

Peter Backhaus

You've got sp@m: a textual analysis of unsolicited Japanese dating invitation mails

Abstract: This paper analyzes the language of Japanese spam mails. Special focus is on one specific type of spam: make-believe dating invitations by women looking for physical relationships with male partners. A corpus of 434 spam mails was compiled between 2009 and 2012. Looking at two of these messages in detail, the paper examines the specific properties of this type of spam. These include linguistic features commonly associated with computer-mediated communication (CMC) and Japanese women's language. Their function is to increase the "authenticity" of the messages. In a second step, I analyze how spam mail writers in their messages deal with the problem of portraying female sexuality and desire, with a special focus on the role of linguistic taboos and transgressions thereof.

Keywords: computer-mediated communication (CMC), Japanese women's language, sexuality, spam

Peter Backhaus: Waseda University, Tokyo, e-mail: backhaus@waseda.jp

You've got sp@m – 出会い系サイト・スパムメールのテキスト分析 –

バックハウス・ペート

本稿は日本語のスパムメールを分析する。特に男性パートナーとの肉体的関係を求めている女性のふりをした招待メールに注目する。434通のスパムメールを含むコーパスは2009年から2012年の間に収集されたものである。本稿はその中から二つのメッセージを詳しく見ることで、この種のスパムメールの特質を考察する。そこには「コンピュータを介したコミュニケーション」(CMC)と女性語に関連した言語学的特性が含まれている。そして、その機能とはメッセージの「信憑性」を増すことにあるのである。第二段階ではスパムメールの書き手がそのメッセージの中で、どのように女性の性と性欲を表現することの問題を取り扱っているのかを、とりわけ言語的なタブーの役割、およびその逸脱に注目しながら分析していく。

1 Introduction

The term *spam* refers to “irrelevant or unsolicited messages sent over the Internet, typically to large numbers of users, for the purposes of advertising, phishing, spreading malware, etc.” (Oxford Dictionaries 2010). It is one of the most common forms of Internet communication, with some estimates suggesting that over 90 % of all e-mails fall into this category (*BBC News*, 4 April 2009). Luckily, only a small fraction of them make it into the inboxes of end users (Goodman et al. 2007: 26), although most people would agree that this is still more than enough. The most frequent types of spam mails offer finance and business opportunities (including the well-known “Nigerian letter”), medical and pharmacological products and services, computer soft and hardware, as well as “adult-oriented” matters (Ahmed and Oppenheim 2006; Cukier et al. 2008).

The type of spam discussed in this paper falls into this last category. It will be referred to as “x-spam,” used here as a general term for various sorts of unsolicited Japanese dating invitations that share a number of characteristics. To start with, they are from females looking for males. The type of relationship that is sought is in most cases sexual, which in the messages is described in varying degrees of explicitness. It is frequently emphasized in the mails that there will be no charge for the male partner. On the contrary, in many cases larger sums of money are offered to the recipient if he is willing to establish a relationship with a female applicant. Alluding to the well-known phenomenon of *enjo kōsai*, a form of “compensated dating” where males pay money to (mostly younger) females to meet them (Leheny 2006: Ch. 3, Ch. 4), the type of relationship that is sought in x-spam has been referred to as *gyaku enjo kōsai* or ‘reverse compensated dating’ (e.g., Sonoda 2007: 124).

With regard to the purported senders, x-spam messages may either be designed as if coming from an individual female looking for a male partner, or as being sent by a matchmaking agency in search of new male members to meet the growing demand from their female subscribers. In the latter case, too, the text frequently contains some shorter messages from female applicants. The agency’s representative who sends the mail in most cases also identifies herself as female, sometimes even admitting that she too is presently on the lookout for a partner.

Needless to say, both types of messages are entirely fake. Saeki (2005) took the trouble to respond to a large number of x-spam messages to see what happens if they are taken at face value. As expected, he found that all of his attempts to get in touch with some real applicant at the other end of the line quickly took him to a stage where he was requested to reveal some confidential information about himself.

Based on these experiences, Saeki (2005: 22–23) identifies three main purposes of x-spam. The first, least harmful, is to prompt the recipient to click a link included in the body of the message, which will take him to some sort of commercial site or worse. The second and more malicious purpose is to make the recipient reply to an x-spam message. This will show the spammers that the address the reply comes from really exists and hence can be used or sold for further spam business. The third purpose goes beyond this by tricking the recipient into giving away sensitive personal information such as his mobile phone number or credit card details in the course of the exchanges that follow.

