SPOKEN LANGUAGE CORPUS FOR MACHINE INTERPRETATION RESEARCH

Yasuyuki Aizawa[1] *Shigeki Matsubara*[2,3] *Nobuo Kawaguchi*[1,3]
Katsuhiko Toyama[1,3] *Yasuyoshi Inagaki*[1]

[1]Graduate School of Engineering, Nagoya University
[2]Faculty of Language and Culture, Nagoya University
[3]Center for Integrated Acoustic Information Research, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
email: {aizawa, matu, kawaguti, toyama, inagaki}@inagaki.nuie.nagoya-u.ac.jp

ABSTRACT

This paper describes a database consisting of speech and language, which we are currently constructing for the purpose of research on machine interpretation. The database contains bilingual data of lectures and dialogues. We have collected of speech about 72 hours in total and transcribed it into the text manually. We have investigated the database in order to acquire empirical knowledge of human interpretation. In this paper, we report the characteristic features of spoken language by Japanese-to-English interpreters.

1. INTRODUCTION

Spoken language translation has been an important research topic in recent years. In fact, some dialogue translation systems have been developed so far. The systems, however, are used in a limited dialogue domain and translate only speech of specific types. Furthermore, the users are expected to speak only in one's own turn. This context suggests that the system is desired to be more natural and intelligent.

It has been proved that spoken language corpora are very useful for many natural language processing tasks during the decade of the 1990s. Especially, parallel corpora have played an effective part in the research on spoken language translation[1][2]. Because they provided the empirical knowledge of human interpreters and have been used as examples in corpus-based translation processing.

This paper describes a database of parallel spoken language, which we are currently constructing. It has the following characteristics:

- Speech is interpreted in a simultaneous fashion to realize natural cross-lingual communication.
- The beginning time and end time are given to each utterance unit of spoken dialogues.
- In addition to cross-lingual dialogues through interpreters, lectures with the interpretation are also contained.
- A spoken lecture read out by one native speaker is interpreted by two or more interpreters whose skill is different from each other.

Our corpus will be useful for the development of an advanced spoken language translation system. Furthermore, it could be used for investigating human interpreting mechanism and building an interpreting theory.

In fact, we have investigated the database to acquire empirical knowledge of Japanese-to-English interpreting. In consequence, we have confirmed frequent utilization of two kinds of techniques: sentence segmentation and phrase-order inversion.

This paper is organized as follows: The next section describes the contents of the database. Section 3 shows how we have constructed the database. Section 4 reports the result of investigating the database.

2. DATABASE DESIGN

We are constructing a parallel spoken language database.

2.1. Contents

Most of the existing parallel corpora supply only cross-lingual dialogue data[1][2]. However, we would like to emphasize that parallel data of lectures is also valuable. Actually, there is a big demand for interpreters of oral presentation in international conferences and of speeches by invited foreign speakers. Constructing such a corpus may enable us to develop an interpretation system for lectures.

Our database supplies not only dialogue data but also lecture data spoken in English and Japanese. As the dialogue task, we adopted "travel arrangement" which have been adopted in a lot of spoken language corpora so far[1][2]. Lectures were done according to the transcriptions of real spontaneous lecture data. Politics, economy, personal reminiscence, and so on have been chosen as themes.

Table 1: Contents of the database

conversational style	: dialogue/lecture
language	: English/Japanese
interpreting style	: simultaneous interpreting

2.2. Simultaneous Interpreting

Interpreting a dialogue in a simltaneous fashion works to avoid interrupting the coherency of communications. In a lecture, a speaker often uses aids which helps the listeners to understand, such as slides and pictures. Listeners can

refer to them simultaneously through simultaneous interpreting. These manners demand that the interpreting be simultaneous. In our corpus, every source of speech has been interpreted by simultaneous interpreter. To explore the relationship between skill in interpretation and interpreting results, a spoken lecture read out by one native speaker is interpreted by two or more interpreters whose skill is different from each other.

3. DATA COLLECTION

3.1. Recording

The recording was carried out in an internal room. Microphones (Sony Dynamic Microphone F-730) were used for stereo DAT recording (Sony Digital Audio Tape-Corder TCD-D100), and one recording channel was assigned to one speaker. Each speech was converted into a wave format, 16 bit, 16 kHz.

3.2. Transcription

Speech was transcribed into texts manually. Then we gave them tags that express linguistic phenomena such as fillers, stagnated utterances and mispronunciations. Regarding dialogue, speech was segmented into some utterance units by pauses. The beginning time and end time are given to every utterance unit of spoken dialogues. It enables us to obtain a time chart. Figure 1 shows an example of the time chart.

