

Спам как объект лингвистического анализа

Ф. О. Смирнов

В докладе рассматривается спам (несанкционированная рассылка сообщений по электронной почте) с позиций лингвистического анализа. Спам вызывает крайне негативную реакцию у большинства пользователей Интернета, в то же время он является одним из самых эффективных способов интернет-рекламы. В докладе представлены результаты исследования, направленного на выявление основных характеристик спама как нового типа рекламных текстов. Выделяются основные компоненты сообщения: тема, тело письма и заключительная подпись. На этой основе строится классификация эмпирического материала исследования – рекламных сообщений на английском и русском языках. В докладе выделены основные типы тем, текстов писем и заключительных подписей, которые характеризуют спам. В рамках сопоставительного анализа сделаны некоторые выводы о коммуникативных характеристиках получателей и отправителей сообщений. Полученные результаты могут стать основой для дальнейших исследований, которые имеют практическую значимость для разработки систем контентной фильтрации.

Слово “спам” (англ. “spam”) в процессе развития Сети претерпело значительную эволюцию значения. Этимология этого слова восходит к агрессивным и навязчивым маркетинговым стратегиям эпохи, предшествовавшей появлению Интернета. “Spam” – марка американских мясных консервов, прославившихся активными рекламными кампаниями. Первоначально это слово использовалось для обозначения широкого спектра явлений, характерных для электронной коммуникации. В настоящее время можно отметить некоторое сужение и конкретизацию значения этого слова. В узком смысле слова, словом “спам” обозначается рассылка нежелательных сообщений рекламного или иного характера по электронной почте (Ашманов, Власова, Зоркий, Иванов, Калинин, 2002; Spam: The Consumer Perspective, 2003). Однако достаточно привычными стали рекламные рассылки посредством других каналов электронной коммуникации (например, с помощью ICQ), а также так называемый “спам поисковых машин”.

В широком смысле слова, спам – это коммуникативная стратегия, которая ориентирована на манипулирование коммуникантом. Она связана с предоставлением нерелевантной для него информации и навязыванием ему интенций манипулятора.

У подавляющего большинства пользователей спам вызывает резко отрицательное отношение, поэтому он воспринимается как нежелательная почта (Ашманов, Власова, Зоркий, Иванов, Калинин, 2002, HTML document). В исследовании, проведенном компанией “DoubleClick” в июле 2003 года, отмечается, что спам является источником крайней фрустрации для пользователей Интернета. Согласно представленным данным, большинство американских пользователей получают более 50 сообщений рекламного характера в сутки, а 21% опрошенных признали, что доля спама превосходит 100 сообщений. Распространенность спама способствовала тому, что пользователи проводят четкое различие между “хорошей” и

“плохой” почтой (The Consumer Perspective, 2003, HTML document). Реализация спама в столь значительных масштабах заставляет все большее количество пользователей прибегать к использованию специальных программ для фильтрации поступающей электронной почты. Большинство из них весьма далеки от совершенства: любой алгоритм фильтрации удаляет вместе с нежелательными сообщениями также некоторое количество сообщений, содержащих значимую для получателя информацию. Сложившаяся ситуация указывает на вполне реальную угрозу целостности Интернета как пространства общения и соблюдению ключевых принципов электронной коммуникации. Восприятие электронной почты, одного из традиционных асинхронных жанров интернет-общения, претерпевает серьезную трансформацию в сознании коммуникантов: максимальное удобство, простота и открытость общения отходят на второй план перед стремлением оградить себя от спама.

Крайне негативное восприятие спама, организация сопротивления распространению нежелательной информации в Сети с помощью программных инструментов и усилий сетевых сообществ представляет собой примечательную тенденцию. Не менее агрессивная традиционная реклама в прессе, на телевидении и радио не встречает сопротивления, сравнимого по масштабам с борьбой со спамом.

