

2008

Genres of Spam

Wendy Cukier

Ryerson University, wcukier@ryerson.ca

Ojelanki K. Ngwenyama

Ryerson University, ojelanki@ryerson.ca

Eva J. Nesselroth-Woyzbun

Ryerson University, enesselr@ryerson.ca

Follow this and additional works at: <http://aisel.aisnet.org/sjis>

Recommended Citation

Cukier, Wendy; Ngwenyama, Ojelanki K.; and Nesselroth-Woyzbun, Eva J. (2008) "Genres of Spam," *Scandinavian Journal of Information Systems*: Vol. 20 : Iss. 1 , Article 1.

Available at: <http://aisel.aisnet.org/sjis/vol20/iss1/1>

This material is brought to you by the Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in *Scandinavian Journal of Information Systems* by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Genres of Spam

Expectations and deceptions

Wendy L. Cukier,
Ryerson University, Canada
wcukier@ryerson.ca

Ojelanki Ngwenyama
Ryerson University, Canada
ojelanki@ryerson.ca

Eva J. Nesselroth-Woyzbun
Ryerson University, Canada
enesselr@ryerson.ca

Abstract. Spam is currently the dominant form of communications on the internet, accounting for most e-mail traffic. Spam is a marketing device, it is also an expensive and time-consuming nuisance for industries as well as a major vehicle for serious internet crimes. While considerable research has focused on the technical aspects of spam, how it works and how it can be blocked, our research aims to better understand why it works. We explore how genre theory can contribute to our understanding of 'spam'. Our study consists of two parts. The first examined the content, form and specific features and considered the manifest relationship to existing genres of communication. The second part of the study focused on a detailed analysis of 111 Nigerian letters, a particularly noxious form of spam. Genre is generally considered useful because it makes communications more recognizable and understandable to recipients, helping readers process information. Our study suggests that spam is not a single genre but adaptations of many recognizable print genres. With spam, genre operates at several levels and is often used to mask rather than reveal intent. The paper concludes that spam exploits genre by conforming to known forms while at the same time breaching those norms.

Key words: genre analysis, spam, internet fraud.

1 Introduction

Spam refers to electronic messages that are unsolicited or unwanted, sent to a large number of recipients (in bulk) without regard to the identity of the individual user, and usually having commercial purposes (Khong 2001). Spam now accounts for the vast majority of communications now on the internet, estimates vary but a conservatively spam is believed to account for two thirds of all e-mail traffic (Moustakas et al. 2005). While spam is not criminal in the US (although spammers often violate policies concerning acceptable use and privacy), it is often a vehicle for perpetrating crime (Chhabra 2005). Spam is not only a problem because it is a vehicle for identity theft and internet fraud but because it consumes significant resources. In 2005, the cost of spam in terms of lost productivity was estimated to have reached \$21.54 billion according to the National Technology Readiness Survey (Claburn 2005). Participants of the 2003 World Summit on the Information Society (WSIS) recognized that spam is a “significant and growing problem for users, networks and the Internet as a whole”.

To date, most research on spam has examined its negative impacts, technical characteristics, regulatory issues and the technologies employed to prevent it from overwhelming ‘legitimate’ communications. Several scholars have focused on developing a spam taxonomy in an effort to identify instances of spam, to prevent spam and to counterbalance the effect of spamming (Drucker et al. 1999; Gyongyi and Garcia-Molina 2004; Pelletier 2004). Spam is typically treated as a single document set, an a single category of communication but there is evidence that spam performs a wide range of functions and takes a variety of forms. Spam is a problem because it works. Understanding how it works, for example, why people comply with spammed requests requires more examination of spam as a form of communication. Our investigation of spam helps shed light on its communicative aspects and, so, some of the reasons why recipients engage in risky behaviours (for example, opening emails from unknown sources) or succumb to messages which are counter-rational (for example, offers of million dollar legacies).

2 Literature Review

2.1 Spam

The literature explores emerging forms of spamming and the ways in which spam is used to collect e-mail addresses, distribute viruses and perpetrate deceptive marketing practices or fraud. For example, spam masquerading as advertising or gibberish is often designed to circumvent filters with the sole purpose of ‘harvesting’ e-mail addresses. In ‘brute force’ and ‘dictionary’ attacks, spam programs send spam to every possible combination of letters at a domain, or to common names and words (Center for Democracy and Technology 2003).

Spammers continually devise new techniques to send spam. Spam may mix content and use orthographic inventions (e.g., “sec’s” in lieu of “sex”) and gibberish (e.g., Subject: “Pittsburgh pullover diorite chimera bray”) to avoid lexical detection by filters. Viruses, worms, and malware, such as Melissa, Love Bug and MyDoom, also use spamming techniques to propagate after a recipient unwittingly activates them. Viruses and worms may install open proxies that can be used to relay spam or to install software which transforms a computer into a ‘zombie’ (i.e., a computer through which spam is sent unbeknownst to the user).

Phishing attacks steal consumers’ personal identity data and financial account credentials. Social-engineering schemes use ‘spoofed’ e-mails to lead consumers to counterfeit websites designed to trick recipients into divulging financial data such as credit card numbers, account usernames, passwords and social security numbers. Hijacking brand names of banks, e-retailers and credit card companies, phishers often convince recipients to respond. Technical subterfuge schemes plant crime-ware onto PCs to steal credentials directly, often using Trojan keylogger spyware. Pharming crimeware misdirects users to fraudulent sites or proxy servers, typically through DNS hijacking or poisoning (Anti-Phishing Working Group 2005). Phishing and scams are distributed as spam, directly leading to identity theft and fraud. Phishing spam increased 52 percent in January, 2004. The statistics show that the response rate to this type of fraud is around five percent (Zeltsan 2004).

While there is a growing body of research on online advertising, surprisingly little attention has focused on examining the types or functions of spam. Most of the research to date on spam has been motivated primarily by an interest in technological and regulatory issues. An early study published in the Association for Computing Machinery (ACM) classified 400 unique messages sent to the AT&T and Lucent sub-domains under study for three months in 1997. The leading categories were money-making opportunities (36%); adult entertainment including singles services;

and sexually oriented products or services (11%) (Cranor 1998). Regulatory issues were the main focus of the article.

