

Леонид АШКИНАЗИ
Мария ГАЙНЕР
Алла КУЗНЕЦОВА

Исследование общества посредством Интернета*

Интернет, как и любой результат деятельности общества, может быть материалом для исследования этого общества. Интернет как носитель информации состоит из текстов, картинок, мелодий, а также комбинаций этих форм (текст с иллюстрациями, песня, видео и т. д.)

Язык Интернета частично является обычным письменным языком (выкладывание «бумажных» текстов), а частично — приближенным к устному языком форумов, блогов, гостевых книг и т. п. В Интернете легко реализуются связи между текстом, картинками и мелодиями, и эти связи также могут быть объектом изучения. Но на данном этапе мы использовали в качестве материала для анализа только тексты и в одном случае — картинки.

При анализе текста можно ограничиться лингвистическим и филологическим уровнями, что само по себе представляет интерес. Но можно попытаться за ними увидеть общество — тогда это и станет социологическим исследованием. Собственно, в этом смысле любое исследование языка и текстов может послужить базой для исследования социологического. Ниже перечислены предпринятые нами попытки получения информации об обществе из Интернета. Чисто лингвистические пункты изложены более кратко, «более социологические» — подробнее.

Чисто лингвистические (пока что) исследования посредством Интернета. Вопросы, базирующиеся на анализе частоты слов. Начнем с анализа грамотности. Проверка всего лишь нескольких десятков слов показала, что уже есть слова, которые в Интернете пишут примерно одинаково часто правильно и неправильно, т. е. по данному слову грамотность утеряна полностью (пишут случайным образом). Некоторые

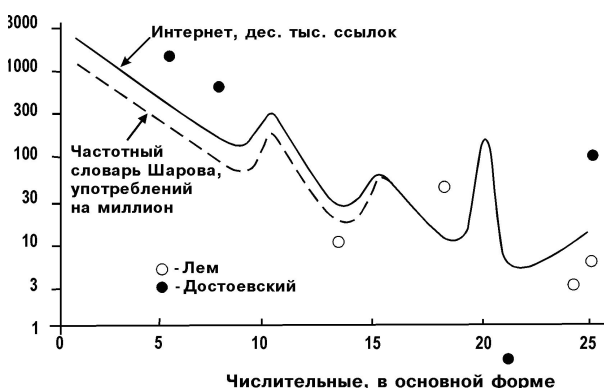
поисковые средства позволяют определять отдельно количество страниц, сайтов и запросов, на которых встретилось то или иное написание. Поэтому доля неправильных написаний может вычисляться по запросам, по страницам и по сайтам. Расчет по запросам дает наиболее чистые данные, не искаженные влиянием редакторов. Доля неправильных написаний, рассчитанная по страницам, существенно меньше, поскольку она «искажена» редакторами. Доля неправильных написаний, рассчитанная по сайтам, также искажена редакторами, но в меньшей степени, поскольку если на сайте встретилось хотя бы одно одно неправильное написание, он попадает и в «правильные», и в «неправильные». Естественным образом отслеживается зависимость грамотности от времени.

Оказалось, что за три года наблюдений грамотность пользователей Интернета не изменилась, в то время как грамотность текстов на сайтах несколько увеличилась. По-видимому, это означает улучшение редактирования. Напрашивается исследование зависимости частоты ошибок от типа ошибки и от контекста, что является уже скорее филолого-педагогической задачей. Интернет позволяет следить за словотворчеством и вообще за изменением языка. Словоупотребление вообще (и связи между словами, в частности) могут зависеть от характера сайта. Например, *data* — это данные, причем формально это множественное число, а есть и архаичное единственное — *datum*. Запросы «*data is*» и «*data are*» дают примерно одинаковое количество ссылок. Но оказалось, что «*data is*» сосредоточено на сайтах *.com* и меньше на *.org*, а «*data are*» на *.edu*, *.uk*, *.org* и немного на *.com*. Нормативная ситуация, согласно «Оксфордскому словарю английского языка», такова: это множественное число, но поскольку в компьютеринге *data* повсеместно употребляется как единственное, то в данной области это считается допустимым.

* За полезные обсуждения авторы благодарны М. Векштейн и С. Чернацкому.

Насколько далек язык Интернета от «русского письменного»? Языки можно сравнивать многими способами, для примера мы сравнили частоты числительных. На рисунке показаны частоты первых 25 числительных (в основной форме — один, два, три...). Видно, что по этому параметру язык Сети близок к письменному русскому. Но отличие заметно — в Сети употребляется относительно меньше больших чисел и больше чисел до 15. Что касается пана Станислава Лема, то он почему-то не любил число 13 и любил число 18, а Достоевский любил числа 5, 8 и 25 и избегал — могли ли мы это предположить? — числа 21 (см. рис 1).

