Г. А. Мартинович, Г. Я Мартыненко. Сравнительный анализ статистических параметров коммуникативно-тематических полей.

В ряде исследований нами было показано, что одноименные ситуативно-тематические и ассоциативные поля, определяемые по данным двух видов ассоциирования (ассоциирование при порождении связных текстов и ассоциирование при порождении реакций в ассоциативном эксперименте), в своих существенных областях оказываются практически идентичными. Следовательно, есть достаточные основания утверждать, что они выявляются по результатам двух видов материализации ***одного и того же*** фрагмента *общественного* языкового сознания [Мартинович 1989а, 1989б, 1993, 1994, 2002]. Обобщение элементов этих полей позволяет с наибольшей полнотой и надежностью определять ***существенные*** (*стандартные, воспроизводимые, социально значимые*) ***части*** этого фрагмента – объединения номинативных единиц, тематически организованных непосредственно в коммуникативных целях, которые, скорее всего, могут быть охарактеризованы как **коммуникативно-тематические поля (КТП)**. Объединения же наименований, реконструируемые на основе данных текстов и ассоциативного эксперимента, можно условно разграничить как, соответственно, **КТП-1** и **КТП-2**.

Из сказанного следует, что исследование индивидуального и общественного лексикона (ассоциативно-семантических сетей, ассоциативных полей, лексического компонента речевой способности и т. п.) принципиально возможно как на основе данных ассоциативного эксперимента, так и на основе данных тематически однородных связных текстов. Необходимо отметить, что рассматриваемые закономерности уже достаточно давно находятся в поле зрения исследователей. Так, А. А. Леонтьев, ссылаясь на работу D. Howes'а [Howes 1957: 75 – 85] отмечает «тот факт, что закономерности распределения слов в свободном ассоциативном эксперименте и вероятность их появления в потоке речи… чрезвычайно близки» [Леонтьев 1957: 10]. В свою очередь, рассматривая способы реконструкции индивидуального лексикона, Ю. Д. Караулов пишет: «Принципиально можно считать установленным, что лексикон организован по сетевому принципу, т. е. основа его структуры задается ассоциативно-семантической сетью. Практика исследований в этой области показывает, что последнюю можно строить двумя способами: либо на основе прямого ассоциативного эксперимента, либо по итогам анализа представительного массива произведенных личностью текстов». И далее обосновывает положение о том, что оба эти способа «находятся в отношении взаимодополнительности друг к другу» [Караулов 1987: 101-102, 107]. Откровенно задачу установления степени соответствия друг другу результатов ассоциативного эксперимента и дистрибу-

тивно-статистического анализа ставят А. Е. Супрун, А. П. Клименко и М. С. Толстая: «Для обоих этих направлений важно, видимо, было бы установление степени соответствия друг другу получаемых результатов, а тем самым приближения к той объективной реальности, поиском которой и занята семасиология» [Супрун и др. 1968: 24 – 29]. Но если исследование явлений ассоциативного характера в отечественной и зарубежной науке экспериментальными методами стало уже довольно обычным делом, то их целенаправленное и планомерное изучение *на основе текстов*, по существу, только начинается, и здесь пока много неясного и непонятного, в чем еще предстоит разобраться. И это, несомненно, необходимо сделать, так как, с одной стороны, текстовые исследования обладают целым рядом весьма существенных преимуществ по сравнению с экспериментальными[1], с другой – они открывают дополнительные, принципиально не осуществимые путем ассоциативного эксперимента, возможности изучения явлений ассоциативного плана. Например, обращение к текстовым источникам позволяет если не реконструировать полностью, то, во всяком случае, хотя бы в значительной мере восстановить ассоциативные вербальные связи не только живых, но и мертвых языков, не только живого языка в его современном состоянии, но и в те или иные исторические периоды его развития и существования и т. п. В последнем случае имеется исключительная и, в общем-то, уникальная возможность реконструкции характерных для тех или иных эпох ассоциативных полей. Причем такое восстановление может осуществляться как в пределах всего национального языка, так и в пределах его различных функционально-социальных вариантов (сфер, стилей, жанров и т.п.) вплоть до индивидуально-авторских, личностных.

Несомненный интерес, по нашему мнению, представляет также специальный сравнительный анализ существенных статистических параметров КТП-1 и КТП-2, не изучавшихся

[1] Данные ассоциативных экспериментов в силу самой своей природы (порождение вне контекста) всегда содержат некоторый элемент неопределенности, не позволяющий в целом ряде случаев интерпретировать их однозначно, что в значительной мере затрудняет классификацию элементов КТП-2. Эта неопределенность практически полностью снимается при использовании аналогичных им данных КТП-1. Возможность обращения к контекстам позволяет в этом случае устанавливать часть речи, значение многозначного слова, лексическую и грамматическую сочетаемость, синтаксическую функцию и т.д. слова, как правило, достаточно определенно (если только сама неопределенность не задана специально автором текста в особых целях).

на предшествующих стадиях исследования[2]. Результаты такого анализа, несомненно, помогут дополнить и уточнить ряд положений, сформулированных главным образом на основе качественно-количественного анализа состава КТП-1 и КТП-2 (см. работы *Мартиновича Г. А.*).

В целях достижения возможно большей однородности исследуемого материала в данном случае было решено ограничиться рассмотрением данных, полученных от испытуемых приблизительно одного и того же культурно-образовательного и возрастного уровня – студентов первого и второго курсов филологических факультетов и факультетов журналистики вузов СПб. В одних группах этих студентов был проведен свободный ассоциативный эксперимент на слово-стимул «Береза», в других – студентам было предложено написать свободный связный текст на тему «Береза». Время ответа в ассоциативном эксперименте ограничивалось 3 минутами, количество реакций не ограничивалось. Время написания экспериментальных текстов было ограничено 10 минутами[3].

В конечном итоге по результатам экспериментальных текстов и ассоциативного эксперимента были получены массивы наименований КТП-1 и КТП-2: соответственно – словари 1195 и 658 наименований, выборки – 3000 и 2500 единиц, которые и используются нами в дальнейшем.

---

[2] Многопараметрический статистический анализ данных КТП-2 см.: [Мартыненко и др. 2003].

[3] Несмотря на то, что исследование номинативных единиц тематически однородных связных текстов небольшого объема, каковыми как раз и являются тексты, написанные испытуемыми, является одним из вариантов дистрибутивно-статистического анализа слова-темы, осуществляемого на основе использования "максимальных интервалов текста" – целых текстов [Шайкевич 1982: 15 – 26], в данных этих текстов (что вполне закономерно) присутствует определенная доля шума, избавиться от которого при помощи собственно формальных процедур практически невозможно (Подробнее об этом см.: [Мартинович 1993: 44-46]). Как показали предыдущие исследования, этот шум («шлак») выявляется **только в результате качественного анализа** состава КТП-1 – функционально-тематической классификации наименований этого поля. Так как проведение такой классификации не входит в задачи настоящего исследования, то с наличием этого шума приходится мириться и лишь по возможности оговаривать его в необходимых случаях.

