

Череповецкий государственный университет  
Кафедра русского языка и общего языкознания

Курсовая работа по русскому языку

# **«Коммуникативно-прагматические признаки спама»**

*Выполнил: студент гр. 2ФР-42 Ненастьев А.*

*Научный руководитель: к.ф.н., профессор Смулаковская Р.Л.*

г. Череповец  
2004 г.

## Содержание

Содержание.....	2
Введение.....	3
Определение спама .....	3
Современное состояние изучения спама .....	5
Соотношение спама и традиционных видов рекламы.....	6
Коммуникативная модель спама.....	8
Особенности коммуникации в среде сети Интернет.....	8
Модель коммуникации спама.....	9
Тематические и структурные разновидности спама .....	14
Тематическая классификация спама .....	14
Композиция рекламного письма .....	14
Лингвистические особенности спама как разновидности рекламного текста .....	19
Лексические особенности спама .....	19
Морфологические особенности спама.....	22
Синтаксические особенности спама .....	24
Вопрос о жанре письма по отношению к спаму.....	26
Лингвистические аспекты фильтрации спама .....	27
Введение.....	27
Алгоритм Байеса.....	29
Шинглы .....	31
Перспективы дальнейших исследований .....	34
Список использованных источников .....	35

# Введение

## Определение спама

Спам как массовое и известное явление появился в довольно короткий срок. Буквально за два-три года потоки нежелательной рекламы увеличились до огромных размеров, и сейчас, по данным исследовательской компании Ferris Research, ежегодно спам обходится американским компаниям в \$9 млрд (для России такой статистики найти не удалось). А по данным консалтинговой компании Radicati Group, в 2003 году 45% всего объема мировой электронной корреспонденции составил спам. К 2007 году этот показатель возрастет до 70%. Однако есть и противоположная точка зрения: по прогнозам компании Brightmail, в 2004 году количество спама начнет уменьшаться и к 2005-2006 году снизится до 10% почтового трафика.

Вместе с тем, до сих пор не существует четкого определения спама. Проблема же такого определения далеко не праздная, поскольку от четкого выявления содержания термина «спам» зависят не только перспективы научного (в том числе и лингвистического) изучения этого явления, но и развитие юридической базы, ставящей целью законодательный запрет рассылки несанкционированной рекламы.

Тем не менее, можно привести мнения независимых экспертов, относительно признаков спама:

Н.Н. Федотов<sup>1</sup>, руководитель группы защиты информации ОАО "РТКомм.РУ":

«"Спамом" следует объявить сообщения, которые обладают четырьмя признаками:

1. рассылаются массово;
2. являются незапрошенными, то есть, рассылаются по каналам e-mail, ICQ, IRC и другим подобным, где пользователь не может исключить получение сообщений, не ограничивая себя существенно, при этом рассылаются без явного или неявного согласия получателя либо после явного выражения им несогласия;
3. содержат рекламу (в том смысле, как она определена в законе «О рекламе», то есть, в широком смысле);
4. являются анонимными, то есть, не содержат идентификации рекламораспространителя, либо содержат неверные данные о рекламораспространителе».

В. Наумов<sup>2</sup>, Санкт-Петербургский институт информатики и автоматизации РАН:

«Для целей правоприменения — чтобы определять, законно или незаконно то или иное действие, связанное со спамом можно привести перечень признаков, исходя из

---

<sup>1</sup> Подробнее см.: Спам обречён // <http://www.spamtest.ru/varticles.html?id=0011>

<sup>2</sup> Там же.

опыта иностранных государств, в которых к данной проблеме подходят по-разному. В этом перечне обязательно должны присутствовать:

1. наличие электронной рассылки по нескольким адресам или неоднократного направления сообщения по одному адресу, проводимых посредством электронной почты или иных служб передачи сообщений (если мы говорим о Сети);
2. отсутствие явного согласия на получение информации;

Имеет также значение:

3. способ получения адресов для рассылки (например, получение информации ограниченного доступа, в частности, персональных данных);
4. содержание сообщения;(В одних штатах США спамом считают только рекламные рассылки; в других к спаму относят и обычную “нерекламную” информацию);
5. наличие достоверной и полной информации об отправителе (например, в Европе планируют заставить в сообщении указывать все координаты и наименование отправителя, вплоть до адреса регистрации. В РФ это не определено.);
6. наличие специальных средств рассылок и/или подделки адресов при осуществлении спама (речь идет не о возможностях, заложенных в обычных почтовых клиентах. Имеются в виду мейл-бомберы, ПО подделки заголовков, etc. Они запрещены в ряде штатов США.)
7. адекватная реализация возможности отказа от рассылки».

Наиболее жесткое определение спама дано в «Нормах пользования сетью» (OFISP-008)<sup>3</sup>, согласно которым спамом считается:

Рассылка (иначе как по прямой инициативе получателя)

1. электронных писем и других сообщений (в том числе единичных) рекламного, коммерческого или агитационного характера;
2. писем и сообщений, содержащих грубые и оскорбительные выражения и предложения;
3. рассылка сообщений, содержащих просьбу переслать данное сообщение другим доступным пользователям (chain letters);
4. использование безличных («ролевых») адресов иначе, как по их прямому назначению, установленному владельцем адресов и/или стандартами.

Таким образом, из всех определений мы можем вычленить основные дифференциальные признаки спама:

1. массовость;

---

<sup>3</sup> Нормы пользования Сетью // <http://www.ofisp.org/documents/ofisp-008.html>

2. анонимность;
3. незапрошенность адресатом.
4. \*рекламный характер.

### **Современное состояние изучения спама**

Следует признать, что на современном этапе спам как явление, оказывающее серьезное влияние на развитие Интернета, изучен недостаточно. В основном исследования касаются технических методов фильтрации спама и лишь в редких случаях затрагивают его языковую сторону. Основным двигателем исследовательской работы является коммерческий интерес компаний-производителей комплексных решений по защите информации, а также держателей сервисов бесплатной электронной почты. Именно они в июле 2003 года основали Национальную коалицию против спама, в которую вошли семь ведущих компаний российской компьютерной индустрии («Ашманов и Партнёры», Голден Телеком, Лаборатория Касперского, Mail.Ru, Московское представительство «Майкрософт», Rambler и Subscribe.Ru).<sup>4</sup> С началом деятельности коалиции можно говорить о начале более-менее системного изучения спама, в том числе и лингвистической его стороны.

Обзор языковых исследований в области спама осложняется также и тем, что компании, профессионально занимающиеся антиспамерской деятельностью, крайне неохотно делятся информацией о тех алгоритмах и способах семантической фильтрации, которые они применяют. Это объясняется как боязнью того, что спамеры, владея подобного рода данными, смогут более эффективно обходить соответствующую защиту, так и несовершенством патентного права РФ в области информационных технологий, которое допускает довольно легкое «заимствование» оригинальных разработок конкурентами.

Главной целью любого исследования спама (и лингвистического в том числе), очевидно, является создание эффективных методов борьбы с ним, позволяющих с высокой долей точности распознавать спам в потоке разнородных писем и каким-либо образом реагировать на его присутствие. Основной трудностью в лингвистическом изучении спама, на наш взгляд, является выделение наиболее общих повторяющихся языковых структур, характерных для спама. Именно такого рода структуры и должны лечь в основу механизма автоматического его распознавания. Однако поскольку речь идет о текстах, создаваемых самыми разнообразными авторами, которые к тому же используют специальные приемы обхода лингвистических фильтров, выделение указанных схем превращается в довольно трудновыполнимую задачу. Создание же некоей единой языковой модели спамерского письма, думается, неосуществимо в принципе. Тем не менее, надо признать, что спам об-

---

<sup>4</sup> Подробнее см.: Национальная коалиция против спама // <http://www.spamtest.ru/coalition.html>

ладает большой степенью стереотипности языкового наполнения что, несомненно, должно лечь в основу его лингвистического анализа.

Данная работа есть попытка дать более-менее целостное языковое *описание* спама. Не претендуя на какие-либо окончательные выводы и не обладая всей полнотой информации (по названным выше причинам), мы ставим в качестве цели исследования изучение наиболее актуальных языковых параметров, которыми характеризуются спамерские письма. В качестве материала для работы послужила выборка из 200 писем, полученных нами на личный электронный ящик за период с апреля 2003 года по апрель 2004-го и отфильтрованных с помощью системы автоматического распознавания спама «SPAMTEST» как стопроцентный спам (добавление технического слова [!! SPAM] в конец темы письма). Цитируемые в работе отрезки рекламных писем выделяются курсивом. Исходная орфография и пунктуация сохраняются в целях повышения достоверности данных.

### **Соотношение спама и традиционных видов рекламы**

Спам — разновидность рекламы и на него распространяются соответствующие законы ее функционирования. Как справедливо замечает Е.А. Терпугова, рекламный текст имеет ярко выраженную прагматическую установку, его цель — побудить адресата совершить какое-либо действие (чаще экстралингвистическое: купить товар или услугу). В то же время «особенность дискурса рекламы состоит в том, что реклама воспринимается в едином информационном потоке, как правило, без четкой установки на восприятие. Рекламный текст принципиально не подлежит глубокому анализирующему осмыслению».<sup>5</sup> Все эти особенности свойственны и спаму.

Если говорить об идентификации спама с каким-либо традиционным типом рекламы, то здесь следует принять во внимание несколько факторов: спам — статическая реклама, очень редко письма содержат какие-либо анимированные элементы, иногда встречается графика, но в подавляющей своей массе это чистый текст. Тем не менее, спамеры для актуализации внимания на определенных отрезках текста применяют целый ряд аттракторов, связанных с широкими возможностями шрифтового варьирования, которые предоставляет электронная почта:

- цветное выделение текста (особенно агрессивно ведет себя красный и смежные с ним цвета: очевидно, в силу своей высокой актуализирующей способности);
- выделение путем увеличения кегля шрифта (сюда же отнесем и выделение путем набора в верхнем регистре);

---

<sup>5</sup> Терпугова Е.А. Рекламный текст как особый тип императивного дискурса // Автореферат на соискание ученой степени кандидата фил. наук. — Кемерово, 2000. — с. 8.

