

Discourse Issues in the Translation of Japanese Email

Laurel Fais and Kentaro Ogura

NTT Communication Science Laboratories

2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan

fais@cslab.kecl.ntt.co.jp Phone: +81 774 93 5316; fax: +81 774 93 5345

Summary

The translation of email presents unique challenges to translation systems. At the same time, features peculiar to email constitute new sources of information for these systems. A corpus of Japanese email is examined and its particular characteristics discussed: visual formatting, use and nature of openings and closings, visual devices for capturing aspects of speech (e.g., nonstandard punctuation, spelling, and use of English), and patterns of discourse marker, formal/informal verb form and verb tense use. It is suggested that author-specific features (use of closings, number of sentences, size of paragraphs, and use of visual devices) allow customization of translation. Implications of these features for modifications to translation systems are discussed.

Keywords: machine translation, discourse, email, Japanese discourse

Introduction

The use of the net is allowing people to be in contact with more people at greater distances than ever before. A robust email translation module is crucial to unimpeded communication in the current information technology scenario. However, systems designed to handle text or speech may not be optimal for email. It is well known that text differs in substantial ways from spoken language (Maynard, 1993; Chafe, 1982); we claim here that the characteristics of email text are significantly different from both. As a result, modifications are required before translation systems designed to handle either text or speech are appropriate for use with email.

The following sections outline the results of the analysis of a corpus of Japanese email and describe visual and discourse-level phenomena unique to email messages which a translation system must accommodate in order

to achieve usable translation. It is further suggested that incorporating knowledge about the habits of the user can make the translation system's performance more effective as well.

Methods

The data consist of a collection of 32 email messages exchanged among five office workers of a Japanese company in June, 1995. The messages constituted a collective attempt to schedule a sports watching outing. Thus, information concerning the corpus is given in Table 1.

The corpus was also labelled for topic-marked phrases, discourse markers, all referring expressions (including full NP's, explicit pronouns and elided pronouns), and tense and politeness level of the verbal elements. Quantifiable measures, including those summarized in Table 1, were subjected to analysis of variance to ascertain if the variability observed among authors is significant.

Table 1. Number of messages, paragraphs, sentences and characters per author.

Authors	Messages	¶'s	Sentences	Characters
H	5	9	20	880
I	10	46	50	1904
M	5	27	38	1268
R	5	29	56	1805
U	7	35	64	2015
Total	32	146	228	7872

Results: Features of the data specific to email

Visual information: organizational

Because email messages, unlike speech, are written, they present a visual aspect that is information-bearing. Authors exploit this to convey a variety of types of information.

Paragraphs. The use of indentation and spacing to mark paragraph shifts is highly idiosyncratic in this corpus. Examples 1-3¹ below show three different styles: no breaks (1), breaks but no indentation (2), and breaks, spaces and indentation (3).

(1)

エさん、こんにちは。

今月のスポーツ観戦ですが、プロ野球の方ですと、17日か18日のスワローズ対ベイスターズ戦にさせていただけると嬉しいです。僕は、今月の23日から27日まで沖縄に行く予定がありますので、ジャイアンツ戦の方ですと参加できません。それからJリーグのカードですが、6月17日からのサントリーシリーズ後半戦には国立で行われるゲームの予定はないようです。近いところでは、24日に市原でジェフ対サンフレッチェ、大宮でレッズ対セレッソ、28日に市原でジェフ対アントラーズ、大宮でレッズ対マリノスの試合があります。他の方がよろしければ、市原でのジェフのホームゲームを見にいきたいと思いますが、どうでしょうか。

Hello, I-san.

About catching a game this month, I will be happy to go to the Swallows versus Baystars baseball game on the 17th or the 18th. Since I have plans to go to Okinawa from the 23rd to the 27th this month, I cannot join you for the Giants games. As for J-league matchups, no games seem to be scheduled at the National Athletic Stadium for the last half of the Suntory Series starting on June 17th. At a nearby stadium, Jef versus Sanfreccia at Ichihara and Reds versus Cerezo at Ohmiya will be held on the 24th, and Jef versus Antlers at Ichihara and Reds versus Marinos at Ohmiya will be held on the 28th. If it is alright with others, I would like to go to see Jef's homegame at Ichihara. What do you think?