В целях устранения влияния возмущающих факторов, порождаемых индивидуальными и коллективными различиями между информантами, все номинативные единицы КТП-1 и КТП-2 были рэндомизированы с помощью таблицы случайных чисел.

1

Сравнение определенных статистических параметров предварим анализом основных особенностей состава рассматриваемых полей.

Элементами КТП являются полнозначные *номинативные единицы* – слова и устойчивые словосочетания[4]. КТП-1 практически идентичны СТП. КТП-2 не совпадают полностью с АП, так как под последними обычно понимают совокупности собственно реакций (словоформ и словосочетаний в той форме, в какой они непосредственно даны испытуемыми), организованных словом-стимулом. Состав КТП-2 определяется в результате незначительного уровня абстракции собственно номинативных единиц от состава ассоциативных полей (от состава реакций).

Ассоциирование в ассоциативном эксперименте является, по нашему мнению, частным случаем ассоциирования в тексте [Мартинович 1997: 24-29]. Поэтому за основание сравнения принимаются элементы КТП-1, на которые накладывались элементы КТП-2, что позволяет выявить степень и характер их соответствия элементам КТП-1. Основные результаты сравнения представлены в упорядоченных по убывающей частоте списках совпадающих и несовпадающих наименований КТП-2, а также в таблицах 1, 2 и 3.

Наименования КТП-2, совпадающие с наименованиями КТП-1 (интервал частот 11 – 129)

Ранг	Наименование	Частота
1	Белый	129
2	Стройный	98
3	Кудрявый	66
4	Зеленый	62
5	Дерево	53
6	Русский	46
7	Роща	45
8	Лес	41

Наименования КТП-2, не совпадающие с наименованиями КТП-1 (интервал частот 3 – 129)

Ранг	Наименование	Частота
1	Трава	19
2	Качаться	11
3	Грустный	10
4	Кривой	10
5	Осина	9
6	Полено	6
7	Ансамбль	5
8	Гнездо	5

[4] Служебные слова, выражающие различного рода отношения, интертематичны и в состав КТП не включаются. Не включаются в него также числительные, местоимения, модальные слова, междометия, звукоподражания, строевые слова (типа ***стал*** *читать*) и слова-форманты (***буду*** *читать*).

9	Тонкий	39
10	Сок	38
11	Весна	30
12	Девушка	30
13	Нежный	30
14	Белоствольный	29
15	Красивый	29
16	Береста	28
17	Веник	28
18	Высокий	27
19	Поле	27
20	Родина	25
21	Лето	23
22	Россия	23
23	Стоять	23
24	Небо	21
25	Плакучий	21
26	Лист	20
27	Подберезовик	20
28	Молодой	19
29	Баня	17
30	Дуб	17
31	Есенин	17
32	Солнце	17
33	Сережки	16
34	Деревня	15
35	Красота	15
36	Кора	14
37	Расти	14
38	Ветвистый	13
39	Зелень	13
40	Светлый	13
41	Тишина	12
42	Гриб	11
43	Дрова	11
44	Красавица	11
45	Одинокий	11

9	Засохший	5
10	Пушистый	5
11	Синий	5
12	Шуметь	5
13	Ансамбль "Березка	4
14	Задумчивый	4
15	Запах	4
16	Изящный	4
17	Корявый	4
18	Луг	4
19	Облетевший	4
20	Стройность	4
21	Сухой	4
22	Гладкий	3
23	Грамота	3
24	Есенинский	3
25	Забор	3
26	Изба	3
27	Лукошко	3
28	Мечта	3
29	Мох	3
30	Наклониться	3
31	Прозрачный	3
32	Раскидистый	3
33	Срубленный	3
34	Тонкоствольный	3
35	Утро	3
36	Фильм	3

Таблица 1

Совпадение наименований КТП-2 с наименованиями КТП-1

по интервалам частот

Частота	Наимен-ования	Совпад. наимен.	%	Выборка	Совпад. выборки	%
1 - 129	658	269	40,88	2500	1923	76,92
6 - 129	85	79	92,94	1623	1558	96,00
11 - 129	47	45	95,75	1337	1307	97,76
16 - 129	34	33	97,06	1173	1154	98,30
21 - 129	25	25	100	1011	1011	100

Таблица 2

Несовпадение наименований КТП-2 с наименованиями КТП-1 по частотам

Частота	Наимен.	%	Выборка	%
1	305	78,41	305	52,86
2	48	12,34	96	16,64
3	15	3,86	45	7,80
4	9	2,31	36	6,24
5	6	1,54	30	5,20
6	1	0,26	6	1,04
9	1	0,26	9	1,56
10	2	0,52	20	3,47
11	1	0,26	11	1,91
19	1	0,26	19	3,29
Всего	389	100,02	577	100,01

Таблица 3

Ранговая корреляция совпадающих наименований КТП-2
с соответствующими наименованиями КТП-1[5]

Частота	Совпад. наимен.	**ρ**
1 - 129	264	0,516
6 - 129	79	0,578
11 - 129	45	0,631
16 - 129	33	0,622
21 - 129	25	0,552

Приведенные в списках и таблицах данные показывают следующее.

В составе КТП-2 обнаружено 269 наименования, общих с наименованиями КТП-1, то есть 41% от всех наименований КТП-2. Эти 269 наименований соответствуют выборке в 1923 словоупотреблений, что составляет 77% всей выборки КТП-2. Другими словами, 77% всех полнозначных наименований, полученных в результате ассоциативного эксперимента со словом-стимулом *береза*, совпадают с наименованиями связных текстов о березе. Весьма существенным является совпадение, прежде всего, наиболее частотных (*социально* зна-

---

[5] Коэффициент ранговой корреляции вычислялся по формуле Спирмана $\rho = 1 - \frac{6 \times \sum_{i=1}^{n} (R_{K_{1i}} - R_{K_{2i}})^2}{N(N^2 - 1)}$, где K_{1i} - частотность лексемы N_i в КТП-1; K_{2i} - частотность лексемы N_i в КТП-2; R_{K1i} - ранг лексемы N_i в КТП-1; R_{K2i} - ранг лексемы N_i вКТП-2; N - количество лексем.

чимых) наименований КТП-2 с наименованиями КТП-1 (см. Табл. 1). Так, совпадают все 25 наименований с частотностью от 21 до 129 (выборка – 1011 единиц), т. е. 100%. Из 34 наименований с частотностью от 16 до 129 (выборка – 1173) совпадают 33 наименования (97%) или 1154 единиц выборки (98%). Из 47 наименований с частотностью от 11 до 129 (выборка – 1337) совпадают 45 наименований (96%) или 1307 единиц выборки (98%). И из 85 наименований с частотностью от 6 до 129 (выборка – 1623) совпадают 79 наименований (93%) или 1558 единиц выборки (96%). Остальные несовпадающие наименования находятся в области частот менее 6, что составляет, по существу, периферию ("хвост") рассматриваемого поля.