While most people would dismiss the idea that they would ever fall for such schemes, the mere fact that x-spam continues to circulate suggests that there is some “success” rate, however small (Sonoda 2007: 124). To succeed, then, the first hurdle the senders of x-spam must overcome is making the recipient believe that the message is genuine. The main linguistic strategies to do so will be discussed in Section 4. A second requirement is that the recipient takes an interest in the woman the message purportedly comes from. This will be examined more closely in Section 5.

2 Previous research

The persistent occurrence of x-spam in the last couple of years has started to attract the attention of Japanese sociolinguists and communication researchers. In a special issue of the journal *Nihongogaku*, Akizuki (2008) analyzes the rhetoric of x-spam messages. Mainly focusing on messages that claim to be sent by matchmaking agencies, he identifies various strategies, and classifies them on the basis of Brown and Levinson’s (1987) politeness theory. Common features he finds include an emphasis on ease of use and the safety of an agency’s services, mention of some extraordinary chance for the recipient being invited to become a member of the organization, and use of casual or female language to enhance the “authenticity” of the messages. This last aspect in particular is of importance for the analysis in this paper.

A second recent study of x-spam has been conducted by Nakamura Momoko (2009; see also 2007: 110–118). Working with a sample of fifty messages received in October 2006, she identifies three basic types: (i) messages by contact agencies, (ii) messages by female individuals, and (iii) messages by male individuals. Special focus is on the role of men’s and particularly women’s language in these messages. Nakamura’s analysis is conducted against the backdrop of a growing number of empirical studies into Japanese language and gender showing that real linguistic practices are far more complex than norma-

tive ideas and stereotypes would suggest (see Okamoto and Shibamoto-Smith 2004). Such stereotypes, however, play an important part in creating gendered personalities in x-spam messages.

As Nakamura Momoko (2009) holds, such messages thus

[...] provide us with good data because they present typical ways of eliciting a receiver's sexual interest using language only [...] As their major purpose is to draw the sexual interest of heterosexual male receivers, it is assumed that the writers choose to present socially appropriate and acceptable female heterosexuality in a caricature manner. Analyzing their use of "women's language" and "men's language", therefore, enables us to understand a normative representation of heterosexuality. (Nakamura Momoko 2009: 2)

A third study of interest is Nakamura Chisako's (2009) research into the language of so-called "bank transfer scams" (*furikome sagi*). In this relatively recent type of fraud, people are tricked into transferring large sums of money to some dubious bank account. Nakamura examines two bogus debt collection reminders that fall into this category. Using relevance theory (Wilson and Sperber 2005), she looks at the interpretative processes that work on the recipient to make them settle a debt that has in fact never been incurred. Though the type of message is different from those discussed in this paper, the make-believe strategies that Nakamura identifies share some striking similarities with those found in x-spam messages. So do, of course, the criminal purposes of these messages.

3 Methodology

Data were collected between April 2009 and March 2012. During these three years, I retained all unsolicited e-mails in Japanese that were identifiable as some sort of dating invitation in an extra folder in my inbox. By the end of the collection period, the total number of these messages exceeded 2,000. Unlike Saeki (2005), I did not engage in any follow-up investigation such as replying to messages or clicking one of the links contained in the mails.

A closer look at the collected messages showed that many were virtually identical. That is why I decided to work with types rather than tokens, discarding all messages that had duplicates with respect to sender and subject line. Although it is possible that some of these messages may have differed slightly in content (see Barron 2006: 887), it was considered necessary to avoid distortions arising from a large number of duplicates. The procedure reduced the final

number of messages to a corpus of 434 items. As it turned out during closer analysis of the data, a small number of almost identical messages had remained. These were retained.

With regard to the different types of x-spam, a quick quantitative analysis showed that 217 messages – exactly 50% of the corpus – took the shape of dating invitations from matchmaking agencies. Another 199 messages (45.9%) were allegedly sent by individual females. Two of these texts will be analyzed in detail in the following sections. Only 18 messages (4.1%) were not classifiable as belonging to one of the two categories, in most cases because they were a non-definable hybrid between agency and individual mails.

In order to make the corpus accessible for further analysis with concordance software, a text file of each message was created, including the sender and subject lines. All files were tokenized using ChaSen (<http://chasen-legacy.sourceforge.jp>), a morphological parser for Japanese texts. Subsequent analysis was conducted with AntConc (<http://www.antlab.sci.waseda.ac.jp/software.html>), a freeware multi-language concordancer developed for corpus linguistic analysis.