Чтобы понять глубинные причины этой реакции, необходимо проанализировать механизм, лежащий в основе рекламного воздействия нежелательного электронного сообщения. Носителем рекламной информации является текст сообщения электронной почты. С лингвистической точки зрения, на организацию текста спама будут оказывать непосредственное влияние две тенденции: соблюдение стилистических рамок электронного письма как жанровой разновидности интернет-коммуникации и использование психолингвистических приемов построения максимально эффективного рекламного текста. Одна из указанных тенденций при построении сообщения может оказаться доминирующей. Их параллельное развитие будет способствовать созданию текста, где обе тенденции гармонизированы. Этот процесс находит свое отражение в структуре электронного сообщения, в его композиции, в выборе лексических, стилистических и прагматических средств, соответствующих интенции автора сообщения.

Особый интерес представляет сопоставительный анализ текстов спама на английском и русском языках. Такой подход дает возможность получить определенные сведения не только о характерных чертах рекламных текстов в Сети, но и об особенностях языковой личности получателя и отправителя сообщений. Эмпирической базой исследования выступили 200 сообщений электронной почты, полученных в ходе несанкционированных рекламных рассылок.

Основными чертами спама являются:

1. анонимный отправитель (указан вымышленный или сгенерированный автоматически адрес);
2. односторонний характер коммуникации, выражающийся в пренебрежении интересами получателя;
3. императивный характер значительного числа сообщений (побуждение к приобретению товаров, услуг, посещению веб-страниц и т.д.);
4. намеренное нарушение орфографии, которое не препятствует адекватному восприятию сообщения, но требует дополнительных усилий со стороны получателя (используется с целью преодолеть программные фильтры).

Каждое из сообщений образовано тремя основными компонентами: темой электронной сообщения, телом письма и заключительной подписью.

Согласно конвенциям интернет-общения электронное письмо в большинстве случаев содержит тему. Тема – это микротекст, выполняющий собственную коммуникативную функцию и обладающий определенной степенью автономности по отношению к основному тексту сообщения. Целью этого текста является привлечение внимания получателя,

побуждение его к восприятию основного текста сообщения, а в последнее время, в связи с развернувшейся борьбой со спамом как с негативным явлением – снижение бдительности получателя, маскировка, стремление ввести бдительного пользователя в заблуждение.

Тело письма несет на себе основную смысловую нагрузку в плане реализации интенции отправителя. При составлении эффективного сообщения необходимо учитывать особенности электронной коммуникации, отличающие ее от традиционного общения. Постоянный рост объемов информации предъявляет повышенные требования к краткости и лаконичности сообщения, а также к его оригинальности и нестандартности. Кроме того, новые возможности письменного текста в Сети подразумевают широкое использование графики, аудио- и видеоматериалов, особенностей гипертекста.

Заключительная подпись содержит информацию об отправителе сообщения (характерно для интернет-общения на английском языке), о возможности отказаться от получения рекламных сообщений в дальнейшем.

Результаты исследования представлены в табл. 1. Полученные данные сгруппированы по основным компонентам сообщения электронной почты.

Табл. 1 Результаты сопоставительного анализа спама на английском и русском языках.

Тема сообщения			
англ. яз. (в %)		рус. яз. (в %)	
Рекламные	50	Рекламные	37
Информационные	20	Информационные	48
Ирритирующие	18	Ирритирующие	6
Бессмысленные (сгенерированные автоматически)	12	Указание на обобщенного адресата сообщения	5
		Тема не указана	4
Тело сообщения			
англ. яз. (в %)		рус. яз. (в %)	
“Сетевой” текст	52	“Сетевой” текст	40
Традиционный рекламный текст	38	Традиционный рекламный текст	33
Ирритирующие тексты	10	Пресс-релизы	27
Подпись			
англ. яз. (в %)		рус. яз. (в %)	
Наличие реального отправителя	27	Извинения перед получа-телем	2
Возможность отказаться от подписки	59	Возможность отказаться от подписки	10
Disclaimer	21		

В рамках анализа темы сообщений можно выделить несколько основных категорий. Самым распространенным типом тем для русскоязычного интернет-общения является информативный, в котором содержится указание на вид товаров или услуг, рекламируемых в сообщении. Часто тема может содержать указание на их производителя. Например: “Кузовной ремонт”, “Заправка картриджей”, “Доставка цветов в любую точку мира!”. В

интернет-общении на английском языке этот тип представлен гораздо меньшим количеством сообщений: “Bachelor's Diploma, Master's, or PhD - No Classes Necessary”, “Diazepam.m Valium.m Xanax.x Vicodin.n”, “Cheapest VIAGRA!”. Представляется, что в отношении рекламного воздействия этот тип будет наименее эффективен: он однозначно маркирует сообщение как рекламное, значительно сужает потенциальную аудиторию.

