

## **A Comparative Study of the Rhetorical Moves in Spam Emails: A Cross-Linguistic Investigation between English and Persian Spammers**

**Dr. Gholam Reza Zarei**

Assistant Prof. in Applied Linguistics

English Language Center, Isfahan University of Technology, Isfahan, Iran

[grzarei@cc.iut.ac.ir](mailto:grzarei@cc.iut.ac.ir)

**Parviz Ahmadi Darani**

Ph.D Candidate in Applied Linguistics

Isfahan Sheikhbahae University

[pahmadi05@gmail.com](mailto:pahmadi05@gmail.com)

### **Abstract**

This paper elaborates a contrastive research on schematic or rhetorical structures in English and Persian spam emails. The rationale behind the present study is to analyze how schematic structures used in spam emails of two languages vary and check the constancy of spam mail conventions across the spammers. Adopting Bhatia's (1993) research on promotional letters and Barron's (2006) on macro-textual analysis of medical spam emails, points of difference and similarity were examined with due attention to the two corpora of spam emails received by the present researchers from on-line medical suppliers over a period of ten weeks in 2013. The results of statistical test (chi-square) employed to different parallel corpora were indicative of regularity in the specific moves and steps, but there were some specific variations in their occurrence. That is, in terms of distribution and sequence of the moves, some discrepancies exist that may be attributed to cross-linguistic differences in English and Persian languages.

**Key Words:** Spam email, Promotional letter, Schematic structures, Macro-textual analysis

### **1. Introduction**

Spam, as “unsolicited e-mail, usually promotional, sent out to multiple recipients” (Crystal, 2001) has become a common figure in our inbox email every day. Though unsolicited or undesirable, most of us on the first impulse or whim are tempted to open them out of curiosity or omit them promptly. Despite this, however, linguistic analyses of spam mail represent a research desideratum since spam accounting for 60% of all internet traffic to them (Gagnon, 2004). A close examination of these emails may afford some insights into computer-mediated communication (CMC) research, particularly in the context of language and culture.

Although from the 1990s onwards, researchers have systematically inspected almost every aspect of computer-mediated communication (CMC); it however seems that very little has been done about unsolicited spam emails. Thus, the aim of this study is to examine the rhetorical moves in the English and Persian corpora in order to unveil their similarities and differences.

This paper first renders an account of spam emails, focusing on its rhetorical structure, communicative purpose and context of use. Subsequently, the corpora as well as the procedures for analysis are then introduced. The investigation itself focuses on the bottom-up macro-textual (move

structure) level of analysis and procedures for analysis are then operationalized. As a final point, it elaborates on the results and discussion of the findings.

Considering the existing paucity in the previous comparative genre-based studies of spam emails, what seems to be the problem now is that we do not know exactly what similarities and differences exist in English and Persian spam emails. More particularly, this research pursues answers to the following question: In what way(s) is the schematic structure of spam emails genre similar or different between English and Persian spammers?

## 2. Background

### 2.1 The definition of spam

*Spam*, as Online Etymology Dictionary (2013) defined it, “is a brand of canned meat, that the British comedy group Monty Python made a sketch about in their second season of their Flying Circus. The scene of the sketch is a dinner where ‘spam’ is an ingredient in all the dishes, basically it is impossible to order anything without ‘spam’. The association with the fact that one cannot receive anything in the mailbox, without also getting something one does not want is not that far-fetched”.

Zeltsan (2004) defines spam as all electronic messages that are unsolicited or unwanted, sent to a large number of users irrespective of the identity of the user, having commercial purposes that can contain viruses that spread by means of e-mail, or deception and cheat mechanism.

Definitions of spam usually comprise the facets that email is unsolicited and sent in mass. Email spam, also notoriously known as junk email or unsolicited bulk email (UBE), is a subcategory of electronic spam consisting of almost indistinguishable messages sent to plentiful recipients by email. One subset of unsolicited bulk email UBE is UCE (unsolicited commercial email). The opposite of "spam", email which one wants, is called "ham", usually when referring to a message's automated analysis (Barron, 2006).

Like all emails, unsolicited promotional emails are divided structurally into two parts, the ‘header’ or ‘heading’ and the ‘body’ or ‘message’ (Crystal, 2001). Details, such as the email address of the recipient (To:), the email address of the sender (From:), a brief description of topic (Subject:), and the date and time of transmission are included in the header.

