Interpreting Unit Segmentation of Conversational Speech in Simultaneous Interpretation Corpus

Zhe DING*, Koichiro RYU*, Shigeki MATSUBARA, Masatoshi YOSHIKAWA***
*****Department of Information Engineering, Nagoya University
**Information Technology Center, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
ding@dl.itc.nagoya-u.ac.jp

Abstract

The speech-to-speech translation system is becoming an important research topic with the progress of the speech and language processing technology. Considering efficiency and the smoothness of the cross-lingual conversation, the simultaneity of the translation processing has a great influence on the performance of the system. This paper describes interpreting unit segmentation of conversational bilingual speech in simultaneous interpretation corpus which has been developed in Nagoya University. By finding the segmentation point of spoken utterances in the speech corpus manually, we identified a clause-unit as a practical interpreting unit. In this paper, we examined the availability of such unit, and segmented spoken dialogue sentences into interpreting units. A large-scale bilingual corpus for which the interpreting units are provided can be used for the simultaneous machine interpretation.

1 Introduction

In these years, with the progress of internationalization, natural and smooth communications on contact with computers in cross-language conversation has been desired. Therefore, the advance of technologies for speech processing and language translation has been highly expected, and the speech-to-speech translation system is becoming one of the most important research topics.

Over the past few years, a considerable number of studies have targeted the conversational speech, and most of them are limited to the estimation of degree of accuracy. But nowadays, considering efficiency and the smoothness of the cross-language conversation, the simultaneity of the translation processing attracts the attention of all many researchers.

As to simultaneous machine interpretation, not only the accuracy of the interpretation but its output timing is also important, although the proper output timing is not well-defined. When a sentence is recognized as an interpreting unit which is said to be a linguistic chunk that could be interpreted separately and simultaneously, the simultaneity will not be satisfied. On the other hand, a small linguistic unit like a word or a phrase, etc. is not an effective interpreting unit either, because it is not necessarily realistic in current technologies of speech recognition (Ryu, 2004). Therefore, in this paper we focused attention on a clause-unit as an interpreting unit.

In this paper, we describe interpreting unit segmentation of conversational bilingual speech in simultaneous interpretation corpus. The effective interpreting unit is identified by finding the segmentation of spoken utterances in bilingual speech corpus. Added to this, we made an investigation into a possibility of simultaneous machine interpretation by extracting such interpreting unit from our bilingual corpus (Tohyama, 2004). A large-scale bilingual corpus for which the interpreting unit is provided can be used for the simultaneous machine interpretation.

This paper is organized as follows: Section 2 explains the concept of the interpreting unit segmentation. Section 3 describes the preliminary investigations. Section 4 describes the technique for annotating the bilingual corpus by the interpreting units. Section 5 provides the result of an experiment and our observations of interpreting unit segmentation.

2 Simultaneous Interpreting Unit

The conversational speech data of the simultaneous interpretation corpus has been developed in Nagoya University (Ryu, 2003). The data consists of the conversational speech between Japanese and English through the simultaneous interpreters in traveling abroad situations such as airport check-in or, booking of a room at a hotel. The speech data of about 60,000 utterances and 420,000 words have been collected. This large-scale bilingual corpus provides the transcribed text between Japanese and English, the bilingual

0011 - 00:50:512-00:52:711 N:
(F えー)七月十五日
0012 - 00:53:223-00:56:287 N:
から二泊御予約の(R マリクローズ)さんでございます<SB>
0013 - 01:00:432-01:00:584 N:
はい<SB>
0014 - 01:00:832-01:04:824 N:
(F え)ただいまお調べをしまして(F え)予約のほうは確かに入っておりますが
0015 - 01:05:152-01:06:095 N:
(F えー)

(a) Japanese native speaker

0011 - 0011-0012 T:
You are Mr. Malik Rose. You reserved a room for two nights from July fifteenth, right?
0012 - 0013-0013 T:
Yes.
0013 - 0014-0014 T:
I checked it. We surely have your reservation.
0014 - 0015-0015 T:
0015 - 0016-0017 T:
You said that the date was from sixteenth, but
0016 - 0018-0018 T:
0017 - 0019-0019 T:
yes.

(b) Japanese-English interpretation

Figure 1: A sample of the transcripts

alignment, the visualization of speaking time, etc. Figure 1 shows a sample of the transcript.

The main difference between consecutive interpretation and simultaneous interpretation would be the beginning time of the interpretation. In general, in order to reduce listener's waiting time, simultaneous interpreters break up the utterance into several meaningful segments, and translate them incrementally. We call such segment interpreting unit. In other words, interpreting unit can be defined as a linguistic chunk that could be interpreted separately and simultaneously.