More recently, the Federal Trade Commission (FTC) analyzed a sample of 1000 items from 11 million messages and again, investment/business opportunities (20%), adult-oriented spam (18%), and finance (17%) were the most common categories (FTC 2003) (See Table 1). Two thirds (66%) of messages contained false “from” lines, “subject” lines, or message text. For example, the FTC study notes, in spite of the Direct Marketing Association guidelines which stipulate that messages must provide instructions for removal, only 36% of messages in the sample provided these (The Direct Marketing Association 2001). Further, comments from e-mail administrators suggest that many of these instructions were likely faulty or deliberately misleading (FTC 2003). The study also notes that fewer than 10% of the sample identified the name, postal address, phone number, and e-mail address of the sender (FTC 2003). Similarly, Jacobsson and Carlsson’s experiment with false e-mail accounts corroborates the failure of most spam e-mails to conform to regulatory requirements such as identifying the sender and providing options to unsubscribe (Jacobsson and Carlsson 2003). Again, the focus of these studies was the regulatory compliance of mass mailings. Despite the preponderance of these messages, research has been almost exclusively forensic.

Spam content, style, and genre have been largely overlooked and few researchers have gone beyond broadly categorizing the types of spam messages. One exception is Orasan and Krishnamurthy’s investigation of the linguistic characteristics of junk e-mail based on an analysis of 673 files which were compared to a corpus of leaflets extracted from the BNC (2002). They noted a number of linguistic differences including shorter sentences, limited vocabulary and increased use of personal pronouns such as “you” in spam (2002). The researchers examined the occurrence of key words such as: free, money, investment, credit, fast, Internet, e-mail, sex, weight and miracle (2002). However, the article is primarily descriptive, drawing few implications.

One form of spam that has attracted specific attention is the ‘Nigerian letter’, an adaptation of a fraudulent scheme whose history dates back to the sixteenth century (Zuckoff 2005). Through the late 1980’s and early 1990’s, these letters evolved to take advantage of newer modes of communication, particularly the fax machine. In the mid-1990’s, there was an explosion of direct mail scams, many originating or purporting to originate in Nigeria. The proliferation of these letters was so great that The United States Secret Service issued an “Advance Fee Fraud Advisory” regarding Nigerian letters (United States Secret Service 2001). The Federal Trade Commission also identified the ‘Nigerian letter’ in its study of false claims in spam (FTC 2003). There is always a promise of reward that is out of proportion to the effort required. Typically, the letters cite an improbable but precise sum of money

and offer as payment for the receiver's services—a percentage ranging from 5% to 40% (Zuckoff 2005).

Nigerian letters operate like many typical advance fee frauds in which a sum of money is promised; however, the victim must first pay a number of 'fees' in order to procure the grand sum. Generally, the target receives an e-mail from an insider or alleged 'official'—representing a present or former foreign government or agency. Often the writer will assert that there are unclaimed funds that are being held in customs or at a bank. To release the funds, the writer asks the victim to declare himself or herself as the rightful inheritor of the funds. For his or her cooperation, the victim is promised a percentage of the funds. For the funds to be released, the victim must provide further fees and payments, usually by wire transfer, for various taxes and expenses to consummate the transaction. The victim must pay these fees (attorney fees, duty, taxes, etc.) to process the transaction, and the sender claims that "just one more" fee/stamp/duty/form, etc. must be processed before the millions can be released. In addition, the victim either sets up an account in Nigeria, or uses his own account to transfer fees from stolen cheques. In a striking double jeopardy, this activity then criminalizes the victim. He or she could be charged with fraud for passing stolen cheques (Zuckoff 2005). Generally, the process is predicated on two factors: 1) escalating commitment to keep the victim on the line—the more the victim invests, the less likely he or she is to walk away; and 2) myth-building—the scams create an illusion of great wealth, or the 'pot of gold' that is just one e-mail away. Initially targeting businessmen, the scam has now expanded to include anyone who will respond, due to the low cost of e-mail transmission in relation to potential gains.

These advance fee schemes are the most lucrative fraudulent activity against individuals in the United States. The average victim loses about \$5000, though some may lose hundreds of thousands of dollars, or even their lives (Dyrud 2005). The scams are highly sophisticated and as victims are drawn in they may be convinced of the authenticity of the proposal by the numerous documents bearing seemingly official looking Nigerian government letterhead, stamps and seals. The scams can become extremely dangerous as the process advances and the stakes escalate (United States Secret Service 2006).

One of the reasons why there are so many Nigerian letters is because, in spite of the improbability of the claims, some people believe them. While there has been considerable investigation of the extent of internet fraud and crime perpetrated with Nigerian letters, our interest is different. We want to understand how Nigerian letters operate communicatively, and understand their main rhetorical features including the use of language and narrative.

2.2 Genre

Northrop Frye, the Canadian literary scholar, is probably the most cited source on genre. In the book, *Anatomy of Criticism*, Frye draws on Aristotle to propose that virtually all literature can be categorized according to universal genres with defined structure, rules and characteristics (Frye 1957). Genres are the literary conventions or ‘codes’ associated with particular forms—for example, epic, tragedy, and allegory. Miller (1994) defines genre as ‘typified rhetorical action based in recurrent situations’. Swales (1990) notes that genres have similar structures, stylistic features, content and intended audiences. The notion of genre as a fixed and decipherable ‘code’ has dominated much of the literature.

Genre has also been used as a way of understanding how business communications work. In advertising, for example, scholars have emphasized the importance of studying the form as well as the content in shaping meaning and effect (Holbrook and Batra 1987; Mick 1987; Mitchell and Olsen 1981; Laskey et al. 1989; Wells et al. 1989). Genre and genre repertoire have been proposed as analytic tools for investigating the structuring of communicative practices within a organizations. For example, Orlikowski and Yates (1992) propose that genres of organizational communications (e.g., memos, meetings, expense forms, training seminars) are habitually enacted by members of a community for particular social purposes. Subsequently, they examined the communication exchanged by a group of distributed knowledge workers in a multiyear, inter-organizational project and suggest that the group’s communicative practices evolved in response to community norms, project events, time pressure and media capabilities (Orlikowski and Yates 1994). Building on structuration theory, Orlikowski and Yates (1994) describe iterative relationships between communications genres and organizational practices.

