

## Статистические характеристики употребления компьютерного сленга на сетевых форумах

Кутузов А.Б. ([tyumenkender@gmail.com](mailto:tyumenkender@gmail.com))

Тюменский Государственный Университет, кафедра перевода и переводоведения

Частота употребления сленговой лексики - это количество появлений сленгоидов в текстах, полученных с сетевых форумов. Мы называем эту частоту синтагматической, поскольку исследуем число употреблений сленгоидов в потоке речи, а не в словаре (парадигматическая частота). Подвергнув синтагматические частоты сленга статистическому анализу, мы получим вероятности выбора продуцентом текста именно сленговой лексики для наименования денотатов, относящихся к компьютерной концептосфере.

Чтобы провести статистический анализ, необходимо параметризовать значимые для нас текстовые данные и трансформировать их в числовой материал, то есть, построить простую стохастическую дискретную модель. В качестве основных моделируемых параметров мы избрали частоту употребления слов с компьютерными денотатами и частоту употребления сленговой лексики.

Первый этап моделирования заключался в наборе корпуса высказываний. Источниками материала нам послужили четыре крупных сетевых форума:

1. Высказывания на форуме сайта [ixbt.com](http://forum.ixbt.com/?id=4), раздел «Техническая поддержка» (<http://forum.ixbt.com/?id=4>). Технически форум относится к сети Интернет.
2. Высказывания на форуме сайта [3dnews.ru](http://forum.3dnews.ru/forumdisplay.php?s=7d27bcde18b6fbb7e82dedf054edc411&forumid=5), раздел «Общие проблемы» (<http://forum.3dnews.ru/forumdisplay.php?s=7d27bcde18b6fbb7e82dedf054edc411&forumid=5>). Технически форум относится к сети Интернет.
3. Высказывания в эхоконференции RU.OS.CMP («Сравнение операционных систем») ([http://www.fido-online.com/x/\\_-0?Msg?5&147&&a13](http://www.fido-online.com/x/_-0?Msg?5&147&&a13)). Технически эхоконференция относится к сети FIDOnet.
4. Высказывания в эхоконференции 1072.COMPNEWS.TALK («Компьютерные новости») ([http://www.fido-online.com/x/\\_-0?Msg?5&1315&&a1](http://www.fido-online.com/x/_-0?Msg?5&1315&&a1)). Технически эхоконференция относится к сети FIDOnet.

Высказывания сохранялись методом сплошного отбора из четырех форумов за 20-21 мая 2004 года каждый. Отбор прекращался по достижении 10 тысяч словоупотреблений. При наборе корпуса из сообщений сразу удалялась вся техническая информация (имя и адрес отправителя и получателя, дата и время публикации), а также цитирование предыдущих писем (во избежание двойной обработки одних и тех же словоупотреблений), листинги программ и автоматические приветствия и прощания (templates).

Таким образом, мы получили четыре текстовых файла (назовем их [ixbt.txt](#), [3d.txt](#), [ruoscmp.txt](#) и [compnews.txt](#)) по десять тысяч словоупотреблений каждый, со «срезами» речевой активности соответствующих сетевых форумов за конкретную дату. Этот корпус уже позволяет приступать к поиску интересующих нас сленговых единиц. Но сначала мы разделили каждый массив на несколько равных фрагментов (выборок). Это необходимо, чтоб иметь базу для статистического анализа, а именно - учесть отклонения выборочных данных от средних.

Файлы с массивами были разбиты на 50 фрагментов с равным количеством строк каждый. Поскольку массивы состоят из 10 тысяч словоупотреблений, каждый фрагмент

содержит около 200 словоупотреблений. Это дало возможность приступить к собственно подсчёту сленговых единиц.

В каждом фрагменте содержатся как просто словоупотребления, обозначающие компьютерные денотаты (далее “компьютерные номинации”), так и сленгоиды. Для нас важны оба слоя лексики, причем второй из них является составной частью первого. Это значит, что денотат любого компьютерного сленгоида по определению принадлежит к компьютерной концептосфере, но не каждая компьютерная номинация является сленгоидом.