Before looking at the data in more detail, a few remarks about terminological issues are in order. As it is clear that the messages contained in the corpus were not sent by the purported senders, there is no need for constant disclaimers about this during the discussion. On the contrary, in order to highlight the make-believe processes at work, a larger part of the analysis refers to the senders as if they were genuine. As to the actual senders, commonly referred to as “spammers,” it needs to be kept in mind that spam messages are frequently generated and disseminated by computers (see Lavergne et al. 2011). Nevertheless, there can be no doubt that the basic text components of the messages in the corpus were produced by human writers. Finally, with regard to the receiving end of the messages, I take the liberty to use generic “he,” assuming that the primary target group of x-spam messages is male.

4 Making it real

As the ultimate goal of x-spam is to elicit some sort of recipient reaction, the messages must create the illusion of some real person waiting for a reply. This section takes a closer look at the various linguistic devices spammers employ to overcome this challenge, analyzing one example from the corpus in detail.

- (1) Message from Yurika (#09.06.17. 0613)
- 1 Subject: 久しぶり～♪
 - 2 From: yuriringo@ezweb.ne.jp
 - 3 Date: 2009/06/17 6:13
 - 4
 - 5 前サイトでアドレス交換したゆりかだけど覚えてるう～???
 - 6 最近アドレス変えたから教えるね&最近モバイル CITY ってゆー SNS
 - 7 はじめたから招待するねえ♪
 - 8
 - 9 変わったアドレスは yuriringo@ezweb.ne.jp だよお♪
 - 10
 - 11 メールくれたら招待するねえ♪
 - 12
 - 13 携帯アドからメールくれるとすぐメール返せるよ o(*^▽^*)o
 - 14
 - 15 ちなみにPCアドだと受信拒否しちゃってるから届かないと思うの (*_ _)

The message presented in example (1) is from a sender named Yurika. The subject line reads, “Long time no see” (line 1). The message starts with a brief self-introduction by Yurika, framed as a question about whether the recipient remembers her from when “we exchanged addresses” on some previous occasion (line 5). Today, she wants to inform us that her contact details have changed and that she has started using a new social networking service (SNS) called “mobile CITY” (*mobairu* CITY; lines 6 and 7). Next she lets us know her new address (line 9), adding that she will invite us to join the SNS once we have sent her a message (line 11). She specifies that she would prefer hearing from us via our mobile phone address, because messages from “PC addresses” are blocked (line 15). As can be understood from the domain name of Yurika’s e-mail address (lines 2 and 9), hers is a mobile phone address, too.

The simple logic behind all this is that if the recipient should indeed respond to Yurika’s mail as specified, the spammers will get hold of his mobile phone address. The question is why someone would want to give away such confidential information to an unknown person on the Web unless there was some potential benefit for him. This must be the prospect of getting connected with Yurika. Even if the recipient cannot remember having ever been in touch with any such person – how could he? – there is still some small chance that Yurika has sent him this message by mistake, which he can now take advantage of to get things going. At least, this is what the spammers would have him believe. As Nakamura Chisako’s (2009) analysis of fraudulent debt collection reminders has shown, such interpretative leeway is a common feature in messages of this type.

The crucial condition for all of this to work is that Yurika is “real,” and the spammers’ main task is to make this credible. In order to create this illusion, they draw on a great repertory of linguistic devices to mimic the register that most people would find appropriate for such types of computer-mediated communication (CMC). As a closer look at example (1) shows, there are indeed a great number of features commonly observed in Japanese CMC (see, e.g., Gottlieb 2010; Katsuno and Yano 2002; Miyake 2005, 2007; Nishimura 2003, 2010).

An important feature of CMC in Japan – as elsewhere – is that it contains various elements characteristic of spoken language. The first thing to note in this respect is the very casual tone of Yurika’s message. Thus, it is written completely in plain style, without a single occurrence of the addressee honorifics *desu* and *masu*. In addition, the text contains a large number of discourse particles (“*ne*” in lines 7 and 11, “*yo*” in lines 9 and 13, and “*no*” in line 15), which are also reminiscent of spoken language. The combination of these two features creates a talk-like register that is a common characteristic of this type of communication.