Scholars have also explored the ways in which genre can illuminate electronic communications (Carliner 2004). “Recognizing genre will facilitate effective user-document interaction” so a particular “genre can be seen as an interface metaphor” (Toms and Campbell 1999). Genre is seen as a means of reducing cognitive processing load and helping readers navigate, process and comprehend both conventional and electronic documents (Roussinov et al. 2001). Åkesson et al. (1998) explore the genre of on-line newspapers examining content, form and functionality as do Ihlstrom and Lundberg 2004. This form of analysis is also applied to computer mediated communication (Bregman and Haythornthwaite 2001) videoconferencing (Pargman and Lantz 2002) and electronic meeting communications (Atunes and Costa 2003). Herring et al. (2004) examine weblogs in the context of traditional and new media. Crowston and Williams concluded that genre was a useful tool for analyzing communications on the internet based on an initial analysis of 100 websites

as a pilot project (1999) followed by an analysis of 1000 websites (Crowston and Williams 2000).

As Chandler (1997) notes in *An Introduction to Genre Theory*, a hierarchical taxonomy of genres is not a neutral or objective procedure. “A genre is ultimately an abstract conception rather than something that exists empirically in the world” (Feuer 1992). Thus, one theorist’s genre may be another’s sub-genre or even super-genre; and what is a technique, style, mode, formula or thematic grouping to one, may be treated as a genre by another. Miller (1994) suggests that “the number of genres in any society ... depends on the complexity and diversity of the society”. The interpretive and cultural-historical aspects of compound mediation are important in understanding the use of documents (Spinuzzi 2003). In other words, in addition to considering the questions, ‘What is the purpose of this genre?’ and, ‘What material goes into one?’ we must also explore the social and political dimensions of the context from which the genre emerged (Agre 1997; Bergquist and Ljungberg 1999). In his later work, Crowston acknowledges the limitations of top-down genre analysis using pre-existing categories and explains that bottom-up analysis allows for multi-dimensional definitions of genre as they become apparent, thus providing more flexibility in the face of limited forms (Crowston and Kwasnik 2004). Our approach to genre combines top down analysis with contextual analysis in order to deal with the complex genre of spam.

3 Our Study

3.1 Purpose

Spam is a major form of electronic communications but has been the subject of relatively little research focusing on its communicative functions and features. As noted previously, most work on spam treats it as a single form of communications. This is because it has been viewed primarily from the perspective of the typical receiver, and in most deliberative instances, as a “constrained organizational actor” (Hasselblad and Kallinikos 2000). Little analysis has been focused on its communicative functions or forms. Our study aims to explore spam from a different perspective as we try to understand more about the way in which it creates meaning. Understanding the characteristics of spam genre is one approach to understanding how it communicates. Our two stage study focuses on:

1. Investigating the characteristics and features of spam from a genre perspective
2. Understanding how spam relates to traditional print document genre
3. Exploring in detail one particular form of spam—the Nigerian letter—which is currently a major vehicle for internet fraud.

Our study includes two parts. The first part examines a sample of 450 spam items received in a single organizational account over a limited period of time. We analyzed the set using a series of levels, including sector, heading, relationship between heading and content, rhetorical purpose, content, structure, tone and other features and forms characteristic of genre analysis. We compare our results to previous studies on electronic media and genre. The second part of the study involves an in depth analysis of a particular genre of spam: the Nigerian letter.

3.2 Data Collection

Spam messages were collected from an individual's university e-mail inbox over a 26 week period (February 21, 2005 – August 20, 2005). A total of 450 messages were collected for analysis. These were messages that had passed through the university's spam filter and represent a tiny fraction of the spam received by the university.

The second set of data was collected over a one year period, from September 15, 2005 to September 15, 2006. Based on the general features identified in the first study—notably the length, the greeting, and the offer, we collected 111 non-duplicated Nigerian letters from three university professors' e-mail accounts.

As the communications collected were those that evaded the university's spam filters, they cannot be considered representative. The university's e-mail server receives approximately 200,000 e-mail messages each day. Of these, two thirds are typically blocked by the server using Postfix and another 10% are flagged as spam. The university uses a multi-layered approach to the spam problem. First, black-lists of known sources of spam or viruses are used to block messages from high risk sites. Next, Postfix, the mail transfer application, applies a basic filter. In cases of viruses or spam attacks, this filter can be used to delete messages of a particular type. Third, a virus check is performed by the server. At the fourth level of filtration, SpamAssassin will perform as a spam filter. If an e-mail fails some of these tests, it may be sent to the recipient's inbox with ****SPAM**** noted in the subject line. However, only one quarter of the spam used in this study was flagged with this subject header. Also, the occurrence of flagged words (e.g., 'penis') will lessen an e-mail's chances of reaching the inbox.

Recognizing the range of factors that affect the amount and type of spam an individual may receive (Internet use habits, newsgroups, technical characteristics of spam filters, etc.), we do not propose that this set of messages is a representative sample of all possible variations of spam sent to a broad range of recipients. However, the researchers find that the overall composition of the spam in the sample roughly reflects the composition of spam recorded in much larger studies (Åkesson et al. 1998). Subsequent work will explore the differences between the messages tracked by the filter and those which were not.

3.3 Data Analysis

The methodology used combined both quantitative and qualitative content analysis. Different approaches to ‘reading’ text are not mutually exclusive and applying multiple perspectives to text may address the limitations of individual techniques in isolation. While critical theorists have tended to reject quantitative strategies for determining the content or meaning of media messages (given the importance of considering both the manifest and latent meanings), even Kracauer grants that quantitative studies might serve as a supplement to qualitative analysis (Kracauer 1952). The sheer volume of mass media texts poses problems in terms of heterogeneity as well as quantity.

As a starting point and for the purpose of coding, we used the predefined categories of spam from the FTC study (2003) combined with Crowston and Kwasnik’s (2004) notions of faceted genre classification. The coders created new categories where none existed and for hybrids (that appeared to be combinations of categories). These messages were coded twice by three coders for sector or subject, source information, function, genre, format, subject line/body relationships, structure, addressee, signer, action desired, tone and regulatory compliance. The coding results were cross-checked for inter-rater reliability. Subsets representing each sector and genre were further analyzed using qualitative discourse analysis to explore recurrent themes and connotative use of language.