Recently, a small unit like word-unit or phrase-unit, etc. has been used as a unit of the simultaneous machine interpretation though it is not efficient and effective adequately, because it is not necessarily realistic in current technologies of speech recognition. Therefore, in this paper we will focus attention on clause-unit as a practical interpreting unit (Kashioka, 2004). The simultaneous interpreting corpus which is segmented into practical interpreting units will be getting valuable in the coming machine interpretation research.

(2.1) ホテルの予約をして来なかったので/こちらでホテルを紹介していただきたいんですけれども
(2.2) I haven't made any hotel reservation /so could you introduce me any nice hotel?

This is an example of bilingual conversational speech with interpreting units. Both Japanese and English consist of two clauses and they are semantically compliant each other. Therefore, we can recognize each clause of Japanese as interpreting units. When "ホテルの予約をして来なかったので" was input, the parallel interpreting "I haven't made any hotel reservation" will be output.

3. Preliminary Investigations

In order to identify interpreting units in Japanese

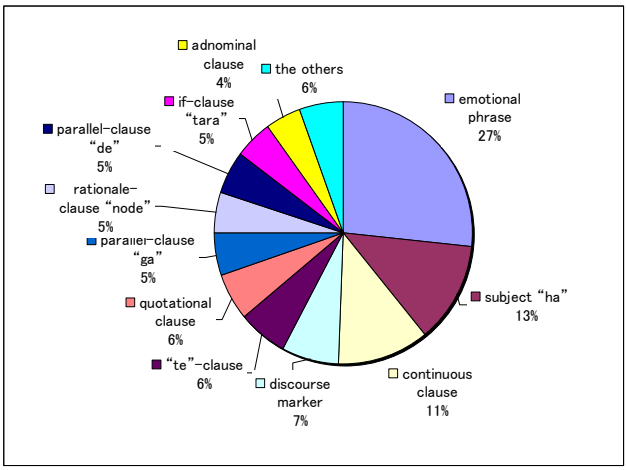

Figure 2: Breakdown of the labels

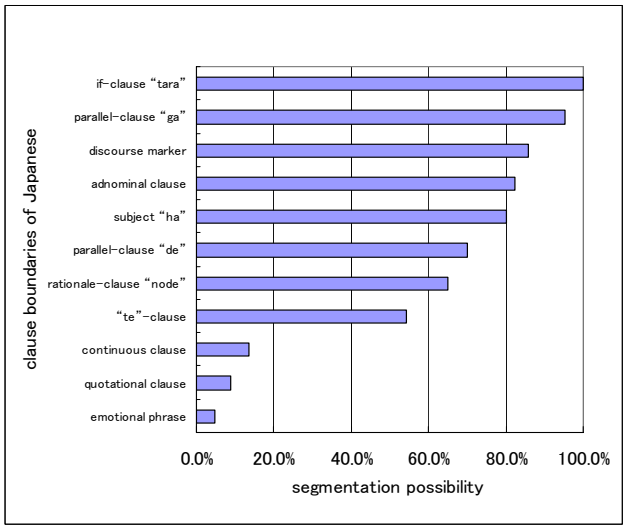

Figure 3: Segmentation possibilities

conversational sentences, we made some provision manually. We used the Japanese-to-English part in conversational speech data of the simultaneous interpretation corpus, which has developed in the Center for Integrated Acoustic Information Research (CIAIR), Nagoya University. We selected 11 dialogues randomly from the corpus. The dialogue data consists of 519 spoken Japanese sentences in total.

At first, we segmented the Japanese sentences into clauses by using a clause boundaries detection program, CBAP (Maruyama, 2004). In the result, 207 sentences were divided into two or more clauses. The clause labels in these sentences are investigated. Figure 2 shows the breakdown of the labels. We can see that the top 11 clause labels of high occurrence rate take over 94% of the total.

Then, we investigated whether these 11 kinds of Japanese clauses can be identified as interpreting units or not. The investigation was done by extracting the segmentation points which satisfy the following two conditions:

- We can recognize the English boundary unit

corresponding to the detected Japanese clause semantically.

- The corresponding boundary units of Japanese and English appear in the same order.

That is, if a Japanese sentence can be segmented into the boundary units A and B, its translation into C and D, furthermore, A and C, B and D can be aligned, respectively, then the boundary between A and B can become a segmentation point. This means that the boundary units A and B can be regarded as interpreting units.

