An Investigation of a Factor That Affects the Usage of Unsounded Code Strings at the End of Japanese and English Tweets

Yasuhiko Watanabe, Kunihiro Nakajima, Haruka Morimoto, Ryo Nishimura, and Yoshihiro Okada

Ryukoku University

Seta, Otsu, Shiga, Japan

Email: watanabe@rins.ryukoku.ac.jp, t13m071@mail.ryukoku.ac.jp, t13m076@mail.ryukoku.ac.jp,
r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

Abstract—In this study, we compare Japanese and English tweets submitted to Twitter and discuss how we use unsounded code strings at the end of online messages. We first define unsounded codes and unsounded code strings used in Japanese and English text. Next, we compare and discuss the usage of unsounded code strings at the end of Japanese and English tweets, especially, to general public and particular persons. Finally, we show the receiver of a tweet, whether general public or a particular person, is a factor that affects the usage of unsounded code strings at the end of Japanese and English tweets. Specifically, Japanese speakers use unsounded code strings at the end of tweets more frequently to particular persons than to general public while English speakers do not.

Keywords–unsounded code string; Twitter; general public; particular persons; non verbal communication.

I. INTRODUCTION

Many of us think that it is easy to understand the meanings of non verbal expressions in online messages even if different language speakers generate them. However, the usage differs between different language speakers and the difference is at risk of bringing unnecessary frictions between them. As a result, it is important to consider the difference, especially, in multilingual computer-mediated communication (CMC) systems. In this study, we show the usage of unsounded code strings, one kind of non verbal expression, differs between Japanese and English speakers and discuss a factor that affects the usage of them.

We often find consecutive unsounded marks and characters are used at the end of online messages, such as mails, chattings, and tweets in Twitter.

(exp 1) I'm freezing!!!!
(exp 2) @*ryuuuuuuuu_2012 soushita hou ga iiyo.......*
     (@ryuuuuuuuu_2012 you had better do it.......)

(exp 1) and (exp 2) are tweets submitted to Twitter. (exp 1) was submitted by an user who chose English as his/her language for tweets. On the other hand, (exp 2) was submitted by an user who chose Japanese as his/her language for tweets. Both (exp 1) and (exp 2) have consecutive unsounded marks at the end of them. These unsounded marks are used for smooth communication. The submitter of (exp 1) is thought to use the three consecutive exclamation marks for expressing his/her impression strongly. On the other hand, the submitter of (exp 2) is thought to use the seven consecutive periods for expressing his/her opinion softly. In this study, we define unsounded marks and characters as *unsounded codes*. Furthermore, we define three or more consecutive unsounded codes as a *unsounded code string*. For example, in Twitter, 14 % of

Japanese tweets and 10 % of English tweets have unsounded code strings at the end of them. Although unsounded code strings are popular, there are few studies on them. As a result, in this study, we investigate how we use unsounded code strings at the end of tweets in Twitter. Especially, we compare Japanese and English tweets in Twitter and discuss a factor that affects the usage of unsounded code strings at the end of tweets. The results of this study will give us a chance to understand the usage of unsounded code strings and improve multilingual CMC systems.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we define unsounded code strings and describes how they are used at the end of tweets in Twitter. Finally, in Section IV, we present our conclusions.

II. RELATED WORKS

There are a considerable number of studies comparing speakers of various languages from various viewpoints, such as, pragmatics, cognitive science, and so on. These studies can be classified into two types:

- studies comparing native speakers of one language to non-native speakers of the same language
- studies comparing native speakers of one language to native speakers of other language directly

This study is classified into the latter. It is because we compare unsounded code strings in Japanese tweets to those in English tweets directly.