In the second phase of the study we refined the coding scheme to examine in more detail the letters we characterized as variants of the Nigerian letter. This process used an open coding scheme that required multiple readings of the letters. First the letters were read and coded for generic categories such as subject, source, function, subject line/body relationships, structure, addressee, signer, action desired, tone, use of language. In the second reading we examined the characters, the narratives and their function. Narrative analysis refers to the examination of those specific structures that are familiar to both the producer and the audience including patterns of character, setting and plot as well as symbolic elements (Hansen et al. 1998). In

this regard, we examined the implied relationship between writer and reader, specifically the devices employed to encourage suspension of disbelief, the setting and characters in the story and plot or story line, the persona of the narrator, the basis of the appeal, the quest or task, the promised reward, and other elements. The letters were coded twice, by two coders and then verified for inter-rater reliability. The coded letters and their elements were then clustered around emerging themes or categories. The coding sheets identified a number of elements linked to the theory of genre, myth, narrative and motifs.

Category	Sub-Category	Our Study (2005)			FTC (2003) N= 1000
		N = 450	%	Total % per cat.	
Financial	Mortgage	29	6.50%	41.78%	20%
	Loans	11	2.50%		
	Other	100	22.20%		
Business Opportunity		48	10.70%		
Adult-Oriented	Adult-oriented	36	8.00%	15.60%	18%
	Male enhancement	34	7.60%		
Health (including prescriptions)		69	15.30%	15.30%	10%
Computer hard/software		45	10.00%	10%	7%
Sales		30	6.70%	6.70%	16%
Other	Unknown	38	8.40%	12.90%	
	Recruiting	5	1.10%		
	Gaming	6	0.20%		
	Misc.	4	0.90%		
	News/sports	3	0.20%		
	Politics	2	0.45%		

Table 1: Common categories of spam (defined by sector or subject)

* Male enhancement advertisements could be classified as adult-oriented or health. We have combined them with adult-oriented for the purposes of comparison.

4 Findings

4.1 Categories of Spam (sector or subject)

As noted, the most common categories of spam in our study paralleled those in the FTC study. These are listed in Table 1. Financial and business opportunities were the most common categories, defined by sector or subject, followed by adult-oriented products and services.

There were differences between the composition of our sample and the much larger random sample used by the FTC. For example, our sample contained considerably fewer product and service advertisements and more health-related spam. This might have been a reflection of individual Internet behavior (i.e., visiting websites that create a ‘cookie’ which will attract specific kinds of spam) as well as the filter used by the organization.

4.2 Manifest Genre

While we initially thought that we would be able to define the spam genre through content and discourse analysis, it became apparent that spam is not a single genre but a composite of genres. Among the spam, we found messages that resembled a wide range of well-established document genres (see Table 2).

The most common genre of spam, at least on the surface, is a personalized memo which includes a description of a product or service with an embedded URL for more information. Sixty-seven percent (66.89%) of the spam was in this form.

The next most common form of spam appeared to be a letter, which might describe a service, but more often appear to be a scam to obtain personal information. For example, in this initial sample we found 14 examples of ‘Nigerian Letters’ which we will discuss further below.

Spam also took the form of ‘confirmations’ of orders or preapproved applications modeled on standard invoice and purchase order forms. For example, “You have been preapproved—your new application number 34”; “Your new application number 56”; or “PGF ALERT: Purchase Order Created for You”, again invoking a well-established business communication form.

Testimonials, a common form of advertising and direct mail, were used in 2.44% of the cases (e.g., “I have always worried about the size of my penis...”). Only 1.56% of the spam analyzed resembled conventional promotional pamphlets and only 1.78% resembled conventional display advertisements.

News bulletins (2.0%) such as announcements on stock prices as well as warnings and announcements (1.11%) such as “Remove all these popup messages today!” and “Microsoft virus warning—September 8th” were found. Newsletters (4.0%) and catalogues (0.67%) also appeared. There were contest winners (e.g., “Winner – winning notification”) reflecting a common form of direct mail.

<i>Types of Genre</i>	<i>No. of data</i>	<i>%</i>
Memo (with URL/link)	301	66.89%
Letter	31	6.89%
Order confirmation, pre-approved, application no.	24	5.33%
Gibberish	17	3.78%
Display advertisement	8	1.78%
News bulletin	9	2.00%
Newsletter	18	4.00%
Pamphlet (HTML)	7	1.56%
Testimonial	11	2.44%
Announcement/warning	5	1.11%
Press release	4	0.89%
Catalogue	3	0.67%
Contest winner	3	0.67%
HTML code	2	0.44%
URL submission form	2	0.44%
Article	1	0.22%
Business card	1	0.22%
Form	1	0.22%
Order forms/price list	2	0.44%
	450	100%

Table 2: Manifest genres

*(8.2% inform of rights to non-contact) of spam

Although the memo was a standard format (and reinforced by the structure of e-mail communications which requires a subject), we found that in two thirds of the cases, the subject line was not indicative of the contents of the message. In slightly more than one third (36.3%) of the messages examined the subject were descriptive: “prescription drugs”, “replica rolex”, “job offer”, “name brand software”. In another 27.6% of cases, the subject line was loosely associated with the content of the message without specifying up front what was being sold. For example, the subject line “men’s silent secret” preceded an e-mail selling impotence drugs. In 21.4%

of the cases, the subject line was a teaser apparently intended to arouse interest: “Good idea”, “Fact or Fiction”, “Power, Possibilities, Opportunities” are examples. In some cases (13.1%), the subject had no relationship to the content (for example, promises of sex were in the subject line while software was in the text) or there was simply no subject line at all (1.6%).

There is other evidence of deceptive and misleading spam. We were unable to accurately categorize the genuine intent of all e-mails, but in 55% of the cases, the provided links did not work when tested after the study period, suggesting that they were in fact not what they purported to be. Approximately 15% of the links within the spam e-mails were live working links, and 30% of the links redirect the user to another site.