- выделение путем модификация начертания шрифта (курсив, жирный шрифт, подчеркивание);
- выделение, связанное с заменой нейтрального фона сообщения (как правило белого) на какой-либо иной.

Разумеется, что все эти средства могут применяться как изолированно, так и в соседстве друг с другом.

Е.Е. Анисимова обращает внимание на цвет и шрифт как на два наиболее важных аттрактора в «креолизованном» тексте и выделяет несколько их функций. Для цвета такими функциями будут являться собственно аттрактивная (привлечение внимания адресата), смысловыделительная (цвет выделяет наиболее важные в смысловом отношении отрезки текста), экспрессивная (способность цвета воздействовать на эмоции человека) и символическая (способность цвета выражать абстрактные понятия). Соответственно, для шрифта будут повторяться все функции, кроме символической, и добавляется характерологическая функция, заключающаяся в том, что шрифт способен вызывать у адресата ассоциации с «определенными историческими эпохами, нациями, социальной средой, отдельными историческими личностями или современными политическими деятелями».<sup>6</sup> Отметим, что характерологическая функция в нашем случае почти незначима, поскольку по техническим причинам адресат должен оперировать лишь небольшим набором стандартных шрифтов, которые довольно нейтральны в своих ассоциациях.

Таким образом, учитывая названные особенности, можно сделать вывод о приближенности спама к некоторым видам традиционной статической рекламы, в частности к рекламе газетной и наружной. Однако при такого рода сопоставлении важно учитывать и еще один принципиальный факт: спам никак не ограничен пространственно. Письмо может содержать одно предложение, а может — целый текст (этим, кстати, во многом объясняется столь широкая популярность спама как способа продвижения товара среди рекламодателей). Следовательно, полностью отождествлять спам и указанные виды традиционной рекламы вряд ли возможно. Несмотря на довольно небогатый набор внешних средств аттракции (по сравнению, скажем с телевизионной рекламой или рекламными щитами, размер которых уже есть средство привлечения внимания), спам обладает большим потенциалом для развития своих языковых свойств, поскольку текст — основное содержание спама, именно благодаря его качеству письмо привлекает или не привлекает внимания и, соответственно, достигает или не достигает своей цели.

---

<sup>6</sup> Подробнее см.: Анисимова Е.Е. Лингвистика текста и межкультурная коммуникация (на материале креолизованных текстов). — М., 2003. — С. 59-62.

# Коммуникативная модель спама

## Особенности коммуникации в среде сети Интернет

Ключевой особенностью, важной для понимания отношений адресата и адресанта в рамках рассматриваемой темы, является анонимная природа Интернета. И.С. Шевченко, исследователь-психолог из Казанского государственного университета, пишет: «Вследствие анонимности и безнаказанности в сети проявляется и другая особенность, связанная со снижением психологического и социального риска в процессе общения, отсутствием внешней невербальной оценки — аффективная раскрепощенность, ненормативность поведения и некоторая “безответственность” участников общения. Человек в сети может проявлять и проявляет большую свободу высказываний и поступков (вплоть до оскорблений, нецензурных выражений, сексуальных домогательств), так как риск разоблачения и личной отрицательной оценки окружающими минимален».<sup>7</sup>

С некоторыми оговорками данный тезис можно применить и к рекламному тексту, функционирующему в среде глобальной сети. Спам действительно носит гораздо более неформальный характер, нежели любой другой вид рекламы: нередко встречается, например, приветствие вроде такого: *«Привет! Как дела?»*. Такие структуры, конечно, не являются фактом ненормативного поведения, скорее они выполняют контактоустанавливающую функцию, но тем не менее присутствие своеобразного «вторжения» в личную эмоциональную сферу адресата очевидно.

В другой своей работе И.С. Шевченко перечисляет некоторые важнейшие особенности сетевого общения, среди которых для нас значимой является только одна: «Добровольность и желательность контактов. Пользователь добровольно завязывает контакты или уходит от них, а также может прервать их в любой момент».<sup>8</sup> Очевидно, что спам, как и всякую рекламу, нельзя подводить под это положение, поскольку он не является межличностным общением в полном смысле этого слова: это особый тип рекламной коммуникации, к которому не всегда можно применить термины, касающиеся психологии сетевых взаимодействий. Тем не менее, соответствующий фрагмент был процитирован не случайно: как ни странно, спамеры стремятся так или иначе создать у адресата впечатление добровольности навязанной коммуникации. Происходит это путем использования в финальной части текста особых «ритуальных формул», например: *«Прошу прощения, если данное сообщение отняло у Вас Ваше драгоценное время и доставило неудобства связанные с*

---

<sup>7</sup> Шевченко И.С. Факторы динамичности самопрезентаций в Интернет-общении // <http://psynet.carfax.ru/texts/shev1.htm>

<sup>8</sup> Шевченко И.С. Некоторые психологические особенности общения посредством Интернет // <http://ncdo.levsha.ru/Pub/030D.htm>



*этой рассылкой!», «Приносим извинения, если наше сообщение доставило Вам неудобства», «Данная рассылка осуществлена в соответствии с ч. 4 ст. 29 Конституции РФ. Ваш электронный адрес был взят из открытых источников. Но тем не менее приносим извинения, если наше сообщение причинило Вам какое-то беспокойство.»* Данные примеры есть формулы речевого этикета, перенесенные из сферы обычного общения в электронную среду. Как правило это сложноподчиненные предложения с придаточным условия, присоединяемым союзом «если», в главной части которых содержится речевой штамп извинения, а в зависимой — лексемы с семантикой «уничижительности» («неудобство», «беспокойство»). На морфологическом уровне главная часть характеризуется использованием глагольных словоформ прошедшего времени, а придаточная — настоящего. Именно это подчеркивает формализованность данной структуры, поскольку какие-либо действия, связанные с неудобствами при получении или прочтении письма уже совершены, а извинения, так как они располагаются в финале текста, являются всего лишь десемантизированным штампом. Примечательно, что указанный элемент письма в последнее время встречается все реже и реже.

В качестве методологически обязательного компонента данного раздела при разговоре о сущности интернет-коммуникации следует упомянуть и довольно очевидное положение о ее текстовой природе, наиболее четко выраженное А.Е. Жичкиной: «...до сих пор большинство сред коммуникации в Интернете — текстовые. Это означает, что единственный источник информации о собеседнике в виртуальной коммуникации - это его текстовые сообщения. В процессе общения в Интернете, в отличие от реального общения, исходно полностью отсутствуют индикаторы социальной позиции человека и невербальное поведение».<sup>9</sup> Этот тезис применим и к спаму, хотя понятно, что особой роли социальная позиция или невербальное поведение адресанта в нашем случае не играют опять же в силу специфического характера рекламной коммуникации.

### **Модель коммуникации спама**

Спам — это опосредованная коммуникация внутри кибернетической модели, предложенной Клодом Шенноном. В данном случае можно выделить присутствие адресата, адресанта, канала связи (Интернет), по которому передается сообщение, и языкового кода, идентичного для обеих сторон.

Обратимся последовательно к каждому компоненту названной коммуникативной модели. Адресант — в нашем случае достоверное определение его довольно сложно, поскольку адресантом письма может являться как заказчик рекламы, так и ее распространитель.

---

<sup>9</sup> Жичкина А.Е. Социально-психологические аспекты виртуальной коммуникации // <http://club.fom.ru/books/socpsy-avc95.rtf>

Текст спама не содержит каких-либо указаний на конкретного автора, что, в принципе, соответствует логике массовой рекламной кампании. Возможно ли сконструировать некий обобщенный языковой портрет спамера? Вряд ли. Слишком большое разнообразие авторов, обладающих совершенно разным уровнем языковой компетенции, не позволяет однозначно выделить некие типизированные черты, характерные для языка спамеров.

Адресат в рамках предложенной коммуникативной модели характеризуется такой чертой, как обобщенность. Спам — массовая реклама, способная охватывать неограниченное количество получателей, что, несомненно, накладывает определенный отпечаток на ее лингвистический строй. Очевидно, что язык такой рекламы должен быть прост и доступен для понимания любому. Так, например, письмо, имеющее целью побуждение адресата к выплате некоего благотворительного взноса на реконструкцию храма, начинается такими словами: *«Руководителям предприятий, Юридическим и Физическим лицам!!!»* Довольно странное и даже комичное обращение, которое, вероятно, возникло в связи с тем, что автор, создавая текст, не смог избавиться от бремени канцелярского языка, при этом учитывая также и то, что в конечном итоге цель послания — получение денег, операции с которыми как раз и характеризуются подобного рода бухгалтерской терминологией. Это подтверждает и финальная часть сообщения: *«Настоящее письмо является основанием для оплаты одноразового целевого благотворительного взноса для Юридических лиц»*. Как видим, автор отрабатывает свой текст в своеобразном «финансовом» дискурсе, тем самым реализуя первичную коммуникативную интенцию (получение денег). На этом фоне резким диссонансом становится стиль ядерной части послания: *«С Благословением от Светлейшего Патриарха Московского и всея Руси Алексия II, к Вам обращается руководитель Благотворительного фонда и Попечительского совета Храма Святителя Николая. <...> В Ступинском районе московской области восстанавливается Иконостас в Храме Святителя Николая. Наш Фонд и Попечительский совет привлек исполнителя и небольшое количество денег. <...> мы обращаемся к Вам с благородной просьбой - пожертвовать на восстановление иконостаса Церкви Святителя Николая, любую доступную сумму денег от 100 рублей. Реквизиты приведены ниже.*

*По окончании всех работ вы сможете присутствовать на Благословении иконостаса Светлейшим Патриархом и торжественном молебне в честь благотворителей».*

В данном случае имеем столкновение двух стилей: официально-делового (начальная и завершающая часть письма) и публицистического, со своеобразным риторическим напластованием. В контексте выяснения особенностей адресанта и адресата названное письмо может быть показателем следующих явлений:

1. Вполне возможно предположить, что данный спам есть плод коллективного авторства. Основания к этому лежат в бросающейся в глаза разностилевой природе текста, что дает повод думать о первичности написания ядерной части непосредственно заказчиком рекламы и добавлении своеобразной «рамки» (неудачной в данном контексте) непосредственно рекламодателем.
2. В то же время есть база для размышлений и над тем, что автор письма предстает в единственном лице, но, как уже было сказано, имеет вполне конкретную коммуникативную цель, которая не дает о себе забыть в выбивающихся из общей тональности отрезках текста. В таком случае разного рода включения финансовой терминологии рассматриваются как псевдослучайные «оговорки по Фрейду».
3. Оставляя в стороне текстологический аспект дела, можно говорить о том, что адресат здесь действительно является обобщенным, что (в довольно специфической форме) заявляет и сам автор письма, обращаясь к юридическим и физическим лицам. Тем самым можно говорить, что письмо, скорее всего, доступно по своему языковому наполнению для всех получателей.