The specific communicative purpose of unsolicited promotional emails and indeed of direct marketing efforts in general, is to persuade prospective customers to engage in immediate interaction with the sender (Connor and Upton, 2003). In other words, in direct marketing there is always an attempt made to establish contact with and get to know the individual prospect—unlike in advertising, where groups of individuals are targeted (Barron, 2006).

From a marketing point of view, an unsolicited promotional email or letter is a direct marketing tool. Direct marketing is similar in status to other elements of promotion, such as advertising, sales promotion (e.g., cut price offers, loyalty points) and public relations (e.g., press releases) (Baron, 2006). Promotional letters are differentiated from ‘mailings’, mailings including any reply cards, catalogues or possibly product samples included with the letter, and also the envelope in which these are delivered. In the context of email communication, internet pages linked via hyperlinks to an email may be understood as included in the term ‘mailing’ (Cukier *et al*, 2006).

Spammers collect email addresses from chat rooms, websites, customer lists, newsgroups, and viruses which harvest users' address books, and are sold to other spammers. They also use a

practice known as "email appending" or "epending" in which they use known information about their target (such as a postal address) to search for the target's email address. Much of spam is sent to invalid email addresses. Spam averages 78% of all email sent. According to the Message Anti-Abuse Working Group, the amount of spam email was between 88–92% of email messages sent in the first half of 2004 (Gagnon, 2004).

## 2.2 Genre approach to spam

Genre has been defined as the staged, structured, communicative events, motivated by various communicative purposes, and performed by specific discourse communities' (Bhatia, 1993, 2004; Flowerdew and Wan, 2010; Swales, 1990, 2004). The sizeable body of research on genres boils down to the two domains: (1) the lexico-grammatical features of a given text, and (2) the identification of their rhetorical structures or 'structural move analysis' (Hyon, 1996). This latter approach is what Nwogu (1997) referred to as "the identification of schematic units or moves." As such, a study of the move structure of spam emails falls in this category.

Three genre analyses of promotional letters and spam emails are of particular relevance for the present study, namely those by Bhatia (1993), Cukier *et al.* (2006) and Barron (2006). Bhatia (1993) investigated the move structure of unsolicited commercial sales promotion letters. Bhatia (1993) describes sales promotion letters in the following way: "a sales promotion letter is an unsolicited letter addressed to a selected group of prospective customers (they may be individuals or companies) in order to persuade them to buy a product or service".

As seen in the quote above, the largest difference between Bhatia's (1993) definition of sales promotion letters and Zeltsan's (2004) definition of spam is the receivers. The spam is sent without regard to the identity of the individual user (Zeltsan's, 2004), whereas the sales promotion letter is addressed to a selected group of prospective customers.

Bhatia (1993) argues that sales promotion letters have five communicative purposes: (1) capture the attention of the potential customer, (2) eliciting a desired response, (3) offer an appraisal of the product, (4) being short enough, not boring the customer but long enough to give details about the product, and (5) encourage further communication. He forewarned, however, that his analysis "has been based on a limited set of data from a specific cultural context. A more informed discussion will require a more comprehensive, rigorous, and sustained analysis of data".

In a pilot study, Cukier *et al.* (2006) have studied 300 different spam messages in order to apply the concept of genre to them. They argue that spam is not a single genre, but several genres. In a study which is very much in line with the present study, Barron (2006) investigated genre analysis using the same corpus of medical spam mails written in English. This study was conducted to contribute to the identification of genre moves in spams. This macro-textual analysis concentrated on describing the schematic structure through which the overall communicative purpose was realized by means of a move analysis whereby a pragmatic function was assigned to particular sections of language.

In another study done by Barron (2006), she tried to characterize a micro-textual analysis of 121 spam emails written by online medical suppliers. She has related spam to its promotional communicative purpose and posed the question as to how spammers exploit language to realize this promotional purpose.

However, no cross-linguistic study has been done comparatively on spam emails with regard to English and Persian spam emails so far. The present study thus aimed at exploring this missing

research in spam emails and sought to find the credible difference in move structures of spams written by English and Persian spammers.