<i>Category of Subject lines</i>	<i>%</i>
Subject line directly reflects content	36.3%
Subject line associated with content	27.6%
Subject line is a teaser	21.4%
Subject line has no relation to content	13.1%
No subject line	1.6%
	100%

Table 3: Different Subject Lines of Spam

Thus the use of a generic structure—the memo with a subject line—is used in much of the spam but in different ways than the traditional print memo.

4.3 Use of Language

One of the most intriguing features of spam as advertisement (rather than spoof, Trojan horse, virus, or phishing device), is its reliance on extraordinary compression and exclusivity of text. Most images are available only sequentially rather than synchronously, accessed through the following steps of text acceptance: 1) subject line, 2) message, and 3) link. Spam advertisers use a multi-layered and gnomic textual approach. They make mimetic use of e-mail as a medium of intimacy, interiorism, and breaking down of social barriers (public/private, conscious/deliberate vs. unconscious/ impulsive domains, etc.). Spammers use words that might be part of the realm of resistance to an inquiry, objection, or serious consideration of a proposal or purchase. They ‘neutralize’ these words/concepts by misappropriating or contextualizing them in the ‘message’ language. By classifying characteristics that have so

far escaped notice, researchers can contribute to the understanding of genre and to the practice of genre analysis.

Many spam messages use language as a technique for thwarting spam filters. We found that 21.3% of messages added language, quotations or ‘alphabet soup’. For example, a spam entitled “Top meds bought online”, with the text “Same medicine, different price!” and a link, also included the following gnomic saying, “television has brought murder back into the home — back where it belongs.” A Google search on these additional texts revealed a wide range of sources from the Bible to quotations from Eleanor Roosevelt.

It is apparent that traditional modes of correspondence have emerged in the electronic medium of spam. Some of these are simply electronic versions of analogue communications. The memo, for example, historically a genre used for efficient factual business communication, has been adapted to sell products, to harvest e-mail addresses, and to distribute viruses. For example, we have memos that are about what they purport to be about. We have memos that use subject lines to gain attention but these do not accurately reflect the actual content of the memo. We have memos that are not actually aimed at informing but are aimed at harvesting e-mail addresses or distributing viruses.

We also find examples of traditional print genre, for example the Nigerian letter. Because of the reach and anonymity afforded by the internet, these have become powerful means of perpetrating fraud and are now rapidly adaptable to prey upon sympathies generated by disastrous events like plane crashes and tsunamis.

Finally, we find examples of new forms of electronic communication which do not resemble print genres. These include, for example, communications which are basically gibberish and are designed exclusively to circumvent spam filters.

4.4 Multiple Layers of Genre

Typical genre analysis may fail to decode the intent of spam messages adequately because spam uses recognizable cultural markers or indices to frame the genre in a way that fools the reader. The memo format is used to imply that the recipient has a business or personal relationship with the sender. Informal subject lines such as “Hi there” and “Haven’t heard from you in awhile” are used to imply the messages are coming from friends, though they may contain adult content or ads for software. Very few of the sampled e-mails actually resemble advertisements. Therefore, the form of one genre is used in place of another as a mask to fool the recipient.

However, the deceptive use of manifest genre goes far beyond the masquerading of advertisements as memos, letters, confirmation forms etc. In many cases, the memo (which is actually an advertisement) is not really intended to sell anything,

but rather is meant to verify an e-mail address, collect personal financial information or distribute a virus. Hence, there are multiple layers of genre, used to mask one deception over yet another. Our study provides further evidence that Crowston and Kwasnik's (2004) notions of the limitations of top-down analysis of spam using pre-existing categories is restrictive and provides further support for bottom-up analysis to account for emerging forms.

Previous studies of spam, as noted above, revealed that a small percentage of spam complies with regulatory requirements found in the United States. In our study, we found that less than 1% of messages informed recipients of their right to non-contact and just over one third (40.9%) included non-contact information, some of it very disguised. Sixty-three percent of the messages we examined did not comply with privacy requirements at all. Again this is, at least in part, because many of the spam communications were not what they initially appeared to be.

4.5 The Nigerian Letter

In our initial study, we found a number of 'Nigerian letters', which were distinguished by other sub-genres in a number of ways. These letters were long, detailed narratives which claimed that the recipient had been selected to receive or earn a very large sum of money typically from a foreign land. These letters tell intensely personal stories. Although the claims in these letters seem highly improbable (in contrast to other forms of spam offering special deals or opportunities), other research indicates that these letters are an effective way of perpetrating fraud. We undertook more in-depth analysis to explore how this genre functions.

Some of the letters are addressed to Dear Sir/Madam or Dear Friend and an equal number are addressed to Brethren, Dear Beloved One, Greetings in the Name of the Lord, Salute you in the name of the Lord. Typically, the letters imply a level of intimacy. The letters are almost never personally addressed and there is rarely any plausible explanation about why the recipient was selected. Often, there is a reference to an unnamed colleague who vouched for the recipient's trustworthiness. As such, the letters evoke a sense of having been 'chosen' and that one has been deemed worthy of inclusion in a secretive, lucrative, and rare opportunity.

Sometimes the person writing claims to be the rightful owner of the money but more often they are an intermediary. The writer may be seeking a business partner: typically there is a situation which has led to a large amount of money in a bank account and, for a wide range of reasons, the writer needs your assistance to access the money in order to circumvent local restrictions or to pose as an heir to the fortune. The writer may be a banker or the personal assistant to a despot such as Charles Taylor. Or, the writer has found a treasure trove—ill gotten gains from a dictator,

treasure from the deposed king, or a deceased and heirless diplomat—and wants your help to access them.

In our analysis we applied broad categories of opportunities presented in these letters: quasi legal but highly rewarding ‘business’ ventures, opportunity to ‘right a wrong’ and claim a share of the reward, letters which plead for help in exchange for a reward or letters which offer a share in an inheritance or treasure. These letters ask the reader to send money or banking information, confidence, discretion, and trust. Some of the situations are straight forward: you are to distribute charitable funds and take a cut. Others skirt the law: you are to act as the ‘front’ or as a means of circumventing arbitrary and negligible national laws. The writer may be a contractor who, for reasons that are often highly complex, has over-invoiced a government sponsored contract and needs your assistance to extract the money. Often they are contracts in the petrol and mining industries or for agriculture. The writer needs the recipient to pose as a contractor. Any discomfort with the notion of skirting the law can be countered or justified as an act of wealth redistribution from a corrupt government, and the fraud victim can believe he has reclaimed funds that did not belong to anyone anyway. Tales of political strife or bloody massacres help to justify the theft of the funds. There are many variations on these themes but all of them promise easy wealth or a windfall of immense proportions.