Приведем теперь пример сознательного отбора языковых средств в зависимости от ориентации на того или иного адресата. Типичным образцом здесь будут являться разнообразные приглашения на финансово-экономические семинары, таргетированные для специалистов соответствующих сфер. Такого рода письма могут содержать, например, такие отрезки текста: *«Целенаправленное управление ассортиментом заключается в переходе от его спонтанного разрастания к разработке ассортимента на основе глубокого анализа исходной ситуации, группировки товаров по назначению и последующей разработкой товарной концепции»*. Или: *«Эффективность деятельности персонала компании повышается за счет оптимизации системы мотивации персонала. Данная программа помогает так отстроить схемы материального и нематериального стимулирования персонала, что при уменьшении финансовых затрат увеличивается результативность и лояльность персонала. Ваш опыт дополняется передовыми технологиями действующих топ менеджеров, консультантов, тренеров»*. Очевидно, что сугубо отраслевая терминология не будет понятна широкому кругу получателей. Тем не менее, такого рода письма попадают в ящики всех пользователей Сети независимо от их профессиональной принадлежности. Здесь стоит сказать также и о том, что адресант спам-посланий не имеет ни малейшего представления о адресате (характере его деятельности, возрасте, увлечениях и т.п.), что обуславливается технической стороной дела (спамер обладает лишь набором электронных адресов, в лучшем случае дифференцируя их по территориально-доменному признаку). Таким образом, адресат в нашем случае есть своего рода абстракция высшего

уровня, обладающая такой степенью обобщенности, которая не имеет места ни в одном другом виде рекламной коммуникации.

Не менее важным при характеристике взаимоотношений адресата и адресанта будет являться мысль о том, что «создатель рекламы не может наблюдать за ответной реакцией потребителя, отслеживать результативность коммуникации».<sup>10</sup> Таким образом, модель коммуникации в данном случае характеризуется отсутствием какой-либо обратной связи между адресатом и адресантом. На практике это приводит к дублированию сообщений в течение определенного периода, видимо, с целью добиться гарантированной доставки послания получателю. Однако спамер не располагает и не может располагать никакими данными относительно восприятия своего сообщения адресатом. Единственным критерием результативности рекламной коммуникации в данном случае будет являться процент совершенных покупок того или иного товара или услуги, спровоцированных спамом, по отношению к общему числу разосланных писем.

Применительно к характеристике канала связи между субъектами коммуникации следует говорить об особенностях Интернета как среды для распространения спама. Спам является тем типом текста, который официально запрещен для функционирования внутри Сети, что накладывает особый отпечаток на построение сообщений. В исходную коммуникативную модель добавляется необходимый компонент — фильтр, или искусственно созданные помехи («шум»), препятствующие установлению взаимосвязи между адресатом и адресантом. В нашем случае роль такого рода помех могут выполнять различного типа технические заслоны и алгоритмы фильтрации, ставящие своей первичной целью недопущение спам-писем для доставки. Таким образом, коммуникативная модель, адекватная заявленной проблеме, будет выглядеть так:



Намеренные помехи, встающие на пути доставки сообщения, закономерно порождают особые модификации текста, связанные с попытками обхода механизмов фильтрации. В рамках коммуникативной модели речь идет об идентичности языкового кода, на котором

<sup>10</sup> Терпугова Е.А. Рекламный текст как особый тип императивного дискурса // Автореферат на соискание ученой степени кандидата фил. наук. — Кемерово, 2000. — с. 9.

общаются отправитель и получатель сообщения. Понятно, что таким кодом в нашей ситуации будет являться русский язык, однако он может претерпевать значительные изменения, затрагивающие его внешнюю форму. Приведем примеры: *«УНИКАЛЬ НАЯ СИСТЕМА АССОЦАТИВНО ОБРАЗНОГО МЫШЛЕНИЯ ВСЕ СТАДИИ ОБУЧЕНИЯ ОТ НУЛЯ ДО ВЫСШЕГО С ПРЕПОДАВАТЕЛЯМИ ИЗ США»*, *«Мебель напрямую от производителя!»*, *«Копировальный центр предлагает высококачественную цифровую печать, от визитки до А3+»*. В процитированных фрагментах сложности в декодировании сообщений возникают как результат сознательного искажения языкового кода с целью проникновения спама сквозь фильтры. В первом случае имеем довольно распространенный прием — добавление пробелов внутри слова, что заставляет систему автоматического распознавания спама неверно членить слова. Во втором отрезке использует другой способ — добавление избыточных буквенных символов, что преследует точно такие же цели. В третьей отрывке находим применение символов неродного алфавита (латинская буква «t» на месте русской «т»). Несомненно, что все эти напластования так или иначе снижают «удобоваримость» сообщения, подчас создавая ситуации коммуникативной неудачи.

Учитывая все сказанное выше становится понятно, что спам представляет собой уникальный способ рекламной коммуникации, характеризуемый специфическими параметрами, несвойственными другим типам рекламы: особыми отношениями адресата и адресанта, специфическим каналом связи и спонтанными флуктуациями языкового кода.

# Тематические и структурные разновидности спама

## Тематическая классификация спама

Классификация спама по темам довольно серьезно разработана ведущими специалистами в области информационной безопасности, мы позволим себе лишь продублировать ее в нашей работе. Е. Касперский подразделяет спам на следующие типы согласно преследуемым спамерами целям:<sup>11</sup>

- рекламный спам (продать товар или услугу);
- мошенничество, финансовые пирамиды, etc (вовлечение в аферу);
- платные порно-сайты и прочие ресурсы сети;
- привлечение к участию в сектах, фондах и просто попрошайничество;
- прочий спам.

В целом такая классификация отражает наиболее распространенные типы спама, бытующие в Сети, хотя, конечно, она может быть в отдельных случаях дополнена другими категориями.

## Композиция рекламного письма

Композиция спама в основном соответствует композиции типичных видов рекламного текста, в частности — газетной рекламе. Н.Н. Кохтев, анализируя рекламу на газетной полосе выделяет следующие структурные элементы:<sup>12</sup>

1. Заголовок, содержащий краткий и запоминающийся рекламный лозунг.
2. Зачин, цель которого заинтересовать читателя в дальнейшем чтении.
3. Основная часть, содержание которой может быть очень разнообразным, но в наиболее общем виде это описание преимуществ рекламируемого товара или услуги.
4. Заключение, призванное закрепить основную мысль текста, окончательно обозначить преимущества товара или услуги.

В отношении спама такая композиция также имеет место, однако есть и определенные трудности. Электронное письмо технически подразделяется на две части: служебные строки (kludges) и непосредственно тело письма (message body), в котором содержится текст. Служебные строки включают в себя, в частности такие, как: отправитель (поле «From»), получатель (поле «To») и тема (поле «Subj»). Таким образом, тема письма существует как бы отдельно от основного текста, что визуально выражается в наличии разного

---

<sup>11</sup> Подробнее см.: Касперский Е. Основные классы угроз в компьютерном сообществе 2003 года, их причины и способы устранения // [http://www.infosecurity.ru/\\_gazeta/content/040330/article01.html](http://www.infosecurity.ru/_gazeta/content/040330/article01.html)

<sup>12</sup> Подробнее см.: Кохтев Н.Н. Психология восприятия и композиция рекламы // Русская речь. — 1991. — №4. — С. 68-72.

рода графических разделителей в почтовых программах. В принципе, никакой существенной роли для структуры текста это не играет, однако уместно предположить, что тема в силу своего отдельного положения, хоть и не имеет совершенно никаких средств для графического выделения, тем не менее обладает довольно высокой «раздражающей» способностью, которую так или иначе стремятся использовать спамеры. Встречаются, например, такие темы: *«Хочешь нанять своего шефа в вахтеры?»*, *«Хочешь нефтяную вышку?»*, *«насчет вашего заказа»*, *«печать от рассвета до заката»*, *«Вам нужны сотни клиентов?»*. Понятно, что в большинстве случаев такие темы соотносятся с содержанием самого письма на довольно абстрактном уровне: так, темы в духе *«Хочешь нанять своего шефа в вахтеры?»* предшествуют письмам, посвященным разным способам заработка в Сети, которые, видимо, должны настолько обогатить адресата, что он сможет на самом деле нанять своего шефа в вахтеры. Заявленные в теме пропозиции могут существовать либо как имплицитная информация, актуализированная лишь единожды (вероятно, предполагается, что она сформирует у адресата определенную пресуппозицию, которая повлияет на восприятие всего письма) и не встречающаяся далее, либо иметь повторы в основной части письма, так это происходит, например, с темой *«Хочешь нефтяную вышку?»*, которая в дальнейшем реализуется так:

*«А еще он приобрел себе дом в самом дорогом месте большого города Большой Земли, и его соседи решили, что Бельдыев владелец сети нефтяных вышек. Как минимум.*

*И теперь для поддержки имиджа, бывший оленевод дарит каждому соседу на День Рождения по нефтяной вышке.*

*И уже никто-никто не позволяет себе дразнить Бельдыева “глупой чукчей”».*

В данном случае нефтяная вышка является своеобразным символом престижа, жизненного успеха, который заявляется с самого начала и поддерживается конкретным примером непосредственно в основной части письма.