### 3. Method

The aim of a genre analysis, as it highlighted by the scholars, is to concentrate on social function and form in a particular recurring rhetorical context and, above all, to examine the link between these. In line with such definition, the present study adopted a quantitative approach to genre analysis of English and Persian spam emails in which primarily comparative and corpus-based methods were deployed. The comparative method was employed in order to locate discrepancies in the utility of rhetorical moves between the English and Persian spam emails.

#### 3.1 Corpus compilation

The corpus of the present research consists of 200 spam emails, 100 English and 100 Persian, received by the researchers in Iran from on-line medical suppliers over a period of ten weeks, during the time period from 11th Jan. to 15th Feb. 2013. The focus was, thus, exclusively on spam medical product emails. To determine how many of spam emails to be included in the study, the researchers employed the Cochran (1977) approach for determining the adequate sample size. Since the margin of error for the current study was .05, it was decided that 200 spam emails, 100 English and 100 Persian, suggested by the Cochran's table. To create a coherent corpus, the most common genre of spam, viz, a personalized *memo* which includes a description of a product with an embedded URL for more information was selected. This sort of spam forms almost sixty percent (59.7%) of the spam emails as Cukier *et al.* (2006) categorized it. This narrow selection of the sample made the data more homogeneous, given possible systematic differences between subtypes of unsolicited promotional communication, particularly given the present sample size. No bias was exercised in the selection of data and they were selected out of convenience.

#### 3.2 Instrumentation

As mentioned earlier, the suggested theoretical framework was based on Bhatia's (1993), research on promotional letters and Barron's (2006) research on macro-textual analysis of service spam emails. Thus, it should be noted that the researchers have made use of the terminologies of relevant studies on genre, such as abovementioned scholars as a point departure for the move nomenclature of the schematic structures both in English and Persian spam emails. The study applied three types of instruments: the AntMover and Wordsmith software and the Anne Barron (2003) move structure framework that are described below.

AntMover is an online text structure analyzer program accessible on the Internet. Once a text file is opened in AntMover, it is introduced into the program for analysis. The user can then select up to four versions of the file. Each spam from the corpus was fed into the AntMover for the identification of the moves and steps in the spam email.

The second instrument utilized in the study was the framework for move analysis developed by Anne Barron (2006). This framework has been designed for human coders, and can be used in manual analyses of the move structures of spam emails. Table 1 depicts Barron's (2006) framework for move analysis for the genre of spam. The framework assumes that a spam email entails 7 moves as it has been illustrated in the Table 1:

Table 1. Prototypical move structure of a spam email

Move	Step	Function/Description	Code
Capture attention		Arousing customer's interest to open a particular spam	M1
Establishing credentials		Highlighting the competence of the organization	M2
Welcome prospect		Inviting refers to previous contact	M3
Introduce the offer		Launching the suggestion	M4
	Step1. Primary Offer		M4S1
	Step2. Secondary Offer		M4S2
Use the pressure tactics		Pushing customers towards purchase	M5
Solicit response		Encouraging the prospective buyer to engage in further communication with the seller	M6
Give a polite way-out		Increasing the credibility of the supplier	M7

### 3. 3 Procedures of analysis

After gathering the required corpus, each spam email in both languages was assigned a unique code (e.g., SPAM # 1, PSPAM # 2 ...). In the next step, a set of analyses were performed. A frequency count was performed to identify the total number of moves and steps in each spam email. Then each spam was saved as a text file to be fed into the AntMover software developed by Anthony (2003).

The present researchers separately coded each spam email and identified the moves; they labeled the moves according to the model proposed by Barron (2006). Then the coders discussed their coding and compared them with the output from AntMover. In the case of disparity in their coding, it was resolved through extensive discussion and mutual consent. The Persian corpus was solely analyzed manually since the AntMover was not applicable to Persian spam emails.

The frequency of each move in each spam was recorded in an Excel file; this was done to verify the extent to which any given move had been used. The recurring patterns or the uses of move cycles were totaled, averaged, and tabulated. This resulted in the identification of general move sequences and patterns. Then, the frequencies and percentages that followed were used as the data that were analyzed qualitatively and quantitatively.