In pragmatics, a considerable number of studies have been made on interlanguage speech acts, such as, expressing compliments [1], apologies [2], gratitude [3], politeness [4], and refusals [5]. In these studies, native speakers of one language were compared to non-native speakers of the same language. Also, there are a considerable number of studies comparing pauses and backchannels of native speakers to those of non-native speakers. Backchannels are listener's responses, such as "uh-huh" and "yeah", given while someone else is talking, to show an interest, attention, or willingness to keep listening. Deschamps investigated how French learners of English made pauses in their English speeches [6]. Bilá and Džambová investigated the function of silent pauses in native and non-native speakers of English and German [7]. Ishizaki investigated how English, French, Chinese and Korean learners of Japanese and native Japanese speakers made pauses in their Japanese speeches [8]. Tavakoli reported that the location of pauses is important in comparisons of native speakers and

TABLE I. The numbers of Japanese and English normal tweets, replies, and retweets (from November/2012 to December/2012).

language	tweet type	normal tweet (%)	reply (%)	retweet (%)	total (%)
Japanese	all	3,813,164 (53.82%)	2,528,642 (35.69%)	743,461 (10.49%)	7,085,267 (100.00%)
	with UCS	356,727 (36.92%)	430,294 (44.54%)	179,166 (18.54%)	966,187 (100.00%)
English	all	16,023,795 (51.27%)	8,267,646 (26.45%)	6,961,800 (22.28%)	31,253,241 (100.00%)
	with UCS	1,121,952 (34.30%)	632,298 (19.33%)	1,516,571 (46.37%)	3,270,821 (100.00%)

"with UCS" means tweets that have unsounded code strings at the end of them

foreign learners [9]. Okazawa reported that native Japanese speakers paused roughly twice as often in their English utterances than in their Japanese utterances [10]. LoCastro reported that Japanese speakers often feel uncomfortable when speaking English because they are unable to use the appropriate backchannels [11]. From the viewpoint of cognitive science, Tera et al. compared and analyzed reading processes of native Japanese speakers and foreign learners of Japanese [12].

On the other hand, there are also a considerable number of studies comparing native speakers of one language to native speakers of other language directly. Especially, many studies have been made on backchannels. It is because backchannels are found in various languages. Maynard showed that backchannel phenomena for Japanese and English differ in terms of type, frequency, and context [13]. Miller reported that Japanese speakers use backchannels more frequently than English speakers [14]. However, in most of previous studies, the frequency of backchannels were observed in various situations while little is known about factors that affect the frequency of backchannels. Chen pointed out that it is important to investigate factors that affect the frequency of backchannels and took White's work [15] for example [16]. White reported that Americans used backchannels more frequently in conversations with Japanese than in conversations with other Americans [15]. In other words, the conversation partner, whether American or Japanese, is a factor that affects the frequency of Americans' backchannels. As a result, when we compare different language speakers, it is important to investigate factors providing different responses between them. In this study, we investigate a factor that affects the usage of unsounded code strings at the end of tweets.

III. UNSOUNDED CODE STRINGS AT THE END OF JAPANESE AND ENGLISH TWEETS IN TWITTER

In this section, we compare and discuss the usage of unsounded code strings at the end of Japanese and English tweets. In Section III-A, we define unsounded code strings and show how they are used at the end of tweets. In Section III-B, we show the investigation object of this study. Finally, in Section III-C, we compare Japanese and English tweets and discuss a factor that affects the usage of unsounded code strings at the end of tweets.

A. The definition of an unsounded code string

First, we define unsounded codes and unsounded code strings. In this study, we define that an unsounded code string is three or more consecutive unsounded codes. In this study, unsounded codes in English text are limited to

- punctuation marks (e.g. !#$%&.,:;<=>?@ (){}).

On the other hand, unsounded codes in Japanese text are limited to the following marks and characters:

- punctuation marks,
- Greek characters,
- Cyrillic characters, and
- ruled lines.

It is because these marks and characters are generally unsounded when they are used at the end of Japanese sentences.

Next, we show how unsounded code strings are used at the end of tweets. Twitter users often use unsounded code strings in order to enable anyone to understand their tweets clearly and avoid unnecessary frictions with others.

(exp 3)   *kadai owattaaaaaaaaaaaa!!!!!!!!!!* (I got my homework dooooooooooone!!!!!!!!!!)
(exp 4)   @jayne_hurley looks amazing !!!
(exp 5)   WEIRDEST dream last night omg...