Конечно, такие «занимательные» темы встречаются далеко не всегда, основная масса тем это:

1. Конкретные указания на тот или иной предлагаемый товар или услугу: *«типография по низким ценам»*, *«Сезонная распродажа домашних соляриев»*, *«Интернет и телефон в Москве и Подмоскowie»*, *«Питьевая вода в офис»* и т.д.
2. Указание на предпочтительного адресата письма: *«Главному бухгалтеру»*, *«В секретариат»*, *«Внимание главного бухгалтера»*, *«НАЧАЛЬНИКАМ»* и т.д. Такие заголовки возникают как следствие уже оговоренного отсутствия у спамера какой-либо информации о получателе.

3. Особый тип тем, призванных формировать у адресата своеобразную ложную presupпозицию, неверное знание о якобы имевшей когда-то место коммуникации: *«RE: Продажа автобусов и запчастей для автобусов, грузовиков и прицепов», «по поводу изготовления дверей QDq!», «По поводу Вашего заказа»* и т.п. Наиболее частотным приемом для достижения указанной цели является добавление в начало темы технического элемента «RE:», обозначающего то, что полученное письмо является ответом на какое-либо ранее отправленное адресатом. Тем самым спамеры с самого раннего этапа восприятия пытаются повысить степень «доверительности» рекламы, увеличить ее психологическую близость для реципиента. Иногда это находит свое продолжение в теле письма: *«После нашего вчерашнего разговора долго вспоминал их названия сайта.....потом полистал свои записи и нашёл! Вот [www.cklimat.ru](http://www.cklimat.ru), а телефончик (095) 506-7-107 Рекомендую..... действительно хорошие цены и есть такое, что не будет у твоего соседа, хотябы ближайшие 3 года!»*

Н.И. Клушина считает, что заголовок (исследователь имеет в виду газетный текст) «обрисовывает тему».<sup>13</sup> В нашем случае мы можем сказать, что тема (в техническом смысле этого слова) обрисовывает тему (в лингвистическом смысле этого слова).

Говоря о темах следует заметить, что спамерам несвойственно использование каких-либо приемов языковой игры, которые встречаются, скажем, в газетной рекламе. Г.С. Шалимова называет такие явления «аллюзиями»: аллюзия — «это стилистическая фигура, которая заключается в соотношении описываемого с устойчивым, общеизвестным понятием или высказыванием литературного, исторического или мифологического характера. В основе такого оформления заглавия лежит стремление пишущего не прямо, открыто выразить содержание публикуемого материала, а сделать его иносказанием, намеком, вызвав у читателя ассоциации с каким-либо явлением, событием или известным произведением и заинтересовать его, заставляя домысливать недосказанное».<sup>14</sup> Подобного рода аллюзии почти не встречаются в случае спама, вероятно, по причине довольно низкой грамотности и общекультурного уровня рекламопроизводителей. С некоторой натяжкой сюда можно отнести темы, содержащие совершенно непонятные для адресата лексемы, значение которых можно выяснить, лишь прочитав письмо до конца. Среди наших писем симптоматичной в этом плане оказалась всего одно: *«Найди в тисьме Японский Хрен и получи скидку»*. Естественно, что полисемия слова «хрен» и широкий круг коннотаций в структуре его лексического значения вводят получателя в некоторое заблуждение и соз-

<sup>13</sup> Подробнее см.: Клушина Н.И. Композиция рекламного текста // Русская речь. — 2000. — №5. — С.85-88.

<sup>14</sup> Подробнее см.: Шалимова Г.С. Форма новая — проблемы старые (об одной модели газетного заглавия) // Журналистика и культура речи. — Вып. 1. — М., 1996. — С. 80-90.



дают у него желание выяснить, что же имелось в виду, особенно учитывая наличие прагматической установки, заданной спамером — получение скидки. Зачин письма не дает никакого ответа на вопрос о «хрене», начинаясь так: *«Знаете ли Вы, что сотрудники офиса засыпают на рабочем месте на 80 % реже, если в помещении есть кондиционер?»* Совершенное противоречие между темой и содержанием письма являются очень своеобразным стимулом для прочтения *всего* сообщения в поисках решения проблемы. Вероятно, составитель спама учитывал это, поместив в финальной части такие строки: *«Наконец, кондиционер General оснащен двойным фильтром от бактерий, один из которых обрабатывают веществом, полученным из японского хрена “васаби”, известного своими бактерицидными свойствами»*. Загадка, наконец, разрешается и далее следуют контактные телефоны фирмы, продающей кондиционеры. Тема письма в анализируемом случае является важнейшим компонентом, приспособленным для «зацепления» внимания воспринимающего и поддержания его (внимания) на высоком уровне в процессе чтения основной части, которая, в принципе, не отличается никакой оригинальностью по сравнению с любым другим спамом.

Таким образом, возвращаясь к проблеме делимитации рекламного письма, попытаемся расчленить его следующим образом:

1. Служебные поля, которые существуют отдельно от тела письма и включают наименование отправителя, получателя и тему, особенности которой оговорены выше.
2. Тело письма, включающее в себя:
  - повтор темы;
  - зачин;
  - основная часть;
  - заключение;
  - ссылка на интернет-ресурс или контактные телефоны (адреса);
  - блок с извинениями за «неудобство».

Данное деление призвано соединить традиционное техническое членение письма как оформленного по строго определенным правилам способа передачи данных и не менее традиционное членение текста на семантической основе.

Некоторые композиционные компоненты могут отсутствовать в каждом конкретном случае: повтор темы и извинения встречаются далеко не всегда и являются факультативными элементами. В то же время может не быть и более важных структурных компонентов, например, в теле письма с темой *«Найди свой...»* содержится всего лишь: *«Путь к Истине <http://www.putkistine.ru>»*. Здесь тема начинает рекламный слоган, который закан-

чивается в теле, однако ни зачина, ни основной части, ни заключения здесь нет. Письмо состоит из двух компонентов: слогана и сетевого адреса. Подобные такому случаи уместно обозначить термином «свернутая композиция».

# Лингвистические особенности спама как разновидности рекламного текста

## Лексические особенности спама

Рассмотрение спама как лингвистического явления уместно сразу начать с лексического уровня, поскольку какую-либо специфику на уровнях фонетическом и словообразовательном вряд ли возможно выявить: в силу письменного характера коммуникации и почти нулевой способности спамеров к словотворчеству соответственно.

Как уже было сказано, спамеры дифференцируют лексику в зависимости от разных условий, среди которых могут быть выделены

- принадлежность спама к той или иной тематической группе;
- предполагаемая сфера деятельности предполагаемого получателя;
- цель, преследуемая письмом;
- языковая личность адресанта (в отдельных случаях);
- прочие условия.

Важнейшим условием для отбора лексики является, на наш взгляд, тематическая группа, к которой принадлежит письмо. Так, наиболее распространенный рекламный спам характеризуется лексическим строем, типичным для любых других видов рекламы. Он, как правило, содержит однозначные лексемы (номиналирующие тот или иной товар), в основном входящие в состав типизированных структур, построенных по модели «наименование товара — цена»:

- «- *Электросушилка для рук VITA (Германия) - 2020 руб.*
- *Электросушилка для рук NOVOTERM (Польша) - 2470 руб.*
- *Одноразовые гигиенические подкладки на унитаз (100 шт.) - 75 руб. (с дозатором)*
- *Туалетная бумага в рулоне 200 м. - 20 руб.*
- *Бумажные полотенца (уп\*250 шт) - от 40 руб.*
- *Пылесосы для профессионалов - от 5166 руб.»*

Или:

- «- *аккредитация представительств (филиалов), иностранных компаний - от 1100 у.е.*
- *лицензии(строительство, алкоголь, туризм, медицина и др.)*
- *бухгалтерия - от 100 у.е.»*

Описание преимуществ той или иной фирмы, как правило, характеризуется использованием абстрактной лексики: *«СЖАТЫЕ СРОКИ НИЗКИЕ ЦЕНЫ ПРОФЕССИОНАЛЬНЫЙ ПОДХОД»*.

Данная тематическая группа может различать лексику в зависимости от других условий, преимущественно — в зависимости от предполагаемых интересов получателя (примеры см. выше). В то же время в спаме, как и в обычной рекламе, неизбежны неудачные попытки представления товара, связанные с лексическим наполнением, к которым можно отнести, например, такое письмо: *«В условиях современных мега полисов домашние животные лишены естественной среды обитания и подвергаются стрессогенной нагрузке. Длительное состояние стресса становится причиной многих заболеваний наших питомцев, в том числе и онкологических. <...> Лигфол (регистрационное удостоверение № P032-2-5.0-2756) представляет собой композицию гуминовые кислот, полученных из природного лигнина. Препарат обладает широким спектром действия, являясь антиоксидантом, адаптогеном стресс - корректором. Лигфол обладает противоопухолевым действием.»* Эффективность такой рекламы весьма сомнительна, ибо избыток терминологии порождает не только непонимание, но и неприятие такого текста. Конечно, можно было бы предположить, что такой спам сориентирован на специалистов в соответствующей области, однако концовка текста опровергает эту гипотезу, четко указывая на желаемого адресата: **«СПРАШИВАЙТЕ В АПТЕКАХ ВАШЕГО ГОРОДА!!!!!!!!!!»**. Логичен будет вывод о том, что специализированная лексика не ведет к коммуникативным неудачам только в том случае, если письмо предназначено для узкого круга «посвященных».

Собственно рекламный спам в подавляющем большинстве случаев содержит лексемы в их исходных значениях, предпочитает положительную (для создания аналогичного образа товара или услуги) или нейтральную их эмоциональную окраску, дифференцируя лексику в зависимости от интересов предполагаемого адресата. Часто используются лексические средства со «стертыми» значениями, рекламные штампы: *«Уникальное предложение!!! Вы покупаете две базы данных по цене одной!!!»* и т.п.