Из данных таблицы 2 видно, что 78,41% несовпадающих наименований (52, 85% выборки) имеют частоту равную 1. 12,34% наименований (16,64% выборки) имеют частоту равную 2 и 3,86% наименований (7,80% выборки) с частотой 3. Следовательно, 94,61% несовпадающих наименований (77,30% выборки) относятся к интервалу частот 1 – 3. В свою очередь в интервале частот 11 – 20 не совпадают всего 2 наименования (0,52%) или 30 единиц выборки (5,20%). Это наименования *трава* 19 и *качаться* 11. Их отсутствие в рассматриваемом нами КТП-1, несомненно, является случайностью, тривиально подтверждающей закономерность совпадения наименований существенной области КТП-2 с наименованиями КТП-1. Это подтверждается и тем, что все наименования КТП-2 в интервале частот 20 – 129 совпадают с наименованиями КТП-1 (см. списки совпадающих и несовпадающих наименований КТП-2).

Отметим также устойчивый положительный (> 50%) коэффициент ранговой корреляции (ρ) совпадающих наименований КТП-2 с соответствующими наименованиями КТП-1 при разных интервалах частот (см. Таблицу 3).

В свою очередь, из 1195 наименований (3000 употреблений) КТП-1 с наименованиями КТП-2 не совпадают 926 наименований (1506 единиц выборки), соответственно 77,40% наименований и 25,20% единиц выборки.

Несовпадение наименований КТП-1 с наименованиями КТП-2 по частотам см. в таблице 4.

Таблица 4

Несовпадение наименований КТП-2 с наименованиями КТП-1 по частотам

Частота	Наимен.	%	Выборка	%
1	679	73,33	679	45,09
2	130	14,04	260	17,26
3	54	5,83	162	10,76
4	20	2,16	80	5,31
5	20	2,16	100	6,64
6	4	0,43	24	1,59
7	6	0,65	42	2,79
8	1	0,11	8	0,53
9	4	0,43	36	2,39
11	1	0,11	11	0,73
12	3	0,32	36	2,39
16	2	0,22	32	2,13
17	1	0,11	17	1,13
19	1	0,11	19	1,26
Всего	926	100,01	1506	100,00

Данные таблицы свидетельствуют о том, что в интервале частот 1 – 3 с наименованиями КТП-2 не совпадает 93% наименований КТП-1 (73% выборки), то есть практически столько же, сколько в процентном отношении не совпадает наименований и употреблений КТП-2 с наименованиями и употреблениями КТП-1 в том же интервале частот. В интервале частот 1 – 5 не совпадает соответственно 97% наименований и 85% употреблений.

С частотностью > 5 не совпадают с данными КТП-2 23 наименования (3%) и 225 единиц выборки (15%): *образ* 19, *очень* 17, *поэт* 16, *приходить* 16, *всегда* 12, *душа* 12, *часто* 12, *делать* 11, *другой* 9, *еще* 9, *народ* 9, *слово* 9, *гулять* 8, *венок* 7, *воспоминание* 7, *говорить* 7, *начинать* 7, *увидеть* 7, *ходить* 7, *весной* 6, *видеть* 6, *думать* 6, *похожий* 6.

Отсутствие в составе КТП-2 девяти высокочастотных глаголов (*приходить* 16, *делать* 11, *гулять* 8, *говорить* 7, *начинать* 7, *увидеть* 7, *ходить* 7, *видеть* 7, *думать* 6), образующих, как известно, предикативную основу предложения, вполне закономерно. Отсутствие 5 наречий (*очень* 17, *всегда* 12, *часто* 12, *еще* 9, *весной* 6) также обусловлено конструктивными особенностями предложения. Остальные 9 наименований (*образ* 19, *поэт* 16, *душа* 12, *другой* 9, *народ* 9, *слово* 9, *венок* 7, *воспоминание* 7, *похожий* 6) действительно являются существенным дополнением данных КТП-2.

Таким образом, рассмотренный материал лишний раз подтверждают основное положение, выдвинутое в предыдущих исследованиях, – положение о существенном совпадении состава КТП-2 с составом КТП-1 в качественном и количественном отношениях.

Перейдем к рассмотрению «поведения» некоторых статистических параметров данных текстов (состав КТП-1) и эксперимента (состав КТП-2).

В статистике упорядочивание значений варьирующего признака осуществляется в одной из трех измерительных шкал: количественной, порядковой (ординальной) и номинальной. Для различных шкал исторически сложились методы обработки данных, пригодные только для работы именно в данной шкале. Самая разработанная система техник сформировалась для количественных переменных. Здесь на основании теории моментов была разработана развитая система средних, показателей вариации, характеристик формы распределения и т. п. Здесь же продуктивно используется система порядковых статистик (мода, медиана, квантили и др.).

Система показателей для номинальной и порядковой шкал существенно беднее. Так, наиболее популярной характеристикой для номинальной шкалы является мода, а для ординальной - медиана.

Ранговые и спектровые распределения обладают всеми внешними свойствами количественной шкалы, но при этом характеризуются патологически большой вариацией признака как на порядковой шкале (ранговое распределение), так и на количественной шкале (спектровое распределение).

На основании литературных данных и собственных изысканий нами был сформирован достаточно полный перечень параметров, с помощью которых может быть описана лексико-статистическая структура коммуникативно-тематического поля.

Подробную характеристику шкал и соответствующих им статистик см. в работе: [Мартыненко и др. 2003: 6 – 11].

Для проведения сравнительного анализа статистических параметров были составлены таблицы зависимостей параметров от объема выборки. См. Таблицы 5 и 6.

Таблица 5

Зависимость параметров КТП-1 (N = 3000) от объема выборки

Объем выборки	Объем словаря	Энтропия	Мера упорядоч.	Средняя частота	Среднее геометр.	Стандарт. отклон.	Среднее лин. откл.	Коэф. вар.по σ	Коэф. вар.по D_f	Коэффиц. разнообр.
N	n	E	O	F_{cp}	F_g	σ	D_f	$V\sigma$	V_{Df}	K
100	90	6,429	0,990	1,111	1,074	0,378	0,202	0,340	0,182	0,911
200	160	7,149	0,976	1,250	1,161	0,602	0,413	0,482	0,330	0,825
300	225	7,621	0,975	1,333	1,202	0,833	0,530	0,625	0,398	0,796
400	281	7,857	0,966	1,423	1,236	1,136	0,660	0,798	0,464	0,779
500	336	8,066	0,961	1,488	1,262	1,318	0,752	0,886	0,505	0,771
600	391	8,242	0,957	1,535	1,276	1,481	0,815	0,965	0,531	0,762
700	433	8,314	0,949	1,617	1,303	1,774	0,931	1,097	0,576	0,755
800	477	8,419	0,946	1,677	1,330	1,964	0,997	1,171	0,595	0,736

