

# A Corpus-based Analysis of Simultaneous Interpretation

A. Takagi<sup>†</sup>, S. Matsubara<sup>‡</sup>, N. Kawaguchi<sup>‡</sup>, and Y. Inagaki<sup>†</sup>

<sup>†</sup>Graduate School of Engineering, Nagoya University

<sup>‡</sup>Information Technology Center/CIAIR, Nagoya University  
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

{atakagi,matu,kawaguti,inagaki}@inagaki.nuie.nagoya-u.ac.jp

## Abstract

This paper provides an analysis of the interpreter’s speeches using an aligned simultaneous monologue interpreting corpus. The following points have been investigated: (1) the interpreter’s speaking speed, (2) the interpreting unit of simultaneous interpretation, and (3) the difference between the beginning time of the lecturer’s utterance and that of the interpreter’s utterance. This paper also describes characteristic features about the timing at which professional simultaneous interpreters start to speak. The analysis will be available for the development of a simultaneous machine interpreting system.

## 1 Introduction

In order to provide an environment supporting natural dialogues between different languages, simultaneous machine interpretation has been studied recently (Amtrup, 1999; Mima, 1998; Matsubara, 1999). Towards a simultaneous interpreting system, not only the quality of the interpretation but its output timing is also important, and it would be effective to investigate and analyze the interpreting process of professional simultaneous interpreters. The Center for Integrated Acoustic Information Research, Nagoya University (CIAIR), is constructing and maintaining the various types of speech and language database for the purpose of the advancement of robust speech information processing technologies (Kawaguchi, 2001). Moreover, a bilingual database of simultaneous interpretation is also constructed as a part of the project (Matsubara, 2001).

Table 1: Outline of simultaneous interpreting corpus

type of speech	monologue(lecture)
type of interpretation	simultaneous
language	Japanese, English
data type	speech, text
topics	computer science, politics, economy, history, etc.
number of interpreters	21
number of English lectures	24 (239 min. in total)
number of Japanese lectures	15 (146 min. in total)

This paper describes an analysis of interpreter’s speeches using the simultaneous interpreting corpus. In the transcript of this corpus, the exact beginning time and ending time are provided for every utterance. Using these time information, we have aligned the interpreter’s utterances with the lecturer’s utterances, and investigated the following points: (1) interpreter’s speaking speed, (2) the interpreting units in simultaneous interpretation, (3) the difference between the beginning time of the interpreter’s utterance and that of the lecturer’s utterance, and so on. This paper also describes the feature about the timing of simultaneous interpretation.

## 2 CIAIR Simultaneous Interpreting Corpus

### 2.1 Outline of the corpus

The project of multilingual data collection in CIAIR is collecting speeches of the simultaneous interpretation in both Japanese-English

Table 2: Statistics of the simultaneous interpreting corpus

item	Eng	E-J	Jap	J-E
recoding time (sec)	14359	57300	8767	35078
speaking time (sec)	11458	40473	5690	22289
morphemes	35474	165127	21708	63872
kinds of morpheme	3506	6115	1457	3130
utterance units	5573	20496	4505	12468
utterance sentences	1263	7423	651	3315
fillers	1251	8067	1482	3261

0001 - 00:05:128-00:06:599 N:  
The topic of this speech  
0002 - 00:07:168-00:08:551 N:  
is Christmas<SB>  
0003 - 00:09:520-00:09:808 N:<noise>  
0004 - 00:09:624-00:10:519 N:  
(noise This) is a little bit  
0005 - 00:10:936-00:12:111 N:  
more of an interesting  
0006 - 00:12:399-00:12:983 N:  
topic<SB>  
0007 - 00:13:520-00:15:448 N:  
It is one that  
0008 - 00:15:711-00:16:664 N:  
we all will be  
0009 - 00:16:672-00:17:328 N:<noise>  
0010 - 00:17:328-00:18:096 N:  
interested in<SB>