Sometimes the letters tell the story of the wrongfully abused or persecuted innocent. In these letters, the writer is a victim of circumstance or misfortune and needs the recipient’s help in accessing a rightful inheritance. A number of writers claim to be white Zimbabweans forced off their land, unable to access their fortunes. Another claims to have access to unclaimed Holocaust funds that will be distributed through a Caribbean bank, but that a national citizen cannot release the funds. The victim is asked to help because the customs or banking officials will not question the legitimacy of a foreigner’s claim to fortune.

Other letters are pure appeals to pathos. In these letters, the writer is in danger and needs help in order to escape. The recipient’s assistance will be rewarded. The sender may even attach a photo of a pretty young woman whom she claims to be. She will claim to be American, and in dire trouble. Either the victim of robbery or some other mishap, she needs your help. She needs to get out of Nigeria and needs cash to do so—either to bribe border officials, pay hotel bills, replace a stolen airline ticket, get out of prison, etc. Or, the writer is dying, usually without heirs, and wants an associate to help distribute an improbably large inheritance to charity. Sometimes, he or she was a rich, bad person until she found God and became a born again Christian. There may be no relatives or relatives may be portrayed as untrustworthy and ungodly. Religion, righteousness, and God are often invoked to build the recipient’s trust. Guilt, pain, agony, helping orphans etc. is usually part of the plea and the victim gets a share of the charity funds.

While the majority of these letters evoke African scenarios based on local events and politics, mutations have emerged where the writer claims to be an Iraqi or American soldier who has come into possession of a stash of treasure in one of Saddam Hussein's former palaces, or a European banker or other financial official who wishes to make a large bank transfer but cannot find a trustworthy partner in his home country. These may claim to be from the United Kingdom, Denmark, Singapore or other developed nations.

Many letters apply current events (mainly political) to establish 'authenticity' or leverage sympathy. There were references to massacres, to cases of corruption, to civil wars, to cases crime or actual plane crashes. On some occasions there are names of real individuals sometimes even coupled with a link to a legitimate news story. For example, one letter referring to a plane crash paired with a bbc.uk link to a story of the crash. In another case, there was the story about treasure being confiscated from the government with a link the news story. References to these incidents appear to be included in an effort to help legitimize the claims but also, presumably, to generate pathos.

Most other forms of spam messages are brief and to the point. But Nigerian letters overflow with detail. For example, one writer pleads:

I had either beautiful daughters and they are all at the junior college level. Owing to our tradition. A spouse without a male child is considered a taboo to the husbands family, Simply put because of this ugly situation my brothers and sisters in law have branded me a useless woman. (sic).

The letters provide a high level of extraneous detail, presumably in order to increase believability and to establish an air of intimacy and encourage a response. The language in this genre of spam is usually ornate and pleading, with archaic expressions and, at the same time, with obvious mishandling of standard locutions, syntax and idioms. The reasons for this are not obvious but according to one author of Nigerian letters, this is entirely deliberate (save for the grammatical and spelling errors). He claims that the verbosity is fashioned to emulate the tone of Nigerian soap operas (partially because many of their targets are Nigerians). The messages are crafted to sound 'out of touch', evoking the tone of a rich, isolated, and sheltered diplomat who is in exile. Another claims he "was told to write like a classic novelist would... Very old world, very thick sentences" to create an excessively gracious and peculiar tone (wiredcomnews, no date). The intent is to make the reader feel sympathetic to the isolation of the sender, but also slightly superior and more competent, so that he assumes he might have the upper-hand in the transaction.

5 Conclusions and Implications

What makes the electronic spam message problematic within the current framework of genre analysis is that its communities of senders are multiple, multi-layered, sometimes overt and covert, and sometimes private and public. As such, a bulk advertisement e-mailed indiscriminately is in some ways similar to print ‘junk’ or direct mail.

Because spam has been defined to include anything from viruses to joke forwards from friends, seeking a rigid definition of it — or, as Zeltsan (2004) suggests, of ‘unsolicited’, or of ‘bulk’ — works against an open inquiry and openness to “the evolution of the phenomenon”. Some bona fide new forms emerge with hidden purposes. For example, “ruse” spam uses lures, misdirection, or a string of nonsense to procure actual e-mail addresses, sorted out behind the ostensible screen.

Our analysis, while preliminary, suggests that spam covers a range of genres, serving a wide range of purposes. While some characteristics are common to many types of spam, there is more evidence to suggest that spam memos, advertisements, letters, and contest announcements are significantly different forms, even though in many cases they may serve similar purposes. Moreover, in the case of spam, genre is not fixed.

Our study shows that genre analysis needs to recognize that one genre can mask another and that it may be necessary to penetrate the multiple layers of a typical spam message to uncover the message’s true intent. Thus, genre becomes a tool for deception. Fraudulent e-mails arrive disguised as ordinary communications which, at first glance, take the shape of a recognizable form. For example, a spam “memo” may evoke the expectations of ‘memo’ in order to catch the reader’s attention and clinch confidence. A memo is a recognizable genre and common format of e-mail messages. The memo may in fact reveal itself to be an advertisement for a product when further analysis of its structural content is made. Yet, at the same time, the memo-advertisement may in fact not be an advertisement at all. Its real purpose may be to harvest e-mail addresses, ‘phish’ for consumer information, distribute a virus, or perpetrate fraud. In other words, we find that simple taxonomies of genre are inadequate to the analysis of spam where a single message may have several layers of genre and where genres quickly mutate.

Spam genres are actually hybrids. For the most part, the messages resemble traditional genres in their manifest form in order to increase the likelihood of eliciting certain behaviors; however, the actual purposes of the spam are often radically different from what they seem to be. While spam clearly embraces a range of genres, these operate on a variety of levels.