Figure 3 shows the rate of segmentation points in the clause boundaries in a label-by-label basis. We can see that the difference between "te"-clause and continuous clause is greater, and therefore, we identify the top eight clauses of this figure: if-clauses "tara", "te"-clause, etc. as interpreting units.

In the result of an examination using the closed data, the accuracy and the recall ratio were 78.9% and 86.7%, respectively, we confirmed our identification method to be effective.

4 Interpreing Unit Segmentation

This section describes a technique for segmenting a spoken Japanese sentence into two or more interpreting units. Figure 4 shows the flow of the interpreting unit segmentation using a Japanese-English conversational speech corpus. The technique consists of three steps: sentence alignment, sentence analysis, and sentence segmentation. Each step will be explained in detail below.

4.1 Data Arrangement

The first step arranges the bilingual data because the original text in the corpus was not separated by sentences. We used DETAG program to break the original text up into sentences and take off fillers which exert a harmful influence on analyzing efficient interpreting units. Every sentence is end up with a punctuation mark.

4.2 Language Analysis

The second step analyzes both Japanese and English sentences linguistically, respectively. In the below, let us use the following pair of aligned sentences (4.1) and (4.2) as an example. This example was extracted from the CIAIR conversational speech corpus in fact.

(4.1) あと日本の流行をお知りになりたいのでしたら若者が多く集まるエリアがございます。

(4.2) And if you want to know about Japanese fashion, there is an area which is crowded with young people.

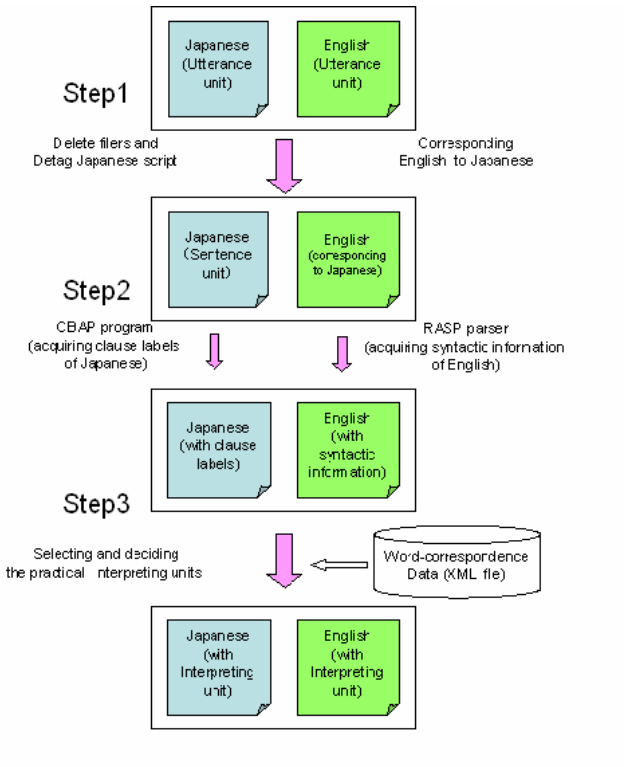

Figure 4: The flow of the interpreting unit segmentation

First, for the Japanese sentence, clause boundaries are provided by CBAP to line up the candidates of interpreting unit segmentations. For example, (4.3) is generated by applying the CBAP to (4.1).

(4.3) あと日本の流行をお知りになりたいのでしたら /if-clause"tara"/
若者が多く集まる /adnominal clause/
エリアがございます /sentence end/

Here, the labels of clause boundary units are wedged between two slash symbols.
The result (4.3) indicates that the sentence (4.1) is divided into three clause boundary units and the above labels are provided for them. Among the labels, both "if-clause" and "adnominal clause" are included in so called eight clause labels, which are defined in the previous section. Therefore, three clause boundary units are all the candidates of interpreting units.

On the other hand, for the English sentences, phrase structures are provided by RASP (Briscoe, 2002), which is one of the context-free parsing program, to define the syntactic fragments of the sentence. Since the RASP parser gives an English sentence to a binary tree, the result is useful for finding the corresponding segmentation points in a top-down fashion. Figure 5 shows the parsing result for the English sentence (4.2).

4.3 Segmentation Into Interpreting Units

The last step extracts the interpreting units of Japanese spoken sentences by considering the word-correspondence between the Japanese and English sentences.