Рисунок 1
ЧАСТОТА СЛОВОУПОТРЕБЛЕНИЯ ЧИСЛИТЕЛЬНЫХ



Исследование частот слов и выражений может применяться для определения отражения в сознании человека любых явлений, в том числе психофизиологических. Например, иногда задается вопрос о соотношении информации, получаемой человеком от зрения и слуха. Вопрос неисчерпаем, но как итоговый результат отражается в речи? Интернет показывает, что как в русском языке, так и в английском формы глагола «видеть» встречаются чаще соответствующих форм глагола «слышать» в среднем в 3-5 раз, что, как ни странно, примерно соответствует расхожему утверждению, что 80% информации человек получает посредством зрения. Но интереснее то, что эта цифра зависит от временной привязки. Если измерить относительные частоты употребления этих глаголов в цепи выражений (давно видел/слышал) — (когда-то видел/слышал) — (видел/слышал) — (в прошлом видел/слышал) — (недавно видел/слышал) — (только что видел/слышал) — (вижу/слышу) — (сейчас вижу/слышу) — (скоро увижу/услышу) — (потом увижу/услышу) — (увижу/услышу) — (когда-нибудь увижу/услышу), то окажется, что в этой цепочке, примерно упоря-

доченной по времени, отношения растут от 2 до 10. Это означает, что человек говорит об ожидании зрительных ощущений в 10 раз чаще, чем об ожидании слуховых, но помнит зрительные лишь в 2 раза чаще, чем слуховые. Действительно, говорят же, что люди чаще думают словами, а не образами. А слово легче представить себе «услышанным», а не написанным. Примерно такой же эффект наличествует и в английском языке.

Вопросы, базирующиеся на «притяжении» слов. Рассмотрим частоту встречаемости двух слов рядом или в одной фразе по отношению к частоте встречаемости этих слов на странице (возможны и другие нормировки). Вот относительные частоты для некоторых пар существительное — прилагательное. Большинство результатов довольно понятно, но нам показалось удивительным, что слова «умная» и «женщина» притягиваются в языке (и, стало быть, в сознании) намного сильнее, чем «глупая» и «баба» и чем «глупая» и «женщина». Вот пары слов, упорядоченные по убыванию притяжения: голая баба — 0,21/0,28, рыжий кот — 0,19/0,27, пьяный мужик — 0,095/0,15, умная женщина — 0,047/0,15, глупая женщина — 0,017/0,10, глупая баба — 0,016/0,040, умная баба — 0,014/0,032, трудолюбивый студент — 0,0070/0,022, добрый экзаменатор — 0,0067/0,018, честный политик — 0,0057/0,035.

Другой вариант пар — это употребление имени собственного вместе с названием класса в качестве пояснения. В качестве объекта мы взяли названия восьмидесяти тысячников и посмотрели, часто ли они употребляются вместе со словами «пик» и «вершина». И тут обнаружилось нечто неожиданное. Понятно, что Джомолунгма употребляется как «вершина Джомолунгма» и «пик Джомолунгма» редко — в 2% случаев, ибо само это название употребляется часто (14 тыс. ссылок) и хорошо известно. Но остальные названия, употребляемые все с примерно одинаковой частотой (от 540 до 970 ссылок) резко разделились на две группы. Аннапурна, Лхоцзе и Чо-Ойю употребляются с «пик» или «вершина» в 31 — 46% случаев, а все остальные — в 6,0—12% случаев.

Посредством Интернета можно изучать и явные связи между словами, например, устойчивые словосочетания, отношения сравнения (быстрый, как свет), уточнения (красный, точнее, малиновый), приближения (худой, почти тощий). Например, «не промахнись» в 10% случаев входит в «не промахнись, Ассунта»,

«промахнулся» – в 8% случаев входит в «Акела промахнулся», причем «Акела» входит в «Акела промахнулся» в 40% случаев. «Быстрый» – это в 9% случаев «быстрый, как ветер», в 2% – как свет, в 1% – как олень.

Третий вариант пары – два существительных. Предположим, что утверждения «А есть В» и «В есть С» верны. Тогда будет верно утверждение «А есть С» – это свойство называется транзитивность. Но транзитивна ли лексика? Понятие транзитивности для лексики не определено, но на интуитивном уровне его можно понять так: если выражения «А есть В» и «В есть С» распространены, то транзитивность означает, что будет распространено выражение «А есть С». Проверим это.

Утверждение «женщина – судьба» встречается в Сети 480+140 раз (здесь и далее вторая, естественно, меньшая цифра – с перестановкой слов), утверждение «судьба – злодейка» встречается в Сети 32.000+9.500 раз, т. е. лимитирующей стадией является первая, но утверждение «женщина – злодейка» – лишь 5+0 раз. Правда, тут еще дело может быть в том, что по отношению к «злодейке» слово «женщина» зачастую избыточно: если понятно, что речь идет о человеческом взрослом существе, то, сказав «злодейка», совсем не надо уточнять «женщина». Не изменяет ситуацию замена «злодейки» на «индейку». Утверждение «судьба – индейка» встречается 9.400+48 раз, а утверждение «женщина – индейка» (в соответствующем смысле) 0+0 раз. То есть лексика в указанном смысле глубоко не транзитивна.

Мы переходим к вопросам, отчасти связанным с социологией.

Чем гордится русскоязычный Интернет? Предположим, нам надо охарактеризовать какую-то вещь, явление или человека. Это можно сделать относительно некой общей шкалы, сказав, что эта вещь красива или что эта вещь уродлива. Реже это делают посредством сопоставления: эта вещь красивее другой, или эта вещь уродливее другой. При этом вещь, о которой человек рассказывает, обычно ставят на первое место. У прилагательных, обозначающих качества, имеются антонимы, и в этом случае отношение частот употребления выражений типа «красивее, чем» к «уродливее, чем» будет корреспондировать с важностью, «силой» этой хвалебной характеристики в группе людей, пишущих в Интернете и для Интернета. Вот спи-

сок отношений частот для распространенных прилагательных:

активнее, чем/пассивнее, чем = 410;
красивее, чем/безобразнее, чем + уродливее, чем = 330;
прикольное, чем/обыденное, чем = 290;
ярче, чем/тусклее, чем = 190;
глаже, чем/шероховатее, чем = 150;
величественнее, чем/низменнее, чем = 69;
смелее, чем/трусливее, чем = 60;
быстрее, чем/медленнее, чем = 6,2;
сильнее, чем/слабее, чем = 5,0;
веселее, чем/скучнее, чем + печальнее, чем = 4,3;
богаче, чем/беднее, чем = 2,7;
подвижнее, чем/медлительнее, чем = 2,7;
умнее, чем/глупее, чем = 2,6;
угловатее, чем/округлее, чем = 2,6;
добрее, чем/злее, чем = 2,5;
мужественнее, чем/женственное, чем = 2,2;
громче, чем/тише, чем = 2,0;
страшнее, чем/безопаснее, чем = 1,6;
оригинальнее, чем/банальнее, чем = 1,6;
больше, чем/меньше, чем = 1,2;
нежнее, чем/грубее, чем = 1,2;
темнее, чем/светлее, чем = 1,2;
сложнее, чем + тяжелее, чем/легче, чем = 1,2;
лучше, чем/хуже, чем = 1,0;
длиннее, чем/короче, чем = 1,0;
могучее, чем/хилее, чем = 1,0.

Мы видим группу основных положительных характеристик с отношением частот более 60 – это «активный», «красивый», «прикольный», «яркий», «гладкий», «величественный», «смелый». Русскоязычный Интернет смотрит на мир как юная девица на ухажеров. Далее – мертвая зона шириной в порядок (от 60 до 6), и потом все остальные в зоне малых предпочтений (от 6 до 1). В эту зону попали «умный», «добрый», «сильный», «мужественный», т. е. существенные характеристики. Форма оказалась важнее содержания. Странно, но сюда попал и «большой» – притом, что некоторые культурологи считают, что в русской культуре существует культ «большого».

Распространение материалов в Интернете: метод анализа «жарености». Часто ли в Интернете «таскают» материалы? Возможность для такого исследования создает, например Google, поскольку он делит ссылки на «наиболее значимые» и «очень похожие на них». Оказалось, что отношение общего количества ссылок (те и те вместе) к количеству оригиналь-

ных ссылок (такowymi мы считали «наиболее значимые») изменяется в широких пределах, по крайней мере — от 1 до 230. Вот некоторые примеры, причем первое число — количество оригинальных ссылок, второе — общее количество. «Квазигруппа» — 20/40, «хиггсовский бозон» — 30/230, «асимптотическая свобода» — 50/150, «поле температур» — 100/140, «тахион» — 250/1.000, «фуллерит» — 250/800, «поле скоростей» — 300/800, «липосакция» — 500/6.000, «нуклон» — 500/3.000, «похудание» — 500/16.000, «гравитационное поле» — 600/7.000, «Майкрософт» — 700/70.000, «кристаллография» — 700/5.000, «ожирение» — 700/46.000, «магнитное поле» — 750/29.000, «электрическое поле» — 800/10.000, «кристалл» — 800/190.000.

Понятно, что с увеличением общего интереса к теме должны расти и количество оригинальных ссылок, и «коэффициент размножения» — отношение общего количества ссылок к количеству оригинальных, поскольку интерес влечет как самостоятельную работу, так и таскание чужого. Но заметен существенный разброс коэффициента размножения при одинаковом или близком количестве оригинальных ссылок, указывающий на некую «жареность» темы. Например, сравните «квазигруппу» с коэффициентом 2 и «хиггсовский бозон» с 8, или «нуклон» с коэффициентом 6 и «похудание» с 30, или «кристаллографию» с ничтожным 8 и одно из любимых заклинаний рекламщиков — «кристалл» — с «коэффициентом жарености» 230.

Исследование самих ссылок показало, что лишь часть из них является ссылками на тот же документ, но лежащий на другом сайте, причем в эту группу входит выкладывание документа второй раз как законное, например, на сайте-зеркале, так и пиратское. Причем вручную провести этот анализ можно только при относительно небольшом количестве ссылок.

Исследование эмоций человека. Например, некоторые слова, оканчивающиеся на гласную букву («дура», «мышка», «убью») при их Интернет-употреблении появляются с повтором оной гласной («дурааа», «мышкааа», «убьюуюю»). Зависимость количества употреблений от количества повторов, очевидно, связана с эмоциями, выражаемыми респондентом в момент ввода. Исследоваться могут как спектр в целом, так и различные функционалы от него — интеграл, среднее количество повторов и другое. Вот что показала, например, элементарная проверка. Слова «кошка», «мышка» и «лошадка» вызывают совершенно разные эмоции — при собственных частотах соответ-

ственно 7.800.000, 3.700.000 и 1.400.000 интегралы повторов будут соответственно 5.000, 200 и 20.000, т. е. относительные интегральные эмоциональности составят соответственно 60,5 и 1400 стотысячных, иными словами «лошадка», эмоциональнее «мышки» в 300 раз.