For example, the submitters used exclamation marks consecutively at the end of (exp 3) and (exp 4) for expressing their feelings strongly. On the other hand, the submitter used periods consecutively at the end of (exp 5) for expressing his/her impression softly.

We may note that there are some submission constraints in Twitter. For example, Twitter users are prohibited to post the same tweets repeatedly. To avoid this constraint, some users use unsounded code strings. For example, (exp 6), (exp 7), and (exp 8) were posted consecutively in a few seconds to a particular user, *ssuzuki16*, beyond the limit of repeated submission.

(exp 6)   @*ssuzuki16 yo...*
(exp 7)   @*ssuzuki16 yo....*
(exp 8)   @*ssuzuki16 yo.....*

B. The investigation object

We obtained tweets by using the streaming API. However, the streaming API allows us to obtain only 1% of all public streamed tweets because of API restriction. We used the streaming API and obtained the following tweets in three weeks in November and December 2012.

- 7,085,267 tweets submitted by users who chose Japanese as their language for tweets. In this study, we call these tweets *Japanese tweets*.

- 31,253,241 tweets submitted by users who chose English as their language for tweets. In this study, we call these tweets *English tweets*.

Table I shows the number of the obtained Japanese and English tweets. These tweets can be classified into three types:
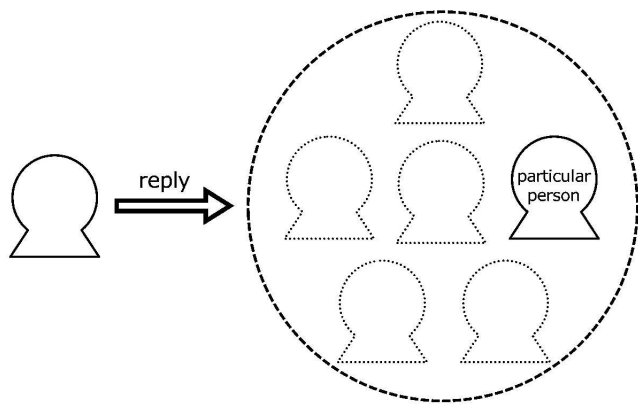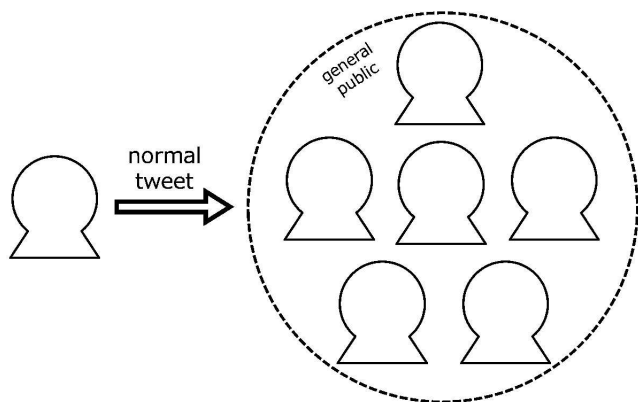
Figure 1. A reply is submitted to a particular user.

Figure 2. A normal tweet is submitted to general public.

- reply
  A reply is submitted to a particular user (Figure 1). It contains "@username" in the body of the tweet. For example, (exp 2) and (exp 4) are replies.

- retweet
  A retweet is a reply to a tweet that includes the original tweet.

- normal tweet
  A normal tweet is neither reply nor retweet. For example, (exp 1), (exp 3), and (exp 5) are normal tweets. Twitter users generally submit their tweets to general public. As a result, most of normal tweets are submitted to general public (Figure 2).

From the obtained Japanese tweets, we extracted 966,187 Japanese tweets that have unsounded code strings at the end of them. These 966,187 Japanese tweets are 13.64% of all the Japanese tweets. On the other hand, from the obtained English tweets, we extracted 3,270,821 English tweets that have unsounded code strings at the end of them. These 3,270,821 English tweets are 10.47% of all the English tweets. Table I shows the number of Japanese and English normal

tweets, replies, and retweets that have unsounded code strings at the end of them.

In this study, we do not discuss unsounded code strings at the end of retweets. It is because, messages in retweets are created not by submitters, but by other users. As a result, retweets are inadequate to investigate how we use unsounded code strings at the end of online messages.