Other features borrowed from spoken language include ellipsis and contraction of (mainly grammatical) morphemes. Most noticeable is the omission of the object particle *o* in no less than six instances (asterisks indicate non-marked accusatives): *adoresu** *kōkanshita* ‘exchanged addresses’ (line 5), *adoresu** *kaeta* ‘changed address’ (line 6), *SNS** *hajimeta* ‘started SNS’ (lines 6 and 7), *mēru** *kuretara* ‘if [you] send [me] mail’ (line 11), *mēru** *kureru* ‘send [me] mail’, and *mēru** *kaeseru* ‘can reply by mail’ (both line 13).

A related example is the contraction of the progressive/resultative verb form *-te iru* into *-teru*. The two examples in the text are *oboete*ru* ‘remember [me]’ (line 5) and *jushin kyōhi shichatte*ru* ‘reject message’ (line 15) (asterisk indicates omitted *i*). Note that the latter example involves further contraction in that the verb construction *shite shimau*, used here to express that the rejection of the sender’s message would be an unwanted state of affairs, is shortened to *shichau*. A similar case is the spelling of the quotative particle *to* as *tte* (line 6), thus orthographically mimicking phonemic processes (gemination, vowel bleaching) that commonly occur in spoken language. The same holds for the deviant spelling of the directly following verb “say” as *yu* rather than *iu*.

Of further note about Yurika’s message is the non-standard marking of vowel lengthening for stylistic purposes, another characteristic of Japanese CMC. The text contains a total of six instances of this feature (asterisks indicate lengthening marks): *hisashiburi** ‘long time no see’ (line 1), *oboeteru** ‘remember [me]?’ (line 5), *tte yu** ‘called’ (line 6), discourse particle *ne** (lines 7 and 11), and discourse particle *yo** (line 9). These cases clearly deviate from lengthening in standard orthography, as found for instance in *mēru* ‘mail’ (lines 11

and 13), not least because they are not normally marked as long vowels at all. In addition, the lengthening is orthographically deviant in that it is indicated by non-standard graphemes such as a wave dash (line 1), a small hiragana letter (lines 9 and 11), or a combination of the two plus a triple question mark (line 5).

This relates to another characteristic feature of CMC, namely the use of pictographic elements. Miyake (2007) makes a basic distinction between pictorial signs, symbols, and emoticons, also known as *kaomaji* ‘face marks’. Symbols and emoticons also occur in Yurika’s message, whose subject line already ends on a note symbol (♪). The same symbol reappears at the end of lines 7, 9, and 11. In quantitative terms, the note is the most frequent symbol in the corpus, with a total of 206 occurrences. As for emoticons, Yurika’s mail contains two such items: at the end of lines 13 and 15. In both cases, as is common practice in Japanese CMC, the face mark is to be read in line with the words rather than, as in most alphabet contexts, perpendicular to them (see Katsuno and Yano 2002: 206). The corpus contains 43 face marks in total.

The analysis so far has identified various linguistic features that spammers draw on in order to fool recipients into taking their messages for real. As has been outlined, they do so by closely imitating the register of Japanese computer-mediated communication. Another crucial point is that the alleged sender must be recognizably female, for which the sender’s name is a first and most obvious index. In the present example, the sender right at the beginning of her message self-identifies by the female name Yurika (line 5). The name is spelled in hiragana, which adds a certain cuteness (Akizuki 2005: 172). In addition, due to its ending in *ka* rather than the more traditional *ko*, it is easily associated with a woman of younger age (see Tomimori, 14 April 2012). A quantitative analysis of the corpus shows that the number of senders whose first name does not end in *ko* clearly exceeds the number of those that do. Of the 260 messages that contain a Japanese female name in the sender line, only 92 (35%) end in *ko*, as opposed to 168 (65%) that end differently. The most frequent names of this type are Miho (8 messages), Emi, and Yumi (both 7 messages).

Another possible device of self-identification is the sender’s address. In the example, the address given in the sender line (line 2) and repeated in the body of the message (line 9) is “yuriringo@ezweb.ne.jp”. The portion to the left of the @ mark contains an abbreviated form of the sender’s name, *yuri*, followed by the word *ringo* ‘apple’. The two terms are connected in a *shiritori way* (i.e., the second word starts with last kana character of first word). The playfulness displayed here again evokes a young female sender.