(2)

Uさん、おもしろそうなネタありがとう。た

¹ In the interests of space, fluent translations only are given. Word-by-word transliteration and translation are given for shorter examples. The author would like to thank an external translator for the translations given here.

まにはマイナーなもの見るのがそれっぽくていいのかな、とも思います。

ところで、ラクロスって何なんでしょう？
tokorode

このところ、「カバティ」と叫ぶやつとか、ほうきで掃くやつとか、見れないスポーツがテレビネタになってますけど、ラクロスもそんなノリなんですか？ 詳しい方いましたら、教えてください、

from M

U-san, thank you for your fun idea. I think that watching minor sports may be good for a change, as it is more our style.

By the way, what is lacrosse?

These days, there are some unusual televised sports where athletes yell, "kabaddi" and others where they sweep with a broom. Is lacrosse the same kind of thing as well? If there is anyone who knows about it in detail, please tell me.

from M

(3) 昨日から、取材で京都にきています。昨日はホテルに泊まったんですが、なんと、いまどきモジュラージャックがない！ まさかと思ってポケットダックをもって来なかったのに…。近くのISDN公衆電話で読むだけ読んだんですが、うしろのおばさんがにらむので、お返事が出せませんでした。今日は友達の家に来てるんで、ゆっくり通信できるというわけです。

さて、テコンドーはちょっとハードそ
sate
うですね。

Hさんの提案の、プロ野球でも日ハム戦がマリノスタジアムというのもいいかな。デーゲームなら、遠出でも東京に戻って来たところで宴会タイムには丁度いいですね。ぴあは東京に置いて来ちゃったんですが、オリックス戦があるといいな。プロ野球はあまり興味はないけど、イチローは好き(^^)。

U

I've been in Kyoto since yesterday to collect materials for business. **(Formal verb)** Last night I stayed at a hotel here, and what surprised me was that the room given to me had NO MODULAR JACK! **(Informal verb)** Since I had never expected it, I didn't bring my acoustic coupler with

me. **(Informal verb)** So I had to go to the nearest ISDN public phone to read your e-mail. **(Formal verb)** I wanted to reply at the time but I couldn't because a middle-aged woman waiting behind me looked irritated with me. Today, however, I'm staying at my friend's home, and don't need to feel rushed about sending e-mail.

Well, Taekwondo sounds a little bit rough.

H-san's proposal to watch a Fighters game or a Marines one at Marine Stadium sounds interesting to me. If we attend a day game held outside Tokyo, we can come back to Tokyo by the evening, just the right time for a drinking party. I haven't brought a *Pia* magazine with me, so I don't know what matchups are scheduled. I hope the Blue Wave will be playing against the Fighters or Marines at that time. I don't have much interest in professional baseball. But Ichiro is my favorite. :)

U

A difference in paragraphing is typically interpreted to cue a difference in topic (Maynard, 1998; Hinds, 1976); paragraph-like structural information has been shown to be important in the interpretation of elided pronouns (Hinds, 1976; Maynard, 1989; Makino, 1979). Both of these types of information can be exploited in translation.

Openings and closings. Email messages have in common with letters that they contain openings and closings (boldface in (1-3)). However, their use is much less constrained in email messages. Closings are typically formalized and devoid of meaning. Openings, on the other hand, contain information about the addressee(s); this is critical to the resolution of second person elided pronouns in translation into English.

The spatial positioning of openings and closings is also less constrained in email messages. Approximately one third of the openings used were incorporated into the opening line of the message (as in (2)). Closings were more consistent in placement, appearing separate from the body of the text (as in (2) and (3)).

Visual information: semantic

Punctuation represents, in text, intonational

and timing information normally associated with speech. In the more casual context of email messages, however, authors use additional visual strategies to capture further aspects of spoken utterances to enrich their textual messages.

Non-standard "punctuation." Center dots are by far the most frequent type of non-standard visual device used in this data (see fourth line, (3)). They are used intra- as well as intersententially and represent a "hanging" intonation which invites the listener/reader to draw inferences, supplementing the explicit meaning in the text. There are some examples of emoticons as well (last line of (3), above, and (5), below), which convey the attitude of the author toward the written content.

Use of katakana. Katakana is the syllabary used primarily to write in Japanese, words or phrases borrowed from foreign languages. It is also used for emphasis, especially in advertising. In email messages, katakana is used to give prominence to expressions normally written in hiragana (the "native" syllabary) or kanji (Chinese characters). In the last line of (4), the author emphasizes her preference by writing "watakushi" in katakana rather than kanji.