### 3.4 Data analysis

Applying the framework outlined the rhetorical moves of the spam email given in Table 1, two raters hand-coded all 200 spams in the corpora. The vast majority of discrepancies that occurred between the two raters resulted from initial disagreement as to where one move ended and the next started, not as to the presence of a particular move. Interrater-reliability was calculated at 81%, with all discrepancies reconciled through discussion. This interrater-reliability is quite good, since, as Bhatia (1993) notes, there are sometimes "cases which will pose problems and escape

identification or clear discrimination, however fine a net one may use. After all, we are dealing with the rationale underlying linguistic behavior rather than its surface form". Once all of the moves were agreed upon and marked, each spam was then tagged to indicate the start and stop of each move in each text.

To estimate the convergent reliability of the data, the frequencies identified by the human raters were totaled and averaged and then correlated with the frequencies obtained from AntMover. This was done through the use of a one-tailed bivariate correlation analysis using Spearman's rho. (Rho = .791) indicated an acceptable level of reliability. As to the reliability of the data, the Inter-rater Agreement was estimated.

The occurrence and cycle of each move for each text was also entered in an Excel file. This allowed keeping track of the total frequency of each move in the corpora, the relative position it occurred in each spam (e.g., first, second, third), what other moves a move most commonly occurred with, how frequently a move was embedded in another move, and how frequently a move occurred in the body of the text.

## **4. Results**

### **4.1 Move Frequencies of the English spam emails**

Table 2 presents information about the moves in the corpus of 100 English spam mails, including the frequency of each move within spam emails. Not surprisingly, the most common move in all of these spams was M6 'solicit response'. This signifies 81% of all the moves occurring in the corpora. As delineated by Barron (2006), "it plays a significant part in realizing the overall promotional communicative purpose of spam emails in that it encourages the individual receiver to engage in further communication with the sender. The analysis reveals that this move is realized semantically in one of three ways. A request to act may be made, contact details may be given or a simple click-through via a hyperlink may be used. The move is regularly combined with other moves, in which case it is less obvious or even somewhat hidden. However, it may also appear as a stand-alone move. Such cases represent the most direct realizations of the move. Interestingly, these stand-alone 'solicit response' move appear towards the end of the spam emails".

The second most frequent move was M1 'capture attention', which occurred 282 times. At the rate of 1.5 times per spam, this move represents 77.6 % of all the moves in this corpus. There are a total of 46 M2 'establish credentials' in the corpus as a whole. The move is found in 32.2% (46) of the mails analyzed, being sometimes used more than once in a particular email (Table 2). As such, it is a frequent but optional move.

Table 2. English move totals, percentages and rate of occurrences

	<b>Move1</b>	<b>Move2</b>	<b>Move3</b>	<b>Move4</b>	<b>Move5</b>	<b>Move6</b>	<b>Move7</b>
<b>Moves</b>							
Total Number	94	46	8	95	8	98	96
% of total	77.6%	32.2%	6.8%	79.5%	6.8%	81%	79.3%
<b>Spams</b>							
W/≥ 1 occurrence	92	15	36	98	51	97	13
% of total	92 %	15%	36%	98%	51%	97%	13%
<b>Words/move</b>	62	39	7	93	7.8	10	9.3
<b>Avg.</b>							

The M3 ‘welcome prospect’ move is realized by the use of welcomes or invitations referring to previous contact. It is only employed in 6.8% (8) of the emails analyzed, and, thus, clearly a highly optional move. In those cases where the move occurs in the body copy, it ranges in length from 5 to 9 words, with an average of 7 words per move (Table 2). Its position, when it occurs, is as the first or second move in the body copy.

M4 ‘introduce the offer’ also clearly seems to be a required move in this genre as it occurs in 79.5% of the spams. As Barron (2006) put it “this move consists of two steps in the present context, the first step being (a) introduce the primary offer by detailing the general offer. The details given in the M4S1 ‘primary offer’ may be general, as for example, in Online Prescription Medications!, given in the first line of the body copy in one particular email. The second step to this move is (b) introduce the secondary offer by detailing secondary aspects of a product purchase, M4S2 (e.g., secure payment, no doctor’s appointment required, free delivery, low prices...)”.