Apart from such straightforward “facts” as name and address, there are a couple of more subtle indications in the body of the message that suggest that this is a message sent by a (young) woman. One of the most commonly men-

tioned features of Japanese women's language is the use of discourse particles. As has been described above, Yurika's message contains five such instances, using the three particles *ne*, *yo*, and *no*. While the first two are commonly ascribed to both women and men, the third one is clearly indexical of a female speaker (e.g., Ide and Yoshida 1999: 465).

Another feature that has already come up in the analysis is the marking of vowel lengthening for stylistic rather than phonemic purposes. As Miyake (2007: 60–61) points out, particularly the use of small hiragana characters to indicate such lengthening “has become increasingly popular recently among female writers, and some seem to use this technique in their messages at every possible point [...] The effect of the deviant small letters is to make the messages soft and ‘*kawaii*’ [cute].” The 56 instances of small hiragana in the corpus testify to the fact that spammers use this device deliberately. Likewise, the use of symbols and emoticons is commonly considered to be particularly frequent in messages sent by females. Recent empirical research supports this (Scott et al. 2009; see also Baron 2008: 64–69).

To summarize, spammers through their use of language try to make their messages appear genuine. To this end, they rely on two linguistic registers: CMC and Japanese women's language. The next section examines how common norms about the expression of female sexuality are reflected in x-spam.

5 Saying the unsayable

One inherent problem of x-spam is how to deal with sexual taboos. By definition, these are not usually spoken, let alone written about. Taboo words are words that are “deemed unfit for normal linguistic usage and by community consensus are banned in everyday language in the public domain” (Apte 2001: 284). On the other hand, x-spam would not be what it is without any content intended to arouse sexual desire. This commonly necessitates the transgression of taboos (Cameron and Kulick 2003: Ch. 5). Things get further complicated by the fact that most x-spam messages pretend to be written by female senders, thus putting the burden of expression on them. This section looks at how (fake) female desire is depicted in x-spam and what transgressions of sexual taboos are involved to do so. To this end, a closer look will be taken at another message from the corpus, presented in example (2) on the next page.

This message purportedly comes from a woman called Misako. Though this is not directly obvious from the sender line, which reads “Secret Affairs Club” (line 2), the sender self-identifies with her name in the last line of the message (line 32). On closer inspection, it can be seen that the name is also contained

(2) Message from Misako (#09.10.30.1943)

1 Subject: こういうエッチって初めてなんですが・・・
 2 From: 秘密の愛人クラブ <misako_rc2459514_in@yahoo.co.jp>
 3 Date: 2009/10/30 19:43
 4
 5 ここはエッチがうまい人が多いって、ママ友の間で話題になってます。
 6
 7 しかも何人もの男性と会えると聞いて、アソコがうずいちゃってます (#^^#)
 8
 9 できれば今からでもお願いしたいくらいです。
 10
 11 よかったらお返事下さい。
 12
 13 <http://rocalobirities.com/neto/?RC2459514>
 14
 15 一応、名前とかは聞かない約束でお願いします。
 16
 17 私も家庭は大切にしたいので・・・
 18
 19 お小遣いはしっかりあげますから、遊び感覚で会ってもらえると嬉しいです。
 20
 21 夫がフェラ好きだったので、舌使いには自信があります♪
 22
 23 <http://rocalobirities.com/neto/?RC2459514>
 24
 25 とか恥ずかしいけど、セックスは好きなほうです・・・。
 26
 27なのに数年前に夫が勝手に現役引退しちゃったので、最近では性欲がたまってるのが
 28自分でも分かります。
 29
 30 お願いです。そんな私を助けてください・・・(;_;)
 31
 32 美佐子

in the sender's address (line 2). The result is a somewhat confusing mixture that leaves it unclear whether the message has been sent by an organization or an individual.

In the subject line, Misako informs us that this is the first time she has become involved in “such kinds of” sexual encounters. What precisely she means is further specified in the body of the message. In line 5, we learn that she heard from her peer “mama friends” that there are many people around “here” who are good at sex. It remains unclear what exactly “here” means, though most likely it is supposed to refer to some virtual community, probably the aforementioned “Secret Affairs Club”. This place, as she tells us in line 7, is

also known for providing opportunities to date more than one man, the mere thought of which she finds sexually stimulating. She describes this in a rather explicit way, to be discussed further below.

Misako goes on by saying that, if possible, she would like to make all of this happen right now (line 9), and therefore would be happy to hear from us (line 11). The Web address that follows in line 13 suggests that she wants us to get in touch with her via this link rather than by replying to her e-mail address. The link reoccurs in line 23.