Мы выделяем словоупотребления с компьютерными денотатами для того, чтобы минимизировать вероятность ошибки при статистическом анализе. Поскольку мы считаем единицами компьютерного сленга лишь слова, называющие реалии компьютерной сферы, мы должны фиксировать не отношение количества сленгоидов к общему количеству слов в выборках, а отношение количества сленгоидов к количеству компьютерных номинаций. В ином случае мы рискуем оказаться в ситуации, когда анализ покажет малый процент сленгоидов, но обусловлено это будет малым количеством компьютерных номинаций вообще. Возможна и противоположная ошибка - анализ показывает высокий процент употребления сленгоидов, но при этом количество компьютерных номинаций может быть так велико, что на самом деле по отношению к этому количеству сленгоиды употребляются не так часто. Без учета количества слов с компьютерными денотатами исследование частотных характеристик сленговой лексики не будет адекватным.

Итак, мы подсчитываем количество компьютерных номинаций и сленгоидов в каждом фрагменте. Поскольку автоматическое выделение интересующих нас лексических единиц невозможно (существующие вычислительные системы не способны проводить полный контекстуальный анализ, необходимый для такой задачи), мы производим его вручную, методом просмотра фрагментов и маркирования компьютерных номинаций и сленгоидов. Например, в следующем фрагменте (из форума [ixbt.com](http://ixbt.com)):

*не понял. При чём тут сканирование USB программой ? А винда на что ? Я так понимаю, прога (типа Nero) общается с сидюками не напрямую, а через драйвер ... или я не прав ? PS: может попробовать перепрошить CD?*

мы отмечаем 12 компьютерных номинаций, а именно: “сканирование”, “USB”, “программой”, “винда”, “прога”, “Nero”, “общается”, “сидюками”, “напрямую”, “драйвер”, “перепрошить”, “CD”. Из них только две не являются одновременно единицами компьютерного сленга - “программой” и “USB”. Первое слово (“программа”) вполне конвенционально и общелитературно, а второе (“USB”) является англоязычной аббревиатурой, употреблённой в русском тексте. Остальные компьютерные номинации обладают такими признаками компьютерного сленга, как субстандартность, неоднозначность, образность и т.д.

По завершении маркирования, в нашем распоряжении оказывается аннотированный корпус. Из каждого аннотированного массива возможно легко извлечь количество компьютерных номинаций и компьютерных сленгоидов в каждом его фрагменте.

Итак, мы располагаем первичными лексико-статистическими описаниями наших массивов. Каждое описание состоит из двух рядов по 50 чисел: первый отражает количество компьютерных номинаций, второй – количество сленгоидов в соответствующих фрагментах. Данные получены эмпирически и составляют объективную параметризованную информацию о нашем корпусе с точки зрения частоты употребления компьютерного сленга. Этот набор данных показывает нам как бы “снимок” определённого аспекта поведения некоего процесса, а именно – дискурса сетевых форумов.

Эти данные можно использовать для построения простейшей модели употребления сленговой лексики. Для этого добавим к описаниям массивов ещё по одному ряду чисел, производному от уже существующих двух. Этот ряд содержит долю сленговых единиц (“процент сленгоидов”) среди компьютерных номинаций данного фрагмента, выраженную

в процентах. Она получается через простейшее выражение: “<число сленгоидов> разделить на <число компьютерных номинаций>/100”.

В качестве среднего балла для каждого из этих рядов мы использовали медианы (поскольку данные находятся в порядковых шкалах), но параллельно приводим и значения средних арифметических. Такая методика находится в согласии с концепцией устойчивости, рекомендующей использовать различные методы для обработки одних и тех же данных с целью выделить выводы, получаемые одновременно при всех методах. [Орлов, 1979]

Вот вычисленные нами средние значения этих параметров для каждого массива (все числа округлены до двух знаков после запятой):

1. Числовые ряды для массива 3dNews			
	<i>Количество компьютерных номинаций</i>	<i>Количество сленгоидов</i>	<i>Процент сленгоидов</i>
Среднее арифметическое частот во фрагментах	31,76	14,63	46,09
Медиана частот во фрагментах	31	14	44,43
Среднее абсолютное отклонение	8,46	5,45	13,1