The M5 ‘use pressure tactics’ represents an effort to push prospective customers towards an immediate purchase, informing them that the present offer is exclusive in some way (Bhatia, 1993). Pressure tactics are only employed in 6.8% (8) of the total 100 English spam emails. The average M5 ‘pressure tactics’ move was 7.8 words in length. Where this move occurred, it was positioned towards the end of the email.

The M7 ‘give a polite way-out’ move is present in 79.3% (96) of the English spam mails analyzed. The average move length here was 9.3 words but the range was very wide stretching from 1 to 72 words per mail. Where this move occurred, it always represented the final move.

Using the concordance program Wordsmith (2012), it is possible to analyze and compare the lengths of each of the moves. M4 ‘introduce the offer’ is by far the longest move in this genre, averaging 93 words per occurrence. M1 ‘capture attention’, the second longest move, is only 2/3 the length at 62 words per occurrence. M3 ‘welcome prospect’ is the shortest (7), with M2 ‘establish credentials’ and M5 ‘use pressure tactics’ averaging 39 and 7.8 words per occurrence, and M6 ‘solicit response’ averaging 10 words per occurrence. Table 2 provides the average words per occurrence for each of the seven moves in English corpus.

#### 4.2 Move Frequencies of the Persian spam emails

As shown in Table 3, the most frequent move of Persian dataset was M6 ‘solicit response’, followed by M4 ‘introduce the offer’. The frequent occurrence of M6 ‘solicit response’ in the present study confirms Barron’s (2006) findings, ‘solicit response’ move was the most substantial and frequent move. M4 ‘introduce the offer’ at 77.6% ranks the second. That is, more than 15% exceeded from its English counterpart.

Table 3 Persian move totals, percentages and rate of occurrences

	<b>Move1</b>	<b>Move2</b>	<b>Move3</b>	<b>Move4</b>	<b>Move5</b>	<b>Move6</b>	<b>Move7</b>
<b>Moves</b>							
Total Number	88	12	3	90	1	95	0
% of total	72.6%	8.4%	2.17%	75%	0.82%	78%	0
<b>Spams</b>							
W/≥ 1 occurrence	82	12	22	97	42	95	0
% of total	82%	12%	10.90%	90%	40.37%	90.7%	0
<b>Words/move</b>	52	24	7	82	9	8	0
<b>Avg.</b>							

M2 ‘establishing credentials’, 8.4%, M5 ‘use the pressure tactics’ and 0.82% occurred at very low rates of frequency across the 100 Persian spam emails. While M7 ‘give a polite way-out’ represented no move in this corpus and its occurrence is equal to zero.

M1 ‘capture attention’ was clearly icing-on-the-cake moves that the Persian spammers could draw upon when desired frequently. M1 represented 77.6% of the moves in this corpus and occurred with same frequency with M4 ‘introduce the offer’ move.

Table 3 provides the average words per occurrence for each of the seven moves in Persian corpus. M4 ‘introduce the offer’ is by far the longest move in this genre, averaging 92 words per occurrence. M1 ‘capture attention’, the second longest move, is only 2/3 the length at 88 words per occurrence. M3 ‘welcome prospect’ is the shortest (7), with M2 ‘establish credentials’ and M5 ‘use pressure tactics’ averaging 39 and 9 words per occurrence, and M6 ‘solicit response’ averaging 10 words per occurrence.

In order to see whether there is a significant difference between the frequencies of move types in the English and Persian spam emails, a chi-square test was run. Table 4 shows the result of chi-square test. Based on the results, the value of significance .00 is less than .05 ( $P = .00$ ,  $X^2 = 55.408$ ), so there is a no significant difference between the frequency of move types in English and Persian spams, particularly M1 ‘capture attention’, M4 ‘introduce the offer’, M6 ‘solicit response’, but there exists significant between M2 ‘establishing credentials’, M3 ‘welcome prospect’, M5 ‘use pressure tactics’.