900	523	8,518	0,943	1,721	1,346	2,133	1,047	1,239	0,608	0,727
1000	567	8,581	0,938	1,764	1,352	2,365	1,107	1,341	0,628	0,725
1100	609	8,670	0,937	1,806	1,374	2,452	1,148	1,358	0,636	0,711
1200	648	8,731	0,935	1,852	1,393	2,622	1,196	1,416	0,646	0,702
1300	682	8,769	0,932	1,906	1,409	2,804	1,257	1,471	0,659	0,694
1400	721	8,826	0,930	1,942	1,419	2,930	1,306	1,509	0,673	0,693
1500	764	8,884	0,928	1,963	1,419	3,063	1,347	1,560	0,686	0,699
1600	795	8,908	0,925	2,013	1,435	3,255	1,415	1,617	0,703	0,697
1700	821	8,926	0,922	2,071	1,455	3,433	1,498	1,658	0,723	0,691
1800	853	8,959	0,920	2,110	1,466	3,569	1,551	1,691	0,735	0,685
1900	885	9,003	0,920	2,147	1,483	3,646	1,597	1,698	0,744	0,678
2000	918	9,035	0,917	2,179	1,489	3,775	1,646	1,732	0,755	0,678
2100	951	9,055	0,915	2,208	1,489	3,952	1,700	1,790	0,770	0,680
2200	987	9,093	0,914	2,229	1,493	4,021	1,728	1,804	0,775	0,677
2300	1011	9,103	0,912	2,275	1,507	4,194	1,792	1,844	0,788	0,674
2400	1044	9,136	0,911	2,299	1,514	4,281	1,828	1,862	0,195	0,673
2500	1066	9,144	0,909	2,345	1,528	4,470	1,883	1,906	0,803	0,665
2600	1094	9,159	0,907	2,377	1,535	4,598	1,935	1,934	0,814	0,666
2700	1118	9,166	0,905	2,415	1,542	4,776	1,989	1,978	0,824	0,663
2800	1149	9,191	0,904	2,437	1,545	4,872	2,021	1,999	0,829	0,661
2900	1173	9,210	0,903	2,472	1,560	4,976	2,065	2,013	0,835	0,656
3000	1195	9,220	0,902	2,510	1,570	5,129	2,109	2,043	0,840	0,650

Продолжение таблицы 5

Объем выборки	Ранговое среднее	Стандарт. откл. по R	Коэффиц. вар. по R	Медиана по R	Золотое сечение	Средн. лин. откл. по R	Коэф. вар. по D	Коэф. рассеян.	Индекс концентр.
N	R_{cp}	σ_r	V_r	Me_r	G_r	D_r	V_{Dr}	Z	γ
100	41	28	0,677	40	28	24	0,577	0,168	1,201
200	67	50	0,746	60	36	44	0,657	0,287	1,403
300	89	70	0,787	75	44	62	0,697	0,350	1,539
400	105	88	0,838	81	48	78	0,743	0,407	1,686
500	121	105	0,858	86	52	93	0,769	0,440	1,785
600	137	122	0,891	92	57	108	0,788	0,463	1,861
700	145	135	0,931	95	53	118	0,874	0,498	1,993
800	156	148	0,949	102	57	129	0,827	0,516	2,064
900	168	162	0,964	108	59	140	0,833	0,528	2,119
1000	178	175	0,983	112	60	151	0,848	0,544	2,191
1100	189	187	0,989	118	63	161	0,852	0,551	2,228
1200	197	198	1,005	121	64	169	0,858	0,564	2,294
1300	203	207	1,020	121	62	177	0,872	0,577	2,365
1400	212	218	1,028	121	64	186	0,877	0,584	2,406
1500	222	231	1,041	122	67	196	0,883	0,591	2,446
1600	227	239	1,053	124	66	203	0,864	0,601	2,507
1700	229	245	1,070	124	66	207	0,904	0,614	2,590
1800	235	254	1,081	124	66	213	0,906	0,620	2,634
1900	241	262	1,087	126	67	219	0,909	0,626	2,676
2000	248	271	1,093	128	67	226	0,911	0,630	2,706
2100	253	280	1,107	127	66	234	0,925	0,638	2,763

2200	261	290	1,111	129	66	242	0,927	0,641	2,785
2300	263	296	1,125	128	66	245	0,932	0,649	2,848
2400	270	305	1,130	130	67	253	0,937	0,652	2,870
2500	272	310	1,140	129	66	256	0,941	0,658	2,923
2600	276	317	1,149	130	65	261	0,946	0,663	2,967
2700	278	322	1,158	129	65	265	0,953	0,669	3,025
2800	284	331	1,165	131	66	271	0,954	0,672	3,049
2900	288	336	1,167	133	66	275	0,955	0,675	3,076
3000	290	341	1,176	132	66	278	0,959	0,680	3,124

Таблица 6

Зависимость параметров КТП-2 (N = 2500) от объема выборки

Объем выборки	Объем словаря	Энтропия	Мера упорядоч.	Средняя частота	Среднее геометр.	Стандарт. отклон.	Среднее лин. откл.	Коэф. вар.по σ	Коэф. вар.по D_f	Коэффиц. разнообр.
N	n	E	O	F_{cp}	F_g	σ	D_f	$V\sigma$	V_{Df}	K
100	71	5,942	0,966	1,408	1,256	0,897	0,610	0,637	0,433	0,746
200	121	6,546	0,946	1,653	1,352	1,498	0,939	0,906	0,568	0,719
300	166	6,897	0,935	1,807	1,403	1,929	1,138	1,068	0,630	0,705
400	200	7,049	0,922	2,000	1,462	2,443	1,380	1,222	0,690	0,690
500	235	7,183	0,912	2,128	1,483	2,887	1,603	1,357	0,753	0,698
600	267	7,289	0,904	2,247	1,510	3,292	1,778	1,465	0,791	0,689
700	296	7,389	0,900	2,365	1,545	3,619	1,937	1,530	0,819	0,679
800	321	7,442	0,894	2,492	1,581	4,060	2,100	1,629	0,843	0,660
900	346	7,538	0,894	2,601	1,629	4,246	2,211	1,632	0,850	0,642
1000	362	7,515	0,884	2,762	1,656	4,832	2,453	1,749	0,888	0,649
1100	391	7,591	0,882	2,813	1,663	5,091	2,531	1,810	0,900	0,639
1200	414	7,639	0,879	2,899	1,679	5,368	2,643	1,852	0,912	0,638
1300	431	7,663	0,876	3,016	1,714	5,700	2,783	1,890	0,923	0,626
1400	454	7,682	0,870	3,084	1,706	6,070	2,930	1,968	0,950	0,648
1500	473	7,708	0,867	3,171	1,726	6,386	3,053	2,014	0,963	0,634
1600	495	7,742	0,865	3,232	1,734	6,661	3,158	2,061	0,977	0,630
1700	517	7,771	0,862	3,288	1,734	6,943	3,252	2,112	0,989	0,629
1800	535	7,799	0,860	3,364	1,758	7,228	3,334	2,149	0,991	0,622
1900	554	7,817	0,858	3,364	1,762	7,478	3,454	2,223	1,027	0,625
2000	572	7,831	0,855	3,497	1,766	7,741	3,566	2,214	1,020	0,622
2100	591	7,839	0,851	3,553	1,766	8,121	3,653	2,286	1,028	0,626
2200	610	7,855	0,849	3,607	1,770	8,400	3,745	2,329	1,038	0,628
2300	628	7,869	0,847	3,662	1,778	8,719	3,834	2,381	1,047	0,623
2400	643	7,883	0,845	3,733	1,791	9,012	3,924	2,414	1,051	0,619
2500	658	7,899	0,844	3,799	1,803	9,241	4,022	2,432	1,059	0,616