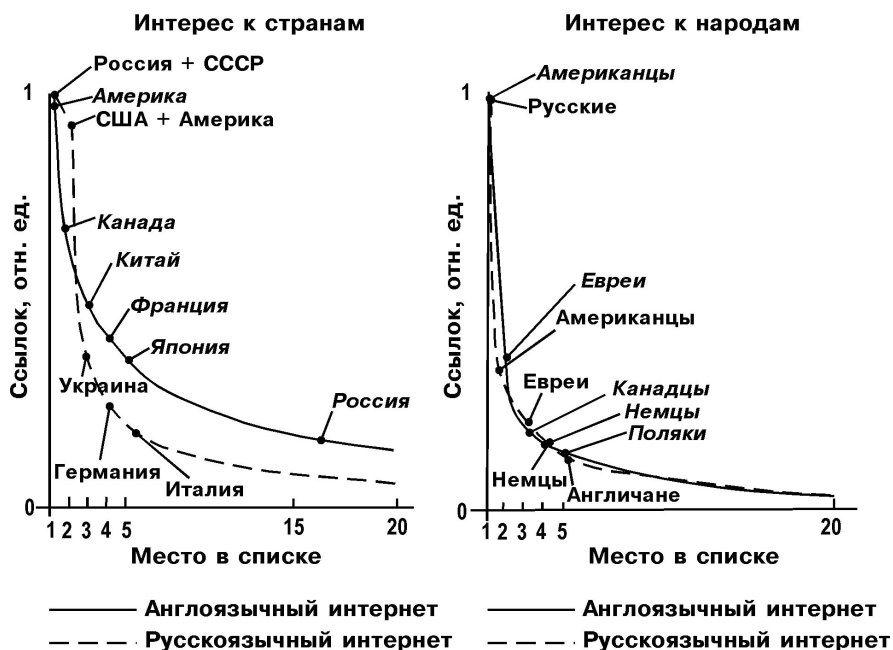
Похожая картина с инвективами: частоты слов «дура», «идиотка», «кретинка» и «дебилка» соответственно 2.500.000, 220.000, 18.000 и 13.000, интегралы повторов соответственно 5000, 40, 1 и 10, т. е. относительные интегральные эмоциональности составят соответственно 200, 20, 5 и 100 стотысячных — «дура» эмоциональнее «кретинки» в 40 раз. При наличии достаточной статистики могут быть исследованы и другие характеристики, но похоже, что имеются два пика — в районе 4-5 и в районе 10 повторов последней буквы.

Мы переходим к вполне социологическим аспектам.

Интерес к странам и народам. Рассмотрим частоты употребления в русскоязычном и англоязычном Интернете названий стран и народов. Если расположить страны в порядке убывания частот упоминания в Интернете и нарисовать зависимость частоты от места в списке (ранга), получатся вот такие кривые (за 1 принята частота упоминания «России» и «русских» в русскоязычном Интернете и «США» и «американцев» — в англоязычном). Глазом, вооруженным Google'ом, видно, что англоязычному населению Интернета (3/4 его — американцы) окружающие страны более интересны, нежели русскоязычному, и народы интересны в той же мере. Так что тупо повторяемая фраза, что американцам ничто не интересно, кроме самих себя, не больше, чем просто тупая фраза. Данные в среднем удовлетворяют закону Ципфа (частота обратно пропорциональна рангу), но в русскоязычном Интернете аномально высок интерес к США — почти так же высок, как к родной стране.

Причем если посмотреть на соотношение интереса к стране и ее народу, то видно, что есть ситуации, когда страна интереснее народа, а бывает и наоборот. Например, в англоязычном Интернете в сторону страны смещен интерес к Китаю, Франции, Японии и Индии (т. е. интерес носит политический характер), а в сторону народа — для Польши, России, Греции (культурный интерес). В русскоязычном Интернете «политичен» (а вернее, «туристичен») интерес к Италии, Кипру, Испании, Турции, Индии, Болгарии, а «культурен» — к Англии, Японии, Китаю, Германии.

Рисунок 2
ЧАСТОТА ССЫЛОК НА СТРАНЫ И НАРОДЫ МИРА



Ощущение своего места в мире может проявляться и иначе. В русскоязычной литературе нам дважды встречались утверждения, что в некоей стране карты мира нарисованы так, что данная страна находится в центре картинке. Мы встречали такие утверждения про США и Новую Зеландию. Простейший анализ не только показывает, что это вранье, но и позволяет обнаружить интересные различия между странами, и возможно, между типами сайтов. Карта мира на сайтах Новой Зеландии расположена Новой Зеландией вперед в 30% случаев, на сайтах Австралии – в 40 (эта разница при объеме выборки 100 сайтов не существенна), а на сайтах Японии – в 80% случаев. В остальных случаях Земля изображена Европой вперед. Поэтому можно предположить, что половина разницы – эффект ментальности, а остальное составлено из ментальности и удобства пользования. Что же касается США, то на сайтах .com, .org, .edu и .gov Земля изображена Америкой вперед в 2, 3, 5 и 8% случаев соответственно (в целом – около 2% случаев), в остальных случаях – Европой вперед.

Отношение к национальностям. Отношение к тем или иным национальностям, складывающееся в обществе, представляет интерес по меньшей мере в трех аспектах. Во-первых, на уровне контактов «общество-общество»: трудно ожидать нормальных отношений между странами А и В, если жители страны А ненавидят на-

селяющий страну В народ. Во-вторых, на уровне контактов «личность-общество»: отношение человека к инициативе, исходящей из страны В, будет зависеть от его отношения к национальности, населяющей страну В. В-третьих, на уровне контактов «личность-личность»: отношение человека к другому человеку будет (по крайней мере, на начальном этапе) зависеть от его отношения к национальности контактанта, который, будучи осведомлен о этом отношении, может знанием воспользоваться, причем несколькими способами.