(4) でも できれば、 個人戦タイプ よ
り、
demo dekireba kojinsentaipu yori
But individual matches

いっせいに大展開する 試合 の
ほうが
isseinidaitenkaisuru shiai no hoga
fast-moving sports
to 好みでありますね、
konomidearimasune prefer

ワタクシは・・・。
watakushiwa I

But I prefer fast-moving sports to individual matches, myself.

Non-standard spelling. In a similar way, non-standard spellings (for example, elongating one sound by repeating the letter

several times, as in “ですうう” just before the emoticon in (5)) are also used to place prominence on a word or to mimic an emphatic pronunciation of the word. In (5), the author is emphasizing the lack of choice, excusing herself from blame for the inconvenience to her correspondents.

(5) で、ここで
 de kokode so here
 おわびしなくてはいけないんですが、
 owabishinakutewaikenaindesuga
 I must apologize but
 先方の都合で アポイント が
 朝
 senpono tsugoode apointo ga asa
 their schedule appointment
 morning からになっちゃったんですうう
 (>_<)
 kara ni nacchattandesuuuu
 had no choice but to set

So please excuse me but considering their schedule, I had no choice but to set an appointment for the morning.

Use of English. The use of English is yet another way to highlight a word. Some of the English used in this corpus is standard in Japanese (“OK”) or simply repeats the use of an English expression from another context (such “ISDN,” fifth line, (3)). English is used to highlight topics and to set off closings as well, as in (2).

Discourse characteristics

Different kinds of texts (including spoken “texts”) differ in the ways they are structured and in the cues, or discourse markers, that accompany those structures (Degand, 1998; Schiffrin, 1987). Because email messages are written text, they contain a number of discourse structures similar to those of newspaper articles or essays. On the other hand, as we have already seen, authors also attempt to capture the flavor of speech, and employ typically spoken discourse markers to do so.

Discourse markers. Words such as conjunctions and adverbial expressions often mark changes in discourse segment or denote

the nature of relationships among pieces of text (Schiffrin, 1987; Maynard, 1993). In this corpus, expressions such as *sate* (3), *tadashi*, *toiuwakede*, and *tokoro de* (2), when occurring sentence-initially, are reliable indicators of new paragraphs.

Table 2. Discourse markers in the corpus.

Discourse marker	#	Discourse marker	#
arui wa	1	sate	5
betsu ni	1	<i>sore de</i>	1/1
chinamini	1	sore dewa	2 ²
<i>de</i>	0/1	sorewasateoki	1
<i>demo</i>	1/5	soso	1
<i>desukara</i>	0/1	sou desu ne	1
douyara	1	sou ieba	1
hitomazu	1	tadashi	3
ichiou	1	toiuwakede	3
koko de	2	tokoro de	3
kono	1	tomaa	1
tokoro			
korekara	1	toriaezu	1
		yappari ³	1

Table 2 gives a partial list of discourse markers for the corpus⁴. Markers in boldface were always found at the beginning of paragraphs. The numbers indicate the frequency of each marker; the first number

² Both of these initiate closings.

³ *Yappari* actually occurs sentence finally, but is given its own line break before and after.

⁴ These data are for sentence-initial occurrences only. It is impossible to say much about the singleton examples, but they are listed because they have been found to be useful discourse markers in other studies (Maynard, 1989; Maynard, 1998; Kawamori et al., 1998; Nakano and Kato, 1998). We do not claim to have made a comprehensive list of the discourse markers in the corpus.

represents the number of times the item occurred paragraph-initially, the second number, represents sentence-initial, but not paragraph-initial occurrences.

Fillers. Utterances such as “um” and “ah” in English are sometimes called fillers or filled pauses, and occur fairly frequently in spoken conversation (Fais and Loken-Kim, 1994; Heeman and Allen, 1997). Although these never appear in more formal written texts, they do appear in email. In this corpus, *a*— or *u*— are used as fillers, sometimes also initiating discourse segments, as in (6).