Продолжение таблицы 6

Объем выборки	Ранговое среднее	Стандарт. откл. по R	Коэффиц. вар. по R	Медиана по R	Золотое сечение	Средн. лин. откл. по R	Коэф. вар. по D	Коэф. рассеян.	Индекс концент.
N	Rcp	σ_r	V_r	Me_r	G_r	D_r	V_{Dr}	Z	γ
100	28	22	0,786	21	14	19	0,679	0,364	1,157
200	41	37	0,902	28	16	32	0,780	0,494	1,976

300	53	50	0,943	32	19	43	0,811	0,535	2,151
400	59	60	1,017	33	19	50	0,847	0,585	2,407
500	65	69	1,062	34	19	58	0,892	0,620	2,631
600	71	78	1,099	35	19	65	0,915	0,640	2,775
700	76	85	1,118	37	20	70	0,921	0,656	2,908
800	79	91	1,152	37	21	74	0,937	0,675	3,076
900	84	97	1,155	39	22	78	0,929	0,681	3,131
1000	84	100	1,190	37	21	80	0,952	0,699	3,321
1100	89	107	1,202	38	21	86	0,966	0,706	3,404
1200	92	113	1,228	39	22	89	0,967	0,715	3,511
1300	94	116	1,234	39	22	91	0,968	0,722	3,596
1400	96	121	1,260	39	21	95	0,990	0,733	3,740
1500	99	126	1,273	39	22	98	0,989	0,736	3,788
1600	101	131	1,297	39	22	101	1,000	0,744	3,911
1700	104	136	1,308	40	22	105	1,010	0,749	3,981
1800	107	140	1,308	40	22	108	1,009	0,751	4,009
1900	108	144	1,333	39	22	110	1,019	0,758	4,139
2000	110	148	1,345	39	22	113	1,027	0,762	4,209
2100	112	152	1,357	39	22	116	1,036	0,767	4,286
2200	114	156	1,368	39	22	118	1,035	0,771	4,360
2300	116	160	1,379	39	22	121	1,043	0,774	4,422
2400	117	163	1,393	39	22	123	1,051	0,778	4,504
2500	119	166	1,395	39	22	125	1,050	0,780	4,535

1. Сравнение зависимости некоторых параметров полей от объема выборки.

Из табл. 3 – 6 видно, что по мере увеличения объема выборки значения большинства параметров неуклонно возрастают. Эта закономерность наиболее типичных зависимостей в наглядной форме показана на рис. 1 – 6.

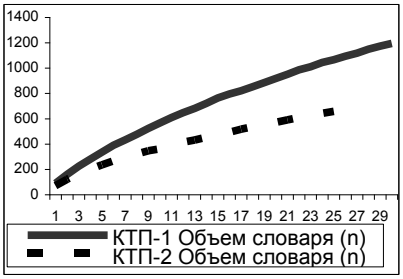

Рис. 1. Зависимость объема словаря от объема выборки.

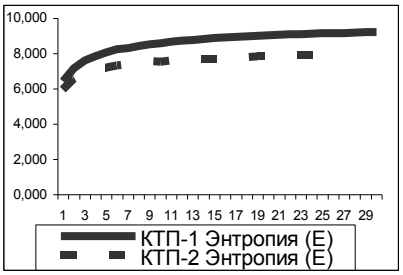

Рис 2. Зависимость энтропии от объема выборки

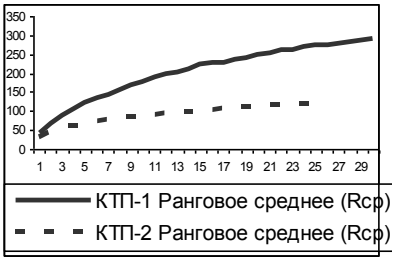

Рис. 3. Зависимость ранговой средней от объема выборки

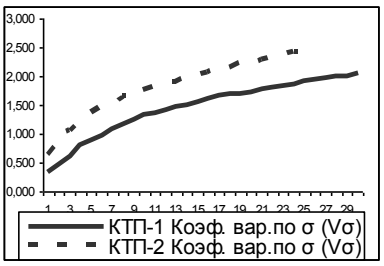

Рис. 4. Зависимость частотного коэффициента вариации по стандартному отклонению от объема выборки

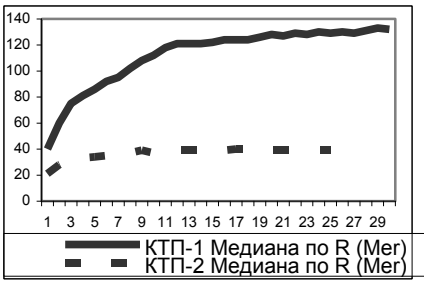

Рис. 5. Зависимость медианы по рангу от объема выборки

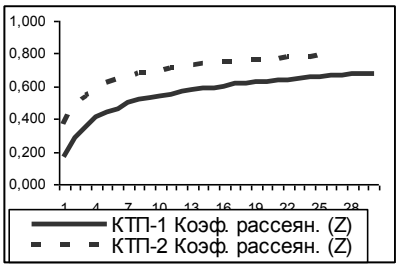

Рис. 6. Зависимость коэффициента рассеяния от объема выборки

В табл. 7 приводятся сравниваемые параметрические данные КТП-1 и КТП-2 при выборке объемом 2500 лексических единиц.

Таблица 7

	Объем выборки	Объем словаря	Энтропия	Мера упорядоч.	Средняя частота	Среднее геометр.	Стандартн. отклон.
	N	n	E	O	F_{cp}	F_g	σ
КТП-1	2500	1066	9,144	0,909	2,345	1,528	4,470
КТП-2	2500	658	7,899	0,844	3,799	1,803	9,241

	Среднее лин. откл.	Коэф. вар.по σ	Коэф. вар.по Df	Коэффиц. разнообр.	Ранговое среднее	Стандарт. откл. по R	Коэффиц. вар. по R
	D_f	V_σ	V_{Df}	K	R_{cp}	σ_r	V_r
КТП-1	1,883	1,906	0,803	0,665	272	310	1,140
КТП-2	4,022	2,432	1,059	0,616	119	166	1,395

	Медиана по R	Золотое сечение	Средн. лин. откл. по R	Коэф. вар. по D	Коэф. рассеян.	Индекс концентр
	Me_r	G_r	D_r	V_{Dr}	Z	γ
КТП-1	129	66	256	0,941	0,658	2,923
КТП-2	39	22	125	1,050	0,780	4,535

На основании приведенных данных можно заключить, что:

– на всем протяжении всех эмпирических кривых сохраняются устойчивые различия между КТП-1 и КТП-2,

– КТП-1 при любых объемах выборки отличается существенно большим объемом (разнообразием) словаря в сравнении с КТП-2 (см. рис.1), причем эти различия между полями стремительно возрастают по мере возрастания объема выборки,

– существенно меньшие различия наблюдаются для параметров, характеризующих внутреннюю структуру полей. Это относится, например, к энтропии, которая характеризует степень равномерности распределения лексических единиц (чем больше энтропия, тем больше равномерность распределения): в КТП-1 она несколько выше (см. рис.2). Примерно такую же информацию содержит график коэффициента вариации (чем больше коэффициент вариации, тем выше неравномерность распределения (см. рис.4): здесь для КТП-2 в сравнение с КТП-1 характерно большее рассеяние частот около средней частоты и следовательно, большая неравномерность (или меньшая равномерность). Аналогичную информацию содержит также и рис.6

– что касается ранговой средней (см. рис.3), то этот параметр чутко реагирует прежде всего на соотношение численностей системно-нормативных (высокочастотных) и окказиональных единиц, указывая при заданном объеме выборки на приблизительную границу между ними. Эта граница

Для КТП-1 на всем протяжении графика располагается существенно выше, чем для КТП-2, т.е. лексическое разнообразие в ядре КТП-1 существенно выше.