Методы исследования отношения можно разделить на три группы – это исследования действий, спонтанных высказываний и ответов на вопросы социологов. Каждая группа методов имеет свои плюсы и минусы. Например, отношения, выраженные в действиях (вандализм на кладбище или усыновление ребенка иной национальности, нежели приемные родители), хороши своей юридической достоверностью, но плохи малой статистикой и низкой разрешающей способностью. Кроме того, в некоторых случаях действия («он на него косо посмотрел») трудно фиксируемы. Ответы на вопросы анкет хороши тем, что сами вопросы могут быть тщательно подготовлены, многократно оттестированы и используются на протяжении длительного времени, но плохи тем, что сама ситуация анкетирования является искусственной и в ряде случаев толкающей респондента на сознатель-

ную или бессознательную корректировку своих высказываний. Анализ спонтанных высказываний хорош тем, что «респондент» находится в естественной обстановке и в меньшей степени цензурирует себя, но плох тем, что сам анализ в этом случае сложнее анализа ответов на анкетные вопросы.

По мере развития Интернета роль анализа спонтанных высказываний будет, как нам кажется, возрастать хотя бы по двум техническим причинам. Во-первых, доступ к Интернет-версиям печатных изданий способен сделать анализ текстов менее трудоемким. Во-вторых, наличие в Сети многочисленных форумов, блогов, гостевых книг и т. п. с их речью, приближенной к живой устной, также увеличивает количество материала для анализа. Каковы вообще могут быть методы анализа текстов?

Первый метод – это традиционный контент-анализ, установление того, с чем ставится в один ряд наш объект или какие его связи рассматриваются, насколько достоверно и глубоко его анализируют, как характеризуется объект и т. п. Опыт применения контент-анализа российской прессы для исследования отношения к национальности имеется¹. Контент-анализ требует содержательного анализа текста человеком, что делает методику трудоемкой. Кроме того, анализ, проводимый человеком, уменьшает объективность, хотя для отслеживания трендов и сопоставления разных изданий метод пригоден.

Второй метод – это вычленение из текстов всех упоминаний заданной национальности с прилагательными (или иными характеристиками) и анализ полученного материала. При этом немедленно возникает вопрос: какие вообще характеристики рассматривать? Можно взять именно те характеристики, которые употреблялись в анкетах социологов (в том или ином исследовании) и сравнить частоты выбора этих характеристик в опросе и названия этих характеристик в Интернете.

Этот подход оказался, однако, неудобен, потому что частота употребления эпитетов в спонтанной письменной речи (речи Интернета) регулируется своими законами: например, язык избегает сложных конструкций и слов, а «социологические» слова при формулировке анкеты с закрытыми вопросами выбираются исходя из требующейся точности зондирования того или иного отношения в условиях опроса. В результате «социологические» слова в Интернет-речи

используются в десятки и сотни раз реже, чем прочие. И, несмотря на необъятность Интернета, сочетание названия конкретной национальности с конкретной характеристикой может встречаться слишком редко для того, чтобы данные оказались репрезентативны.

Можно было бы использовать для поиска те слова, которыми респонденты характеризуют различные национальности в открытых вопросах анкет, но такие данные известны только учащимся московских национальных школ (В.Д. Шапиро, М.Г. Герасимова) и жителям Перми (О.Л. Лейбович, В.Н. Стегний, А.Н. Кабацков, Н.В. Шушкова)². Кроме того, спонтанная письменная речь Интернета отличается от спонтанной устной речи ответов на открытые вопросы анкет.

Можно пойти не по социологическому, а по психологическому пути: взять какой-нибудь стандартный список характеристик человека и собрать статистику в Интернете. Но проблема, аналогичная той, что возникает при анализе «социологических характеристик», возникла и здесь.

Поэтому мы выявили все характеристики, которые использует Интернет-речь для характеристики четырех национальностей (англичане, азербайджанцы, русские, евреи). На данном этапе мы игнорировали наличие синонимичных названий национальностей, как нейтральных, так и окрашенных, хотя в ряде случаев они составляют заметную долю от основного словоупотребления. Так, около 50% составляют «британцы», «британец» и «британка» от «англичан», «англичанина» и «англичанки» соответственно, около 25% – составляют «чурки» от «азербайджанцев» и «жид» от «еврея». Разделение высказываний на оценки и самооценки не делалось, не анализировались демографические характеристики авторов высказываний, хотя считается, что состав пользователей Интернета сейчас несильно отличается от всего населения, но характеристики авторов текстов и, тем более, авторов высказываний на форумах могут отличаться сильнее.

Измеренные частоты были нормированы на общую частоту употребления наименования данной национальности в Интернете. Таким образом получились относительные частоты употреблений всех эпитетов (случаи использования слова «русские» как прилагательного исключались). Для примера приведем по семь наиболее частых характеристик каждой национальности (в скобках – относительные частоты).

¹ Ашкинази Л.А., Гайнер М.Л. // Вестник Еврейского университета в Москве. 1995. № 1; www.jewish-heritage.org/kipi2.htm

² См.: http://www.hse.ru/journals/wrldross/vol04_2/ogl.htm

ты × 106). Англичане – «сдержанные» (2300), «странные» (1000), «хитрые» (630), «богатые» (540), «умные» (290), «глупые» (270), «наивные» (240); азербайджанцы – «забавные» (470), «богатые» (380), «хитрые» (280), «умные» (130), «культурные» (120), «наивные» (67), «честные» (56); русские – «тупые» (750), «умные» (470), «горячие» (420), «надежные» (360), «глупые» (300), «богатые» (280), «честные» (170); евреи – «религиозные» (4600), «умные» (900), «хитрые» (550), «богатые» (390), «честные» (290), «жадные» (230), «глупые» (210). Данные списки не могут считаться исчерпывающими в силу некоторых особенностей работы поисковых программ.