Table 4 Chi-Square results of English and Persian spam emails

Moves	Steps	Frequency		X <sup>2</sup>	P-Value
		ES	PS		
1. Capture Attention		94	88	0.356	0.551
2. Establishing Credentials		46	34	1.856	0.143
3. Welcome Prospect		8	3	2.273	0.132
4. Introduce the Offer		95	90	2.222	0.136
	1. Primary Offer	94	88	0.356	0.551
	2. Secondary Offer	84	68	1.684	0.194
5. Use the Pressure Tactics		8	1	4.5	0.934
6. Solicit Response		98	95	0.02	.887
7. Give a Polite Way-Out		96	0	0.00	0.102

As it stands in Figure 1, obligatory moves include M1 ‘capture attention’, M4 ‘introduce the offer’ and M6 ‘solicit response’ since they appeared in all spam emails in the both corpora. Moreover, both English and Persian spammers treated M2 ‘establishing credentials’, M3 ‘welcome prospect’, and M5 ‘use the pressure tactics’ as optional. The most central discrepancy was observed in relation to M7 ‘give a polite way-out’ that was non-existent in the Persian corpus. Furthermore, a comparison showed similarities in the obligatory nature of the M6 ‘solicit response’. The move was, however, considerably more frequent in the online context, not least, it was suggested, due to the possibilities offered by hypertext. The M6 ‘solicit response’ move was found to make spam mails clearly promotional by encouraging further communication (Barron, 2006).

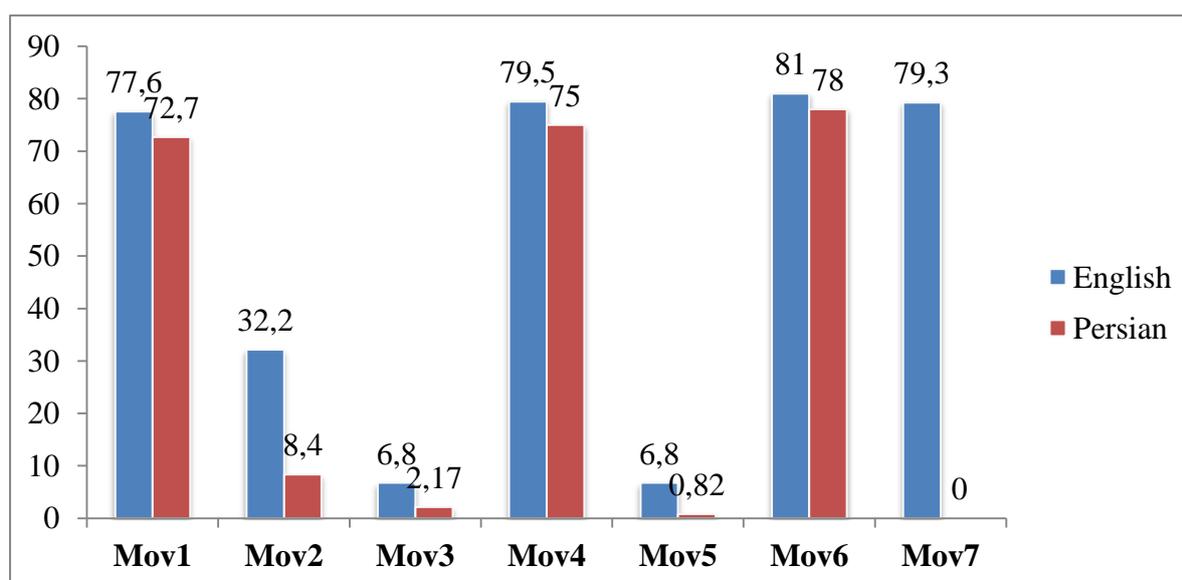


Figure1. Percentages of M1 to M7 by English and Persian spammers

### 4. 3 Move Positions in the English and Persian spam emails

Of equal interest to how frequently the individual moves occurred in the genre of spam mail and their relative lengths are where they occurred relative to each other in the spam. Thus, the location of two of the moves turns out to be quite predictable. One first observation about the position of the moves within spam mails is their presence in cyclic fashions. For instance, though M1 ‘capture attention’ and M7 ‘give a polite way-out’ did occur very frequently, when M1 ‘capture attention’ occurred as the initial move in the English spam emails 97% of the time, and M7 ‘give a polite way-out’ occurred as the final move before closing 100% of the time. A M7 ‘give a polite way-out’ followed half of the ‘pressure tactics’ moves whereas the other half was followed by a M6 ‘solicit response’ move. 75% (6) were preceded by an M4 ‘introduce the offer’ move.