– наиболее существенны различия в поведении кривых для медианы по рангу (см. рис.5), характеризующей уровень стабилизации высокочастотной (в данном случае ме-

дианной) зоны словаря на постоянном уровне при нарастании объема выборки. Этот уровень для КТП-1 существенно выше.

Приведенные соображения имеют сугубо эмпирический характер. Попытаемся их усилить с помощью теоретических моделей.

2. Сравнение аппроксимаций и прогнозы

В качестве теоретических моделей нами были использованы две функции: функция Вейбулла $y = k - ke^{-ax^b}$ и степенная логиста $y = \frac{kx^b}{a + x^b}$. Метод аппроксимации и его экспериментальная проверка описаны в работе [Мартыненко и др. 2003]. Значения постоянных коэффициентов для обеих функций и результаты аппроксимации для объема словаря, энтропии, ранговой средней и медианы по рангу представлены в табл. 8 – 12.

Таблица 8

Постоянные коэффициенты аппроксимирующих функций

ОБЪЕМ СЛОВАРЯ				
	КТП-1		КТП-2	
	Вейбулл	Логиста	Вейбулл	Логиста
К (асимптота)	2600	4500	2250	4000
a	0,036	48,34	0,037	52,83
b	0,837	0,842	0,725	0,728

ЭНТРОПИЯ				
	КТП-1		КТП-2	
	Вейбулл	Логиста	Вейбулл	Логиста
К (асимптота)	9,65	10,35	8,15	8,5
a	1,099	0,611	1,320	0,427
b	0,306	0,48	0,298	0,530

РАНГОВАЯ СРЕДНЯЯ				
	КТП-1		КТП-2	
	Вейбулл	Логиста	Вейбулл	Логиста
К (асимптота)	400	600	175	240
a	0,111	13,20	0,184	7,244
b	0,722	0,745	0,564	0,607

МЕДИАНА по R				
	КТП-1		КТП-2	
	Вейбулл	Логиста	Вейбулл	Логиста
К (асимптота)	135	150	40,1	40,5
a	0,353	2,860	0,869	0,928
b	0,696	0,913	0,512	1,150

Результаты аппроксимации зависимости объема словаря, энтропии, ранговой средней и медианы по рангу

Таблица 9

Результаты аппроксимации объема словаря

N	Объем словаря					
	КТП-1	Теория		КТП-2	Теория	
		Вейбулл	Логиста		Вейбулл	Логиста
1	90	91	91	71	74	74
2	160	161	161	121	122	122
3	225	223	223	166	162	162
4	281	280	281	200	197	198
5	336	334	334	235	230	230
6	391	385	385	267	261	261
7	433	433	433	296	290	290
8	477	479	479	321	317	317
9	523	523	523	346	343	343
10	567	566	566	362	368	368
11	609	607	607	391	392	392
12	648	647	646	414	415	415
13	682	685	685	431	437	437
14	721	722	722	454	458	458
15	764	758	757	473	479	479
16	795	793	792	495	499	499
17	821	827	826	517	519	519
18	853	860	859	535	538	538
19	885	892	891	554	556	556
20	918	923	923	572	575	575
21	951	954	953	591	592	592
22	987	983	983	610	610	610
23	1011	1012	1012	628	626	626
24	1044	1040	1040	643	643	643
25	1066	1068	1068	658	659	659
26	1094	1094	1095		***675***	***675***
27	1118	1121	1122		***690***	***691***
28	1149	1146	1148		***706***	***706***
29	1173	1171	1173		***721***	***721***
30	1195	1196	1198		***735***	***736***
П Р О Г Н О З						
50		***1589***	***1612***		***981***	***985***
100		***2119***	***2250***		***1377***	***1405***
500		***2596***	***3578***		***2142***	***2545***
1000		***2600***	***3934***		***2235***	***2973***

Таблица 10

Результаты аппроксимации энтропии

N	Энтропия					
	КТП-1	Теория		КТП-2	Теория	
		Вейбулл	Логиста		Вейбулл	Логиста
1	6,429	6,435	6,423	5,942	5,972	5,941
2	7,149	7,170	7,191	6,546	6,541	6,553
3	7,621	7,577	7,598	6,897	6,843	6,861
4	7,857	7,850	7,866	7,049	7,041	7,055

5	8,066	8,051	8,061	7,183	7,183	7,193
6	8,242	8,208	8,212	7,289	7,292	7,298
7	8,314	8,334	8,334	7,389	7,378	7,381
8	8,419	8,439	8,435	7,442	7,448	7,448
9	8,518	8,528	8,521	7,538	7,507	7,505
10	8,581	8,604	8,596	7,151	7,557	7,554
11	8,670	8,671	8,661	7,591	7,600	7,596
12	8,731	8,729	8,719	7,639	7,638	7,633
13	8,769	8,781	8,770	7,663	7,671	7,666
14	8,826	8,828	8,817	7,682	7,700	7,695
15	8,884	8,870	8,860	7,708	7,727	7,722
16	8,908	8,908	8,898	7,742	7,750	7,746
17	8,926	8,943	8,934	7,771	7,772	7,768
18	8,959	8,975	8,967	7,799	7,791	7,788
19	9,003	9,004	8,998	7,817	7,809	7,807
20	9,035	9,031	9,026	7,831	7,825	7,824
21	9,055	9,056	9,053	7,839	7,840	7,840
22	9,093	9,079	9,078	7,855	7,854	7,855
23	9,103	9,101	9,101	7,869	7,867	7,869
24	9,136	9,121	9,123	7,883	7,879	7,883
25	9,144	9,140	9,144	7,899	7,890	7,895
26	9,159	9,158	9,164		***7,900***	***7,907***
27	9,166	9,175	9,183		***7,910***	***7,918***
28	9,191	9,190	9,200		***7,919***	***7,928***
29	9,210	9,205	9,217		***7,927***	***7,938***
30	9,220	9,219	9,234		***7,935***	***7,948***
П Р О Г Н О З						
50		***9,395***	***9,454***		***8,032***	***8,073***
100		***9,542***	***9,690***		***8.105***	***8,201***
500		***9,644***	***10,033***		***8,148***	***8,371***
1000		***9,649***	***10,120***		***8,150***	***8,411***