In the remaining part of the message, Misako confesses that she is married and wants us to respect each other's privacy (lines 15, 17). She also describes and further expands on her sexual skills (lines 21) and sexual desire (line 25), which has become increasingly strong since her husband "retired from his duties" (lines 27, 28). In addition, she assures us that there will be some "pocket money" for us if we meet up (line 19). Misako's message closes with a request to "help" her (line 30).

The example contains many of the characteristic CMC features outlined in the previous section. Most noteworthy is the occurrence of spoken language elements such as the contraction of *te iru* (lines 5, 7, 27) and *te shimau* (lines 7 and 27) forms, as well as the use of emoticons (lines 7 and 30) and the music note symbol (line 21).

On the other hand, there are also a couple of features that distinguish Misako's message from Yurika's. One of them is the speech style, which in the second example is much more formal, with virtually all affirmative sentences ending on *desu* or *masu*. One likely reason for this is the age differences of the two personae that the spammers are trying to create. While the casual style in example (1) is intended to imitate the speech of a relatively young and most likely single female, the formal language in example (2) is to evoke a more mature woman, in line with her implicit self-description as a wife (lines 17, 21, 27) and a mother (lines 5, 17). This coincides with the fact that Misako's name has the more traditional *ko* ending.

A second major difference between the two examples is that Misako's message is far more explicit. While in the case of Yurika, anything supposed to happen beyond the (more or less innocent) exchange of mail addresses is left entirely to the recipient's imagination, Misako's message leaves no doubt whatsoever as to the sexual nature of the encounter she is after. This becomes apparent already in the subject line (line 1) with her use of the term *etchi*, an all-encompassing expression for various kinds of sexual states of minds and activities (see West 2011: 1–2). It reoccurs in the first sentence of the message's body (line 5). Other terms explicitly relating to sexual activities are *fera*, a clipped

form of the loan word *ferachio* ‘fellatio’ (line 21), *sekkusu* ‘sex’ (line 25), and *seiyoku* ‘sexual desire’ (line 27).

Despite the explicitness of these terms, Misako’s message is clearly designed to display an awareness of common social norms according to which it is not appropriate for females to express sexual desire. Thus, in terms of speech style, one more motivation for her use of the formal style may be to counterbalance the sexual explicitness of her message. As Nakamura Momoko (2009: 13) has suggested, polite and super-polite language features in x-spam may be employed “to avoid appearing too sexually aggressive.”

On the lexical level, Misako’s awareness of sexual taboos becomes obvious from her use of euphemisms. For example, she circumscribes the sexual encounter she is looking for by the rather innocent verb *au* ‘meet’ (lines 7 and 19); the remuneration she is willing to pay to her male partners is referred to as *okozukai* ‘pocket money’ (line 19); and her husband’s recent sexual indifference is paraphrased as *geneki intai* ‘retirement from active duty’ (line 27).

The most obvious euphemism occurs in line 7, where Misako describes sexual arousal by an itchy feeling in her pubic region. She refers to this part of her body as *asoko*, a deictic expression that literally just means ‘(over) there’. Though there is no doubt as to what this term is supposed to mean here, a more explicit expression is avoided. This taboo consciousness is a general feature of the corpus, which contains 45 references to female genitals in total. With 28 occurrences (62%), *asoko* is the most frequent option. In 24 of the cases, the term is spelled in katakana, probably to graphically distinguish it from ordinary usage. Of further note is that there are no other occurrences of *asoko* apart from the 28 cases just mentioned. In other words, there is no such thing as ordinary usage of the term in the corpus at all.

Misako’s orientation to normative ideas about female sexuality also becomes apparent on the rhetorical level. To start with, she considers it necessary to mention right in the subject line that this is her first time (*hajimete*) to look for such encounters. Furthermore, her rather straightforward affirmation that she is a person who likes sex needs to be mitigated by an expression of embarrassment (*hazukashii*) about being this way (line 25). This is graphically reinforced by the face mark closing the *asoko* line (line 7), which represents blushing. The second face mark at the end of her message (line 30) even shows her in tears, as though to acknowledge that there is something substantially wrong with her.