As it is expected to a very great extent, are the positions of M2 ‘establishing credentials’ and M3 ‘welcome prospect’. If one ignores the presence of M1 ‘capture attention’, M2 ‘establishing credentials’ occurs as the initial move in the spam mails 74% of the time. And M2 ‘establishing credentials’, regardless of its position in the spam, is immediately followed by M3 ‘welcome prospect’ 87% of the time. Another interesting feature about M3 ‘welcome prospect’ is that 25% of these moves had either a M4 ‘introduce the offer’ or a M5 ‘use the pressure tactics’, or both, embedded in them, which represented 50% of the combined occurrences of these two moves in the corpus.

Of the remaining 132 combined occurrences of M4 ‘introduce the offer’ and M5 ‘use the pressure tactics’, they immediately followed M3 ‘welcome prospect’ 60% of the time. M6 ‘solicit response’ is probably the most flexible of the moves occurring, when it does, with nearly equal frequency towards the beginning, middle, and end of the both English and Persian spam emails.

### 4. 4 Move structures of the spam emails from the two corpora

Based on the analysis, there was almost straightforward linear structure (M1-M2-M3-M4-M5-M6-M7) appearing in either set of data. Of these, only two patterns of move structure (M4-M6, M1-M4-M6) were shared by at least more than half of the English and Persian spammers.

Most spam emails in both sets of data were constructed cyclically (86.66 % of the English corpus and 76.66 % in the Persian corpus). M4 ‘introduce the offer’ and M6 ‘solicit response’ were the most cyclical moves in both datasets. Four moves, including M1 ‘capture attention’, M2 ‘establishing credentials’, M3 ‘welcome prospect’, and M5 ‘use the pressure tactics’ were non-cyclical moves in the Persian corpus, whereas only M3 ‘welcome prospect’ was a stable move in the English corpus. M7 ‘give-polite-way-out’ was also of a cyclical nature, particularly in the English corpus. It always re-occurred in a sequence with either M4 ‘introduce the offer’ or M6 ‘solicit response’ while it was not detected in the Persian corpus.

As it is observable in both of the summary figures (Tables 2 and 3), it should be noted that it is quite difficult and challenging to propose a fixed structural framework, and that there is no such thing as an absolutely established organizational pattern in this genre. Therefore, according to Bhatia (2004), we should not expect to be able to analyze genre with a high degree of predictability and certainty.

## 5. Discussion

Based on the results of the genre analysis of the 200 spam emails in the English and Persian corpora, a couple of observations can be made about how moves are used within this genre. First of all, the results has revealed that some of the moves originally identified by Bhatia (1998) appear to

be obligatory in the genre, while others are seen to be merely optional. Secondly, it seems clear that the juxtaposition of the moves relative to each other shows meaningful patterns.

M1 'capture attention', M4 'introduce the offer' and M6 'solicit response' appear to be required moves in this genre. The preeminence of these three moves can be discerned by the fact that not only do they occur in nearly every English and Persian spam mail in the corpora, but they generally occur more than once, they usually occur as the first, fourth and sixth moves in the spam, they are by far the longest of the moves.

That M1 'capture attention', M4 'introduce the offer' and M6 'solicit response' are the most significant – in frequency, size, and position in the spam emails – is not surprising. At its most basic level, the purpose of the spam email is to tell the readers what the organization is, thereby leading to establish or enhance new relations. These functions are accomplished in these three moves.

The other four moves serve, then, as optional tools that individual spammers in this genre can incorporate in various ways to tailor the effect of the spam on the recipient. For example, M2 'establishing credentials', M3 'welcome prospect', M5 'use the pressure tactics' and M7 'give-polite-way-out' clearly play a secondary role in the spam mails as they tend to be quite short in length and as often as not are embedded in another move, usually M3 'welcome prospect'. Essentially, it seems their function is to serve as a reminder. In the case of M2 'establishing credentials', the recipients most often are reminded of organization's competence. With M3 'welcome prospect', the function of this move is simply to remind the recipients to look at invitation concerning previous contact that has been included with the spam.