2. Числовые ряды для массива ixbt			
	<i>Количество компьютерных номинаций</i>	<i>Количество сленгоидов</i>	<i>Процент сленгоидов</i>
Среднее арифметическое частот во фрагментах	48,22	21,72	44,65
Медиана частот во фрагментах	47,5	19	46,66
Среднее абсолютное отклонение	19,83	10,59	10,86

Отметим, что фрагменты номер 35 и 36 из массива ixbt не содержали компьютерных номинаций (они содержат длинное обсуждение отличий зануления и заземления, без упоминания компьютеров). Поэтому эти фрагменты не используются в статистическом анализе: они не влияют на синтагматическую частоту сленгоидов. Интересно, что ixbt.com - единственный из форумов, содержащий фрагменты, где отсутствуют сленгоиды и компьютерные номинации. Это косвенно свидетельствует о том, что область интересов его аудитории несколько шире, чем у аудиторий других форумов.

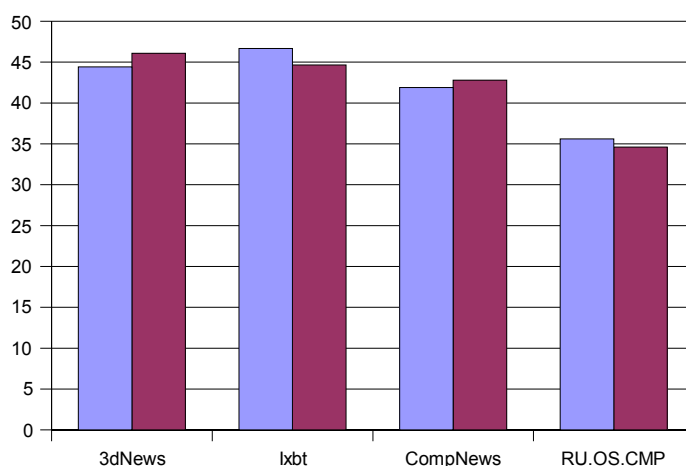
3. Числовые ряды для массива 1072.COMPNEWS			
	<i>Количество компьютерных номинаций</i>	<i>Количество сленгоидов</i>	<i>Процент сленгоидов</i>
Среднее арифметическое частот во фрагментах	36,18	15,78	42,8
Медиана частот во фрагментах	35,5	16	41,89

Среднее абсолютное отклонение	9,51	6,39	12,36
-------------------------------	------	------	-------

4. Числовые ряды для массива RU.OS.CMP			
	<i>Количество компьютерных номинаций</i>	<i>Количество сленгоидов</i>	<i>Процент сленгоидов</i>
Среднее арифметическое частот фрагментах во	31,94	11,52	34,62
Медиана частот во фрагментах	31	9,5	35,62
Среднее абсолютное отклонение	11,69	6,22	12,39

На графике 1 мы наглядно показали соотношение средних синтагматических вероятностей употребления сленга по четырем форумам.

*График 1. Средняя вероятность употребления сленгоидов. По оси Y показан процент сленгоидов, синий столбик – значение по медиане, фиолетовый – по средней арифметической.*



Видна хорошая сходимость результатов по среднему арифметическому и по медиане. Незначительно расходятся лишь данные по проценту употребления сленговой лексики для форумов ixht.com и 3dnews.ru, но в любом случае эти форумы делят между собой первое и второе места. Таким образом, моделирование двумя различными методами вычисления средней частоты даёт очень схожие результаты, а это свидетельствует об адекватности полученных данных.

На графике 2 отражена динамика процента употребления сленговой лексики по фрагментам каждого из массивов. По оси X отложены номера фрагментов, а по оси Y – доля сленговых единиц среди компьютерных номинаций данного фрагмента (компьютерные номинации, манифестированные сленговой лексикой).