Эти списки можно использовать разными способами. Например, можно для каждой пары национальностей из списка вычислить «расстояния» как сумму модулей разностей относительных частот употреблений всех эпитетов. Эти расстояния оказались следующими (в порядке возрастания «расстояния», × 106). Азербайджанцы – русские: 3000; англичане – азербайджанцы: 5600; англичане – русские: 6300; азербайджанцы – евреи: 7200; русские – евреи: 7700 и, наконец, англичане – евреи: 10000. Такая же

процедура может быть проделана для наименований не народов (англичане, русские...), а лиц (англичанин, русский... англичанка, русская...).

Мы использовали также следующий, чисто качественный, подход. Для названных выше национальностей определялись наиболее часто употребляемые прилагательные и глаголы и делалась попытка составить примитивный текст на основе только этих прилагательных и глаголов, причем с соблюдением ранга характеристики. Вот что получилось.

Богатый и старый, но неутомимый англичанин изобрел и выставил на аукционе нечто такое, за что получил много денег и решил – вот моя любимая молодая и типичная англичанка. Еще вчера она родила, отбилась от грабителя и отомстила. Тут, откуда ни возьмись, молодой, рослый, типичный азербайджанец, который зарезал и изнасиловал, несмотря на присутствие рядом молодой, настоящей, русскоязычной азербайджанки, а потом взял что-то и говорит – я продал это старому, бедному, нормальному ортодоксальному еврею, который купил, сел и сказал – один я остался! Где моя старая, чистокровная, галахическая и красивая еврейка, которая когда-то пришла, сказала, вышла замуж и родила, где великий

Таблица 1

ЧАСТОТНОСТЬ ИНВЕКТИВНЫХ НАИМЕНОВАНИЙ В ИНТЕРНЕТЕ (приведены данные только для национальностей с максимальным инвектированием более 0,01)

Национальность	Инвективные наименования*	Более редкие инвективы	Относительная частота употребления инвективного наименования			
			Ед. ч. Муж. р.	Ед. ч. Жен. р.	Мн. ч. Муж. р.	Мн. ч. Жен. р.
Украинец	Хохол		0,58	0,39	0,3	0,42
Американец	Америкос Пиндос	Янки	0,074	0,001	0,085	0,001
Азербайджанец	Азер	Айзер	0,012	0,007	0,05	0,7
Еврей	Жид	Пархатый, Жидомасон	0,22	0,083	0,21	0,1
Эстонец	Чухна		0,013	0,016	0,14	0,044
Армянин	Хачик		0,092		0,033	
Итальянец	Итальяшка	Макаронник	0,025	0,0008	0,033	0,001
Японец	Узкоглазый Япошка		0,012	0,0016	0,06	
Чеченец	Чех	Чичик	0,032		0,2	
Немец	Фрицы Немчура Гансы				0,029	
Латиноамериканец	Латинос		0,06		0,026	
Китаец	Китаез		0,02		0,007	
Негр	Ниггер черномазый	Нигер	0,1	0,01	0,19	
Афганец	Дух		0,12		1,0	

* Более 80% всех случаев.

и одинокий русский брат? Тот ответил – вот он я: родился, женился, сел и написал. А еврей – нет, это я сделал! Русский – зато вот моя молодая, новая и простая русская. Когда она вышла за меня, она сказала мне кое-что, но не бросила, как азербайджанка, и не инсценировала – хотя, как и та, родилась.

Далее наша задача состояла в определении частот инвективных наименований национальностей в Интернет (см. табл. 1). Под относительной частотой мы понимаем отношение частоты употребления всех инвективных наименований к частоте употребления названия национальности. Если конкретное слово употребляется и как инвективное наименование, и в каком-либо ином смысле («япошка» – как название ресторана, «фриц» – как имя и др.), вносилась соответствующая коррекция, «сила инвективы» не анализировалась. Отсутствие значения в таблице означает отсутствие соответствующей формы в языке или очень малую относительную частоту употребления.

Из данных таблицы следует, что, во-первых, мужчины служат объектом инвектирования почти всегда чаще женщин (7 случаев против 1), во-вторых, народ – чаще отдельных лиц (8 случаев против 3); в-третьих, в приведенном списке первые места по инвектируемости завоевали: в личном зачете – «украинец», «еврей», «афганец» и «чеченец», в командном – «афганцы», «украинцы», «чеченцы» и «евреи».

По абсолютной частоте инвектирования первые три места завоевали украинцы («хохлы» – 1.200.000 упоминаний), американцы («америкосы» и «пиндосы» – 1.100.000), евреи («жиды» – 1.000.000) Эти данные менее показательны, так как отражают одновременно и склонность к инвектированию, и общий интерес к данной национальности и, кроме того, относительно быстро меняются со временем.

Подобное исследование производилось и ранее, в 2002 г. (А.В. Моченов, С.С. Никулин, А.А. Вахин)¹. Тогда в командном зачете (личное первенство не проводилось) первые три места в Интернете взяли: украинцы, американцы, евреи, а в бумажной прессе – американцы, украинцы, евреи.