(6) あー、Iさんメールに会名しか
 Ah I-san meiru ni kaisha mei
 shika
 well I--san mail on company name
 only
 入って いなかったんで、
 haitte inakattande
 written since
 一瞬 仕事の話かと思っちゃいましたよん。
 isshun shigotonohanashi kato
 for a moment work
 omocchaimashitayon think

Well, I-san, since you only put your company name on your mail, for a moment I thought you wrote me about work.

Formality. Shifts from the *-masu* (formal) form of the verb to the *-ru* (informal) form of the verb signal shifts in author point of view and thus indicate how adjacent portions of text should be organized and understood.

It has been suggested (Hinds, 1976) that formal verb forms are used where there is physical separation between two interactants (e.g., over the telephone or in letter writing). Since email communication involves such separation, we might expect to see predominantly formal verb forms used. Contrary to this expectation, however, there is a wide variety in usage of formal/informal verb forms in this corpus (Table 3). Just as with openings, some authors, such as R, model their email behavior on speech, using both formal and informal forms, while others,

such as H, model their behavior on letter writing, using primarily formal forms.

Table 3. Numbers of formal/informal verb forms per author.

	H	I	M	R	U
Formal(- <i>masu</i>)	20	39	31	36	45
Informal (- <i>ru</i>)	0	4	5	5	4
Informal (∅)	0	7	3	12	11

The use of formal/informal verbs may also signal information not about the relative social or physical positions of the email participants, but about the status of the information being conveyed. The use of the informal form can indicate the subordinate nature of that portion of the message, or a change in addressee (Makino, 1983; Maynard, 1998). Some of the verbs in (3) have been marked for formality. Note that the informal verbs deal with background material having to do only with the situation of the author; once the author returns to the subject of dealing with others' email, she returns to a formal verb form.

Tense. While formal/informal and tense forms of verbs are generally considered to be syntactic phenomena, they are discussed here because they serve a discourse function as well. Like the formal/informal distinction, the past/nonpast tense distinction can be used as a signal of shifting perspective and change in discourse segment (Makino, 1981; Iida, 1998; Walker, 1998; Maynard, 1998). Even in an example like (7), where *mo* (“too”) might indicate that the second sentence is adding information to the first sentence, the author signals his shift in perspective by inserting a paragraph break between the two sentences and by couching the first in the past tense, and the second in the present.

(7) ちょっとメールを覗くのを
 chotto meruwo nozokunowo
 for some time mailbox checked サボっ
 ていたら, sabotteitara
 not having

だいぶ話しが 進んでしまっていました。
 daibu hanashiga susundeshimattemashita
 a lot matter has
 progressed

Not having checked my mailbox for some time,
 the matter has progressed.

New ¶
 私も急な話して 申し訳けないんですが、
 watashi mo kyunahanashide
 moshiwakenaindesuga
 I too an urgent matter have to tell I'm
 sorry

I'm sorry; I have to tell you something urgent, too.

Sentence-final particles.

Sentence-final particles such as *yo*, *ne*, etc., are a prominent feature of spoken Japanese. They have been called particles of “interaction” (Maynard, 1990) or “rapport” (Maynard, 1989). As such, they are not found in formal written texts. However, certain particles are present in this corpus; Table 4 lists the most frequent sentence-final particles.

Table 4. Frequent sentence-final particles.

Particle	#	Particle	#
<i>ne/na</i>	26	<i>kedo</i>	5
<i>ka</i>	22	<i>de</i>	3
<i>yo</i>	7	<i>kana</i>	3
<i>ga</i>	5		

The use of sentence-final particles is another way in which authors model their representation of the meaning in their messages on strategies used in speech.

Results: Author-specific features

Speech recognition components of spoken language translations systems benefit from training to a particular's speaker's voice (Ström, 1994). In a similar way, certain aspects of email messages are typical of particular authors, and could be used to “train” an email translating system. It is easy to identify the author of an email message from the mail header or closing (if one is present).

Message closings

How a message is closed is significantly dependent upon the author of the message. Significant p-values were obtained for analyses of variance (ANOVA) for both the number of characters used in closing ($p = 0.0007$; Figure 1) and the number of lines used ($p < 0.0001$; Figure 2). The latter result reflects the fact that authors differed distinctively in the formatting of their paragraphs and especially of their closings. Fifteen messages had no closing at all. Of the remaining 17 messages, seven used the name of the author only (as in (3)); nine used some sort of closing expression (for example, *dewa mata*, “see you later”) and a name (as in (2)); and one used only a conventional closing expression (*dewa*).