Другим распространенной категорией тем являются ирритирующие. Они изначально ориентированы на донесение до пользователя ложной информации с целью побуждения его к восприятию основной части письма. Для этого авторы текстов активно используют общепринятые конвенции интернет-общения, знакомые каждому пользователю. Самым распространенным типом является размещение префикса Re: в начале фразы, образующей тему. Это указывает получателю, что он имеет дело с ответом на запрашиваемую информацию. Авторы сообщений спама широко используют любопытство, повышенный интерес к новой информации и нестандартный характер текста. Вот лишь некоторые примеры ирритирующих заголовков: “От Екатерины”, “Я же говорил, что это произойдет”, “Я нашел то, что ты просил”. Аналогичные приемы можно отметить и в электронной коммуникации на английском языке: “Why you forgot me? (Alice X. z.)”, “Respond by Tuesday 9am”, “Re_ Monday”.

Другим достаточно распространенным типом тем являются тексты, содержащие указание на обобщенного адресата сообщения. С точки зрения эффективности коммуникации, это тоже попытка сузить аудиторию получателей нерелевантной для них текстовой информации. Это же побуждает неопытных пользователей действительно перенаправлять, передавать сообщения указанным в теме адресатам, то есть письмо обеспечивает собственное тиражирование самими получателями рекламы. Обычно это набор стандартных адресатов: “Главным бухгалтерам”, “Секретарю, помощнику руководителя” или даже просто “Начальникам”.

Бессмысленные темы представляют собой автоматически сгенерированные тексты, которые направлены на то, чтобы рекламное сообщение успешно преодолело фильтры и не было удалено как спам.

Основную долю тем электронных сообщений в рамках реализации спама составляют рекламные. Как показывают результаты исследования, они могут базироваться на традициях внесетевого рекламного дискурса или строиться с учетом особенностей электронной коммуникации. Тексты первой группы насыщены штампами, часто используемыми в рекламе оборотами; вторую группу отличает нестандартность и новизна текстов. Например: “Хотите жить вечно?” (реклама лекарственного средства), “Как победить холод” (реклама отопительных систем), “Как бороться со спамом?” (реклама услуг по рассылке), “Бесплатная доставка”, “Лучшие предложения отдыха!”, “VIP класс почти даром”. Темы сообщений, составляющих данную категорию в англоязычном общении, представляются более стереотипными: они содержат, как правило, указания на низкую цену, возможность получить скидку, экономию или получить товар бесплатно. Можно привести значительное количество примеров: “Camel, LM, Kent 1.50 per pack!”, “Up to 50% Savings Brand name Products”, “Innovative Gifts For Everyone On Your List - PLUS Free Shipping”, “Up to 5 Audiobooks for only 99 cents! Plus a Gift!”. Таким образом, слова “gift”, “free”, “save” выступают в качестве ключевых для привлечения внимания участников коммуникации.

Говоря о теле письма, можно выделить две основные категории текстов. Речь идет о текстах, составленных с учетом особенностей восприятия информации в Сети, и о традиционных рекламных текстах. “Сетевые” тексты отличает предельная лаконичность в подаче информации, активное использование возможностей гипертекста (ссылки на более подробную информацию), стилистическая неоднородность. Не менее важной для эффективного восприятия информации является оформление текста: короткие абзацы (2-3

предложения), использование списков, выделение значимой информации с помощью графических средств. Автор такого типа текстов полагает, что языковая личность пользователя Интернета обладает своеобразием, которое необходимо учитывать при порождении высказываний. Представляется, что ключевой особенностью является скептицизм участника электронной коммуникации в отношении традиционных, конвенциональных текстов. “Сетевой” человек мало доверяет рекламным лозунгам, ценит свое время, негативно воспринимает сообщения, перегруженные художественными средствами. Оригинальность в подаче информации, поиск новых комбинаций языковых средств самого различного уровня (лексика, фразеология, стилистика), избегание штампов способны привлечь внимание пользователя Интернета.