Относительные данные, т. е. частоты инвектирования, зависят не только от собственно «отношения», но и от других факторов. Например, от благозвучия и краткости как нормативного, так и инвектированного наименования. Влияет и поливалентность терминов – напри-

мер, термины «дух» и «чех» могут нести не только инвективную нагрузку, но и милитаристскую, причем частота второго употребления будет зависеть от состояния военных действий, а для исторических событий – от наличия, например, юбилеев или иных празднеств. Влияет на частоты и сопутствующая инвектированию агрегация как нескольких национальностей, так и их инвективных наименований («урюки», «чурки», «черные»).

Естественным следующим шагом являлся поиск корреляции наименований с характеристиками, т. е., например, определение, насколько чаще американцы «являются» тупыми, нежели умными, насколько чаще «америкосы» тупы, нежели умны, и главное – отличаются ли эти отношения для американцев и «америкосов». Наличие корреляции укажет на «осмысленное» употребление инвективного наименования, а отсутствие – на привычное, клишированное употребление, не связанное с реальными чувствами говорящего.

Для анализа корреляции между употреблением инвективных наименований национальностей и приписываемых им характеристик были взяты наиболее часто (как в абсолютных, так и в относительных величинах) инвектируемые национальности и определены частоты употребления с положительными и отрицательными эпитетами. Проверено 7 национальностей, 6 пар эпитетов, прямой и обратный порядок слов, контекст контролировался, и соответствующая поправка вводилась. К сожалению, достаточная для каких-либо выводов статистика набралась только для американцев, украинцев и евреев. Полученные данные приведены в таблице 2.

Из таблицы видно, что некоторые характеристики не зависят от употребления инвективного наименования самой национальности: например, мнение о трусости и злобности американцев не зависит от того, назвал их говорящий «американцами» или «америкосами». Некоторые же характеристики зависят очень сильно. Например, убежденность в хитрости украинцев возрастает в 400 раз, когда их называют «хохлами», в лживости американцев в 70 раз, когда их называют «америкосами», в тупости и хитрости американцев, хитрости и злобности евреев, наглости украинцев – в соответствующих случаях – в 20 раз. В среднем же отношение частоты употребления уничижительных характеристик к неуничижительным возрастает при переходе к инвективным наименованиям в 4–5 раз.

¹ См.: www.smi.ru/02/06/28/548258.html

Таблица 2

ЧАСТОТА УПОТРЕБЛЕНИЯ УСТОЙЧИВЫХ ЭПИТЕТОВ, ХАРАКТЕРИЗУЮЩИХ «НАЦИОНАЛЬНЫЕ ОБРАЗЫ»

Характеристики	Американцы	Америкосы, пиндосы	Украинцы	Хохлы	Евреи	Жиды
Тупые/умные	5,2	100	1,4	14	0,08	1,2
Трусливые/смелые	16	12	1,6	>5	2,2	>10
Злые/добрые	1,2	1,9	0,34	1,5	0,48	>10
Наглые/культурные	6,4	22	0,44	>100	1,0	3,0
Лживые/честные	0,012	0,8	0	0,33	0,41	1,2
Хитрые/простые	0,073	1,6	0,3	98	0,87	19

Исследование доверия к валютам. В период с ноября 2008 г. по апрель 2009 нами было осуществлено исследование степени относительного доверия граждан к валютам (доллару США, евро и рублю). Как источник информации использовался рынок сдачи в аренду и продажи жилого и нежилого фонда, т. е. четыре рынка. Индикатором – «индексом доверия к доллару (евро)» – было отношение количества предложений по той или иной цене в долларах (евро) к количеству предложений по близкой цене в рублях. Поскольку количество предложений немонотонно зависит от цены, имея выбросы на круглых цифрах предложений (например, отношение количеств предложений по \$ 10.000 и \$ 9.000, или \$ 10.000 и \$ 11.000 составляет около 2, то же – 200.000 руб. и 180.000 руб. или 200.000 руб. и 230.000 руб.), для увеличения статистики использовались круглые цифры. Анализ накопленных данных показал следующее.

Рынки аренды и продажи как жилого, так и нежилого фонда инерционны – быстрая реакция на колебания курса валют отсутствует.

Тренд индексов доверия к валютам за указанный период невелик – например, на рынке аренды жилого фонда (этот рынок мы изучали наиболее долго) индекс доверия к доллару в нижней части ценового диапазона возрос с ноября 2008 г. к январю 2009 в 2-3 раза, оставаясь много меньше 1, далее ситуация не изменялась. Иными словами, эти индексы отражают долговременные (характерное время – месяц) тренды социальных настроений.

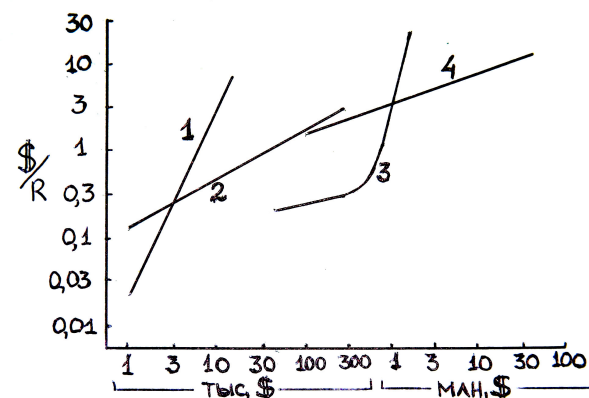
Индексы доверия сильно зависят от суммы сделки. На эти зависимости влияют два фактора – рациональный и иррациональный. Рациональный – это удобство дальнейшей работы с вырученными средствами (вложения или накопления), иррациональный – психология, привычки субъекта.