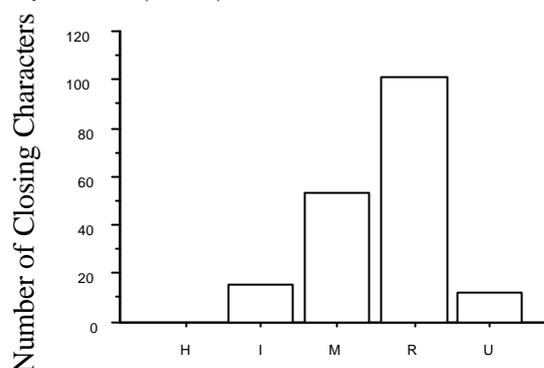


Figure 1. Number of characters used in closings for each author.

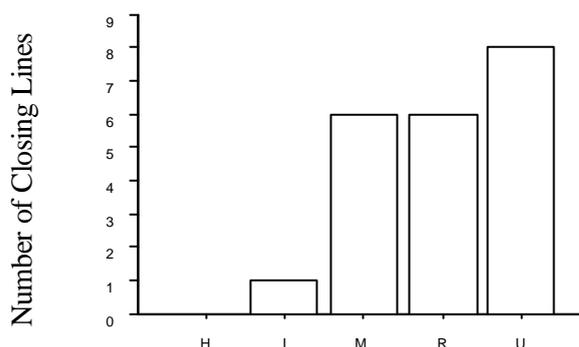


Figure 2. Number of lines used in closings for each author.

Number of sentences and paragraphs

The size of the message as measured in paragraphs (Figure 3) and sentences (Figure 4) was also significant. Differences for

number of paragraphs were significant at $p < 0.03$; for number of sentences, $p < 0.052$. Authors averaged from about two to about six paragraphs per message and from about four to about 11 sentences per message.

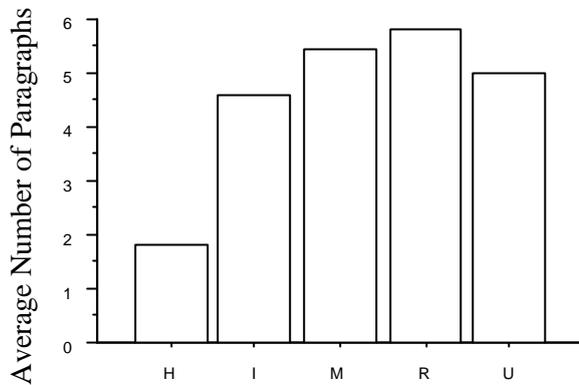


Figure 3. Average number of paragraphs per author.

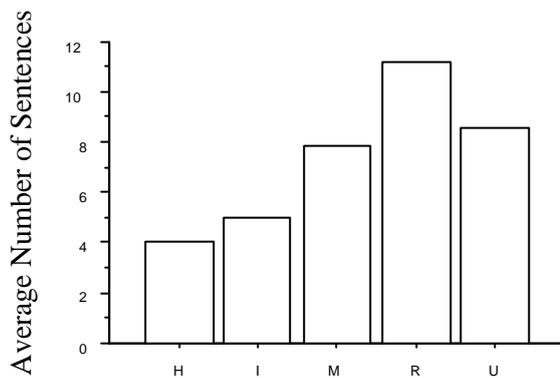


Figure 4. Average number of sentences per author.

Size of sentences and paragraphs

While the average size of sentences did not differ by author, it is still an important piece of information for a machine translation system. As can be seen in Figure 5, author H tended to write the longest sentences as measured in characters, while R wrote the shortest, with the average ranging from about 28 to 43. The range of characters per sentence taken over all the messages was 14 to 66.

The amount of text in any given paragraph did differ significantly by author ($p < 0.01$; Figure 6). Paragraph size ranged on average from nearly 40 characters to nearly 100 (compare the second paragraph in (3) and the

main paragraph in (1)), with H writing the longest paragraphs, and I and M the shortest.