Вторая группа спама, ставящая целью вовлечение в аферу или финансовую пирамиду куда более интересна в лексическом отношении. Приведем один отрывок: *«Недавно меня уволили из WebMoney за то, что я случайно узнал один их секрет: оказывается, существуют такие кошельки (их всего около 5), когда на них посылаешь любую сумму - позже, через 3 дня тебе приходит сумма в 2 раза больше! Я им решил отомстить... Сейчас я накрутил таким образом около 5870 \$!!! Создано было для поддержки своих работников !!! Работники кормятся за счет фирмы очень неплохо!!»* Уже здесь очевидно откровенное смешение стилей, широкое использование разговорных и жаргонно-блатных слов («накрутил», «кормятся»). Причина этого, вероятно, коренится с одной стороны в том, что такие письма, как правило, ставят целью крайне агрессивное вторжение в эмоциональную сферу получателя, для которого форма своеобразного «диалога» (следует упомянуть и

очень серьезную степень «личностности» таких писем — почти всегда используется 1-е лично ед.ч., в данном случае еще и четкая оппозиция «Я — ТЫ») более понятна и интересна. С другой стороны, автор спама зачастую не рефлексировал по поводу лингвистического качества продуцируемого им текста, неосознанно используя привычные для него языковые средства.

Эта группа спама имеет и своеобразный «канон», к которому можно отнести, например, такой отрезок текста: *«В чем секрет такого головокружительного успеха?», - спросите вы. Дело в том, что в программе заложена секретная формула, которая обеспечивает 100%-ный успех всем участникам бизнеса за счет учета таких тончайших факторов, которых человеческий мозг просто не способен охватить. Что это за формула? Это секрет легендарного создателя Миямото Ичикава. Но главное, что программа работает! И работает блестяще! Это чудо-система, в которой успех безумно заразен! Спешите заразиться и вы!»* На этом материале мы можем наблюдать обширное использование рекламных штампов, идущих еще от советских политических лозунгов: «головакружительный успех», «легендарный создатель», «мозг не способен охватить», «секретная формула» и т.д. Тем самым создается определенный миф, в данном случае — миф о быстром заработке. Согласно К.Э. Штайн, такие тексты следует воспринимать как разновидность идеологического дискурса, особенностью которого является то, что «если пропозициональное содержание не соответствует ситуации, то имеет место “непрозрачное” относительно ситуации употребление того или иного выражения. При этом поведение говорящего детерминировано уже не ситуацией, а языком (речью). В результате — конфликт речевой деятельности и действительности...»<sup>15</sup> В глаза бросается очень активное использование лексем с абстрактными значениями, слов, которые нерелевантны по отношению к реальной действительности. Нередко и характерное для идеологических текстов «противопоставление ментальных миров»<sup>16</sup> и «негативная характеристика альтернативной картины мира с позиций исповедуемой иерархии ценностей»<sup>17</sup>: *«Те, кто утверждает, что сегодня в Рунете можно зарабатывать не более 120\$ в месяц, причем единственный способ - , так же глупы как и лжецы, сулящие Вам миллионы и миллиарды долларов в день. ЗАРАБАТЫВАТЬ В РУНЕТЕ МОЖНО!!! И МОЖНО БЕЗ ОСОБЫХ ПРОБЛЕМ ЗАРАБАТЫВАТЬ ПО 400, 500, 600: ДОЛЛАРОВ В МЕСЯЦ И БОЛЕЕ!!! (Причем совер-*

<sup>15</sup> Штайн К.Э. Заумь идеологического дискурса в свете лингвистической относительности // Текст. Узоры ковра. Сб. статей научно-методического семинара «TEXTUS». — Вып. 4. — Ч.2. Общие проблемы исследования текста. — СПб-Ставрополь, 1999. — С. 115.

<sup>16</sup> Ворожбитова А.А. Тоталитарный дискурс советского периода: опыт лингвориторической интерпретации // Текст. Узоры ковра. Сб. статей научно-методического семинара «TEXTUS». — Вып. 4. — Ч.2. Общие проблемы исследования текста. — СПб-Ставрополь, 1999. — С. 130.

<sup>17</sup> Там же.

шенно законно не платить с этих денег ни копейки налогов!) Говорю это не с позиций теоретика, но опираясь исключительно на свой собственный опыт! Причем ни баннеры, ни спонсоры не имеют к этому практически никакого отношения!» В целом можно сказать, что эта тематическая группа спама может быть квалифицирована как своеобразная трансформация советского идеологического дискурса.

Оставшиеся группы спама в лексическом отношении либо ничем не примечательны, либо могут быть соотнесены с какой-либо из уже проанализированных групп.

### Морфологические особенности спама

В области морфологии спам характеризуется доминированием глаголов и отглагольных существительных, что находит свое объяснение в самой сути рекламы, которая является двигателем для продажи и активно побуждает адресата к тем или иным действиям. Е.А. Терпугова считает, что «реклама не может требовать и приказывать, поэтому при функциональной ориентировке на побуждение грамматический показатель волеизъявления — повелительное наклонение — отсутствует в большинстве рекламных сообщений».<sup>18</sup> Действительно, в традиционном рекламном тексте в силу разности позиций говорящего и слушающего, наличием между ними иерархии, императив встречается довольно редко. Спам вполне может функционировать в рамках такой модели, побуждая к действию лишь косвенно, скажем, через перечисление преимуществ той или иной фирмы: «В своей работе компания стремится к максимальному удовлетворению потребностей своих Заказчиков и Деловых партнеров, старается учитывать все нюансы производства работ и пожеланий Заказчика». Или через глаголы в форме 1-го лица мн.ч.: «Предоставляем площадки для мини-футбола», «приглашаем Вас принять участие в тренинге», «Мы поможем родиться Вашему бизнесу!». Однако в спаме, по сравнению с традиционной рекламой, доля императивных форм значительно выше: «Праздник Великой Октябрьской Социалистической революции, или День Примирения и Согласия, прошел, разбередив патриотические чувства. Не дай им поогнать! Укрась свой телефон революционной символикой, специально для тебя мы собрали коллекцию лучших логотипов, картинок и рингтонов», «Для получения подробной информации свяжитесь с нами по телефону: 9.9.5 - 3.9.9.4», «Приезжайте к нам! Мы будем рады видеть Вас!», «Сократите расстояние между нами! Звоните по тел. 939-04-02, 8-916-6566494». Увеличение доли императивов связано с тем, что они зачастую служат как дополнительные связующий элемент между говорящим и слушающим, как прямое обращение одного к другому с той или иной целью. Большое количество таких форм является из-

---

<sup>18</sup> Терпугова Е.А. Рекламный текст как особый тип императивного дискурса // Автореферат на соискание ученой степени кандидата фил. наук. — Кемерово, 2000. — с. 10.

вестным приемом в традиционной рекламе, ориентированной на молодежную среду (достаточно вспомнить «Не тормози! Сникерсни!») — в этом случае повелительное наклонение создает атмосферу интимного общения, стирая иерархичность. Такие же цели, вероятно, преследует и авторы спама.

Если говорить о морфологических особенностях различных тематических групп спама, то наиболее четко здесь можно дифференцировать собственно рекламный спам и спам, ставящий целью вовлечение в аферу. Разделение можно провести в первую очередь на основе уже упоминавшего присутствия или отсутствия личностного начала, формально выражаемого через формы 1-го лица ед.ч. Для традиционного рекламного спама более характерны формы 1-го лица мн.ч.: «предлагаем», «поставляем», «обеспечиваем» и т.п. Их использование вполне логично, ибо адресант здесь обобщен: это некая компания или группа лиц. Другое дело вторая группа рекламных писем, которые зачастую начинаются так: *«Пол года назад наткнулся на кое-что ОСОБЕННОЕ. Если ты сейчас потратишь несколько минут своего времени, то в результате получишь Кучу ДЕНЕГ!»*, *«Я отлично понимаю, что на Ваш почтовый ящик ежедневно валится огромное количество “серой”, бледной и никому не нужной “электронной макулатуры” рекламного типа»*. Причина именно такого выбора форм также очевидна: говорящий выступает от своего собственного имени, пытаясь сделать общение психологически комфортным для слушающего. Взаимоотношения адресата и адресанта не мыслятся последним как деловое партнерство, рекламная коммуникация, скорее это непринужденная беседа между «давно знающими» друг друга людьми, что подтверждается и такого рода обращениями: *«Дорогой друг!»*, *«Привет, дружнице!»* и т.п. Контакт между говорящим и слушающим зачастую поддерживается и за счет создания коммуникативных оппозиций, формально выражаемых через противопоставление местоименных и глагольных форм: «Я — ТЫ», «Я — ВЫ», «МЫ — ВЫ», «МЫ — ТЫ».

В области словоизменения существительных рекламный спам характеризуется доминированием форм именительного падежа, которые служат для оформления названий товаров или услуг. Такие формы частотны как в теме, так и в теле письма: *«Сезонная распродажа домашних соляриев»*, *«Стратегическое и оперативное управление ассортиментом предприятия»*, *«аренда офисных помещений»*. Группа тем, призванная обозначить предполагаемого адресата, очевидно, определяется использованием форм дательного падежа: *«Главному бухгалтеру»*, *«Начальникам»* и т.п. «Мошеннический» спам не столь типизирован в области падежной парадигмы и может использовать любые формы в зависимости от воли говорящего.

Что касается прилагательных, то их характерологическая функция в основном дает о себе знать, когда речь идет об описании тех или иных преимуществ товара, услуги или фирмы: *«Богатый опыт, великолепная репутация», «На сегодняшний день рекламная рассылка по e-mail является одним из самых эффективных, доступных и дешевых способов рекламы», «Восхитительная SPA-процедура с камнями и эфирными маслами. Быстрое восстановление сил и энергии»*. Прилагательные активно участвуют в образовании уже отмеченных рекламных штампов, что чувствуется и в приведенных отрезках текста.