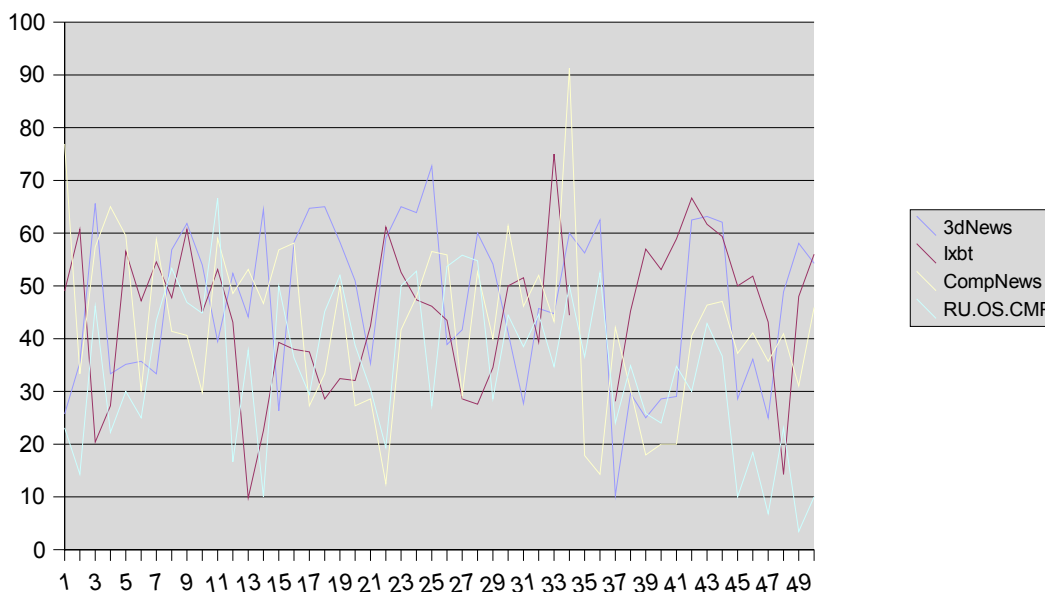


График 2. Динамика частоты употребления сленга

Как можно видеть, в целом динамика процента употребления сленгоидов принципиально не меняется от форума к форуму. Частота употребления сленга нормально распределена в границах 20 и 60 процентов от количества компьютерных номинаций. Динамика колебаний очень слабо изменяется от форума к форуму, и лишь незначительное количество фрагментов выходят за эти рамки в ту или иную сторону.

Таким образом, мы получили четыре модели форумов, каждая из которых характеризуется 52 числами – выборочными частотами употребления сленгоидов во фрагментах этих форумов, средней частотой и средним абсолютным отклонением. Средний процент употребления сленгоидов и его отклонение отражают синтагматическую вероятность употребления сленговой единицы для компьютерной номинации в каждом конкретном форуме. Например, посетители эхоконференции ixbt.com, испытывая потребность назвать какой-либо объект из компьютерной концептосферы, примерно в 47 процентах случаев (с абсолютным отклонением в 11 процентов) назовут его при помощи какой-либо сленговой единицы (а в остальных 53 процентах – используют термин или общелитературное слово). Теперь в нашем распоряжении есть вероятности употребления сленга для каждого из форумов.

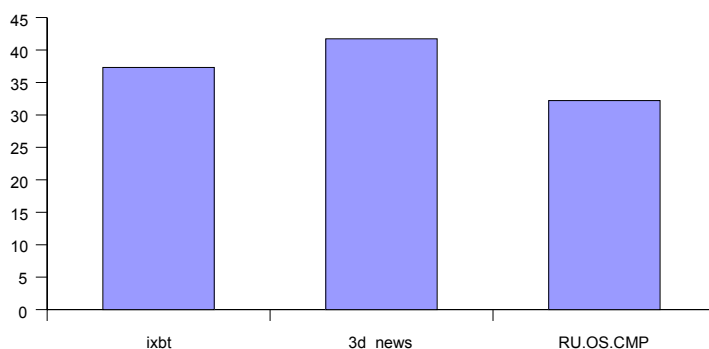
Определяя вероятность употребления сленга, не следует забывать, что мы экстраполировали данные некоторой выборочной совокупности на генеральную. Вычислим для каждого из массивов абсолютную и относительную ошибку определения средней частоты при 95-процентной надежности выкладок:

Массив	Абсолютная ошибка	Относительная ошибка, в процентах
3dNews	0,87	1,88
ixbt	0,89	2
1072.COMPNEWS	0,96	2,25
RU.OS.CMP	0,73	2,11

Как можно видеть, ни в одном из массивов абсолютная ошибка не превысила одного процента вероятности употребления сленга, а относительная - 2,5 процентов от этой вероятности, что является свидетельством хорошей сгруппированности выборочных

частот вокруг средней. Ошибка наблюдения в данном случае существенно меньше измеряемой величины.

Для проверки наших моделей мы провели повторную пилотную выборку высказываний из форумов 3dnews, ixbt и RU.OS.CMP за 18 декабря 2005 года. Выборка из эхоконференции 1072.COMPNEWS.TALK не производилась, поскольку поток сообщений за указанный период в этой эхоконференции отсутствовал. Средние проценты вероятности употребления сленга в повторной выборке приведены в следующем графике:



*График 3. Средняя вероятность употребления сленгоидов при повторной выборке.*

Можно видеть, что сохранились прежние различия в вероятностных характеристиках употребления компьютерного сленга: форумы ixbt и 3dnews делят между собой первое место, а RU.OS.CMP по-прежнему обладает наименьшей вероятностью употребления сленговых единиц. Динамика количества слов с компьютерными денотатами в зависимости от форума во второй выборке также идентична первой: наибольшее количество таких слов употребляется на форуме ixbt, а наименьшее – в 3d\_news. Таким образом, мы можем предварительно заключить, что построенные нами модели обладают некоторой предсказательной силой и действительно отражают объективные характеристики исследуемых сетевых форумов.

Отметим, что, хотя динамика частоты употребления сленгоидов распределена во всех форумах похоже, тем не менее, мы предполагаем существование объективных различий по абсолютной вероятности употребления сленга по каждому форуму. Это предположение подтверждается малыми значениями абсолютной ошибки определения средней частоты. Разности наших средних по форумам больше, чем значения соответствующих абсолютных ошибок. Следовательно, каждому форуму присуща своя вероятность употребления сленга.

Сразу обращает на себя внимание отличие RU.OS.CMP от других массивов: вероятность употребления сленга практически на 10 пунктов меньше, чем в других форумах. Этот феномен требует объяснения, которое мы дадим далее при помощи метода дисперсионного анализа.

В целом, можно констатировать, что статистические законы, внешним проявлением которых являются выборочные частоты в вышеприведенных рядах, действуют достаточно надёжно. Синтагматическая частота употребления компьютерного сленга на сетевых форумах устойчива и является характерным признаком форума и компьютерного дискурса.

Итак, мы выяснили, что между исследованными сетевыми форумами существуют определенные количественные отличия (неоднородности) по признаку вероятности употребления сленгоидов. Но это арифметические различия; необходимо определить, существенны ли они статистически (объективизировать их), то есть, действительно ли они

вызваны именно качественной неоднородностью сравниваемых форумов или же различия объясняются естественными статистическими колебаниями, а выборки на самом деле однородны. Существует также вероятность, что они (различия) вызваны какими-либо другими факторами, отличными от принадлежности выборок к разным форумам. Следует проверить нашу изначальную гипотезу о том, что компьютерный дискурс можно рассматривать как генеральную совокупность, единую по параметру вероятности употребления компьютерного сленга.

Для объективизации неоднородностей мы используем метод дисперсионного анализа, проводя его при помощи соответствующего модуля из статистического языка R-base, включённого в состав табличного процессора Gnumeric 1.2.13.

Прежде, чем выявлять различия между форумами по параметру вероятности употребления сленга, следует выяснить, можно ли рассматривать их, как принадлежащие одной и той же генеральной совокупности (то есть, компьютерному дискурсу) по параметру количества компьютерных номинаций. Ведь не исключено, что форумы существенно отличаются по этому признаку.