In this study, we compare unsounded code strings at the end of (1) normal tweets and (2) replies. It is because we intend to compare and discuss the usage of unsounded code strings at the end of tweets to (1) general public and (2) particular persons. As mentioned, normal tweets are generally submitted to general public. On the other hand, each reply is submitted to a particular person.

Figure 3 and Figure 4 show the cumulative relative frequency distribution of

- the length of all the Japanese and English tweets (excluding retweets),
- the length of Japanese and English tweets (excluding retweets) that have unsounded code strings at the end of them, and
- the length of unsounded code strings at the end of Japanese and English tweets (excluding retweets).

As shown in Figure 3 and Figure 4, the distribution of the length of Japanese tweets shifts to shorter ranges than the length of English tweets. On the other hand, the distribution of the length of unsounded code strings at the end of Japanese tweets shifts to longer ranges than the length of those at the end of English tweets.

C. The comparison of the usage of unsounded code strings at the end of Japanese and English tweets

Next, we compare the length of unsounded code strings at the end of normal tweets and replies. Figure 5 and Figure 6 show the cumulative relative frequency distribution of the length of unsounded code strings at the end of

- Japanese normal tweets and replies, and
- English normal tweets and replies.

As shown in Figure 5 and Figure 6, the length of unsounded code strings at the end of Japanese and English normal tweets have a similar distribution pattern to those of Japanese and English replies, respectively. As a result, it may be said that the length of unsounded code strings at the end of tweets are less affected by whether the tweets are normal tweets or replies.

Next, we compare the length of normal tweets and replies that have unsounded code strings at the end of them. Figure 7 and Figure 8 show the cumulative relative frequency distribution of

- the length of all the Japanese and English normal tweets, and
- the length of Japanese and English normal tweets that have unsounded code strings at the end of them.

On the other hand, Figure 9 and Figure 10 show the cumulative relative frequency distribution of

- the length of all the Japanese and English replies, and
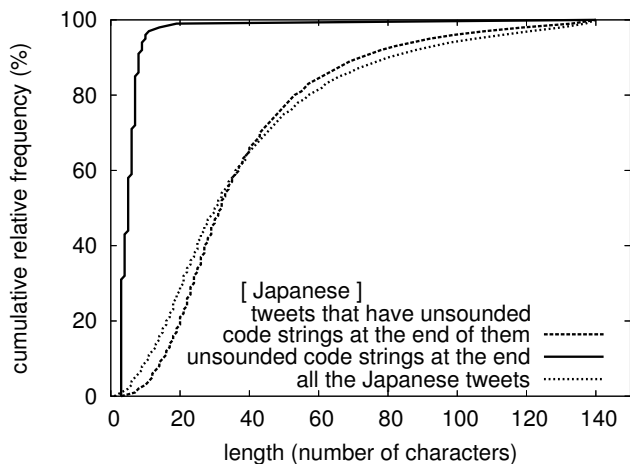- the length of Japanese and English replies that have unsounded code strings at the end of them.

Figure 3. The cumulative relative frequency distribution of the length of (1) all the Japanese tweets, (2) Japanese tweets that have unsounded code strings at the end of them, and (3) unsounded code strings at the end of them.
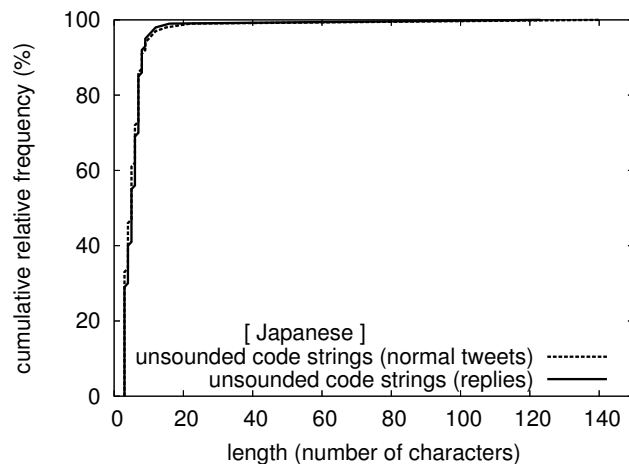