В качестве недоступной нам на данный момент перспективны исследование морфологического строя спама следует отметить необходимость проведения статистического анализа рекламных писем с целью точного выяснения процентов использования разных частей речи в их разнообразных формах. Не обладая такой статистикой мы позволили себе сделать чисто умозрительные выводы, основанные на субъективных ощущениях, оставшихся после прочтения всей анализируемой выборки писем. Достоверность таких выводов может быть легко подвергнута критике.

### **Синтаксические особенности спама**

Синтаксический строй спама определяется ярко выраженным господством восклицательных предложений, призванных акцентировать внимание адресата на важности сообщения. Нередко встречается избыточное повторение восклицательных знаков или даже их препозиция по отношению к актуализируемому участку текста: *«удивительное предложение по аренде офиса около !!! Министерства Иностранных Дел !!!»*. Образованная таким образом восклицательная «рамка» выполняет выделительную функцию, акцентируя внимание на важном, по мнению адресанта, участке текста: на том, что офис находится именно возле названного министерства. Таким образом, в данном случае восклицательные знаки являются еще одним аттрактором наряду с шрифтовым и графическим варьированием. Примечательно и то, что письмо, содержащее процитированный фрагмент, создано в формате plain text<sup>19</sup>, который не предполагает никак изысков в области модификаций шрифта, цвета, фона и т.п., которые предоставляет противоположный ему формат — HTML. Вероятно, с помощью использования семантики знаков восклицания автор пытался компенсировать отсутствие каких-либо других средств актуализации.

Среди частотных на синтаксическом уровне явлений следует отметить и агрессивное использование вставочных конструкций, оформляемых в подавляющем большинстве случаев с помощью скобок. Вставки в спамах, как правило, выполняют поясняющую функцию: *«Стоимость участия - 18'900 руб. (НДС не облагается)», «Недвижимость в Украи-*

---

<sup>19</sup> В связи с отсутствием адекватного и устоявшегося перевода на русский язык этого термина оставляем только англоязычный вариант.



*не (купля-продажа, инвестиции аренда: коммерческая, жилая)», «Инвестирование средств пенсионных накоплений (выбор управляющих компаний)», «Согласование (СЭС, УГПС, БТИ, Перепланировки и др)».* Скобочное оформление вставок может быть учтено при фильтрации спама как достаточно типизированная и частотная конструкция, которая довольно редко используется в нерекламных письмах.

Что касается предпочтений спамеров в области использования предложений различных структурных типов, то спам в этот плане очень разнороден, это могут быть и сложные структуры с несколькими компонентами и разными типами связи: *«Если ваша невзрачная серая простыня давно уже не радует тело и вы слишком заняты, чтобы искать ей достойную замену, заходите на [www.mypresent.ru](http://www.mypresent.ru)»,* и простые номинативные предложения: *«ДОСТАВКА И ТАМОЖЕННАЯ ОЧИСТКА», «МЕТАЛЛИЧЕСКАЯ МЕБЕЛЬ».*

На синтаксическом уровне спам выявляет, пожалуй, наименьшее число конструкций, которые могли бы служить в качестве маркера для его точно опознавания.

## Вопрос о жанре письма по отношению к спаму

Возвращаясь к проблеме интимизации коммуникации между автором спама и его получателем закономерно поставить вопрос: влияет ли сам жанр письма на большую степень доверия к спаму по сравнению с другими способами рекламы? Действительно, даже традиционная адресная реклама (рассылка писем в обычные почтовые ящики) всегда эффективна и дает приличное число откликов (15-20%) на нее. Логично предположить, что и спам должен быть не менее эффективен, поскольку адресация письма *лично* получателю, должна быть серьезным фактором, влияющим на вероятность прочтения такого письма.

Вероятно, такая ситуация действительно имела место в то время, когда поток спама находился в разумных пределах и не превышал долю личной корреспонденции. Однако на данный момент электронная почта уже совсем не то средство общения, которым она являлась изначально. С. Мухин, соавтор «Норм пользования сетью» так говорит об этом: «Почта предназначалась (как это ни банально) для того, чтобы люди могли общаться между собой. Так же как раньше переписывались по обычной почте, но гораздо быстрее и удобнее. Те, кто работают в сети давно, помнят эти времена, когда почтовый ящик, образно говоря, был больше похож на холодильник с продуктами, чем на помойный бачок с объедками. Дело вкуса, конечно, но я подозреваю, что многие из “нового поколения” просто не знают, как выглядела почта раньше и как должна бы выглядеть».<sup>20</sup> На наш субъективный взгляд почта все более приближается к другим формам сетевого общения: чатам, форумам, где обмен сообщениями идет в довольно динамичном режиме и восприятие и продуцирование текстов не отличается высокой степенью подготовленности. Прежде всего сами спамеры виновны в том, что рекламные письма все реже и реже читают (из-за слишком большого их числа). Исследование, проведенное сайтом *silicon.com*, показало, что 90% непрошенных писем отправляются в корзину сразу после получения. Всего пять процентов пользователей Интернета, принявших участие в опросе, отметили, что не удаляют спам сразу, а хранят его в папке *Inbox*, которую чистят по мере наполнения. Более того — 2,3% опрошенных *silicon.com* заявили, что они пересылают непрошеную почту интернет-провайдерам отправителей в надежде на то, что те закроют ящики спамеров, а 0,2% пересылают спам в полицию или специальные ведомства.

Таким образом, можно заявить, что жанр письма не играет большой роли в отношении адресата к письму. Спамеры пытаются эксплуатировать этот жанр, намеренно создавая подделки под частные письма, которые стремятся любым способом войти в доверие получателя (примеры см. выше), однако такого рода способы эффективны далеко не всегда.

---

<sup>20</sup> Интервью с соавтором OFISP-008 Сергеем Мухиным // <http://www.spamtest.ru/varticles.html?id=0026>

# Лингвистические аспекты фильтрации спама

## Введение

Необходимость фильтрации, базирующейся не на технических параметрах письма, а на его содержании возникает как следствие анонимной природы Интернета и несовершенства сетевых протоколов, допускающих легкое подделывание служебных полей. Основные критерии эффективности контент-фильтров это полнота и точность фильтрации: «Полнота — это количество распознанных программой сообщений заданного типа (например, спам или конфиденциальные письма), отнесённое к общему числу обработанных сообщений этого типа. Точность — количество правильно классифицированных сообщений, отнесённое к их общему числу тех, которые программа признала относящимися к заданному типу».<sup>21</sup> Реализация типичного модуля фильтрации спама описана в статье И.С. Ашманова и др. «Технология фильтрации содержания для Интернет».<sup>22</sup> Такой модуль подразумевает следующие компоненты:

- лингвистический процессор;
- семантический процессор;
- база ключевых слов и выражений;
- автоматизированное рабочее место редактора / лингвиста;
- некоторые другие технические блоки.

Рассмотрим наиболее важные блоки подробнее. Лингвистический процессор выполняет следующие операции:

- разбиение текста документа на слова;
- распознавание языка документа;
- лемматизация и морфологизация;
- индексация слов и словосочетаний;
- некоторые другие задачи.

Роль семантического процессора заключается в сравнении исходного письма с массивом имеющихся текстов и присвоении этому письму той или иной категории. При этом семантический компонент может действовать двумя способами:

- сравнение с эталонными текстами;
- «построение семантического образа текстового документа и сравнение его с семантическими образами списка категорий».<sup>23</sup>

---

<sup>21</sup> Ашманов И.С., Власова А.Е., Зоркий К.П. и др. Технология фильтрации содержания для Интернет // [http://www.dialog-21.ru/archive\\_article.asp?param=7519&y=2002&vol=6078](http://www.dialog-21.ru/archive_article.asp?param=7519&y=2002&vol=6078)

<sup>22</sup> Там же.

<sup>23</sup> Там же.

В первом случае семантический модуль работает по принципу антивируса, сопоставляя содержание предложенного ему письма с набором сигнатур, уже имеющихся в базе. Под сигнатурой понимается «уникальное число, поставленное в соответствие некоторому тексту и/или функция его вычисления».<sup>24</sup> Каждая сигнатура маркирована как принадлежащая к определенной категории. При совпадении с ней письма оно получает аналогичную категорию. «При сравнении текстов хорошие результаты дает использование для создания сигнатур алгоритма нечёткого сравнения, основанного на статистическом сопоставлении лексического состава текстов с учётом морфологических форм слов».<sup>25</sup> Необходимость в нечетком сравнении возникает как результат использования спамерами случайных комбинаций символов, которые вставляются в письма, тем самым делая каждый спам уникальным по своему содержанию.

Второе направление работы семантического модуля описывается так: «При построении семантического образа текстового документа происходит поиск и подсчет всех слов и словосочетаний, характерных для той или иной семантической категории, определение веса (основного и дополнительного) различных семантических категорий в применении к конкретному текстовому документу. На основании всех данных формируется список релевантных семантических категорий».<sup>26</sup>

Третий компонент фильтра — база данных. Это иерархизированный набор семантический категорий и их признаков. Интересный подход к работе с такой базой демонстрирует система «Галактика-Зум», предназначенная для статистического анализа текстов в литературоведческих целях: «Разработанный нами статистический метод позволяет выделять характерные для выборки слова и словосочетания, основываясь на сравнении конкретной выборки с общим фоном базы. Выражаясь лингвистическими терминами, наша система показывает тематические единицы базы данных, которые являются тематическими для конкретной выборки. Это позволяет как осуществлять скорочтение, так и выделять актуальные для запроса темы.»<sup>27</sup> Однако применимость такого подхода для детектирования спама, наверное, невелика: в нашем случае требуется находить максимальное количество соответствий между базой данных и конкретным текстом, «Галактика-Зум» же ищет *нетипичные* для базы данных элементы письма, что вряд ли поможет определить спам.