Figure 1: Sample of the transcript (English native speech)

and English-Japanese, and constructing a large-scale bilingual speech corpus(Aizawa, 2000). In 2001, the monologue speeches of English and Japanese and their simultaneous interpretation have been collected. As English speeches, the lectures about politics, economy, history, etc., and, as Japanese speeches, those about the themes relevant to computer science have been adopted. Moreover, each monologue speech is interpreted by four professional interpreters of which the degree of experience differs one another. In a Japanese lecture, since the lecturer uses the presentation slide, and the interpreter can also see the slide. Speech data is transcribed into ASCII text files according to the transcription criteria of “Corpus of Spoken Japanese(CSJ)”(Maekawa, 2000). Figure 1 and Figure 2 show the sample of the transcript. The discourse tags were provided

0001 - 00:06:928-00:09:423 I:  
このスピーチのテーマはクリスマスです<SB>  
0002 - 00:11:848-00:15:031 I:  
これは少しおもしろいテーマかもしれません<SB>  
0003 - 00:17:488-00:19:431 I:  
私達みんな関心のある  
0004 - 00:19:832-00:20:575 I:  
トピックです<SB>  
0005 - 00:21:312-00:23:183 I:  
私はヨーロッパで育ちました<SB>  
0006 - 00:23:664-00:25:943 I:  
私の親は宣教師でした<SB>  
0007 - 00:26:792-00:29:855 I:  
したがっていつもあちこちと引越していました<SB>  
0008 - 00:30:072-00:31:336 I:  
ヨーロッパ中<FV>  
0009 - 00:32:208-00:34:399 I:  
も行きましてアフリカにも行きまして<SB>  
0010 - 00:35:248-00:36:639 I:  
クリスマスの(W たんびに;たびに)

Figure 2: Sample of the transcript (English-Japanese interpreting speech)

for the language phenomena characteristic of spoken language, such as fillers, corrections, and misstatements. Moreover, the transcript is segmented into utterance units by the pause for 200ms or more, or that for 50ms after the sentence break, and the starting time and ending time have been provided for every utterance unit. Table 1 shows the outline of the simultaneous interpreting corpus.

## 2.2 Statistics of the Corpus

As fundamental statistics of the simultaneous monologue interpreting corpus, we have examined the recording time, the number of utterance units, the number of morphemes(words), the number of different morphemes(words), the speaking time, and the number of discourse tags. The result is shown in Table 2. Eng, E-J, Jap and J-E in the table mean English lec-

#	lecturer's utterance	interpreter's utterance
0	0001 - 00:05:128-00:06:599 N: The topic of this speech 0002 - 00:07:168-00:08:551 N: is Christmas<SB>	0001 - 00:06:928-00:09:423 I: このスピーチのテーマはクリスマスです <SB>
1	0003 - 00:09:520-00:09:808 N:<noise>	
2	0004 - 00:09:624-00:10:519 N: (noise This) is a little bit 0005 - 00:10:936-00:12:111 N: more of an interesting 0006 - 00:12:399-00:12:983 N: topic<SB>	0002 - 00:11:848-00:15:031 I: これは少しおもしろいテーマかもしれません <SB>
3	0007 - 00:13:520-00:15:448 N: It is one that 0008 - 00:15:711-00:16:664 N: we all will be 0009 - 00:16:672-00:17:328 N:<noise> 0010 - 00:17:328-00:18:096 N: interested in<SB>	0003 - 00:17:488-00:19:431 I: 私達みんな関心のある 0004 - 00:19:832-00:20:575 I: トピックです<SB>

Figure 3: Sample of the aligned corpus

ture, English-Japanese simultaneous interpretation, Japanese lecture, and Japanese-English simultaneous interpretation, respectively. In this paper, a morpheme in English means a word, and the number of morphemes in Japanese was calculated on the basis of the result of a Japanese morphological analyzer called ChaSen(Matsumoto, 2000). The number of kinds of morphemes in English is the number of words whose notations differ, and that in Japanese is the number of morphemes whose basic forms differ.

The interpreter's speaking time is about 4 times as much as the lecturer's time because the simultaneous interpretations by four interpreters were collected for one lecturer's speech. On the other hand, the number of utterance sentences of English-Japanese interpreters is about 6 times as much as that of the English lecturers and that of Japanese-English interpreters is about 5 times as much as that of the Japanese lecturers. This fact substantiates the empirical knowledge supported by an interpreting theory that a professional interpreter may segment a lecturer's utterance sentence into 2 or more sentences, and interpret it (Mizuno, 1995).

