Analysis of Emphasis on Japanese-English Bilingual Corpora *

◎ Do Quoc Truong, Sakriani Sakti, Graham Neubig,
Tomoki Toda, Satoshi Nakamura (NAIST)

1 Introduction

Conventional speech-to-speech (S2S) translation [1] systems only translate the content of the utterance and ignore paralinguistic information included in the input speech. As a step towards addressing this limitation, in this paper we analyze paralinguistic information across languages. Among the various types of paralinguistic information, we focus on emphasis, a type of information that is used to convey the focus of the sentence. Emphasis is an important factor especially when repeating an initially misheard sentence (a situation that occurs often when using less-than perfect speech translation systems). This paper describes an analysis regarding how emphasis is expressed across two languages: Japanese and English.

2 Construction of a corpus of emphasized speech

The construction of this corpus is described in the follow procedure:

First, we selected 16k pairs of sentences from the Basic Travel Expression Corpus (BTEC) [2] which consists of sentences from travel domain, and performed part-of-speech tagging on both languages. We used the NLTK [3] toolkit for English sentences and Mecab [4] for Japanese sentences.

Next, we performed word alignment between the sentences using the pialign tool [5] to ensure that the same words were analyzed in both sentences. This setting allows us to analyze how emphasis can be translated between languages. After this, we had 2500 sentences. These sentences were verified manually to ensure the correctness and naturalness of emphasized units. After manual verification, a total of 1015 pairs of sentences remained, the details of which are shown in Table 1.

There were 2 speakers, one Japanese and the other American. They can speak both Japanese and English fluently. The speakers were asked to emphasize some particular words intentionally.

Table 1 The conversation corpus materials

Utterances		1015
Emphasized units		1305
Utterances has X emphasized units	1	776
	2	193
	3	41
	4	5

Table 2 Recorded speech data for the conversation corpus

Speaker	**Utterances**	**Emphasized words**
Male Japanese	1015	1305
Male American	1015	1305

3 Analysis

We did an analysis of the recorded data to see the differences between Japanese and English in terms of amplitude, duration and changes in fundamental frequency (F0). All recorded data were used for evaluation.

First, taking a look at amplitude in Figure 1 we can see that the amplitude of Japanese emphasized words are almost equal to English emphasized words, the different between median of amplitute distribution of emphasized words and normal words is 10 dB in English and 9 dB in Japanese. However, the Figure 2 shows that English speakers tend to emphasize words by using more duration than Japanese speakers. The median of duration for emphasized words is 400 ms higher than normal words in English and 250 ms in Japanese. The study in [6] said that Japanese emphasis uses less power and duration than English. This is consistent with our duration's observation, but inconsistent regarding amplitude's observation. The reason is in our recording conditions, the speakers were asked to speak the Japanese sentences right after English sentences, so the Japanese emphasized words may have been influenced by the English emphasized words.

Figure 3 shows the average standard deviation

* 日英対訳コーパスにおける強調音声の分析

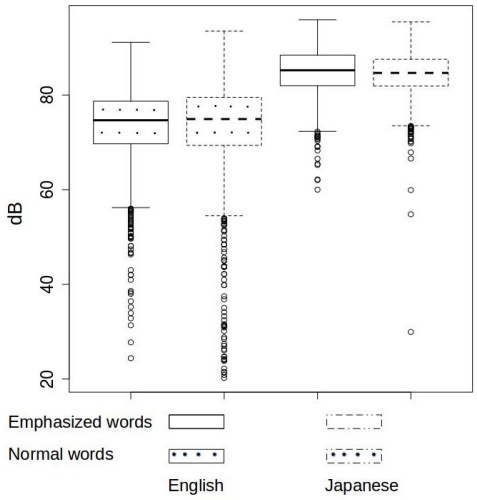

Fig. 1　Amplitude distribution

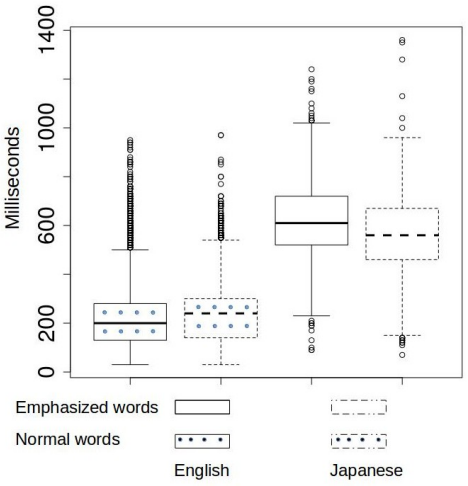

Fig. 2　Duration distribution

(std) distribution of F0 for a single word, representing the amount of frequency change within words. In this work, we calculate the standard deviation for emphasized and normal words. The higher average std, the more fluctuation of F0. We can see from the result that the emphasized words have higher fluctuation of F0 compared to normal words. The different between median of average std distribution of emphasized words and normal words in Japanese is 14.4 Hz and 14.8 Hz in English.

4　Conclusion

This paper described the collection of an emphasized speech corpus and the analysis of emphasis factor across languages. We found that our analysis is consistent with the prior knowledge that in English people often use more power and duration

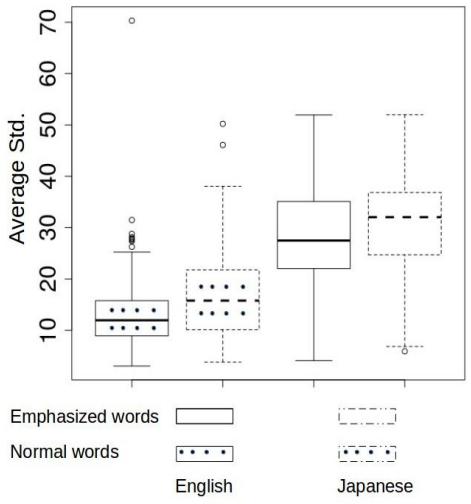

Fig. 3　Average standard deviation distribution

than in Japanese to emphasize words.

5　Acknowledgements

Part of this work was supported by JSPS KAKENHI Grant Number 24240032.

参考文献

[1] S. Nakamura, "Overcoming the language barrier with speech translation technology," *Science & Technology Trends - Quarterly Review No.31*, April 2009.

[2] T. Takezawa G. Kikui, E. Sumita and S. Yamamoto, "Creating corpora for speech-to-speech translation," in *Proceedings of EUROSPEECH*, 2003, pp. 381–384.

[3] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, Inc., 1st edition, 2009.

[4] T. Kudo, *MeCab: Yet Another Part-of-Speech and Morphological Analyzer*, http://mecab.sourceforge.jp, 2006.

[5] G. Neubig, T. Watanabe, E. Sumita, S. Mori, and T. Kawahara, "An unsupervised model for joint phrase alignment and extraction," in *Proceedings of ACL*, 2011, pp. 632–641.

[6] Mary E. Beckman and Janet B. Pierrehumbert, "Japanese prosodic phrasing and intonation synthesis," in *Proceedings of ACL*, 1986, pp. 173–180.