Для этого проведем дисперсионный анализ всех четырех массивов по ряду “количество компьютерных номинаций”. Мы формулируем “нулевую гипотезу” о том, что различие между сравниваемыми форумами отсутствует (они принадлежат к одной генеральной совокупности), наблюдаемые отклонения объясняются случайными колебаниями энтропийной выборки. Нулевая гипотеза утверждает, что в системе наших выборок отсутствует упорядочивающий негэнтропийный фактор принадлежности к разным совокупностям, что данные в выборках распределены нормально (хаотично).

При пятипроцентном пороге ошибки модуль дисперсионного анализа выдает следующие результаты в стандартной таблице R-base:

<i>Дисперсионный анализ: однофакторный по количеству компьютерных номинаций в 4 массивах</i>				
		<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
		12,59	0,00%	2,650677

Мы видим, что полученное экспериментально отношение средних квадратов эффекта и ошибки ( $F=12,59$ ) значительно больше табличного для таких выборок (2,65). Это значит, что с 95-процентной вероятностью расхождение средних статистически значимо и независимый фактор (фактор принадлежности к определенному форуму) оказывает существенное влияние на разброс средних значений количества компьютерных номинаций.

Оценим этот результат. По всей видимости, какой-то (какие-то) из массивов существенно отличается от других по количеству компьютерных номинаций. При внимательном рассмотрении статистических данных заметно, что в массиве *ixbt* очень высоки как среднее количество компьютерных номинаций во фрагменте и общее их количество в массиве (50,23 и 2411 соответственно), так и дисперсия (613,97, почти в 4 раза выше, чем у остальных групп). Можно предположить, что именно аномально высокая сумма отклонений выборочных данных от среднего в массиве *ixbt* оказывает влияние на дисперсионный анализ и приводит к такому большому значению  $MS_{\text{эффект}}$ . Мы можем проверить эту теорию, проведя процедуру дисперсионного анализа для количества компьютерных номинаций в остальных трёх массивах, исключая *ixbt*. Нулевую гипотезу о принадлежности форумов к одной генеральной совокупности оставим той же. Результаты анализа таковы:

<i>Дисперсионный анализ: однофакторный по количеству компьютерных номинаций в 3 массивах</i>				
		<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
		1,452767	23,00%	3,057621

Как видим, здесь отношение между MS эффекта и MS ошибки гораздо менее значимо.  $F=1,45$  меньше, чем  $F_{крит}=3,0576$ . Это означает, что нулевая гипотеза верна с вероятностью более 20 процентов ( $P=23\%$ ). Различия в количестве компьютерных номинаций в данных трёх массивах статистически совершенно не значимы; в среднем можно найти около 1600-1700 таких слов на 10000 словоупотреблений.

Следовательно, неоднородность действительно была вызвана форумом ixbt и относительно большим количеством компьютерных номинаций, используемых на нём (при том, что по вероятности употребления сленга, как будет показано ниже, он существенно не отличается от других форумов).

Аналогично ряду словоупотреблений с компьютерными денотатами, подвергнем дисперсионному анализу все четыре форума по параметру “процент сленгоидов”:

*Дисперсионный анализ: однофакторный по вероятности употребления сленгоидов*

	<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
	6,460937	0,03%	2,651153

В исследуемых форумах наличествуют некоторые неоднородности по вероятности употребления сленгоидов, ведь  $F=6,46$  больше чем  $F_{крит}=2,65$ . Необходимо найти источник неоднородности (фактор отклонения).

Нам известна по крайней мере одна возможная причина неоднородности: два из исследуемых форумов технически относятся к сети Интернет (ixbt и 3dNews), и два - к сети FIDOnet (RU.OS.CMP и 1072.COMPNEWS). Возможно, вероятность употребления сленгоидов - это параметр, специфичный для сети в целом. Проверим это предположение. Объединим числовые ряды соответствующих форумов, получив две выборки по 100 вхождений каждая<sup>1</sup>. Проведём по ним дисперсионный анализ:

<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
10,94953	0,11%	3,889341

В данном случае  $F$  также значительно больше  $F_{крит}$ . Это может означать, что вероятность употребления сленгоидов действительно зависит от сети. Однако, мы анализировали синтетические выборки. Каждая из них в свою очередь состояла из двух других, однородность которых мы ещё не доказали. Нужно подтвердить (или опровергнуть) однородность форумов внутри каждой сети. Сравним при помощи дисперсионного анализа попарно выборки ixbt-3dNews и RU.OS.CMP-1072.COMPNEWS.