Figure 5. The cumulative relative frequency distribution of the length of unsounded code strings at the end of Japanese normal tweets and replies.
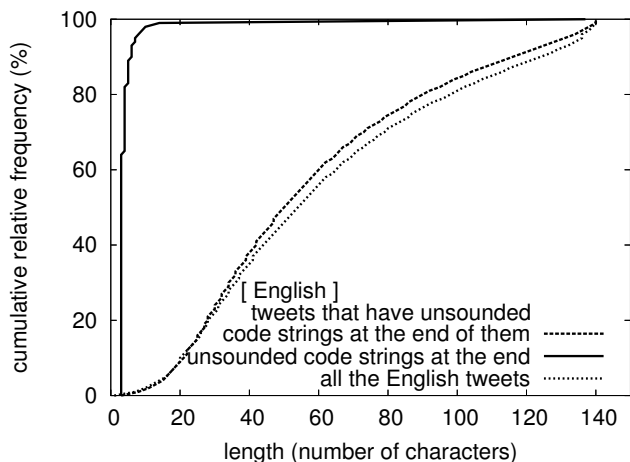

Figure 4. The cumulative relative frequency distribution of the length of (1) all the English tweets, (2) English tweets that have unsounded code strings at the end of them, and (3) unsounded code strings at the end of them.
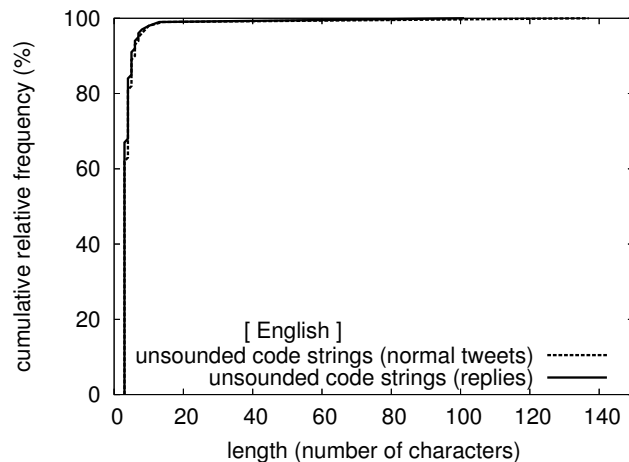

Figure 6. The cumulative relative frequency distribution of the length of unsounded code strings at the end of English normal tweets and replies.

As shown in Figure 7–10, only the distribution of the length of Japanese replies that have unsounded code strings at the end of them shifts to longer ranges than the length of all the Japanese replies. On the other hand, the distribution of the length of the other tweets (Japanese normal tweets, English normal tweets and replies) that have unsounded code strings at the end of them do not shift to longer ranges when they are longer than 30 characters. It may be said that the length of Japanese tweets that have unsounded code strings at the end of them are affected by whether the tweets are normal tweets or replies. On the other hand, the length of English tweets that have unsounded code strings at the end of them are not affected.

Next, we investigate that the percentages of tweets that have unsounded code strings at the end of them are affected by whether the tweets are normal tweets or replies. Figure 11 shows the percentages of Japanese normal tweets and replies