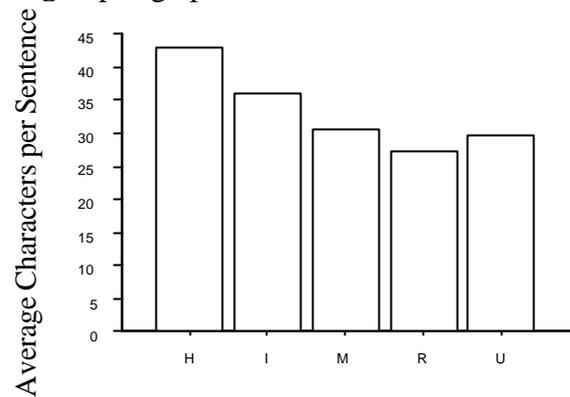


Figure 5. Average number of characters per sentence, by author.

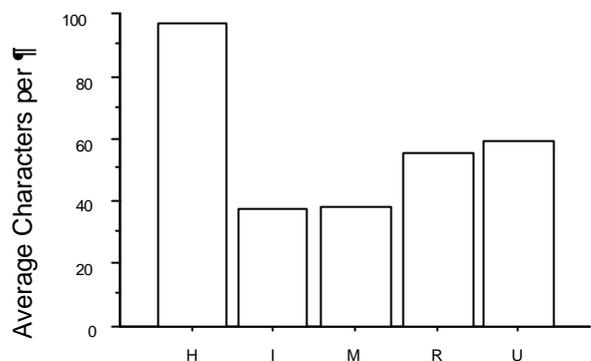


Figure 6. Average number of characters per paragraph, by author.

Use of graphics, katakana and English.

Authors showed consistent patterns in variables that were not particularly measurable as well. There were four distinct patterns of the use of graphics, katakana and English among the five authors (Figure 7). Author H used very little of any of these devices for enhancing textual meaning. Authors I and M used very little graphics or katakana but a relatively large amount of English. Author R showed the opposite pattern: quite a bit of graphics and katakana and less English. Author U used relatively more graphics than katakana or English.

Discussion

Since email is a text medium form of communication, it is likely that first attempts at translating email are made using translation

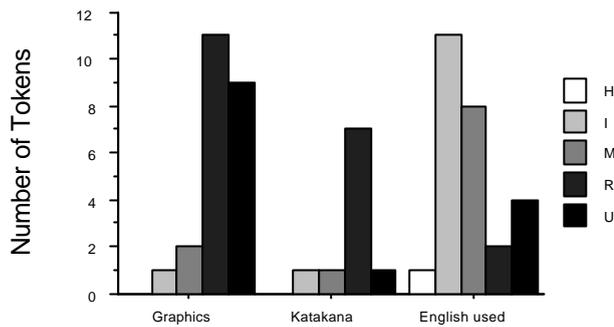


Figure 7. Number of tokens of graphic devices, katakana and English used per author.

systems developed to handle other types of text: newspaper articles, instruction manuals and the like. However, these systems cannot adequately deal with the specific challenges that the particular characteristics of email text pose.

These challenges are varied. Email does provide visual information for the determination of paragraphs (or discourse segments), but the nature of this information is much more variable in email than it is in more formal text genres. Similarly, the variability in format for openings and closings also makes their recognition a more difficult problem than the recognition of, say, headlines or section titles. Authors of email make conscious attempts to supplement text communication with features of speech as well, using for this purpose nonstandard forms of punctuation and spelling, nonstandard uses of katakana, substitution of English words for Japanese ones, and an extra-linguistic system of graphic devices (emoticons). In addition, authors may use fillers and a higher frequency of sentence-final particles typical to speech.

The strategies that authors use to capture speech characteristics require particular attention. The use of center dot strings constitutes an explicit cue for semantic inference systems to be used, as well as a cue that the text span of which they are a part may be only a clause and not a well-formed sentence. The use of (non-standard) katakana and non-standard spellings may be rendered as the English equivalent with underlining to show the implied emphasis. English words may simply be left, of course,

although the use of English in some cases is equivalent to placing the word in quotation marks. Fillers differ from language to language and so the fillers written in these messages as *a*—or *u*— require association with appropriate English lexical entries to generate “ah” or “well.” Likewise, Japanese emoticons differ from English ones and require an appropriate table of correspondence

The use of discourse markers illustrates the hybrid (written/spoken) nature of email. Maynard (1993) notes that the same discourse marker may be used differently in speech and in text. The example that she cites, *dakara*, is used only once in our email corpus, but it is used in the primary sense for written texts (“because”). Similarly, Maynard (1998) lists *sate* and *tokoro de*, which are frequent and consistent discourse markers in our corpus, as important “topic shifting” markers in text. On the other hand, other markers found in the email corpus, such as *sou ieba* and *sore de wa*, are cited as prominent discourse markers for speech (Nakano and Kato, 1998; Maynard, 1989). Thus the types of discourse markers used are typical of both speech and text.