Nigerian letters, which make up a disproportionate number of internet frauds, are a genre of particular interest. In spite of the obvious improbability of many elements of the stories presented—that the recipient was specifically chosen to receive a large sum of money for very little effort—a proportion of Nigerian letters do achieve their intended result. Our analysis suggests that this particular genre of spam, employs a range of common devices designed to appeal to emotions. From the greeting, to the characters, the setting, the ‘plot’, and the use of language, this genre of spam works to erode skepticism and appeal to deep rooted emotions, hopes and dreams. Taken together the elements in the Nigerian letter strongly appeal to emotions. Almost all letters make reference to a tragedy,—a war, political event, tsunami or plane crash. In the cases of the charitable appeals the writer is generally dying of esophageal, breast, or unspecified cancer. This coupled with the other conventions described above—the large treasure, the exotic locale and characters, even the use of archaic language all would appear to reinforce the appeal and lead readers to suspend disbelief. The excessive detail and the implied intimacy all appear calculated to establish trust. Nigerian letters are by no means new, as we noted at the outset. But the internet, with its potential for mass mailings, for anonymity and global reach, has transformed what was a minor source of fraud to a significant criminal threat.

This paper is the only a starting point for further work exploring spam and why it works. The analysis of the use of genre in spam to shape expectations is a starting point. Interesting further work could include doing longitudinal studies. As previously noted, our impression is that spam adapts quickly to changing world events, to new forms of spam filtering and patterns of consumer behaviour. Studying the ways in which spam evolves over time could shed light on the complex race between spammers and antispammers. We also recognize that an individual’s internet behaviours as well as spam filters affect the types and genre of spam received and a larger scale study with more participants would allow us to generalize more about the patterns of spam and genre. In addition, there is evidence that new forms of spam are emerging to take advantage of new types of communications including texting and mobile email and it would be interesting to examine how spam genre are adapting. The reason why spam continues to dominate internet communications is, in part, because it works. We believe that certain forms of spam, for example, Nigerian letters, tap into deeply rooted hopes and dreams to produce behaviours that are counter-rational. A rich area for further examination, drawing on the extensive research in consumer behaviour, is exploration of the ways in which spam persuades and manipulates recipients. Focusing more attention on the behavioural aspects of spam recipients will also shed light on ways to prevent spam fraud and crime.

References

- Agre, P.E., "Designing Genres for New Media: Social, Economic, and Political Contexts," 1997, <http://dliis.gseis.ucla.edu/people/pagre/genre.htm>, accessed on March 30, 2007.
- Åkesson, M., Ihlström, C. and Svensson, J., "Genre Structured Design Patterns – The Case of Online Newspapers," *Proceedings of IRIS 27*, Falkenberg, Sweden.
- Anti-Phishing Working Group, "What is Phishing and Pharming?" 2005, <http://www.anti-phishing.org>, accessed on March 30, 2007.
- Askehave, I. and Nielsen, A. E., "Digital genres: a challenge to traditional genre theory", *Information Technology & People*, (18: 2), 2005, pp.120-141.
- Atunes, P. and Costa, C. J., "From Genre Analysis to the Design of Meetingware", the *International ACM SIGGROUP Conference on Supporting Group Work*, International ACM SIGGROUP. 2003, pp. 302.
- Bergquist, M. and Ljungberg, J., "Genres in Action: Negotiating Genres in Practice", The 32nd *Hawaii International Conference on System Sciences*, Hawaii, 1999.
- Carliner, S. and Bosworth, T., "Genre: A Useful Construct for Researching Online Communication for the Workplace," *Information Design Journal*, (12:2), 2004, pp. 124–136.
- Center for Democracy & Technology, "Why am I Getting All this Spam? Unsolicited Commercial E-mail Research Six Month Report." www.cdt.org/speech/spam/030319spamreport.shtml, accessed on March 30, 2007.
- Chandler, D., "An Introduction to Genre Theory," August 11, 1997. www.aber.ac.uk/media/Documents/intgenre, accessed on March 30, 2007.
- Chhabra, S., Fighting Spam, Phishing and E-mail Fraud. Unpublished Masters Thesis, University of California, Riverside, 2005.
- Claburn, Thomas, The cost of spam in terms of lost productivity has reached \$21.58 billion annually. *InformationWeek*, February 3, 2005. <http://www.informationweek.com/story/showArticle.jhtml?articleID=59300834>. accessed on October 15, 2007.
- Cranor, L.F. and La Macchia, B.A., "Spam!" *Communications of the ACM*, (41:8), August 1998.
- Crowston, K. and Kwasnik, B.H., "A Framework for Creating a Facetted Classification for Genres: Addressing Issues of Multidimensionality," *The 37th Hawaii International Conference on Systems Science*, Hawaii, January, 2004.
- Crowston, K. and Williams, M., "Reproduced and Emergent Genres of Communication on the World Wide Web," *The Information Society*, (16:20), 2000, pp. 1–215.
- Crowston, K. and Williams, M., "The Effects of Linking on Genres of Web Documents," *The Hawaii International Conference on Systems Science*, Hawaii, January 1999.
- Davidson, E. J., "Analyzing Genre of Organizational Communication in Clinical Information Systems," *Information Technology & People*, (13:3), 2000, pp. 196–209.
- Doring, N., "Personal Home Pages on the Web: A Review of Research," *Journal of Computer Mediated Communication*, (7:3), 2002.
- Drucker, H., Wu, D., and Vapnik, N., "Support vector machines for spam categorization," *IEEE Trans. Neural Networks*, (10), 1999, pp. 1048–1053.