---

<sup>24</sup> Сегалович И. Детектирование массовых рассылок на Яндекс.Почте // <http://www.spamtest.ru/varticles.html?id=0032>

<sup>25</sup> Ашманов И.С., Власова А.Е., Зоркий К.П. и др. Технология фильтрации содержания для Интернет // [http://www.dialog-21.ru/archive\\_article.asp?param=7519&y=2002&vol=6078](http://www.dialog-21.ru/archive_article.asp?param=7519&y=2002&vol=6078)

<sup>26</sup> Там же.

<sup>27</sup> Антонов А.В., Курзинер Е.С. Автоматическое определение тематики большого необработанного текстового массива // [http://www.dialog-21.ru/archive\\_article.asp?param=7516&y=2002&vol=6078](http://www.dialog-21.ru/archive_article.asp?param=7516&y=2002&vol=6078)

Оставляя в стороне теоретический аспект фильтрации, о котором нет более смысла говорить из-за ограниченного объема работы, попытаемся рассмотреть конкретные практические реализации способов маркирования спама по его языковому наполнению.

## Алгоритм Байеса

«Принцип работы байесовского фильтра весьма прост. На основании анализа двух массивов текстов, один из которых представляет собой спам, а второй — обычные тексты, вычисляются относительные частоты, с которыми разные слова встречаются в этих массивах. Из этих частот вычисляются вероятности того, что слово встретится в тексте при условии, что этот текст принадлежит массиву “спам”. На этом этапе появляется таблица, которая говорит, что, если текст — спам, то, например, вероятность того, что в нем встретится слово “sexu” равна 0.96. Смысл “формулы Байеса” в том, чтобы “обратить” построенную таблицу вероятностей и говорить о вероятности того, что текст представляет собой спам при условии, что в нем встретилось данное слово. Особенностью “байесовских” фильтров является и то, что они могут автоматически учитывать и наиболее распространенные способы “зашумления” текста».<sup>28</sup> Технология фильтрации спама на основе алгоритма Байеса была предложена не так давно Полом Грехемом (Paul Graham), который применял ее для фильтрации собственной корреспонденции и добился следующих результатов: «When I tried writing a Bayesian spam filter, it caught 99.5% of spam with less than .03% false positives. It's always alarming when two people trying the same experiment get widely divergent results. It's especially alarming here because those two sets of numbers might yield opposite conclusions. Different users have different requirements, but I think for many people a filtering rate of 92% with 1.16% false positives means that filtering is not an acceptable solution, whereas 99.5% with less than .03% false positives means that it is».<sup>29</sup> В основе байесовского фильтра лежит, как уже указывалось, построение двух массивов текстов, на основе сопоставления с которыми письмо будет квалифицироваться как спам или не-спам. Ключевым понятием здесь является слово, которое маркировано принадлежностью к массиву спама, оно называется «токеном». Грехем дает такие признаки токена:

- «Case is preserved.
- Exclamation points are constituent characters.
- Periods and commas are constituents if they occur between two digits. This lets me get ip addresses and prices intact.
- A price range like \$20-25 yields two tokens, \$20 and \$25.

---

<sup>28</sup> Кошкин С., Пашко Д. Ответы на вопросы читателей в электронном журнале «Спамтест» // <http://win.subscribe.ru/archive/inet.safety.spamtest/200310/17185223.html>

<sup>29</sup> Graham P. Better Bayesian Filtering // <http://www.paulgraham.com/better.html>

- Tokens that occur within the To, From, Subject, and Return-Path lines, or within urls, get marked accordingly. E.g. “foo” in the Subject line becomes “Subject\*foo”. (The asterisk could be any character you don't allow as a constituent.)»

На основе такого понимания он дифференцирует работу своего алгоритма в зависимости от наличия токена в той или иной части письма. Так токен «free» может иметь следующие рейтинги:

Subject*FREE	0.9999
free!!	0.9999
To*free	0.9998
Subject*free	0.9782
free!	0.9199
Free	0.9198
Url*free	0.9091
FREE	0.8747
From*free	0.7636
free	0.6546

Таким образом, это слово, встречающееся в поле «Тема» (служебные поля в данной матрице отделяются от токена астериском) имеет рейтинг выше, чем если бы оно встретилось непосредственно в теле письма или в поле «Отправитель», что вполне логично: принадлежность письма к спаму вполне очевидна, если такой токен указан в теме и совсем не факт, если он находится в теле письма.

Наиболее очевидным недостатком фильтра Байеса является проблема ложных срабатываний. А. Черезов отмечает два типа таких срабатываний: «false positive — неверное зачисление письма в спам, т.е. собственно ложное срабатывание, и false negative — неверное причисление письма к не-спаму, т.е. “ложное не срабатывание”. В случае байесового фильтра “ложное не срабатывание” — не проблема — один раз указать фильтру, что это спам, и больше вы подобных писем не увидите. А вот false positive практически сводит на нет эффект борьбы со спамом: приходится просматривать папку “спам” в поисках возможно ошибочно попавших туда важных писем».<sup>30</sup> Ложные срабатывания могут возникать как следствие намеренной стилизации письма под не-спам, включения отрезков художественных произведений или, например, такого текста: *«Здравствуй, дорогой внук! Пишет тебе твоя бабушка...»* Далее следует реклама нужного товара. Такие тексты сбивают с толку фильтр, поскольку содержат токены, не относящиеся к массиву спама. Более того, единожды проиндексировав такое письмо как спам, мы повышаем вероятность дальнейших ложных отнесений действительно личных писем к спаму.

Пол Грехем сам указывает на такого рода недостатки своего алгоритма: «There are two kinds of spams I currently do have trouble with. One is the type that pretends to be an email from

<sup>30</sup> Черезов А. Статистические (вероятностные) методы фильтрации спама // <http://www.spamtest.ru/varticles.html?id=0036>

a woman inviting you to go chat with her or see her profile on a dating site. These get through because they're the one type of sales pitch you can make without using sales talk. They use the same vocabulary as ordinary email.

The other kind of spams I have trouble filtering are those from companies in e.g. Bulgaria offering contract programming services. These get through because I'm a programmer too, and the spams are full of the same words as my real mail».<sup>31</sup>

Фильтр Байеса идеально подходит для нужд отдельных пользователей, единственным условием для его функционирования является постоянное обновление базы данных для успешного отлова новых типов спама. Тем не менее, он не может быть использован в крупных почтовых системах из-за «большого разнообразия словарного состава клиентских ящиков».<sup>32</sup>

Перспективы такой фильтрации намечены самим создателем. Одной из таких перспектив он считает использование на отдельных токенов, а их групп: «One idea that I haven't tried yet is to filter based on word pairs, or even triples, rather than individual words. This should yield a much sharper estimate of the probability. For example, in my current database, the word “offers” has a probability of .96. If you based the probabilities on word pairs, you'd end up with “special offers” and “valuable offers” having probabilities of .99 and, say, “approach offers” (as in “this approach offers”) having a probability of .1 or less».<sup>33</sup> Вполне возможно, что эта идея уже реализована (статья написана в августе 2002-го года), но мы не располагаем информацией об этом.

## Шинглы

Как уже было сказано, спамеры всеми способами стремятся обеспечить уникальность каждого своего письма, включая в него случайные последовательности символов или, например, приветствие, в которое входит электронный адрес получателя. Распознавание такого рода дубликатов и есть задача метода «шинглов», предложенного в 1997 году Андреем Бродером. Этот метод не является чисто лингвистическим, но мы посчитали нужным включить его описание в этот раздел, поскольку он также работает непосредственно с содержанием письма, то есть текстом.

«Очевидно, чтобы повысить вероятность того, чтобы в результате небольших изменения текста контрольная сумма не изменилась, можно попытаться выбрать из текста несколько подстрок. Шингл (от английского shingle — чешуйка, черепичка) это и есть подстрока текста, по которой происходит вычислений контрольной суммы.

<sup>31</sup> Graham P. Better Bayesian Filtering // <http://www.paulgraham.com/better.html>

<sup>32</sup> Подробнее см.: Сегалович И., Тейблум Д., Дилевский А. Принципы и технические методы работы с незапрашиваемой корреспонденцией на «Яндексе» // <http://www.spamtest.ru/varticles.html?id=0028>

<sup>33</sup> Graham P. A Plan for Spam // <http://www.paulgraham.com/spam.html>

Выбирать такие подстроки можно по-разному. Во-первых, можно брать разный шаг, например: символ, слово, предложение. Во-вторых, решить, как они должны идти — внахлест (как раз так и получаются именно “шинглы”) или встык. В-третьих, следует понять, какого размера должны быть подстроки (выбранный размер должен свести к минимуму случайные повторы, то есть должен быть достаточно большим, но при этом оставаться достаточно малым, чтобы типичные изменения текста не разрушили все сигнатуры, конкретные цифры я здесь не привожу, по понятным причинам они не должны афишироваться), и делать ли их фиксированного размера. И, в-четвертых, поскольку возможных подстрочек в тексте чересчур много, надо решить — какие запоминать, а какие выбрасывать».<sup>34</sup>

В общем виде механизм действия шинглов описывается так: «Scan each document, compute hash values (“fingerprints”) at every, possibly overlapped, fragment of the document (“a shingle”), then sample the obtained set with some reasonable criteria and use this reduced set of checksums (“a sketch”) in all the following set operations. The fragments could be sentences or sequences of bytes, consonants, or words. Fixed and variable size sketches were suggested: a number of smallest checksums and all the checksums divisible by some integer».<sup>35</sup>

Для каждого письма может быть создано несколько шинглов, и в таком случае мы сталкиваемся с проблемой того, что не все из них будут совпадать с анализируемым документом. В таком случае вводится дополнительный «супершингл» — контрольная сумма всех шинглов письма. В таком случае идентичными будут считаться письма с совпавшими супершинглами, то есть со *всеми* совпавшими шинглами. Создатели шинглов так описывают этот механизм: «Similarly, we can estimate the resemblance of two sketches by computing the meta-sketch (sketch of a sketch). We compute super shingles by sorting the sketch's shingles and then shingling the shingles. The document's meta-sketch is then determined by its set of super shingles. If two documents have even one super shingle in common, then that means their sketches have a sequence of shingles in common».<sup>36</sup>