Таблица 11

Результаты аппроксимации ранговой средней

N	Ранговая средняя					
	КТП-1	Теория		КТП-2	Теория	
		Вейбулл	Логиста		Вейбулл	Логиста
1	41	42	42	28	29	29
2	67	67	68	41	42	42
3	89	88	88	53	51	51
4	105	105	105	59	58	58
5	121	120	120	65	64	64
6	137	134	134	71	69	70
7	145	146	146	76	74	74
8	156	158	158	79	78	79
9	168	168	168	84	82	82
10	178	178	178	84	86	86
11	189	187	187	89	89	89
12	197	196	195	92	92	92
13	203	204	203	94	95	95
14	212	211	211	96	98	98
15	222	218	218	99	100	100
16	227	225	224	101	102	102
17	229	231	231	104	105	104
18	235	237	237	107	107	107
19	241	243	243	108	109	108
20	248	249	248	110	110	110

21	253	254	253	112	112	112
22	261	259	259	114	114	114
23	263	264	263	116	116	115
24	270	268	268	117	117	117
25	272	272	273	119	119	118
26	276	276	277		***120***	***120***
27	278	280	281		***121***	***121***
28	284	284	285		***123***	***123***
29	288	288	289		***124***	***124***
30	290	291	293		***125***	***125***
П Р О Г Н О З						
50		***339***	***350***		***142***	***143***
100		***382***	***420***		***160***	***166***
500		***400***	***531***		***175***	***206***
1000		***400***	***557***		***175***	***216***

Таблица 12

Результаты аппроксимации медианы по рангу

N	Медиана по R					
	КТП-1	Теория		КТП-2	Теория	
		Вейбулл	Логиста		Вейбулл	Логиста
1	40	40,109	38,866	21	23,292	20,994
2	60	58,739	59,555	28	28,503	28,536
3	75	71,694	73,211	32	31,394	32,058
4	81	81,477	83,029	33	33,272	34,051
5	86	89,182	90,473	34	34,593	35,319
6	92	95,413	96,335	35	35,569	36,190
7	95	100,548	101,082	37	36,315	36,822
8	102	104,838	105,012	37	36,900	37,301
9	108	108,463	108,324	39	37,368	37,674
10	112	111,552	111,157	37	37,748	37,973
11	118	114,204	113,608	38	38,062	38,218
12	121	116,496	115,752	39	38,323	38,422
13	121	118,487	117,645	39	38,542	38,593
14	121	120,225	119,329	39	38,729	38,740
15	122	121,749	120,837	39	38,888	38,867
16	124	123,089	122,196	39	39,024	38,977
17	124	124,272	123,428	40	39,142	39,074
18	124	125,320	124,550	40	39,245	39,160
19	126	126,250	125,576	39	39,334	39,237
20	128	127,078	126,518	39	39,413	39,305
21	127	127,817	127,387	39	39,482	39,367
22	129	128,478	128,191	39	39,543	39,423
23	128	129,071	128,936	39	39,597	39,474
24	130	129,602	129,630	39	39,645	39,520
25	129	130,080	130,277	39	39,688	39,563
26	130	130,511	130,882		***39,726***	***39,602***
27	129	130,900	131,450		***39,761***	***39,638***
28	131	131,251	131,983		***39,791***	***39,671***
29	133	131,569	132,484		***39,819***	***39,702***
30	132	131,857	132,957		***39,844***	***39,731***
П Р О Г Н О З						
50		***134,369***	***138,837***		***40,061***	***40,056***
100		***134,977***	***143,856***		***40,121***	***40,282***
500		***135,000***	***148,540***		***40,125***	***40,440***
1000		***135,000***	***149,221***		***40,125***	***40,457***

Примечания.

1. Полужирным курсивом отмечены прогнозируемые значения.

2. N – объем выборки (в сотнях словоупотреблений)

На данном экспериментальном материале ни одна из аппроксимирующих функций не обнаруживает очевидного преимущества перед другой в отношении адекватности эмпирическим данным. Проверка по критерию согласия Колмогорова показала, что на эмпирически достигнутых объемах выборки вероятность расхождений между теоретическими и экспериментальными данными по четырем параметрам практически нет. Причем этот вывод справедлив как по отношению к КТП-1, так и по отношению к КТП-2.

Расхождения между теоретическими функциями становятся заметными лишь при объемах выборки, которые примерно в два раза превышают объем, достигнутый в данном исследовании (см. табл. 9 – 12). По мере увеличения объема выборки это расхождение увеличивается все более, достигая максимума на асимптотических уровнях. Это означает, что если оставаться на уровне разумных выборок и не увлекаться статистической гигантоманий, то можно пользоваться с одинаковой степенью практической результативности одной из рассматриваемых функций. Некоторым преимуществом при этом обладает функция Вейбулла, так как она значительно быстрее, чем логистическая функция устремляется к предельным величинам.

Что же касается теоретической стороны дела, то ее прояснение с неизбежностью обрекает исследователя на увеличение объема выборки. Только таким способом можно получить более четкую ориентацию в том, по какому пути («вейбулловскому» или «логистическому») пойдет эмпирическая кривая. Если траектория этого пути при наращивании объема выборки прорисуется достаточно откровенно, то можно не только сделать жесткий выбор в пользу одной из функций, но и воспользоваться некоторыми замечательными свойствами каждой из них. Однако не исключается, что поиски варианта аппроксимирующей функции, более адекватно отражающих организацию данных полей, могут быть продолжены.

Тем не менее, некоторые полезные выводы можно сделать и не выходя за пределы наших наблюдений. Обратимся к анализу максимальных (асимптотических) значений трех переменных. Сравним асимптотические уровни как для двух полей, так и для двух вариантов аппроксимации (см. табл. 9 – 12).

Таблица 13

Асимптотические уровни переменных в КТП-1 и КТП-2 при аппроксимации с помощью функции Вейбулла

	КТП-1	КТП-2	КТП-1/ КТП-2
Объем словаря	2600	2250	1,16
Энтропия	9,65	8,15	1,18
Ранговое среднее	400	175	2,28
Медиана по рангу	135	40,1	3,37

Таблица 14

Асимптотические уровни переменных в КТП-1 и КТП-2 при аппроксимации с помощью логистической функции

	КТП-1	КТП-2	КТП-1/ КТП-2
Объем словаря	4500	4000	1,13
Энтропия	10,35	8,50	1,22
Ранговое среднее	600	240	2,50
Медиана по рангу	150	40,5	3,70

Таблица 15

Асимптотические уровни переменных в КТП-1 при разных способах аппроксимации

	Вейбулл	Логиста	КТП-1/ КТП-2
Объем словаря	2600	4500	1,73
Энтропия	9,65	10,35	1,07
Ранговое среднее	400	600	1,50
Медиана по рангу	135	150	1,11

Таблица 16

Асимптотические уровни переменных в КТП-2 при разных способах аппроксимации

	Вейбулл	Логиста	КТП-1/ КТП-2
Объем словаря	2250	4000	1,78
Энтропия	8,15	8,50	1,04
Ранговое среднее	175	240	1,37
Медиана по рангу	40,1	40,5	1,01

Из данных табл. 13 и 14 видно, что асимптотические уровни по всем переменным для КТП-1 выше, чем для КТП-2, но их соотношение при двух способах аппроксимации примерно одинаково. Примечательно, что наибольший перепад между асимптотами имеет место для рангового среднего и медианы. Эти переменные, по-видимому, очень чутко, в отличие от объема словаря и энтропии, реагируют на концентрацию частот в верхней зоне частотных словарей. Эта концентрация существенно выше для КТП-2 в сравнении с КТП-1.