Для рынка нежилого фонда характерный диапазон цен для аренды от \$1 тыс. до \$300

тыс. в месяц, для продажи от \$100 тыс. до \$30 млн. Диапазоны цен перекрываются, а зависимость оказывается общей, т. е. субъекту (фирме, частному лицу), с точки зрения выбора валюты сделки, не важно, сдает он или продает. Индекс доверия изменяется во всем диапазоне (4,5 порядка) в двойных логарифмических координатах линейно, увеличиваясь примерно от 0,1 до 10. Арендодатели, сдающие в аренду нежилой фонд за \$1 тыс. в 9 случаях из 10 предпочитают рубль, в районе \$100 тыс. любовь к рублю и доллару оказывается одинаковой, а в верхней части ценового диапазона, продавая недвижимость за \$30 млн., в 9 случаях из 10 цену выставляют в долларах. Поскольку потери при конвертации и относительно невелики и могут быть заложены в цену, эта разница, по-видимому, отражает долю тех, кто предполагает деньги пускать в оборот не сразу же, а в течение какого-то времени и прогнозирует устойчивость соответствующих валют на этот срок.

Рисунок 3

ИНДЕКС ПРЕДПОЧТЕНИЯ ВАЛЮТ (рубли, доллары)



1. жилой фонд, аренда.
2. нежилой фонд, аренда.
3. жилой фонд, продажа.
4. нежилой фонд, продажа.

Для рынка жилого фонда ситуация иная. Во-первых, основные диапазоны цен на аренду

и продажу в общем не перекрываются: цены на аренду лежат в диапазоне от \$1 тыс. до \$20 тыс. в месяц, для продажи — от \$50 тыс. до \$2 млн. Во-вторых, индекс доверия к доллару при аренде изменяется существенно сильнее — от 0,02 до 10, а при продаже, хоть и в меньших пределах, — от 0,2 до 10, но зато нелинейно. Возможно, что более сильная зависимость объясняется более простой моделью работы физических лиц с деньгами — они либо предполагают на это жить, либо хранить, причем хранить предполагают в долларах.

По индексу доверия к евро (Е) статистики мало. При аренде жилья в диапазоне цен от 1.000 Е до 15.000 Е он растет примерно на два порядка — от 0,008 до 0,8.

Исследование сайта знакомств www.mamba.ru.

Была исследована связь возраста субъекта объявления и запрошенных параметров объекта. Основное поле наблюдений — Москва, объявления как с фото, так и без.

Некоторые наблюдения сделаны для «не Москвы» и объявлений с фото. На основном поле обнаружены следующие зависимости.

С увеличением возраста субъекта доля объявлений, в которых вообще оговорен возраст объекта, изменяется немонотонно. Для самых молодых он составляет 30% для девушек и 10 для юношей, достигает максимума в районе 30–50 лет — 75% для девушек и 45 для юношей — и немного уменьшается для больших возрастов. Такая зависимость объясняется меньшей предусмотрительностью молодых («А, на месте разберемся!») и некоторой долей платных предложений (см. ниже).

Минимальный и максимальный приемлемый возраст (естественно, не сами значения, а средние по многим объявлениям) растут с возрастом респондента, но, естественно, медленнее, причем для респондентов-мужчин медленнее, чем для респонденток-женщин. Возраст «кроссовера», когда середина запрошенного диапазона совпадает с возрастом респондента,

для женщин составляет около 50 лет, и лишь более пожилые соглашаются на сверстников, тогда как для мужчин «возраст кроссовера» — около 25 лет.

С изменением возраста субъекта изменяется диапазон приемлемых возрастов, при этом для субъектов-женщин диапазон устраивающих их возрастов мужчин несколько сокращается (от 14 до 10 лет), а для субъектов-мужчин диапазон приемлемых возрастов женщин несколько расширяется (за счет сохраняющегося согласия на молодых партнеров).

Функция распределения субъектов по диапазону приемлемых возрастов для самых молодых девушек 20 и 25 лет бимодальна: один максимум приходится на диапазон 4–9 лет, второй — на 24–29. Пик, относящийся к большим возрастам, собирает около 20 % всех субъектов — т. е. 20 % молодых девушек декларируют, что для них возраст партнера малозначителен. Среди девушек больших возрастов и всех мужчин такие ситуации тоже имеют место, но там их значительно меньше.

Для запроса «с фото» функция распределения (и интеграл) для младших возрастов увеличивается тем сильнее, чем младше возраст: для девушек 20 умножается на 2, для женщин 30 лет — на 1,3, для женщин 40 лет не меняется; для юношей 20 лет умножается на 4,5, для мужчин 30–60 лет — в среднем на 1,5.

Для «не Москвы» отличия даже для девушек 20 лет невелики — возможно, несколько возрастает доля субъектов, предпочитающих узкий диапазон возрастов.

В заключение отметим, что возможности использования Интернета для получения информации об обществе (здесь мы не касаемся очевидных возможностей по нарушению приватности, отслеживанию контактов отдельных лиц и т. п.) практически не осознаны. Серьезные исследования в этом направлении вел, кажется, только Даниил Шанцев¹.

¹ См.: folk.uio.no/dansh/real