At first, the keywords in the sentences extracted using the word-corresponding data. As a keyword, the word whose part-of-speech are any one of noun, adjective, and adverb, was extracted. The POS tagging for Japanese sentences and English sentences are executed by Chasen (Matsumoto, 1999) and Brill's tagger, respectively. The result for (4.3) is (4.4), and for (4.2) is (4.5).

(4.4) あと(NN_1 日本)の(NN_2 流行)をお知りになりたいのでしたら /if-clause"tara"/ (NN_3 若者)が多く集まる/adnominal clause/ (NN_4 エリア)がございます/sentence end/

(4.5) And if you want to know about (NN_1 Japanese) (NN_2 fashion), there is an (NN_4 area) which is crowded with (NN_3 young people)

Here, keywords are expressed as the bracketed word with part-of-speech, the numbers shows the word correspondence.

Next, the keyword sequence are generated and the segmentation points are extracted. For example, the keyword sequences of (4.4) and (4.5) are as follows:

(4.6) (NN_1 日本) (NN_2 流行) /if-clause "tara"/ (NN_3 若者) /adnominal clause/ (NN_4 エリア) /sentence end/

(4.7) (NN_1 Japanese) (NN_2 fashion) (NN_4 area) (NN_3 young people)

By considering the appearance order of the keywords between Japanese and English, the boundary between the 1st and 2nd clauses in the Japanese sentence is extracted as interpreting unit segmentation.

Finally, the segmentation points are provided for the English sentence. It is required to find the segmentation points in the sentence since those in the keyword sequence are already decided. We utilize the result of phrase structure parsing for that. For example, there exists a segmentation point between (NN_2 fashion) and (NN_3 area) in (4.7). This means that any one of four word segmentations in "(NN_2 fashion), there is an (NN_4 area)" is the segmentation point. It can be extracted based on the fragment segmentation in the binary tree of Figure 5 because this tree shows that this sentence (S) can be divided into "And if you want to know about Japanese fashion" as a prepositional phrase (PP) and "there is an area which is crowded with young people" as a sentence (S).

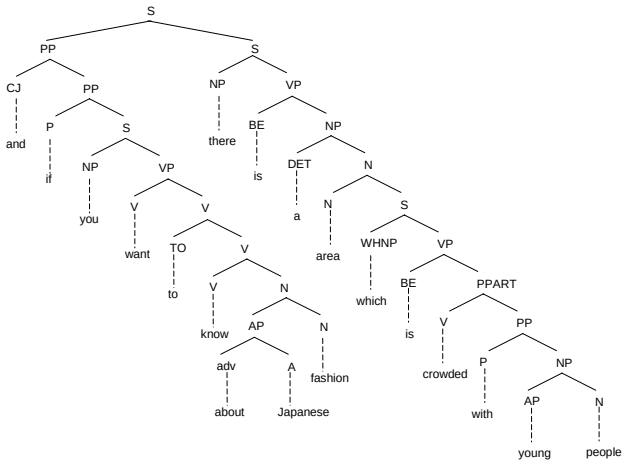

Figure 5 : Binary tree by RASP

5 Segmentation Experiment

In order to evaluate the effectiveness of interpreting unit segmentation of conversational sentences and the feasibility of the technique which has been explained in the previous section, we have made a segmentation experiment. An experimental data, we used the Japanese-to-English part in conversational speech data of the simultaneous interpretation corpus. The data has 216 spoken dialogues and 8721 sentences.

First, we tried to segment these sentences. There existed 5019 clause labels with the exception of "sentence end". The total of the clause labels which were matching the top eight were 3846, and the sentences including at least one clause label in the top eight is 2375. After applying the method of Step 3 described in section 4, we found 1005 clause labels which can be recognized as interpreting unit candidates, and 677 sentences which are including such interpreting unit segmentation.

After examining the 1005 clause labels further, we found there are some characters of them. Figure 6 shows the relation between the amount of the sentences with interpreting unit segmentation and the amount of the interpreting segmentation in such sentences. We may, therefore, reasonably conclude that there are not a few sentences should be segmented even in conversational speech.