Обращает на себя внимание также то, что асимптотические уровни для энтропии и медианы по рангу при разных способах аппроксимации очень близки друг к другу и в КТП-1 и в КТП-2. По-видимому, это объясняется тем, что данные переменные являются интенсив-

ными характеристиками, описывающими внутреннюю структуру полей, тогда как объем словаря является экстенсивной, объемной характеристикой. Что касается ранговой средней, то она занимает промежуточное положение, являясь, с одной стороны, *интенсивной (объемной)*, а другой – *интенсивной (структурной)* характеристикой, коррелирующей с объемом словаря в целом и с его окказиональной («хвостовой») частью одновременно.

Сравнив данные табл. 15 и 16, находим, что различия между соотношениями асимптотических уровней по всем параметрам для КТП-1 и КТП-2 весьма незначительны (см. данные, относящиеся к столбцам КТП-1/КТП-2).

Значительный интерес, на наш взгляд, представляет собой сравнительный анализ не только максимальных, но и околомаксимальных значений параметров в обеих полях при различных вариантах аппроксимации..

Попытаемся оценить скорость сходимости обеих теоретических моделей путем вычисления объемов выборок, необходимых для достижения 90% и 99% от асимптотического уровня, характерного для конкретного параметра в одном из двух полей. 90%-ный и 99%-ный уровни вычисляются путем подстановки в уравнения (3), (4) величины одного из двух уровней и решения этих уравнений относительно объема выборки. Результаты вычислений представлены в табл. 17 – 20.

Таблица 17

Объемы выборки при околомаксимальных значениях объема словаря (90%-ный и 99%-ный уровни для КТП-1 и КТП-2 по отношению к асимптотическому)

	КТП-1		КТП-2	
Уровни	Вейбулл	Логиста	Вейбулл	Логиста
90%	136	1361	335	4770
99%	310	23624	871	112420

Таблица 18

Объемы выборки при около максимальных значениях энтропии (90%-ный и 99%-ный уровни для КТП-1 и КТП-2 по отношению к асимптотическому)

	КТП-1		КТП-2	
Уровни	Вейбулл	Логиста	Вейбулл	Логиста
90%	8,69	34,9	7,3	12,7
99%	13,2	5203	68,2	1179

Таблица 19

Объемы выборки при околомаксимальных значениях ранговой средней (90%-ный и 99%-ный уровни для КТП-1 и КТП-2 по отношению к асимптотическому)

	КТП-1		КТП-2	
Уровни	Вейбулл	Логиста	Вейбулл	Логиста
90%	66,7	609	88,2	975
99%	174	15230	302	50615

Таблица 20

Объемы выборки при околомаксимальных значениях медианы по рангу (90%-ный и 99%-ный уровни для КТП-1 и КТП-2 по отношению к асимптотическому)

	КТП-1		КТП-2	
Уровни	Вейбулл	Логиста	Вейбулл	Логиста
90%	14,8	35,1	16,7	6,3
99%	40,0	485,6	26,1	51,0

Проинтерпретируем данные табл. 17- 20.

Функция Вейбулла обнаруживает более быструю сходимость к предельным величинам по всем переменным. Для этой функции на 90%-ном уровне среднее значение объема выборки по четырем параметрам в 9 раз превышает соответствующее значение для логистической функции. Что касается КТП-2, то здесь мы имеем 13-кратное превышение. Еще более контрастные результаты получены для 99%-ого уровня (83-кратное и 130-кратное превышение, соответственно). Важно, однако, при этом отметить, что для энтропии и медианы приведенные соотношения не столь разительны. Кроме того, при 90-%-ном уровне значения обеих функций находятся в пределах эксперимента, представленного в данной работе, или близки к ним. Это означает, что обе переменные обладают весьма высоким уровнем состоятельности независимо от варианта модельных интерпретаций. Это утверждение справедливо по отношению к обеим аппроксимациям.

Если же сравнивать эффективность предложенных моделей с точки зрения их сходимости в КТП-1 и КТП-2, то на 90%-ном уровне в среднем по четырем параметрам функция Вейбулла обнаруживает в среднем двукратное увеличение объема выборки при движении от КТП-1 к КТП-2 (57 и 112). При 99%-ном уровне имеем 3-кратное увеличение (510 и 1452). Что касается логистической функции, то здесь на 90%-ном уровне имеем трехкратное (510 и 1452) увеличение, а на 99%-ном – 4-кратное (11134 и 41066).

В целом проведенное исследование подтверждает отмеченное выше положение о взаимодополнительности КТП-1 и КТП-2, а также возможность их обобщения в КТП. Изучение статистических параметров обобщенного коммуникативно-тематического поля является предметом специального рассмотрения.

ЛИТЕРАТУРА

Howes 1957 – D. H o w e s. On the relation between the probability of a word as an association and in general verbal usage. “Journal of Abnormal and Social Psychology”, 1957, v. 54.

Караулов 1987 – К а р а у л о в Ю. Н. Русский язык и языковая личность. – М., 1987.

Леонтьев 1977 – Л е о н т ь е в А. А. Общие сведения об ассоциациях и ассоциативных нормах // Словарь ассоциативных норм русского языка. – М., 1977. С. 5 – 16.

Мартинович 1989а – М а р т и н о в и ч Г. А. Вербальные ассоциации и организация лексикона человека // Филологические науки. – 1989, № 3. С. 39–45.

Мартинович 1989б – М а р т и н о в и ч Г. А. К проблеме онтологии языка // Вестник Ленинградского университета. Сер. 2, 1989. В. 3. С. 54 – 60.

Мартинович 1993 – М а р т и н о в и ч Г. А. Текст и эксперимент. – СПб, 1993.

Мартинович 1994 – М а р т и н о в и ч Г. А. Текст и эксперимент (исследование коммуникативно-тематического поля в русском языке). Автореф. дис. ... докт. филол. наук. – СПб., 1994.

Мартинович 1997 – М а р т и н о в и ч Г. А. Вербальные ассоциации в ассоциативном эксперименте. СПб, 1997.

Мартинович 2002 – М а р т и н о в и ч Г. А. О новом типе учебного словаря русского языка // Лексикография: Информационный бюллетень – СПб, 2002. С. 4 – 23.

Мартыненко и др. – М а р т ы н е н к о Г. Я., М а р т и н о в и ч Г. А. Многопараметрический статистический анализ результатов ассоциативного эксперимента. – СПбГУ, 2003.

Супрун и др. 1968 – С у п р у н А. Е., К л и м е н к о А. П., Т о л с т а я М. С. К сопоставлению психолингвистического и дистрибутивного анализа семантики // Методы экспериментального анализа речи. – Минск, 1968. С. 122 – 128.

Шайкевич 1982 – Ш а й к е в и ч А. Я. Дистрибутивно-статистический анализ текстов: Автореф. дис. ... докт. филол. наук. Л., 1982.