M5 'use the pressure tactics', even occurring with low frequency in English and Persian spam emails, also plays an important role of informing the recipient how much the organization pushes them towards their purchase. Nevertheless, this role is noticeably a secondary one when the frequency, number of occurrences and length of this move are considered in relation to M2 'establishing credentials', M3 'welcome prospect' and M7 'give-polite-way-out'. M7 is clearly optional move, occurring in 100% of English spam emails since English spammers are obliged to include this move in their spams by strict rules imposed by government, whereas it is non-existent in the Persian corpus due to lack of binding rules for spammers.

## 6. Conclusion

A crucial goal for this study was to compare the schematic structures of spam emails from genre analysis vantage. This orientation appears to be one that will prove to be very fruitful, providing much of the qualitative detail common to genre analysis – allowing us to answer questions about how spams are structured and organized and why – while at the same time offering the reliability that is best assured by the quantitative power of research.

This study proved that the genre of spam mail does indeed live up to this billing as representing a very dynamic form of language use offering many creative options. Nevertheless, like all genres, there is clearly an inherent, coherent structure that defines it, although this genre may offer more flexibility in its structure than many. Significance of this flexibility appears to be the fact that there may be inconsistency in move sequencing, frequencies and positions between spams written by the English spammers representing differently from the Persian spammers. For instance, preliminary analysis seems to indicate M2 'establishing credentials' not only occurs more frequently, but tends to be longer in spams written by English spammers than for those written by Persian ones. Future research on the corpora can be undertaken to scrutinize into the ways those

different marketing domains may organize common moves in peculiar ways to appeal to their specific audiences.

Although it appears that the schematic structure for spam emails that has been proposed here is a robust one that readily accounts for a wide variety of spams that make up the corpora, this analysis offers limited insight into how these moves are realized linguistically. For example, are there common superficial linguistic features common to this genre? While this structural analysis is an important first step toward understanding what makes a spam email, this knowledge will benefit practitioners most when we can also discuss and provide examples of the common linguistic features of these spams.

## References

- Anthony, L. (2003). *AntMover*. Available online—Google Search
- Barron, A. (2006). Understanding spam: A macro-textual analysis. *Journal of Pragmatics*, 38(6), 880-904.
- Barron, A. (2006). 'Click Here to Shop Now! A micro-linguistic analysis of medical spam mails'. In Boenigk (ed.), *Innovative Wirtschaftskommunikation: Interdisziplinäre Problemlösungen für die Wirtschaft. Europäische Kulturen in der Wirtschaftskommunikation*, vol. 9, Deutscher Universitäts-Verlag, Wiesbaden, pp. 209-226.
- Bhatia, V. K. (1993). *Analyzing genre: Language use in professional settings*. New York: Longman.
- Bhatia, V. K. (2004). *Worlds of Written Discourse. A Genre-Based View*. London: NY Continuum.
- Cochran, W. G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.
- Connor, U. and Upton, T. (2003). Linguistic dimensions of direct mail letters. In: Meyer, C.F., Leistyna, P. (Eds.), *Corpus Analysis. Language Structure and Language Use*. Rodopi, Amsterdam, pp. 71-86.
- Crystal, D. (2001). *Language and the internet*. Cambridge: Cambridge University Press.
- Cukier, *et al.* (2006). Genres of spam: Expectations and deceptions. Proceedings of the Hawaiian international conference on system sciences—2006.
- Flowerdew, J. and Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific purposes*, 29(2), 78-93.
- Gagnon, G. (2004). Detecting spam. Available online at <http://www.pcmag.com>.
- Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16, 119-138.
- Spam. *Online Etymology Dictionary*. Available online at Dictionary.com website: <http://dictionary.reference.com>.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.

Wordsmith (2012). Available online at <http://www.lexically.net/wordsmith/index.html>.

Zeltsan, Z.(2004). General overview of spam and technical measures to mitigate the problem. ITU-T SG 17. Interim rapporteur meeting, November, 2004.

### **Biographical Note**

Dr. Gholam Reza Zarei received his Ph.D from Isfahan University in 2001. He is currently the director of English Language Center, Isfahan University of Technology, Iran. He has taught general and specialized English courses at both undergraduate and graduate programs. Also, He has published some skill books as well as many articles in national and international journals.

### **Biographical Note**

Parviz Ahmadi Darani is a PhD candidate and is currently teaching at Sheikhbahae University of Isfahan, Iran. His research interests include ESP, assessment, material development, discourse analysis, and translation. He has published some articles in both national and international journals.