36	35	21	16	16	22
жить	48	31	24	20	27	26	день	36	28	19	19	23	19
любовь	48	43	35	29	30	28	время	35	25	26	20	11	5
обман	48	32	32	27	30	35	дядя	35	20	21	15	14	21
гора	47	36	22	24	22	22	кричать	35	27	32	21	25	21
деньги	47	42	33	27	32	20	мать	35	29	27	15	18	18
дурак	47	27	24	25	25	26	свободный	35	23	21	18	17	21
жена	47	29	26	23	25	22	стол	35	21	21	18	16	21
зло	47	26	38	29	23	31	брат	34	27	17	12	16	6
добро	46	22	20	15	17	14	глаза	34	25	18	24	20	22
муж	46	26	26	19	25	31	новый	34	25	26	19	20	13
семья	46	25	21	24	14	25	река	34	31	24	17	17	27
больной	45	37	25	23	22	21	терять	34	23	16	23	23	20
встреча	45	33	34	26	17	23	черный	34	19	20	12	12	19
думать	45	36	27	33	17	23	маленький	33	28	25	17	24	24
жизнь	45	36	25	28	33	29	много	33	24	26	23	23	13
лес	45	35	28	15	20	22	слабый	33	25	19	23	17	17
ненавиде	45	25	26	23	18	22	умный	33	30	24	20	18	25
							ть						
красивый	44	27	27	30	23	20	веселый	32	24	27	17	23	18
разговор	44	35	30	31	25	26	ветер	32	33	21	25	16	19
война	43	19	25	26	23	23	друг	32	25	24	24	16	23
женщина	43	27	32	23	19	33	надеяться	32	24	14	18	19	10

быстро	42 18 25 22 19 20	начало	32 25 24 13 15 19
вечер	42 20 24 25 26 24	палец	31 27 18 13 17 18
жадный	42 33 23 21 15 24	бабушка	30 29 29 23 22 23
машина	42 29 31 26 22 11	хлеб	30 20 16 15 21 15
мужчина	42 33 36 22 29 28	искать	29 30 25 24 27 26
ходить	42 26 24 23 23 23	красный	28 19 22 19 17 17
голова	41 27 17 18 18 20	путь	28 25 20 22 21 21
успеть	41 40 28 22 22 21	зеленый	27 21 16 15 17 20
город	40 26 20 24 23 13	белый	26 12 13 13 13 17
гость	40 32 30 23 30 24	ночь	26 16 18 23 19 18
дверь	40 17 25 23 19 18	родина	25 20 18 15 14 21
деревня	40 30 27 26 31 32	пить	20 19 15 13 15 26
дом	40 29 25 26 28 21	родной	15 13 9 12 8 12

Подведем некоторые итоги количественных исследований ассоциативных словарей, которые позволили получить следующие основные результаты и выводы:

1. Увеличение на 100 человек количества респондентов, отвечающих на стимулы, увеличивает в среднем на 36.41% число разных реакций в словарных статьях прямого словаря.

2. Реакции ассоциативных полей стимулов можно разделить на две группы: “постоянные” и “вероятностные”. “Постоянные” реакции (их не более десятка) имеют относительную частоту встречаемости более 5% в любых выборках (опросах) объемом не менее 100 респондентов.

3. При увеличении объема выборки (опроса) со 100 до 200 респондентов относительные частоты встречаемости (ранги) “постоянных” реакций стабилизируются.