### 3 Parallel Corpus Alignment

In order to analyze the process of simultaneous interpretation in detail, it is required to align utterances by the interpreters with those of the lecturers in a possibly small unit. An alignment support tool (Figure 4) which work on the internet have been developed using CGI script. The users can perform alignment work by carrying out the mouse click on the bilingual text displays. The alignment data can be used for analysis of interpreting units and timing. We have aligned the corpus by this tool according to the following conditions:

- The smallest unit of alignment is a utterance unit.
- The corpus is aligned as small as possible.
- There is no counterpart for the lecturer's utterance unit such as a filler, an abbreviation and a supplement, etc.

The alignment data is made by hand for 16 lectures presently in both Japanese-English and English-Japanese. Figure 3 shows a part of an aligned corpus. The left-hand side expresses Japanese lecturer's utterance units, and right-hand side expresses Japanese-English interpreter's utterance units. Each line means an alignment pair.

Table 3: Statistics of the aligned corps

item	E-J: 1717 pairs		J-E: 1441 pairs	
	Sum	Ave	Sum	Ave
overlapping time (sec)	1845	1.07	1232	0.86
difference of beginning time (sec)	5482	3.19	5955	4.13
difference of ending time (sec)	5383	3.13	5829	4.04
lecturer’s speaking time (sec)	6593	3.84	5077	3.52
interpreter’s speaking time (sec)	6214	3.62	5388	3.74
lecturer’s morphemes	22360	13.02	18572	12.89
interpreter’s morphemes	25926	15.10	15554	10.79

#	lecturer’s utterance	interpreter’s utterance
0	0001 - 00:05:128-00:06:599 N: The topic of this speech 0002 - 00:07:168-00:08:551 N: is Christmas<SB>	0001 - 00:06:928-00:09:423 I: このスピーチのテーマはクリスマスです<SB>
1	0003 - 00:09:520-00:09:808 N: <noise>	
2	0004 - 00:09:624-00:10:519 N: (noise This) is a little bit 0005 - 00:10:936-00:12:111 N: more of an interesting 0006 - 00:12:399-00:12:983 N: topic<SB>	0002 - 00:11:848-00:15:031 I: これは少しおもしろいテーマかもしれません<SB>
3	0007 - 00:13:520-00:15:448 N: It is one that 0008 - 00:15:711-00:16:664 N: we all will be 0009 - 00:16:672-00:17:328 N: <noise> 0010 - 00:17:328-00:18:096 N: interested in<SB>	0003 - 00:17:488-00:19:431 I: 私達みんな関心のある 0004 - 00:19:832-00:20:575 I: トピックです<SB>

Figure 4: Alignment support tool

For each alignment pair in 16 sets of the aligned corpus, we examined the following items:

- The time overlapping with the lecturer’s utterance in the interpreter’s utterance
- The difference between the beginning time of the lecturer’s utterance and that of the interpreter’s utterance.
- The speaking time of lecturers and interpreters.
- The numbers of the morphemes of lecturer’s utterances and interpreter’s utterances.

Table 3 shows the result. Sum and Ave in the table mean the total of the time for 16 lectures and the average of the time per alignment pair, respectively.

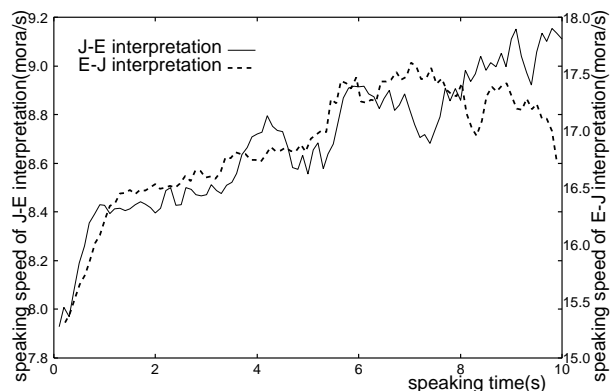


Figure 5: Interpreter’s speaking speed

## 4 Analysis of Simultaneous Interpreting Corpus

Simultaneous interpretation may overlap with the corresponding native speech. It is expected that an interpreter recognizes a part of a lecturer’s utterance as an interpreting unit, and interprets it at an early stage. We have investigated the interpreting units of professional interpreter by analyzing the aligned corpus.