In terms of frequency, too, the use of discourse markers in email text falls somewhere between that of more formal texts such as newspaper articles and that of speech. Maynard (1998) and Soria and Ferrari (1998) note that the frequency of the use of what we have been calling discourse markers is much higher in speech than in text. For English, Dahlgren notes that discourse markers occur at discourse segment boundaries in 16% of the set of *Wall Street Journal* articles she examined, but in 41% of the discourse boundaries in *Wheels*, a novel written in the narrative style. Kawamori et al. (1998) found that nearly half of the turns in their spoken Japanese corpus began with a discourse marker. In our corpus of email texts, discourse markers occur at the beginnings of 39 paragraphs, that is, about 27% of discourse boundaries, placing the frequency of the use of discourse markers somewhere between that of newspaper texts and written narratives. This hybrid nature of email text is an indication that modifications to existing

Table 5. Comparison of discourse marker frequencies across text genres: numbers are percent of discourse segments initiated with discourse markers.⁵

Genre	<i>Wall Street Journal</i>	Japanese email	<i>Wheels</i> , narrative prose	Japanese conversation
%	16%	27%	41%	50%

systems are required in order both to represent accurately features of email messages that differ from speech and/or text, and to take advantage of the unique sources of information in email.

Some of these additional sources of information particular to email are the aspects of the messages that are author-specific. An email translation system can be trained for a particular author, just as speech recognition systems are trained for particular speakers. Once the author of a message is known (information which can be obtained from the message header or the message closing), additional information becomes available to the translation system. Author-specific paragraphing and closing conventions provide discourse and visual structure information. Because the number of sentences and paragraphs and the size of the paragraphs are author-specific, the system can “know” where to expect likely paragraph boundaries. Similarly, knowledge of characteristic patterns of the use of nonstandard elements in text can make the translation of those features more successful. As we have seen, authors differ in the extent to which they signal relationships among pieces of text using changes in formal/informal and nonpast/past verb forms, discourse markers, and sentence-final particles. An understanding of these different styles can make the translation of these relationships more faithful to the author’s original intent.

⁵ Of course, this is only a rough indication, given that we are comparing across languages.

Let us consider one specific area in which the particular characteristics of email have bearing. Anaphora resolution for both explicit and omitted pronouns is a traditional problem for the translation of Japanese into any language which requires overt expression of verbal arguments. A number of different approaches to this problem have been suggested (Iida, 1998; Walker et al., 1994; Nakaiwa and Shirai, 1996; Nakaiwa et al., 1996; Yoshimoto, 1988). What these approaches have in common is the need for information about how the discourse is structured. Email messages contain information about discourse structure in a variety of forms: some authors use explicit visual formatting to show paragraph breaks, and discourse markers and patterns of the use of formal/informal and nonpast/past tense forms also cue changes in discourse segment. Further, author-specific information pertaining to typical size of paragraphs can be used in conjunction with these features to guide discourse segmenting decisions and help to disambiguate the functions of discourse markers which are not uniquely discourse-initial. Thus, features particular to email messages can contribute to the accurate translation of (omitted) pronouns.

Future Directions

The corpus under examination here is small and restricted to one topic and thus representative of only one level of interaction (i.e., casual) among the participants in the correspondence. Work is already underway to perform a similar analysis on a larger corpus of email correspondence among virtually the same group of researchers on a more formal, work-related topic. It will be instructive to determine if the characteristics observed in the current corpus are valid for a corpus concerned with a more formal topic as well.

In addition, other issues require further investigation. The extent to which author-constructed paragraphing on the one hand, and reliable discourse markers on the other, format the text into the same structure is an important question. The extent to which both match

native speaker intuitions about discourse segmenting is another interesting issue. Different results from different segmentation methods may then make different predictions about the interpretation of elided pronouns across discourse segment boundaries. The authentic data provided by this email corpus is a rich ground for exploring the issue of the resolution of elided pronouns as well.