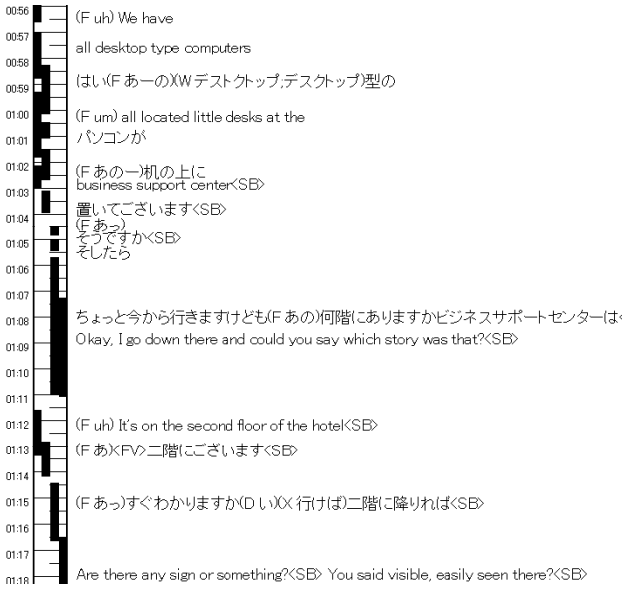

Figure 1: An example of the time chart. The leftmost number indicates the time from the beginning of the dialogue (*minute*:*second*). The Bar on the scale exhibits the duration of each utterance unit (in order from left: English speaker, English-to-Japanese interpreter, Japanese speaker, Japanese-to-English interpreter). The rightmost text is utterance unit.

3.3. Current Status

Currently, we have collected the speech of about 72 hours in total and transcribed them into texts of 22,229 sentences, as Table 2 shows.

Table 2: Current status

recording time (hours)	dialogue	32
	lecture	40
	total	72
utterances (sentences)	dialogue	8,676
	lecture	13,553
	total	22,229

4. CORPUS INVESTIGATION

Simultaneous machine interpreting is one of the most ambitious applications of spoken language processing. It requires incremental analysis and transfer of source language. There exists a difference in word-order between source language and target language, especially between linguistically distant languages such as English and Japanese. Generally, a verb is located on the earlier part of an English sentence. On the other hand, a verb often appears at the last part of a Japanese sentence. Therefore, it is difficult to decide the English verb before a whole Japanese sentence is uttered.

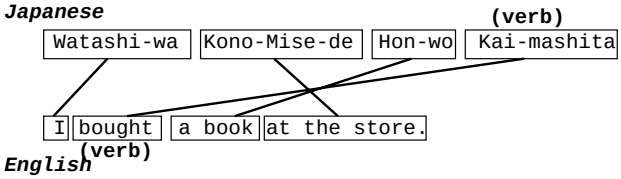

Figure 2: Different word-order between Japanese and English

Figure 2 shows the example. A Japanese verb "Kai-mashita" appears at the last part of the sentence "Watashi-wa Kono-Mise-de Hon-wo Kai-mashita". So, an interpreter can utter only "I" before "Kai-mashita" is uttered.

Not only verb, but also negating word and interrogative are located on the later part in a Japanese sentence. Besides, Japanese sentences are allowed to be uttered in various word-order. Thus, boundaries of spoken Japanese sentences are very ambiguious. As a result, it is difficult to determine the timing to output for plausible translation.

They may inhibit the simultaneity of Japanese-to-English interpretation.

4.1. Interpreting Technique

Human interpreters generally use particular empirical know-how to overcome the difficulties[3][4][5]. Therefore, we can say that it is effective for a simultaneous interpreting system to utilize such the know-how[6][7]. We found some kinds of techniques for Japanese-to-English

interpretation from our database. These are explained below. Each explanation is followed by an example. **J)**, **E)** and **E∗)** denotes a source sentence, the interpreting result and the normal translation respectively.

1)sentence segmentation

One Japanese sentence is translated into two or more English sentences by using demonstrative pronoun such as "that" and "it". Using this technique, an interpreter becomes receptive for the later part of a Japanese sentence.

J) ソフトウェア工学というものは実験的な分野だという言い方もできるでしょう.

E) The software engineering is the experimental field. You may be able to say that way.

E∗) You may be able to say that the software engineering is the experimental field.

2)counterchanging phrases

Prepositional phrases or adverbial phrases are usually on the last part in English sentences. However, if a phrase is uttered in an early stage of a Japanese utterance, an interpreter can translate it into English as a prepositional phrase or an adverbial phrase before the English sentence.

J) そこに並べられた車の中で 私が選んだのはスバル R-2 という車でした.

E) Among those cars what I chose was Subaru R-2.

E∗) What I chose was Subaru R-2 among those cars.

3)transformation into the passive voice

In a Japanese sentence, an objective word is allowed to be uttered before a subject word. In such a sentence, to translate it into a sentence of passive voice is helpful for simultaneous interpretation. This transformation means an exchange of the structural subject and object. Thus, it enables interpreters to utter in advantageous word-order.

