Written or spoken language?

— English email features found in mailing list postings —

Keiko HAYASAKA

O Introduction

Spoken and written language cannot be simply distinguished by the medium, (I use this term here following McCarthy and Carter (1994) instead of mode or channel by others) but are assumed to be socio-culturally relative. Intuitive judgement is not always correct. McCarthy and Carter (1994: 5) showed informants ten different kinds of texts and asked them to judge if each was either spoken or written. Some common criteria, for example, contraction and second person pronoun for spoken and passive sentence for written texts, as well as intuitive judgement put the informants on the wrong track. Real language use is a mixture of written and spoken language in many cases. Church sermons are more like written language, although they are spoken, while newspaper advertisements are like spoken language, although they are written. The medium, whether the text is spoken or written, cannot determine the language by itself.

In contrast to the discourse of conversation, advertisements, academic lectures or church sermons, the language used in email messages has not been discussed as much as other styles. The rapid development of technology and the popularity of email correspondence among the EFL learners can lead us to further study of the language used in this new type of communication. This paper attempts to find and examine the uniqueness of the language used in email, if there is any.

The organization of the current paper is: first, some previous studies are reviewed, then research of the subjects' 191 email messages, approximately 18,000 words, are qualitatively and quantitatively analyzed and the results are discussed. Finally, the implications of the findings relevant to EFL programs will be discussed.

Research questions the author raises for the current study are:

1. Is email a spoken or written discourse? How is it measured?
2. Is there any unique vocabulary to characterize the feature?
3. Are there any implications in email writing for the teaching and learning of English?

I Review of some previous studies and their criteria of written/spoken language

Lexical density (LD), the proportion of lexical words in the text, is a characteristic of written/spoken language. Ure (1971), after examining 34 spoken and 30 written texts, approximately 21,000 words each, showed that lexical density of spoken texts tends to be lower than that of written texts. One of her findings was, a lexical density is bordered by 40%, higher than that for written texts and lower for spoken ones. According to her study, the lexical density of written

texts ranged between 36-57%, while the lexical density of spoken texts ranged between 24-43%.

Stubbs (1986) challenged some of Ure's assumptions and refined the definition of lexical/grammatical words. Accordingly, Stubbs measured the lexical density of 6 different texts from the London-Locus Corpus. His results showed different percentages of lexical density with all 6 texts being higher than 56%. However, the results supported Ure's assumptions, and monologues showed higher LD than dialogues. Computational linguistics has enabled more explicit calculations and analyses.

Chafe and Danielewicz (1987) studied the properties of spoken and written language using conversations, lectures, letters and academic papers. Twenty adults, professors or graduate students at the University of California at Berkeley or the State University of New York at Albany, were the subjects. The data of four styles of communication from the same group of subjects were analyzed to see their 'spokenness' and 'writtenness'. The criteria they applied were: level and variation of vocabulary and grammatical structures such as clauses and sentences. Spoken and written language features were more readily distinguished between conversations and academic papers, with a type-token ratio of 18 vs 24 words per intonation unit and 6.2 vs 9.3 respectively, which is not surprising. The former style had shorter units and they were inexplicit. Messages were implicit and the speakers showed more involvement with the reality. The latter, on the other hand, featured 'writtenness' due to the writer's detachment from the audience. Written messages were more transactional than interactional. The other two styles, academic lectures and letters, fell between conversation and academic writing in terms of spoken/written language.

Carter, R. and McCarthy, M. have conducted numerous studies on spoken and written discourse, grammar and vocabulary, both jointly and individually. Among which, Carter & McCarthy (1995) deals with the unique features of spoken grammar and its pedagogic implications. They advocate the importance of exposing learners to the data of spoken language and observing the grammar of speech in natural contexts and in different genres. Carter & McCarthy (1997) examined vocabulary in written and spoken texts. Corpus-based research of vocabulary in spoken texts and its comparison with written texts exemplify the unique features of vocabulary for spoken language. Corpus data reveals the different usage of *start* and *begin,* for example, the former being far more frequently used in spoken texts. The quantitative data show the difference between the two types of discourse: the top 50 written word-forms cover 38.3% of all the text, while the equivalent in spoken word-forms is 48.3%. This argues that functional words occupy nearly half of the text in spoken language, and thus requires contexts to be fully understood.

II　Current Study

Email postings and academic research papers by the same subjects were examined to see the differences in the language of the students' writings, if any. Research question *1. Is email a spoken or written discourse?* was raised under the hypothesis that email writing shows more spoken features because of its interactiveness with the recipients, spontaneity in writing and its informality on topics.

The subjects joined the Student SL-Lists (for the details of these lists, see Ueno, et al., 1999).