- Dyrud, M., "I Brought You a Good News: An Analysis of Nigerian 419 Letters," *The Association for Business Communication Annual Convention*, California, October 2005. <http://www.businesscommunication.org/conventions/Proceedings/2005/PDFs/07ABC05.pdf> accessed on March 30, 2007.
- Federal Trade Commission, "False Claims in Spam, A Report by the FTC's Division of Marketing Practices," April 30, 2003. <http://www.ftc.gov/reports/spam/030429spamreport.pdf>, accessed on March 30, 2007.
- Feuer, J., "Genre Study," in *Channels of Discourse, Reassembled: Television and Contemporary Criticism*, (2nd ed), University of North Carolina Press, Chapel Hill, 1992, pp.138–157.
- Firth, D. and Lawrence, C., "Genre Analysis in Information Systems. Research", *Journal of Information Technology Theory and Application*, 5:3, pp. 63-77.
- Frye, N., *The Anatomy of Criticism*, Princeton University Press, Princeton, New Jersey, 1957.
- Gyongyi, Z., and Garcia-Molina, H., "Web Spam Taxonomy," *Technical Report TR 2004-25*, Stanford University, 2004.
- Hansen, A., Cottle, S., Negrine, R., and Newbold, C., *Mass Communication Research Methods*. New York University Press, New York, 1998.
- Hasselbladh, H., and Kallinikos, J., "The Project of Rationalization: A Critique and Reappraisal of Neoinstitutionalism in Organization Studies," *Organization Studies*, (21:4), 2000, pp. 697- 703.
- Hawkins, R.P., Pingree, S., "Television and Behaviour: Ten years of scientific progress and implications for the eighties: Vol. 2," National Institute of Mental Health, 1982, pp. 224-247.
- Herring, S.C., Scheidt, L.A., Bonus, S. and Wright, E., "Bridging the Gap: A Genre Analysis of Weblogs," *The 37th Hawaii International Conference on System Sciences*, 2004.
- Holbrook, M.B. and Batra, R., "Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising," *Journal of Consumer Research*, December 14, 1987, pp. 404–420.
- Holbrook, M.B. and Hirschman, E.C., "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings and Fun," *Journal of Consumer Research*, September 9 1982, pp. 132–140.
- Jacobsson, A., and Carlsson, B., "Privacy and Unsolicited Commercial E-mail," *The Seventh Nordic Workshop on Secure IT Systems*, Gjøvik, Norway, 2003.
- Kracauer, S., "The Challenge of Qualitative Content Analysis", *Public Opinion Quarterly*, (16), 1952, p. 631.
- Khong, W.K. "Spam Law for the Internet," *Journal of Information Law and Technology*, 2001 http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2001_3/khong, accessed October 15, 2007.
- Laskey, H.A., Day, E., and Crask, M.R., *Social Communication in Advertising: Persons, Products, and Images of Well-Being*, Methuen, Toronto, 1989.
- Mick, D.G., "Toward a Semiotic of Advertising Story Grammars," in *Marketing Signs: New Directions in the Study of Signs for Sale*, J. Umiker-Sebeok (ed.), Mouton de Gruyter, Berlin, 1987.

- Miller, C.R., "Genre as Social Action", in *Genre and the New Rhetoric*, A. Freedman and P. Medway (eds.), Taylor & Francis, London, 1994.
- Mitchell, A.A. and Olsen, J.C., "Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude?" *Journal of Marketing Research*, August 18, 1981, pp. 318-332.
- Moustakas, E., Ranganathan, C. and Duqenoy, P., "Combating Spam Through Legislation: A Comparative Analysis of US and European Approaches," CEAS: *Second International Conference on E-mail and Anti-Spam*. Stanford University, Stanford, California, 2005.
- Orasan, C. and Krishnamurthy, R., "A Corpus-based Investigation of Junk E-mails," *Language Resources and Evaluation Conference*, Las Palmas, Spain, 2002.
- Orlikowski, W.J. and Yates, J., "Genre Repertoire: The Structuring of Communicative Practices in Organizations," *Administrative Science Quarterly*, (39:4), 1994, pp. 541-574.
- Orlikowski, W.J. and Yates, J., "Genres of Organizational Communication a Structural Approach to Studying Communication and Media," *The Academy of Management Review*. (17:2), Briarcliff, 1992, pp. 299-326.
- Orlikowski, W., Yates, J., and Yoshioka, T., *Community-based interpretive schemes: Exploring the use of cyber meetings with a global organization*, MIT, Cambridge, 2000.
- Pelletier, L., Almhana, P., and Choulakian V., "Adaptive Filtering of Spam," *Communication Networks and Services Research*, 2004, pp. 218-224.
- Radicati Group, "E-mail Hygiene Survey Results," 2005. www.radicati.com/e-mail-survey2005.shtml. 2005, accessed on March 30, 2007.
- Roussinovm D., Crowston, K. Nilan M. Kwasnik, B., Liu X & Cai, J. "Genre-based navigation on the web", *The 34th Hawaii International Conference on Systems Science*, 34, Maui, 2001.
- Spinuzzi, C., *Tracing Genres through Organizations: A Sociocultural approach to Information Design*, MIT Press, Boston, 2003.
- Swales, J.M. *Genre Analysis. English in Academic and Research Settings*, Cambridge University Press, Cambridge, 1990.
- The Direct Marketing Association. Online Guidelines. 2002. <http://www.thedma.org/guidelines/onlineguidelines.shtml>, accessed on March 30, 2007.
- Toms, E.G. and Campbell, D.G., "Genre as Interface Metaphor: Exploiting Form and Function in Digital Environments," *The 32nd Hawaii International Conference on System Sciences*, January, Hawaii, 1999.
- Ultrascan, "Advance Fee Fraud in 37 Countries", 2006. http://www.ultrascan.nl/html/aff_37_countries.html, accessed on March 30, 2007.
- United States Secret Service, "Public Awareness Advisory Regarding '4-1-9' or 'Advance Fee Fraud' Schemes," <http://www.secretservice.gov/alert419.shtml>, accessed on March 30, 2007.
- Wells, W.D., Burnett, J. and Mortuary, S., *Advertising Principles and Practice*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- World Summit on Information Society Declaration, "Declaration of Principles: Building the Information Society" A Global Challenge in the New Millenium," World Summit on the Information Society, December 12, 2003, www.itu.int/wsis/docs/geneva/official/dop.html, accessed on March 30, 2007.

- Zeltsan, Z., "General Overview of Spam and Technical Measures to Mitigate the Problem," ITU-T SG 17 Interim Rapporteur Meeting, Nov. 20, <http://strategis.ic.gc.ca/app/sitt/spamforum/thread.jspa?threadID=282&tstart=0> accessed on March 30, 2007.
- Zuckoff, M., "Annals of Crime: The Perfect Mark," *The New Yorker*, (82:13), 2005, pp. 36-42.