Тексты, составленные с опорой на традиции внесетевого рекламного дискурса, напоминают рекламные модули или статьи, которые публикуются в печатных изданиях (См. Приложение 2). Эти тексты отличает: значительный объем, наличие малоинформативных фраз (штампов), оформление текста без учета особенностей восприятия информации пользователем Интернета. “Распространенный недостаток рекламных текстов – многословие, что связано с неумением или нежеланием сократить текст...” (Кохтев: 1991: с. 12).

Высокий процент англоязычных текстов, составленных с учетом особенностей электронной коммуникации (более 50%), свидетельствует о том, что отправители сообщений признали неэффективность традиционных рекламных текстов в Сети. Для интернет-общения на русском языке характерна иная картина: основную массу сообщений (60%) составляет реклама, построенная на традициях внесетевого рекламного дискурса. В качестве объяснения можно предположить, что такое положение объясняется различной степенью актуализации ценностей интернет-общения для американских и российских участников коммуникации. Вероятно, ситуация будет меняться с ростом популярности Сети как пространства коммуникации.

Обращает на себя внимание также тот факт, что большинство сообщений (59%) на английском языке предоставляют получателю возможность отказаться от дальнейшей рассылки подобных сообщений. 27% сообщений содержат в подписи к сообщению реальные координаты отправителя, а 21% - информацию, регламентирующую взаимную ответственность отправителя и получателя (англ. “disclaimer”). В 85% рекламных сообщений на русском языке отсутствуют фразы, содержащие извинения, и не предусмотрена возможность отказа от рассылки. Если американцы ценят соответствие принятым соглашениям, которые регламентируют рассылку рекламных сообщений по электронной почте (“We are fully compliant with the CAN-SPAM Act 2003”), то спам на русском языке отличается принципиально иной установкой. Достаточно часто можно встретить ссылки на узаконенный характер несанкционированных рекламных рассылок: “Целью передачи данного материала в Ваш адрес является потенциальная полезность для Вас, а не причинение беспокойства. Данная рассылка осуществлена в соответствии с ч.4 ст.29 Конституции РФ. Ваш электронный адрес был взят из открытых источников. Приносим свои извинения, если наше сообщение отняло у Вас время и причинило Вам неудобства”. Безусловно, необходимо учитывать, что в США с недавнего времени существует правовая база для борьбы со спамом. Тем не менее, представляется, что в основе такого расхождения лежит различное отношение к понятию приватности, которое находит свое отражение в электронной коммуникации. Американцы трепетно относятся к защите сферы частных интересов, куда входит также и электронная почта, в то время как русские, вероятно, значительно больше обеспокоены другими аспектами спама.

Спам как коммуникативную стратегию интернет-общения отличает высокий динамизм. Ориентация на привлечение внимания участников интернет-общения заставляет спам эволюционировать вместе с коммуникативными представлениями пользователей Интернета,

их опытом и предпочтениями. Осторожное или даже негативное отношение большинства пользователей к несанкционированным рассылкам заставляет его авторов учитывать характеристики языковой личности получателя, которые обусловлены, в том числе, национально-культурными особенностями. Лингвистический анализ спама в последнее время приобретает значительную практическую ценность. Для разработки эффективных программных средств фильтрации необходимы адекватные данные о структуре и содержании текстов, а также о коммуникативных особенностях отправителей и получателей сообщений.

Литература

1. Ашманов И. С. Власова А. Е. Зоркий К. П. Иванов А. П. Калинин А. Л. Технология фильтрации содержания для Интернета // Материалы “Диалога – 2002”. Том 2. Прикладные проблемы, 2002. URL: http://www.dialog-21.ru/archive_article.asp?param=7519&y=2002&vol=6078
2. Кохтев Н. Н. Стилистика рекламы: Учебно-метод. пособие для студ. факультетов и отд. журналистики гос. университетов. – М.: Изд-во МГУ, 1991. – 92 с.
3. Спам способствует взаимной изоляции пользователей Интернета // Компьюлента, 10.06.2003 URL: <http://www.compulenta.ru/2003/6/10/39920/>
4. Spam: The Consumer Perspective, 2003. URL: <http://www3.doubleclick.com/market/2003/10/dc/email1.htm>