Although the subjects were obliged to post a certain number of letters, depending on their interests, they were free to choose, subscribe to and unsubscribe from the 10 lists offered at the time this research was conducted. The topic ranged from their daily activities, such as school life and part-time jobs (CHAT-SL, MUSIC-SL, SPORT-SL, etc), to discussions on their major field of psychology and social works (DISCUSS-SL). The subjects' postings were always read by the subscribers of the lists and responded to by them, if the topics attracted the readers. The subjects also responded to the postings and sent their opinions and comments. The writers knew that their postings were always read by someone, which illustrates the interactive nature of mailing lists. For the present research, only the postings by the subjects of the research were analyzed due to copyright issues.

The survey reports, on the other hand, were the end-of-the-term assignments of the class. The subjects worked on the project and wrote the report at the end of the term (see Hayasaka, 2000 for this project in detail). They referred to academic articles, scaffolded useful phrases when necessary and had more time to edit. However, no errors were corrected by the instructor of the class. Topics of the reports were related to their major field such as *stress, the effect of non-verbal communication and volunteering activities.*

Total number of words in email writing summed up to 17,896 and that in academic reports summed up to 7,011. Although they are not equivalent in quantity, the author believes it is worth comparing because those two styles of writing were collected from the same subjects.

Collected data were analyzed quantitatively and qualitatively with WordSmith 3.0 (Oxford, 1998). Type-token ratio and lexical density were applied to measure spoken/written language quantitatively, which will address the second half of research question 1. Some features of vocabulary were examined to measure spoken/written language qualitatively, which will address research question 2.

III   Results and discussion

A.   Quantitative Analysis.

1.   Type-token Ratio

Email messages have an audience, are interactive and not graded, while academic reports have only one reader, the teacher and are usually graded. Thus their styles are readily expected to appear different. In order to test the hypothesis, the author examined type-token ratio and lexical density of these two styles.

Table 1   Type-token ratio

email writing	academic writing
.13	.16

Table 1 shows that type-token ratio is different in these two styles. Email writing shows a lower type-token ratio than academic reports. Type-token ratio is calculated by dividing the number of different words by the total number of words and expresses the range of lexical choices. Spoken language is assumed to have a lower ratio because words are used repeatedly in a narrower range in

a limited time.　However, as Chafe and Danielewicz point out, the longer the text, the lower the
ratio becomes because in longer texts words are more likely to repeat.　However, the spokenness or
writtenness should not be judged by the ratio per se.　Rather the ratio could be used to compare the
relative spokenness or writenness of the texts.　The current subjects did not write email 'on the fly',
as many native speakers usually do, but occasionally used dictionaries and took time to compose
sentences, as revealed in a subsequent survey.　The result illustrates the different ratios between the
two styles of writing: emails being lower and academic reports being higher.　Type-token ratio
revealed that email writing is relatively more 'spoken' than academic writing.

2.　Lexical Density (LD).

Table 2　Lexical density

email writing	academic writing
.63	.72

Table 2 shows the lexical density of the two styles.　Lexical density analysis is a means to
measure the spoken/written quality of a text and is calculated thus: $LD=100L/T$, where L is the
number of lexical words and T is the total number of words in the text.　Since grammatical words
are limited in number, L is calculated by deducting the grammatical words from the total.　A list of
grammatical words, in Stubbs (1982) (see Appendix), was applied to measure the lexical density of
the two styles because it gives a more accurate idea of what grammatical words are.　As Stubbs
points out, there are some ambiguities in the words listed.　For example, the use of *is* in the
following sentences is different: the first being a lexical word and the second being a grammatical
word.

He *is* a singer.

He *is* singing.　　　　　　　　　　　　　　　　　　　　　　　　　　　　(p.36)

This kind of ambiguity is difficult to clarify with a computer.　Another example of ambiguity
lies in quantity as well as quality.　Below are three examples of *man's* and how it/they are counted
by the program Stubbs applied.

The man's hat　　　　　　one Lexical word

The man's gone　　　　　　one Lexical word and one grammatical word

The man's a student　　　two Lexical words　　　　　　　　　　　　　　　　(p.37)

Computation is not perfect in distinguishing differences like these.　It is assumed, however,
that the occurrence of this type of ambiguity occupies a small portion of the whole text.

As a whole, Stubbs' examples of lexical density in the sub-texts of the London-Lund Corpus are
summarized as follows (40-41).

Text number	Description		Density
1	10-5	radio commentary: state funeral	56%

2	10-1	radio commentary: cricket	54%
3	11-4	House of Commons	48%
4	12-1	church sermons	47%
5	07-1	phone conversations: friends	45%
6	08-1	phone conversation: business	44%

A more refined tabulation of lexical and grammatical words revealed a higher lexical density in his study than Ure's, written and spoken bordered by .40. However, Stubbs' findings support Ure's, where monologues (1,2 and 4) show a higher lexical density than dialogues (3,5 and 6) as expected.