Сравнение ixbt-3dNews:

<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
0,65882	41,00%	3,940163

Как видим, внутри Интернет-форумов наблюдается хорошо выраженная однородность (вероятность подтверждения нулевой гипотезы - 41 процент). Вероятность употребления сленгоидов в них распределена нормально.

Сравнение RU.OS.CMP-1072.COMPNEWS:

<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
7,030764	0,93%	3,938111

Эхоконференции FIDOnet оказались неоднородны с вероятностью более 99 процентов. Поскольку это так, результаты нашего анализа по сетям не должны приниматься во внимание, ведь сравнивались не две однородные совокупности, а одна

<sup>1</sup>На самом деле, в выборке по Интернет-форумам будет 98 вхождений, поскольку, как указывалось ранее, в двух фрагментах из форума ixbt не оказалось ни сленгоидов, ни компьютерных номинаций вообще.



однородная (интернет-форумы) и одна неоднородная. Следовательно, мы не можем утверждать, что вероятность употребления сленгоидов однозначно зависит от сети. Фактор, оказавший влияние на результаты анализа по четырем выборкам, нужно искать глубже, на уровне отдельного форума.

Исходя из данных, полученных в предыдущем разделе, можно также предположить, что источником неоднородности является массив RU.OS.CMP, с его низким абсолютным количеством сленгоидов (1731 сленгоид, единственный форум с этим параметром меньше 2000) и невысокой вероятностью их употребления (35 процентов, на 10 процентов меньше, чем у остальных форумов). Проверим эту гипотезу, сравнив массивы без RU.OS.CMP. Получаем:

<i>F</i>	<i>Значение P</i>	<i>F критическое</i>
0,996365	0,371733	3,058486

*F* эмпирическое значительно меньше табличного, нулевая гипотеза (о принадлежности выборок к одной совокупности) верна с вероятностью 37 процентов. Поскольку форумы, проанализированные без RU.OS.CMP, оказались однородны, предположение о том, что неоднородность вносит именно RU.OS.CMP, получило подтверждение.

Причина такого поведения этого массива, вероятно, лежит в экстралингвистической плоскости и состоит в том, что посетители RU.OS.CMP – в основном профессионалы, хоть и продолжающие получать удовольствие от неформального общения. Аудитория же остальных форумов по большей части состоит из непрофессиональных, хотя и активных пользователей. Интересно, что хотя в самой этой эхоконференции сленг употребляется реже, чем в других массивах, её правила (выраженные в виде текста) значительно более неформальны, чем в остальных трёх форумах. «Устав конференции» изобилует шутками, просторечиями, сленгом, окказионализмами и даже параграфемикой в виде эмодзи («смайликов»). Например:

*Ну, как обычно: прошу дискутирующих не завязывать в конференции личной переписки... (просторечное вводное слово "ну")*

*Категорически не рекомендуется писать на ОПХБЕР-ском языке. В случае использования китайского языка постарайтесь придерживаться BIG5 кодировки. (шутливое сленговое выражение "опХБЕР-ский язык", означающее неправильную кодировку кириллицы)*

*В своей любимой ОС Вы, в этом случае, тоже плохо разбираетесь, скорее всего. :)* (эмодзи)

С другой стороны, правила форума 3dNews (у которого самая большая абсолютная и средняя частота сленгоидов) тяготеют к стандартному, даже канцелярскому стилю. Редкие сленгоиды сопровождаются пояснениями в скобках, например:

*На форуме запрещается искать кряки (crack)...*

Можно сделать вывод, что на частоту употребления компьютерного сленга гораздо сильнее влияет состав аудитории форума, нежели принятые на нём эксплицитные правила, и тем более их стиль. Настоящие «мета-правила», стоящие за высказываниями на форумах составляют дискурс по Фуко или Фрау, «предконтекст», включающий в себя, в том числе, и порождающую и целевую аудитории. Эти правила имплицитны, и выражаются в различных аспектах структуры того или иного форума – например, в вероятности употребления сленгоида вместо конвенционального термина (эта вероятность ниже на RU.OS.CMP, как показал дисперсионный анализ) или в допустимости обсуждения внекомпьютерных тем (вероятность этого выше на форуме ixbt, как показывает наличие в его массиве двух фрагментов, лишенных компьютерных номинаций).

Таким образом, мы описали статистические характеристики употребления единиц компьютерного сленга на форумах ixbt.com, 3dnews.ru, 1072.COMPNEWS и RU.OS.CMP и

показали, что на форуме 3dnews компьютерный сленг для номинации компьютерных денотатов употребляют с вероятностью около 44 процентов, на форуме ixbt – 47 процентов, в эхоконференции 1072.COMPNEWS – 42 процента, и, наконец, в эхоконференции RU.OS.CMP – с вероятностью 36 процентов, причём абсолютная ошибка составляет не более одного процента. Предсказательная сила этой модели была проверена на повторной выборке, которая показала схожую динамику вероятностных характеристик употребления компьютерного сленга на исследуемых форумах.

Также был применен метод дисперсионного анализа для выявления неоднородностей в распределении вероятности употребления сленга по форумам. Мы определили, что частота употребления сленга по исследованным форумам однородна, за исключением эхоконференции RU.OS.CMP, которая из-за своей специфической аудитории обладает пониженным значением параметра «вероятность употребления сленга». Следовательно, сетевые форумы можно рассматривать как единую субкультуру с примерно одинаковым отношением к соответствующему подязыку. В целом, исследуемые форумы однородны по параметру вероятности употребления компьютерного сленга для компьютерных номинаций. В среднем этот параметр составляет около 44 процентов. Другими словами, из ста произвольных слов с компьютерными денотатами, взятых на форумах, около 44 будут принадлежать сленговой лексике, а остальные - общелитературному языку или конвенциональной терминологии. Кроме того, анализ показал, что на форуме ixbt.com употребляется значительно больше слов с компьютерными денотатами (компьютерных номинаций), чем на других форумах. Это может стать темой для дальнейших исследований, посвященных статистическим характеристикам компьютерных номинаций на сетевых форумах.

Мы делаем вывод, что компьютерному дискурсу (как и дискурсу в целом) объективно присущи количественные признаки, поддающиеся вычислению и анализу. Эти признаки устойчивы и характерны. Они могут быть использованы в прикладных целях: для атрибуции или адаптивного транскодирования (реферирования со сменой стилового регистра) текстов, полученных с сетевых форумов

### **Литература**

1. Структурная и прикладная лингвистика, вып.2 // Межвузовский сборник. Изд-во ЛГУ, 1983.
2. Структурная и прикладная лингвистика, вып. 4 // Межвузовский сборник. Изд-во СПбГУ, 1993.
3. Статистика речи и автоматический анализ текста: сборник статей. Л.: Наука, 1980.
4. Андреев, Н.Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. - Л.: Наука, 1967
5. Беликова, И.А. Особенности образования терминов-неологизмов в подязыке компьютерной техники: автореферат диссертации на соискание ученой степени к.ф.н. - Омск, 2004
6. Головин, Б.Н. Язык и статистика. - М.: Просвещение, 1971
7. Малаховский, Л.В. Теория лексической и грамматической омонимии. - Л.: Наука, 1990
8. Моргун, Н.Л. Научный сетевой дискурс как тип текста: автореферат диссертации на соискание ученой степени к.ф.н. - Тюмень, 2003
9. Носенко, И.А. Начала статистики для лингвистов. М.: Высшая Школа, 1981.
10. Орлов А.И. Устойчивость в социально-экономических моделях. - М.:Наука, 1979.
11. Фуко, М. Слова и вещи. Археология гуманитарных наук. - М.: Прогресс, 1997.