References

- Chafe, W. L. 1982. Integration and involvement in speaking, writing and oral literature. In D. Tannen (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, 3. Norwood, NJ: Ablex Publishing Corporation, 35-53.
- Degand, L. 1998. On classifying connectives and coherence relations. In M. Stede *et al.*, 1998, 29-35.
- Fais, L., and Loken-Kim, K. 1994. Effects of communicative mode on spontaneous English speech. *Technical Report of the Institute of Electronics, Information, and Communication Engineers*, NLC94-22 (1994-10).
- Heeman, P. and Allen, J. F. 1997. Intonational boundaries, speech repairs and discourse markers: modelling spoken dialogue. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid*, 254-261.
- Hinds, J. 1976. *Aspects of Japanese Discourse Structure*. Tokyo: Kaitakusha Co., Ltd.
- Iida, M. 1998. Discourse coherence and shifting centers in Japanese texts. In M. A. Walker, A. K. Joshi, and E.F. Prince (eds.), *Centering in Discourse*, Oxford UK: Oxford University Press.
- Kawamori, M., Kawabata, T., and Shimazu, A. 1998. Discourse markers in spontaneous dialogue: a corpus based study of Japanese and English. In M. Stede *et al.*, 1998, 93-99.
- Makino, S. 1979. Paragraph, is it a legitimate linguistic unit?—a case study from English and Japanese. In R. L. Brown, Jr., and M. Steinmann, Jr. (eds.), *Rhetoric 78: Proceedings of Theory of Rhetoric: An Interdisciplinary Conference*, Minneapolis, MN, 283-296.
- Makino, S. 1981. Tense switching in Japanese written narrative discourses. In S. Makino (ed.), *Papers from the Middlebury Symposium on Japanese Discourse Analysis*, 125-150.
- Makino, S. 1983. Speaker/listener orientation and formality marking in Japanese. *Gengo Kenkyuu* 84, 126-145.
- Maynard, S. K. 1989. *Japanese Conversation: Self-contextualization through Structure and Interactional Management*, Vol. XXXV in the series *Advances in Discourse Processes*, R. Freedle (ed.), Norwood, NJ: Ablex Publishing Corporation.
- Maynard, S. K. 1990. *An Introduction to Japanese Grammar and Communication Strategies*. Tokyo: The Japan Times, Ltd.
- Maynard, S. K. 1993. *Discourse Modality, Subjectivity, Emotion and Voice in the Japanese Language*. Philadelphia, PA: John Benjamins Publishing Co.
- Maynard, S. K. 1998. *Principles of Japanese Discourse*. Cambridge, UK: Cambridge University Press.
- Nakaiwa, H., Bond, F., Uekado, T., and Nozawa, Y. 1996. Resolving zero pronouns in texts using textual structure. *Proceedings of the Second International Conference on New Methods in Language Processing (NeMLaP-2)*, Ankara, Turkey, 25-36.
- Nakaiwa, H., and Shirai, S. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, Copenhagen, Denmark, 812-817.
- Nakano, Y. and Kato, T. 1998. Cue phrase selection in instruction dialogue using machine learning. In M. Stede *et al.*, 1998, 100-106.
- Schiffrin, D. 1987. *Discourse Markers*. Cambridge, UK: Cambridge University Press.
- Soria, C. and Ferrari, G. 1998. Lexical marking of discourse relations—some experimental findings. In M. Stede *et al.*, 1998, 36-42.
- Stede, M., Wanner, L., and Hovy, E. 1998. *Discourse Relations and Discourse Markers, Proceedings of the Workshop, COLING-ACL'98*, Montreal, Canada
- Ström, N. 1994. Experiments with a new algorithm for fast speaker adaptation. *Proceedings of the 1994 International Conference on Spoken Language Processing*, Yokohama, Japan, 459-462.
- Walker, M. A. 1998. Centering, anaphora resolution, and discourse structure. In M. A. Walker, A. K. Joshi, and E.F. Prince (eds.), *Centering in Discourse*, Oxford, UK: Oxford University Press.
- Walker, M., Iida, M., and Cote, S. 1994. Japanese discourse and the process of centering. *Computational Linguistics* 20(2), 193-231.
- Yoshimoto, K. 1988. Identifying zero pronouns in Japanese dialogue. *Proceedings of*

the International Conference on Computational
Linguistics, Budapest, Hungary, 779-784.