B.   Qualitative Analysis

Table 3   Frequency of first and second person pronouns

	email writing		academic writing	
	frequency	rank	frequency	rank
I	914(5.1%)	1	57(0.8%)	9
we	152(0.8%)	13	95(1.4%)	7
you	331(1.8%)	5	19(0.3%)	23

Table 3 shows the frequency of the first and second person pronouns, *I, we* and *you* in the two styles. In email writing, both *I* and *you* are frequently used and show interactive communication between the writer and the reader. Frequent use of *you* also indicates that the writers actively involved the readers in their messages. On the other hand, the use of these two personal pronouns in academic writing is very limited.

McCarthy and Carter (1997) list the 50 most frequent words from Cambridge International Corpus (CIC) written data and CANCODE spoken data. First and second person pronouns *I, we* and *you* rank 7th, 46th and 16th in written data and 2nd, 26th and 3rd in spoken data. These three pronouns appear more frequently in spoken texts. The frequent use of the first person pronoun is reported by Chafe and Danielewicz (1987), in which they also reveal the frequent use of the first person pronouns in certain other styles. Among the four styles they examined, conversations, lectures, letters and academic papers, the first person pronouns were far more frequently used in conversations and letters. As far as the frequency of first and second person pronouns, both McCarthy & Carter (1997) and Chafe & Danielewicz (1987) agree that these pronouns can be an indicator of the 'spokenness' of a text.

3.   Samples

Below are extracts from the texts: one from email messages and the other from academic report abstracts.

<Extract 1>

Hi! My Name is Mika Sato. I am a student of university in Sapporo.

Do you have much free time? For me, from spring to fall I was busy by club activity. I belong to American football club.

But in winter, because we can't play football I can have free time.

It makes me very happy and I am looking forward to do something fun

If you are me, do you do? Please tell me your advice to spend wonderful time.

I want your message! See you. (sic)

〈Extract 2〉

　　This study was conducted to find relation between way of interpersonal communication (face-to-face, telephone, and letter) and human relationship orientation.　So we use IOS-V in order to measure human relationship orientation.　This survey was sent worldwide through the Internet and letter.　102 subjects responded this survey.　They were divided into three groups by IOS-V score level.　Statistical analysis of the collected data showed that middle scored group used more face-to-face communication than low group when they talk private topics.　And the subjects got high score, tend to more use face-to-face when they speak ether private or public. (sic)

Table 4　Sample texts: TTR, LD and 1st and 2nd person pronoun

	Extract 1 email writing	Extract 2 academic writing
Total words	87	105
Type-Token Ratio	.64	.69
Lexical Density	.65	.70
1st & 2nd person pronouns	11	1

　　Extract 1 is an example of a posting to CHAT-SL and Extract 2 is an abstract attached to an academic report.　Although the length of the two texts is not very different, all three criteria show some different results.　Email writing produces a lower TTR and LD, but a higher frequency of first and second person pronouns than an academic report.　Each category illustrates that email writing has more features of spoken language compared with academic writing.

　　Extract 1's interactive nature is seen in the frequent use of personal pronouns.　Questions and requests for advice from the reader using *you* are examples of interactiveness.　To make the exchange reciprocal, ego involvement is expressed by frequent use of first pronoun *I*.　The closing phrase, *"See you"*, suggests that the subject feels he/she is having a face-to-face conversation with the person on-line, although the message is read asynchronously.

　　Extract 2, in contrast to Extract 1, represents typical academic writing.　The passive voice structures and impersonal subject matter are indicative of its writtenness.

IV　Implication

　　The answer to the research question 3, "Are there any implications of email writing for the teaching and learning of communicative English?", is affirmative.　An email exchange supplies

students with more opportunities to practice communicative interaction in English. As discussed previously, email writing has features of spoken language and students, whose exposure to the target language is limited, will benefit from an email exchange in English. The transition of classroom activities to real use of language is more readily realized by utilizing email writing via pedagogically well-schemed sites and programs.

V Limitations and further study

The author has dealt only with asynchronous communication for the current study. The subjects edited their writing before posting and thus made it less spontaneous and impromptu than ordinary email postings. If data were collected from synchronous email-exchanges, more spoken features should have been found.

For further study, more criteria for measuring spokenness need to be added to grammar and vocabulary. Some suggestions for this are given by McCarthy & Carter (1997) and Chafe & Danielewicz (1987). The criteria they use for measuring spoken/written language range widely from phrase/sentence construction to vocabulary based on spoken corpus data and analysis.