Описание шинглирования в нашей работе, вероятно, представлено довольно сбивчиво, что объясняется тем, что шинглы все же более математический, нежели лингвистический способ обработки документов, требующий представления конкретных фор-

---

<sup>34</sup> Сегалович И. Детектирование массовых рассылок на Яндекс.Почте // <http://www.spamtest.ru/varticles.html?id=0032>

<sup>35</sup> Ilyinsky S., Kuzmin M., Melkov A., Segalovich I. An efficient method to detect duplicates of Web documents with the use of inverted index // <http://company.yandex.ru/articles/article7.html>

<sup>36</sup> Broder A., Glassman S., Manasse M. Syntactic Clustering of the Web // <http://decweb.ethz.ch/WWW6/Technical/Paper205/Paper205.html>



мул и алгоритмов. Не обладая знаниями в данной области отсылаем к статье Бродера «Syntactic Clustering of the Web», содержащей все необходимые выкладки.<sup>37</sup>

В то же время шинглы далеко не единственный подход к определению дубликатов писем. Здесь мы должны снова вернуться к фильтрам Байеса: «Совсем необязательно искать очень похожие документы по контрольным суммам и хитрым подстрочкам. Вполне успешно <...> работают и лексические (основанные на словах) методы. Все разнообразие этих методов сейчас разбивают на два класса, локальные и глобальные лексические сигнатуры.

Если локальные сигнатуры рассматривают документ изолированно от коллекции и пытаются извлечь несколько характерных слов, основываясь только на их статистике в самом документе — TF (характерный пример: взять 5 самых частотных слов в документе длиннее пяти букв и упорядочить их по убыванию частоты), то глобальные либо пытаются при анализе документа учитывать информацию о глобальной статистике слова — IDF, либо вообще выбирают опорные слова, опираясь исключительно на уже существующий инвертированный индекс (см. метод Яндекса). Для работы глобальных методов необходимо как-то считать глобальную статистику слов, что в интенсивной антиспамовой системе вполне возможно, например в рамках байесовского подхода».<sup>38</sup>

Таким образом, можно сказать, что механизмы фильтрации спама, основанные на контент-анализе, на данный момент довольно эффективны, хотя и не задерживают всего объема рекламной корреспонденции. Одной из главных проблем здесь возникающих становится оперативное пополнение баз данных почтовых систем новейшими образцами спама. Именно от этого зависит, как уже говорилось, эффективность байесовского подхода. В дальнейшем речь следовало бы вести о распределенных способах обнаружения спам-атак, позволяющих в онлайн-режиме выявлять массовую рассылку и блокировать ее. Однако этот вопрос не относится к теме нашей работы.

---

<sup>37</sup> Broder A., Glassman S., Manasse M. Syntactic Clustering of the Web // <http://decweb.ethz.ch/WWW6/Technical/Paper205/Paper205.html>

<sup>38</sup> Сегалович И. Детектирование массовых рассылок на Яндекс.Почте // <http://www.spamtest.ru/varticles.html?id=0032>

## Перспективы дальнейших исследований

Подведем итог проделанной работе. Понятно, что одна из ключевых целей, поставленных в начале, достигнута не была: выявление типизированных структур на всех языковых уровнях оказалось довольно сложным занятием, и можно с уверенностью сказать, что это одна из очень тяжелых задач, решение которой, вероятно, сделало бы фильтры во много раз эффективнее. Проблема определения конструкций, однозначно указывающих на принадлежность письма к спаму, осложняется не только разнородностью его языкового строя, но и человеческим фактором, а именно намеренными искажениями текста рекламного письма в целях обхода фильтров. Четкая классификация таких искажений помогла бы отсеивать спам с большей эффективностью.

В области лингвистики интересно было бы исследование механизмов, которые влияют на поведение воспринимающего. Очевидно, что спам существует, поскольку он способен реально влиять на сознание получателя и заставлять его совершать то или иное действие, тем самым производя ощутимую экономическую прибыль рекламодателю и рекламораспространителю. При рассмотрении данной проблемы, наверное, уместен был бы психолингвистический подход. Помимо этого следовало бы изучить и концептосферу рекламных писем, так как материал для этого представлен в большом количестве: «мошеннический» спам очень сильно концептуализирован, а некоторые концепты даже выделяются формально — путем набора в верхнем регистре.

Наиболее очевидным недостатком этой работы является отсутствие основательной лингвистической базы для исследования, на основе которой можно было бы делать далеко идущие выводы. Подведение такой базы явилось бы началом настоящего научного лингвистического изучения спама как рекламного текста.

Конечно, отдельного внимания заслуживают различные структурные компоненты спама: тема, инициальные и финальные строки и т.п. Их детальный анализ, возможно, позволил бы выявить некоторые особенности, повторяющиеся в разных письмах.

Немало внимания следует уделить и существующим технологиям фильтрации спама на основе контент-анализа. Подробное изучение лингвистических аспектов фильтрации необходимо в качестве базы для модификации существующих фильтров с целью повышения полноты и точности их срабатываний.

В целом можно сказать, что если спам на самом деле будет существовать еще несколько лет, а объемы его будут только возрастать, материала для исследований станет все больше и, соответственно, необходимость в его лингвистическом изучении будет осознаваться все более ясно.

## Список использованных источников

1. Broder A., Glassman S., Manasse M. Syntactic Clustering of the Web // <http://decweb.ethz.ch/WWW6/Technical/Paper205/Paper205.html>
2. Graham P. A Plan for Spam // <http://www.paulgraham.com/spam.html>
3. Graham P. Better Bayesian Filtering // <http://www.paulgraham.com/better.html>
4. Ilyinsky S., Kuzmin M., Melkov A., Segalovich I. An efficient method to detect duplicates of Web documents with the use of inverted index // <http://company.yandex.ru/articles/article7.html>
5. Анисимова Е.Е. Лингвистика текста и межкультурная коммуникация (на материале креолизованных текстов). — М., 2003. — С. 59-62.
6. Антонов А.В., Курзинер Е.С. Автоматическое определение тематики большого не-обработанного текстового массива // [http://www.dialog-21.ru/archive\\_article.asp?param=7516&y=2002&vol=6078](http://www.dialog-21.ru/archive_article.asp?param=7516&y=2002&vol=6078)
7. Ашманов И.С., Власова А.Е., Зоркий К.П. и др. Технология фильтрации содержания для Интернет // [http://www.dialog-21.ru/archive\\_article.asp?param=7519&y=2002&vol=6078](http://www.dialog-21.ru/archive_article.asp?param=7519&y=2002&vol=6078)
8. Ашманов И.С., Власова А.Е., Зоркий К.П. и др. Технология фильтрации содержания для Интернет // [http://www.dialog-21.ru/archive\\_article.asp?param=7519&y=2002&vol=6078](http://www.dialog-21.ru/archive_article.asp?param=7519&y=2002&vol=6078)
9. Ворожбитова А.А. Тоталитарный дискурс советского периода: опыт лингвориторической интерпретации // Текст. Узоры ковра. Сб. статей научно-методического семинара «TEXTUS». — Вып. 4. — Ч.2. Общие проблемы исследования текста. — СПб-Ставрополь, 1999.
10. Жичкина А.Е. Социально-психологические аспекты виртуальной коммуникации // <http://club.fom.ru/books/socpsy-avc95.rtf>
11. Интервью с соавтором OFISP-008 Сергеем Мухиным // <http://www.spamtest.ru/varticles.html?id=0026>
12. Касперский Е. Основные классы угроз в компьютерном сообществе 2003 года, их причины и способы устранения // [http://www.infosecurity.ru/\\_gazeta/content/040330/article01.html](http://www.infosecurity.ru/_gazeta/content/040330/article01.html)
13. Клушина Н.И. Композиция рекламного текста // Русская речь. — 2000. — №5. — С.85-88.
14. Кохтев Н.Н. Психология восприятия и композиция рекламы // Русская речь. — 1991. — №4. — С. 68-72.

15. Кошкин С., Пашко Д. Ответы на вопросы читателей в электронном журнале «Спамтест» // <http://win.subscribe.ru/archive/inet.safety.spamtest/200310/17185223.html>
16. Национальная коалиция против спама // <http://www.spamtest.ru/coalition.html>
17. Нормы пользования Сетью // <http://www.ofisp.org/documents/ofisp-008.html>
18. Сегалович И. Детектирование массовых рассылок на Яндекс.Почте // <http://www.spamtest.ru/varticles.html?id=0032>
19. Сегалович И., Тейблём Д., Дилевский А. Принципы и технические методы работы с незапрашиваемой корреспонденцией на «Яндексе» // <http://www.spamtest.ru/varticles.html?id=0028>
20. Спам обречён // <http://www.spamtest.ru/varticles.html?id=0011>
21. Терпугова Е.А. Рекламный текст как особый тип императивного дискурса // Автореферат на соискание ученой степени кандидата фил. наук. — Кемерово, 2000.
22. Черезов А. Статистические (вероятностные) методы фильтрации спама // <http://www.spamtest.ru/varticles.html?id=0036>
23. Шалимова Г.С. Форма новая — проблемы старые (об одной модели газетного заглавия) // Журналистика и культура речи. — Вып. 1. — М., 1996. — С. 80-90.
24. Шевченко И.С. Некоторые психологические особенности общения посредством Интернет // <http://ncdo.levsha.ru/Pub/030D.htm>
25. Шевченко И.С. Факторы динамичности самопрезентаций в Интернет-общении // <http://psynet.carfax.ru/texts/shev1.htm>
26. Штайн К.Э. Заумь идеологического дискурса в свете лингвистической относительности // Текст. Узоры ковра. Сб. статей научно-методического семинара «TEXTUS». — Вып. 4. — Ч.2. Общие проблемы исследования текста. — СПб-Ставрополь, 1999.