This second example, therefore, depicts female sexuality as something problematic that requires a larger repertory of linguistic devices to abate its explicitness. When composing their messages, x-spammers thus face a dilemma. On the one hand, their fake female authors need to include explicitly

sexual content in order to stimulate the sexual interest of the male recipients and make them react in some way. On the other hand, the texts need to reflect social norms according to which it is inappropriate for a woman to openly express sexual desire.

6 Conclusions

This paper has dealt with Japanese spam messages containing sexual content. Analyzing two such messages in detail, we see how spammers try to elicit a reply from male recipients by creating make-believe dating invitations by make-believe female persons. They do so by imitating the linguistic characteristics of genre (computer-mediated communication) and gender (women's language). With regard to the former, it is clear that spoken language features and the use of emoticons and other graphic characters is a common feature of x-spam messages. As to women's language, gender-indicative items such as discourse particles and the use of small hiragana characters have been identified.

How realistic the final product eventually is differs from message to message, as well as from reader to reader. My personal impression from working through the 434 mails of the corpus has been that there are indeed differences with regard to the "authenticity" of a message. While some could easily be (mis)taken for real, others exaggerate both genre and gender features, a phenomenon Kinsui (2003) has referred to as "virtual Japanese."

Making the messages look "real" largely means conforming to commonly held stereotypes about the language of women and the language of cyberspace. As Nakamura Momoko (2009; see the quote in Section 2) has emphasized, Japanese women's language can be a handy resource to create fake female senders. A similar point could be made with regard to computer-mediated communication, too. Corpus linguistic research may help to further test these two claims on an empirical basis (see, e.g., Nishimura 2010).

Section 5 has looked at the problem of having the make-believe female authors of x-spam messages make their indecent proposals. This has been analyzed in detail for the text in example (2). On the lexical level, various euphemisms are used to avoid direct reference to sexual concepts (though the example also contains a couple of rather explicit terms). A more technical factor, which has not been part of the analysis but should be taken into consideration, is the deliberate use of non-explicit language in order to avoid getting caught in spam filters (see Crystal 2011: 71–72).

Display of taboo awareness, however, has also been identified beyond the lexical level, in terms of speech style and rhetoric. Most notable features dis-

cussed are the use of the formal style to mitigate sexual aggressiveness and the expression of embarrassment, both through verbal and “facial” (*kaomoji*) expressions.

Spam messages like those discussed in this paper thus provide an interesting topic for linguistic research. The very fact that they are entirely spurious makes them hold many insights as to the role of language in making people and performing gender, both fictive and non-fictive.

Bionote

Peter Backhaus is Associate Professor at Waseda University, Tokyo. His main research fields are sociolinguistics, conversation analysis, and corpus linguistics.

References

- Ahmed, Tanzila & Charles Oppenheim. 2006. Experiments to identify the causes of spam. *Aslib Proceedings* 58(3). 156–178.
- Akizuki, Kōtarō. 2005. *Arienai Nihongo* [Impossible Japanese]. Tokyo: Chikuma Shinsho.
- Akizuki, Kōtarō. 2008. Supamu mēru no retorikku [The rhetoric of spam mails]. *Nihongogaku* 26(7). 36–47.
- Apte, M. L. 2001. Taboo words. In Rajend Mesthrie (ed.), *Concise encyclopedia of sociolinguistics*, 283–287. Amsterdam: Elsevier.
- Baron, Naomi. 2008. *Always on: Language in an online and mobile world*. Oxford: Oxford University Press.
- Barron, Anne. 2006. Understanding spam: A macro-textual analysis. *Journal of Pragmatics* 38. 880–904.
- BBC News*. 4 April 2009. Spam overwhelms e-mail messages. <http://news.bbc.co.uk/2/hi/technology/7988579.stm> (accessed 18 July 2012).
- Brown, Penelope & Stephen Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Cameron, Deborah & Don Kulick. 2003. *Language and sexuality*. Cambridge: Cambridge University Press.
- Crystal, David. 2011. *Internet linguistics: A student guide*. London & New York: Routledge.
- Cukier, Wendy L., Ojelanki Ngwenyama & Eva J. Nesselroth-Woyzbun. 2008. Genres of spam: Expectations and deceptions. *Scandinavian Journal of Information Systems* 20(1). 69–92. http://iris.cs.aau.dk/tl_files/volumes/Volume20/no1/20-1%20Cukier.pdf (accessed 18 July 2012).
- Goodman, Joshua, Gordon V. Cormack & David Heckerman. 2007. Spam and the ongoing battle for the inbox. *Communications of the ACM* 50(2). 25–33. http://cs.unc.edu.ar/~cic/dm2009/downloads/papers_de_interes/spam_cacm2007.pdf (accessed 18 July 2012).