VI Conclusion

The subjects' email writing was compared to their academic reports and spoken features were investigated by using three criteria: Type-Token Ratio, Lexical Density and first and second person pronouns. Email writing showed lower TTR and LD, but higher frequency of the pronouns, all of which are characteristic of spoken language.

The author discussed implications of email writing for enhancement of students' communicative skills. In an environment where face-to-face communication is limited, this media-aided communication makes it easier for learners to find partners to exchange opinions and discuss current or social issues. Since the language they use is found to have spoken features, this type of communication is believed to promote students' oral skills.

Moreover, email exchange has recently become a favored method of communication by students. "Email has extended the language's stylistic range in interesting and motivating ways. In my view, it is an opportunity, not a threat, for language education." (Crystal, 2001: 128)

[References]

Carter, R & M. McCarthy. (1995) Grammar and the Spoken Language. *Applied Linguistics*, 16, 2. 141-158.

Chafe, W. and Danielewicz, J. (1987) Properties of Spoken and Written Language. In R. Horowitz & S. J. Samuels (Eds.) *Comprehending Oral and Written Language*. San Diego, CA: Academic Press. 83-113.

Crystal, D. (2001) *Language and the Internet*. Cambridge: Cambridge University Press.

Hayasaka, K. (2000) "Content-based English Learning and Communication (Content-based Gakushu to Communication)," *Hokusei Ronshu*. Hokusei Gakuen University. 31-41.

McCarthy, M. and Carter, R. (1997) Written and spoken vocabulary. In N. Schmitt & M. McCarthy (Eds.) *Second language vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University

Press.   20-39.

McCarthy, M & R. Carter. (1994) *Language as Discourse*.   New York: Longman

Stubbs, M. (1986). Lexical Density: A Technique and Some Findings.   In Coulthard, M. *Talking about Text*.   Birmingham English Language Research.   27-42.

Ueno, Y., Hayasaka, K., Sasaki, M., Yoshida, M. Abe, A., Oda, T. &   Yokoyama, K. (2000) Corpus-Based Analyses of Email by Japanese College Students, *JACET BULLETIN*.   137-149.

Ueno, Y., Abe, A., Oda, T., Hayasaka, K., and Yoshida, M. (1999) "Gogaku Kyoiku ni Ikasu Email Koryu (2)" "Email Exchange in Language Learning (2)." *Hokkai Gakuen Daigaku Gakuen Ronshu* (The Journal of Hokkai-Gakuen University).   101.   45-83.   Sapporo: Hokkai Gakuen University.

Ure, J. (1971) Lexical density and register differentiation.   In Perren, G. E. & Trim, J. L. M (Eds.), *Applications of Linguistics: Selected Papers of the Second World Congress of Applied Linguistics, 1969*.   Cambridge, London: Cambridge University Press.   443-452.

Appendix   List of G (grammatical) words, Stubbs (1982:36-37)

a above across after against all along alongside although amid amidst among amongst an and any anybody anything anywhere apropos as at atop because before behind below beneath beside besides between beyond both but can can't cos could couldn't dare daren't despite dhi doesn't don't during each ei either every everybody everyone everywhere except few for from he he'd he'll he's her hers herself him himself his how however if in inside into much must mustn't my myself needn't neither never nevertheless no no-one nobody none nonetheless noone nor not notwithstanding of off on or ought oughtn't our ours ourselves out outside over per plus shall shan't she she'd she'll she's should shouldn't since so some somebody someone than that that'd that'll that's the thee their theirs them themself themselves then there there'd there's there've these they they'd they'll they're they've thine this those thou though through throughout thy till to toward towards uhuh under underneath until up upon us via we we'd we'll we're we've what what'd what's what've whatever when whenever where wherever which whichever while whilst who whom whose why will with within without won't would wouldn't ye yeah yes yet you you'd you'll you're you've your yours yourself yourselves I I'd I'll I'm I've (207 words)

(*dhi* and *ei* are transcriptions used in the data for stressed definite and indefinite articles.)

[Abstract]

Written or Spoken Language? :

English E-mail Features Found in Mailing List Postings

Keiko HAYASAKA

The present study examines the e-mail writing of Japanese college students and compares it with academic writing by the same subjects to find its degree of 'spokenness'. The type-token ratio and the lexical density of these two styles revealed that e-mail writing has more features of spoken language than academic writing. Moreover, qualitative analysis of the subjects' e-mail writing revealed its spokenness through its vocabulary. By using the criteria of previous studies, the author argues there is a unique mode of communication.

Since e-mail writing is found to utilize features of spoken language, active use of pedagogically well-schemed e-mail correspondence will enhance EFL students' interactive communication, in tandem with face-to-face communicative oral English practice, in circumstances where their exposure to the target language is limited.