that have unsounded code strings at the end of them. As shown in Figure 11, 9.4 % of Japanese normal tweets have unsounded code strings at the end of them while 17.0 % of Japanese replies have unsounded code strings at the end of them. As a result, the percentages of Japanese normal tweets and replies that have unsounded code strings at the end of them differ considerably from each other. In other words, Japanese replies have unsounded code strings at the end of them more frequently than Japanese normal tweets. On the other hand, Figure 12 shows the percentages of English normal tweets and replies that have unsounded code strings at the end of them. As shown in Figure 12, 7.0 % of English normal tweets have unsounded code strings at the end of them while 7.6 % of English replies have unsounded code strings at the end of them. As a result, the percentages of English normal tweets and replies that have unsounded code strings at the end of them differ little from each other. In addition, the percentages of English normal tweets and Japanese normal tweets that have
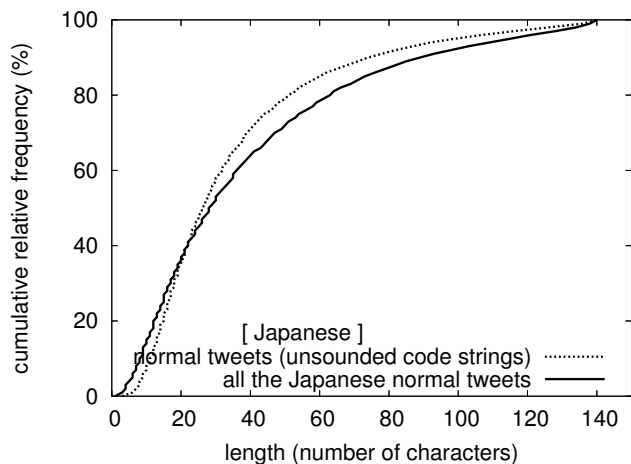
Figure 7. The cumulative relative frequency distribution of the length of all the Japanese normal tweets and Japanese normal tweets that have unsounded code strings at the end of them.
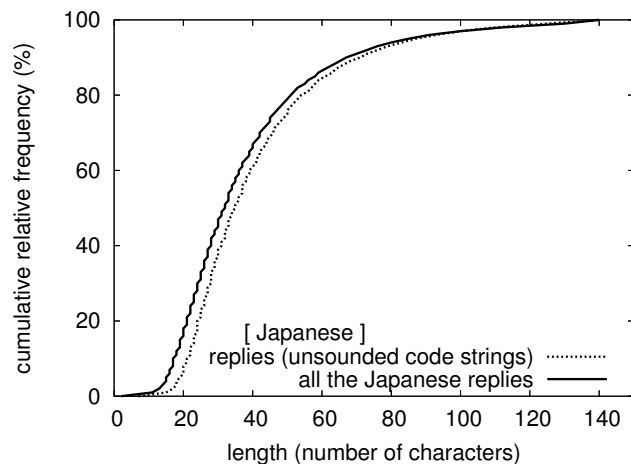
Figure 9. The cumulative relative frequency distribution of the length of all the Japanese replies and Japanese replies that have unsounded code strings at the end of them.
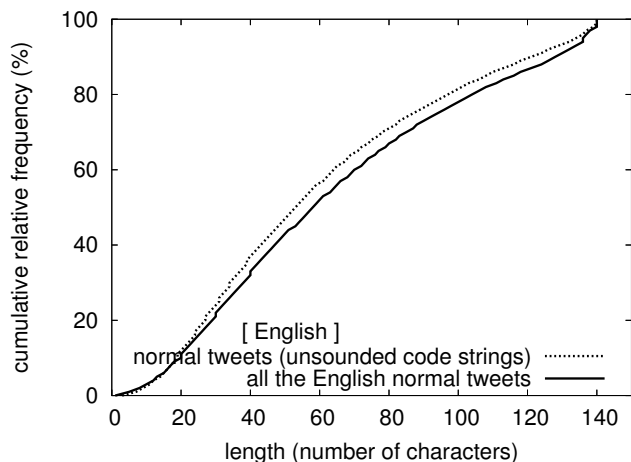
Figure 8. The cumulative relative frequency distribution of the length of all the English normal tweets and English normal tweets that have unsounded code strings at the end of them.
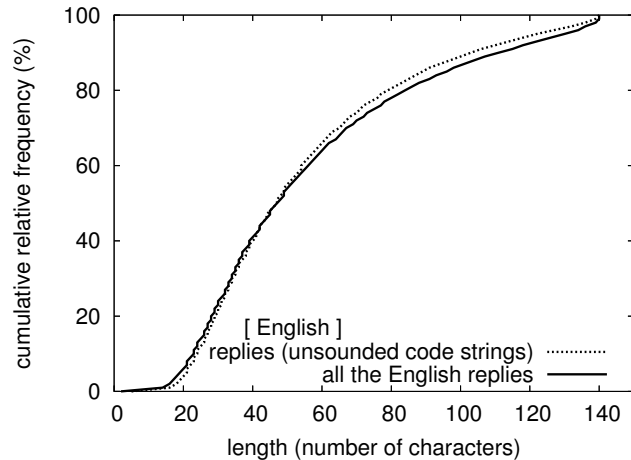
Figure 10. The cumulative relative frequency distribution of the length of all the English replies and English replies that have unsounded code strings at the end of them.