J) 初めてあのアーバンリゾートという言葉を使い始めました.

E) At that time for the first time ever, the word urban resort was used.

E∗) At that time, we used the word urban resort for the first time ever.

4)summarization

Depending on the information density of the source utterance, an interpreter summarizes it in an appropriate degree. Summarization helps an interpreter to catch up with a speaker.

J) 二月から四月と八月から十月までがこの島の雨期にあたります. 乾期は四月から八月、十月から二月までの期間.

E) One from February through April and from August to October are rainy season. And the remaining periods are in the dry season.

E∗) One from February through April and from August to October are rainy season. And dry season is from April to August and from October to February.

Table 3: Investigated sentences

	sentences
speaker	1,617
interpreter	2,034

We investigated how often these techniques are utilized by simultaneous interpreters.

12 sessions of Japanese lectures composed of 1,617 Japanese sentences and 2,034 English sentences, which are aligned manually, are used. Then, the appearances of 1)sentence segmentation and 2)counterchanging phrases in these sessions have been counted.

Table 4 shows the numbers. "Sentence segmentation" appears in 58 appearances, and the 28 Japanese sentences contained an expression "という"[1], and are segmented there.

Table 4: Sentence segmentation

key word	number
という[1]	28
others	30
total	58

We found several keywords in the Japanese sentences containing "counterchanging phrases". Table 5 shows the key words and the numbers of their appearance.

Table 5: Counterchanging phrases

key word	number
年 (year)	59
時 (time)	34
中 (in/on)	22
において (in)	17
頃 (at about)	10
others	279
total	421

4.2. Towards an Interpreting System

As described at the beginning of this section, a Japanese sentence has a predicate in the latter part. An interpreting system has two options. One is to wait until the whole sentence is inputed. The other is to decide a plausible aspect at the middle of the input sentence, and output the partial interpretation. In this case, if the output is revealed to be wrong by the latter part, the system rephrases to modify the intent of the output. The former spoils the simultaneity, and the latter should take a risk to rephrase.

If the system has "sentence segmentation" rules, the system will be able to modify the intent of the output without rephrasing for certain types of Japanese sentences. The key word "という" will play an important role to construct "sentence segmentation" rules.

[1] A phrase "という" expresses a kind of quotation.

"Counterchanging phrases" means that specific kinds of phrases can be interpreted immediately. Our investigation will help to learn what kinds of Japanese phrases are allowed to be interpreted immediately, and how they should be interpreted into English.

5. CONCLUSION

This paper have described a spoken language database which we are constructing for research on spoken language translation. The database consists of parallel corpus of lectures and dialogues spoken in English and Japanese. Also, we have investigated the database and confirmed some kinds of techniques for simultaneous interpreting to be utilized frequently. We will continue collecting parallel data, and introduce the know-how acquired through more investigation in-depth into a simultaneous interpreting system[7].

6. ACKNOWLEDGEMENTS

The authors are grateful to Dr. Toshiyuki Takezawa and Dr. Eiichiro Sumita for their helpful comments. The collection and transcription of the spoken language data have been carried out cooperatively with Inter Group Corporation. The authors wish to thank specially Mr. Masafumi Yokoo for his contribution. This work is partially supported by the Grant-in-Aid for COE Research of the Ministry of Education, Science, Sports and Culture Japan and by The Telecommunications Advanced Foundation.

7. REFERENCES

1. Ehara,T., Ogura, K. and Morimoto, T.: ATR Dialogue Database, In *Proc. of ICSLP-90*, pp.1093-1096, (1990).
2. Morimoto, T., et al.: Speech and Language Database for Speech Translation Research, In *Proc. of ICSLP '94*, pp.1791-1794 (1994).
3. Mizuno, A.: On Simultaneous Interpretation from Japanese into English, In *Journal of the Interpreting Research Association of Japan*, Vol.5, No.5, pp.4-21 (1995). (in Japanese)
4. Yang, C.: Segmentation in Slight Translation from Japanese into Chinese, In *Journal of the Interpreting Research Association of Japan*, Vol.7, No.1, pp.4-17 (1997). (in Japanese)
5. Kamei, C.: A Case Study of Japanese-English Simultaneous Interpreting: In *Journal of the Interpreting Research Association of Japan*, Vol.7, No.2, pp.96-103 (1998). (in Japanese)
6. Mima, H., Iida, H. and Furuse, O.: Simultaneous Interpretation Utilizing Example-based Incremental Transfer, In *Proc. of COLING-ACL '98*, pp. 855-861 (1998).
7. Matsubara, S., Iwashima, K., Kawaguchi, N., Toyama, K. and Inagaki, Y.: Simultaneous Japanese-English Interpretation based on Early Prediction of English Verb, In *Proc. of SNLP-2000*, pp. 268-273 (2000).