- Gottlieb, Nanette. 2010. Playing with language in E-Japan: Old wine in new bottles. *Japanese Studies* 30(3). 393–407.
- Ide, Sachiko & Megumi Yoshida. 1999. Sociolinguistics: Honorifics and gender differences. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 444–480. Malden, MA: Blackwell.
- Katsuno, Hirofumi & Christine Yano. 2002. Face to face: On-line subjectivity in contemporary Japan. *Asian Studies Review* 26. 205–232.
- Kinsui, Satoshi. 2003. *Vācharu Nihongo: Yakuwarigo no nazo* [Virtual Japanese: The puzzle of role language]. Tokyo: Iwanami Shoten.
- Lavergne, Thomas, Tanguy Urvoy & François Yvon. 2011. Filtering artificial texts with statistical machine learning techniques. *Language Resources and Evaluation* 45. 25–43.
- Leheny, David. 2006. *Think global, fear local: Sex, violence, and anxiety in contemporary Japan*. Ithaca: Cornell University Press.
- Miyake, Kazuko. 2005. Kētai mēru no hanashi kotoba to kaki kotoba: Denshi media jidai no vijuaru komyunikēshon [Spoken and written features of mobile phone messages: Visual communication in the electronic media age]. *Media to Kotoba* 2. 50–75.
- Miyake, Kazuko. 2007. How young Japanese express their emotions visually in mobile phone messages: A sociolinguistic analysis. *Japanese Studies* 27(1). 53–72.
- Nakamura, Chisako. 2009. Furikome sagi higai ni okeru ukete no kaishaku purosusu [Receivers' interpretation processes in bank transfer scams]. *Media to kotoba* 4. 128–155.
- Nakamura, Momoko. 2007. *'Sei' to Nihongo* [Sexuality and the Japanese language]. Tokyo: NHK Books.
- Nakamura, Momoko. 2009. Language as heterosexual resource. *Shizen Ningen Shakai* 47. 1–23. http://opac.kanto-gakuin.ac.jp/cgi-bin/retrieve/sr_bookview.cgi/U_CHARSET.euc-jp/NI20000519/Body/01_1.html (accessed 18 July 2012).
- Nishimura, Yukiko. 2003. Linguistic innovations and interactional features of casual online communication in Japanese. *Journal of Computer-Mediated Communication* 9(1). <http://jcmc.indiana.edu/vol9/issue1/nishimura.pdf> (accessed 18 July 2012).
- Nishimura, Yukiko. 2010. A corpus-based comparison among CMC, speech and writing in Japanese. In Michaela Mahlberg, Victorina González-Díaz & Catherine Smith (eds.), *Proceedings of the Corpus Linguistics Conference CL 2009*. <http://ucrel.lancs.ac.uk/publications/cl2009/> (accessed 18 July 2012).
- Okamoto, Shigeko & Janet S. Shibamoto-Smith. 2004. *Japanese gender, language, and ideology: Cultural models and real people*. Oxford: Oxford University Press.
- Oxford Dictionaries. 2010. Spam. Oxford: Oxford University Press. <http://oxforddictionaries.com/definition/english/spam> (accessed 4 September 2012).
- Saeki, Kenzo. 2005. *Netto gentei renai kakumei: Supamu mēru taishō* [Net-only love revolution: Spam mail grand prix]. Tokyo: Tatsumi Publishing.
- Scott, Douglass J., Yuuki Kato & Shogo Kato. 2009. Comparing cultural and gender differences in the informal mobile telephone text messages of Japanese and American college students. *Waseda Journal of Human Sciences* 22(2). 71–86.
- Sonoda, Michio. 2007. Pawā appu kōzu: Kochira sekyuriti sōdanshitsu [Power up course: This is security consultation room]. *Nikkei Network* 5. 124–129.
- Tomimori, Hinako. 14 April 2012. Kodomo no nandokumei ga kyūzōchū [Rapid increase of children with names that are difficult to read]. *Asahi Shimbun*. 4.
- West, Mark D. 2011. *Lovesick Japan: Sex, marriage, romance, law*. Ithaca & London: Cornell University Press.

Wilson, Deirdre & Dan Sperber. 2005. Relevance theory. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 607–632. Malden MA: Blackwell.