unsounded code strings at the end of them differ little from each other. From these points, it may be said that Japanese speakers use unsounded code strings at the end of tweets more frequently to particular persons than to general public while English speakers do not. In other words, the receiver of a tweet, whether general public or a particular person, is a factor that affects the usage of unsounded code strings for Japanese speakers, however, not for English speakers. It is not clear whether this phenomenon is specific for Japanese speakers.

IV. CONCLUSION

Unsounded code strings, in other words, consecutive unsounded marks and characters are frequently used at the end of online messages. However, there were few studies on them. In this study, we investigated unsounded code strings at the end of Japanese and English tweets in Twitter. Then, we showed that Japanese speakers use unsounded code strings

at the end of tweets more frequently to particular persons than to general public while English speakers do not. It may be said that the receiver of a tweet, whether general public or a particular person, is a factor that affects the usage of unsounded code strings for Japanese speakers, however, not for English speakers. In order to discuss whether this phenomenon is specific for Japanese speakers, we intend to analyze tweets in various languages. The results of this study will give us a chance to understand the usage of unsounded code strings and improve multilingual CMC systems.

REFERENCES

[1] N. Wolfson, "The social dynamics of native and nonnative variation in complimenting behavior," in The Dynamic Interlanguage: Empirical Studies in Second Language Variation. Springer US, 1989, pp. 219–236.

[2] M. L. Bergman and G. Kasper, "Perception and performance in native and nonnative apology," in Interlanguage pragmatics. Oxford University Press, 1993, pp. 82–107.
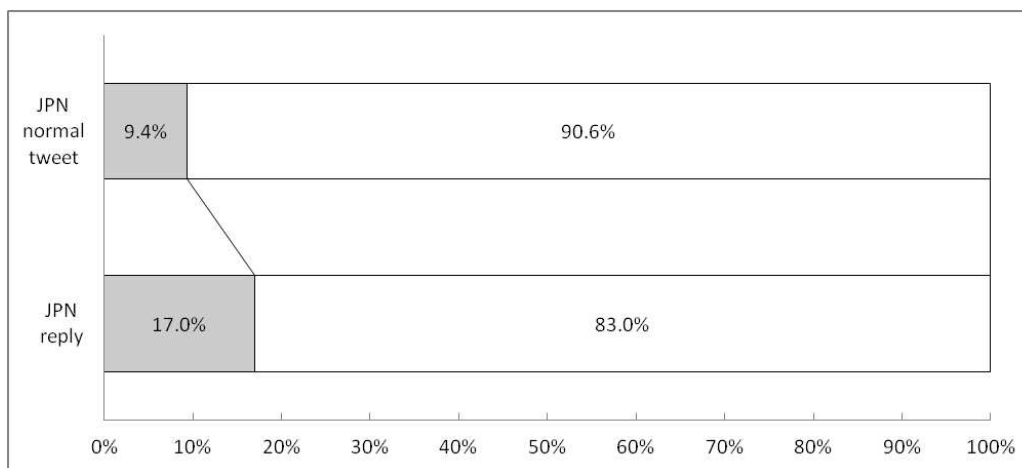
Figure 11. The percentages of Japanese normal tweets and replies that have unsounded code strings at the end of them
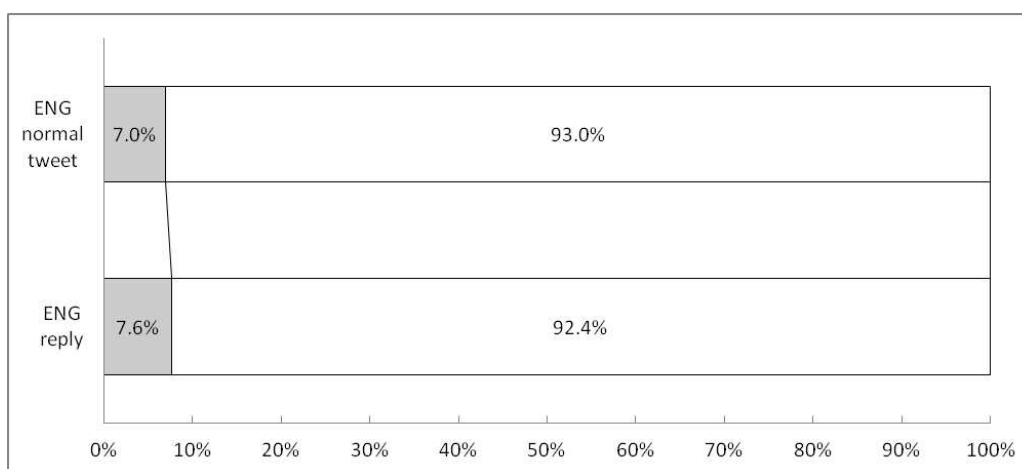