Figure 7 shows the rate of segmentation possibility of the eight clause labels by using the method of section 4 automatically. Comparing Figure 4 with Figure 7, we may conclude that the segmentation possibility of the eight clause labels acquired by hand is differing greatly from the result acquired by machine. From Figure 7, we can also see that specific clause such as "discourse marker" is the most difficult clause label to extract. The reason may be thought as the amount of the keywords

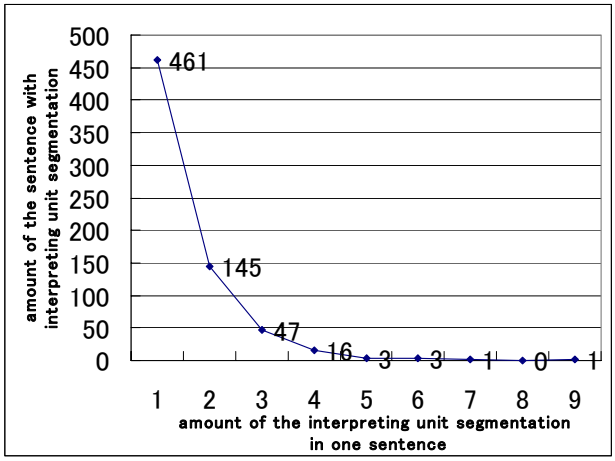

Figure 6: Relation between sentences and interpreting unit segmentation

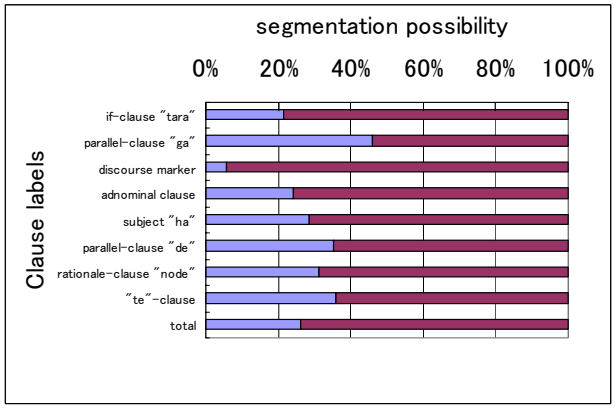

Figure 7: Rate of segmentation possibility of the eight clause labels

which can be aligned from the word-correspondence data is not enough. For example, if verbs can be extracted as keywords, more practical interpreting unit may be extracted.

6 Concluding Remarks

This paper has described a method for interpreting unit segmentation of conversational speech in CIAIR simultaneous interpretation corpus. The segmentation is executed by extracting specific clause boundaries in Japanese sentences and by finding the segmentation points in the corresponding English sentences based on word alignment. We have made a segmentation experiment using the conversational bilingual speech. The result shows the possibility that the top eight Japanese clause labels can be identified as interpreting units. That is, when these clause labels appear at Japanese speech, a simultaneous machine interpreting system can break up the spoken sentences into two or more segments and translate them incrementally. The practical interpreting unit segmentation would play an important role for supporting natural and smooth cross-lingual machine-mediated speech communication.

7 Acknowledgements

The authors would like to thank their colleague Mr. Kazuya Tanaka for his valuable contribution in implementation. They also wish to express their gratitude to Dr. Hideki Kashioka and Dr. Takehiko Maruyama for their helpful suggestions. This research was partially supported by the Grant-in-Aid for Young Scientists (No. 17700148) of JSPS.

References

K. Ryu, S. Matsubara, N. Kawaguchi, and Y. Inagaki, 2003. "Bilingual Speech Dialogue Corpus for Simultaneous Machine Interpretation Research", Proceedings of Oriental COCOSDA-2003, pp. 217-224.

H. Tohyama, S. Matsubara, K. Ryu, N. Kawaguchi, and Y. Inagaki, 2004. "CIAIR Simultaneous Interpretation Corpus", Proceedings of Oriental COCOSDA-2004, Vol. II, pp. 72-77.

T. Kashioka and T. Maruyama, 2004. "Segmentation of semantic unit in Japanese monologue", Proc. of O-COCOSDA-2004, pp. 87-92.

T. Maruyama, T. Kashioka and H. Tanaka, 2004. "Development and evaluation of Japanese clause boundaries annotation program", Journal of Natural Language Processing, 11(3):39-68, 2004.(In Japanese)

E. Briscoe and J. Carroll, 2002. "Robust accurate statistical annotation of general text", Proc. of the 3rd International Conference on Language Resources and Evaluation, pp.1499-1504.

K. Ryu, A. Mizuno, S. Matsubara and Y. Inagaki, 2004. "Incremental Japanese Spoken Language Generation in Simultaneous Machine Interpretation", Proc. of Asian Symposium on Natural Language Processing to Overcome language Barriers, pp. 91-95.

Y. Matsumoto, A. Kitauchi, T. Yamashita and Y. Hirano, 1999. "Japanese Morphological Analysis System ChaSen version 2.0 Manual", NAIST Technical Report, NAIST-IS-TR99009.