Figure 12. The percentages of English normal tweets and replies that have unsounded code strings at the end of them

[3]   M. Eisenstein and J. Bodman, "Expressing gratitude in American English," in Interlanguage pragmatics. Oxford University Press, 1993, pp. 64–81.

[4]   S. Tanaka and S. Kawade, "Politeness strategies and second language acquisition," Studies in Second Language Acquisition, vol. 5, no. 1, 1982, pp. 18–33.

[5]   L. M. Beebe, T. Takahashi, and R. Uliss-Weltz, "Pragmatic transfer in ESL refusals," in Developing communicative competence in a second language. Newbury House Publishers, 1990, pp. 55–73.

[6]   A. Deschamps, "The syntactical distribution of pauses in English spoken as a second language by French students," in Temporal variables in speech. De Gruyter Mouton Publishers, 1980, pp. 255–262.

[7]   M. Bilá and A. Džambová, "A preliminary study on the function of silent pauses in L1 and L2 speakers of English and German," Brno Studies in English, vol. 37, no. 1, 2011, pp. 21–39.

[8]   A. Ishizaki, "How does a learner leave a pause when reading Japanese aloud?: A comparison of English, French, Chinese and Korean learners of Japanese and native Japanese speakers (in Japanese)," Japanese-Language Education around the Globe, vol. 15, 2005, pp. 75–89.

[9]   P. Tavakoli, "Pausing patterns: differences between L2 learners and native speakers," ELT Journal, vol. 65, no. 1, 2011, pp. 71–79.

[10] S. Okazawa, "Pauses and fillers in second language learners' speech,"

Studies in Language and Culture, vol. 23, 2014, pp. 52–66. [Online]. Available: http://opac.library.twcu.ac.jp/opac/repository/1/5697/ [retrieved: August, 2015]

[11] V. LoCastro, "Aizuchi: A Japanese conversational routine," in Discourse Across Cultures: Strategies in World Englishes. Prentice Hall, 1987, pp. 101–113.

[12] A. Tera, K. Shirai, T. Yuizono, and K. Sugiyama, "Analysis of eye movements and linguistic boundaries in a text for the investigation of Japanese reading processes," IEICE Transactions on Information and Systems, vol. E91.D, no. 11, 2008, pp. 2560–2567.

[13] S. K. Maynard, "On back-channel behavior in Japanese and English casual conversation," Linguistics, vol. 24, no. 6, 1986, pp. 1079–1108.

[14] L. Miller, "Verbal listening behavior in conversations between Japanese and Americans," in The Pragmatics of Intercultural and International Communication. John Benjamins Publishing, 1991, pp. 110–130.

[15] S. White, "Backchannels across cultures: A study of Americans and Japanese," Language in Society, vol. 18, no. 1, 1989, pp. 59–76.

[16] C. Tzuching, "Existing research of Japanese backchannels : An overview for the future (in Japanese)," Japanese language education, vol. 2002, 2002, pp. 222–235.