### 4.1 Analysis of Speaking Speed of Interpreter

We have investigated the speaking speed of the interpreter. The speed was calculated with the number of moras per utterance unit. We assigned one mora to a short vowel and a consonant behind vowel, and two moras to long vowel in English. And we calculated the number of moras by using the result of ChaSen(Matsumoto, 2000) in Japanese. Figure 5 shows the result.

Although the interpretation begins at an early stage in order that an interpreter

may raise simultaneity, however, interpreter's speaking speed is slow because the role of the interpretation is not fixed at the beginning of the utterance. As the lecturer's utterance is progressing, the interpreter can grasp a lecturer's intentions gradually. Therefore, it is considered that the interpreter makes the speed gather gradually.

## 4.2 Japanese-English Simultaneous Interpretation

16 sets of the aligned data are used for the investigation. An interpreting unit may become a key phrase for determining at which timing the system starts to generate the result. In order to clarify such a unit, we have analyzed the feature of each alignment pair. Figure 6 shows the distribution of the number of morphemes in lecturer's utterances by the solid line. The number of the average morphemes is 12.89, and the range of 20 morphemes occupies 83.1% of the whole.

Among these, the utterance consisting of a few morphemes could be regarded as an interpreting unit. Because the system can interpret it immediately when such utterance is detected.

We have extracted the alignment pairs in which lecturer's utterances consist of four or less morphemes. There exist 231 pairs equivalent to 16.0% of the whole. The main features are listed below: The frequency is shown in a parenthesis.

- Lecturer's utterances consisting of a conjunction.(84 pairs)

**lecturer:** *soshite*  
**interpreter:** Then

- Lecturer's utterances consisting of a subject.(42 pairs)

**lecturer:** *yuza-ga*  
**interpreter:** the user,

- Lecturer's utterances to be interpreted as a prepositional phrase.(65 pairs)

**lecturer:** *koubunkaiseki-dewa*  
**interpreter:** in syntactic parsing,

Next, we have examined the difference between the beginning time of the interpreter's utterance and that of the lecturer's utterance.

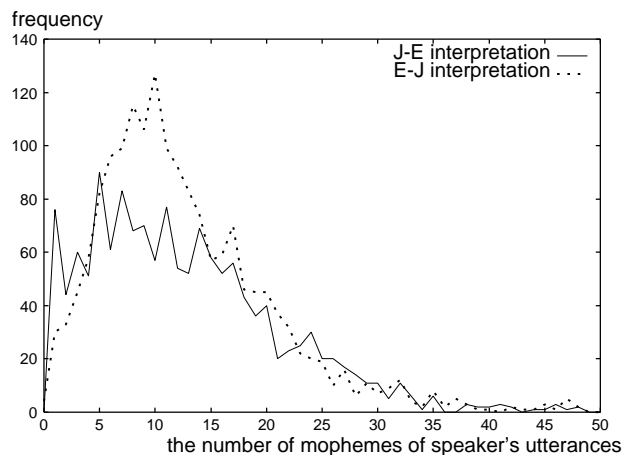


Figure 6: Distribution of the length of the lecturer's utterance

If lecturer's utterance can be interpreted by small difference, it is expected that an appropriate reason exists and therefore it may be able to be used as a technique for starting to make an interpretation at an early timing. The solid line of Figure 7 shows a distribution of the difference between the beginning time of the interpreter's utterance and that of the lecturer's utterance. The average of the difference of the beginning time is 4.1 seconds.

We have extracted the factors that the interpreter can follow the lecturer by picking up 30 alignment pairs of smaller difference of the beginning time. The main features observed in these alignment pairs are as follows:

- Predicting of the next phrase by referring the information on the slide of the lecture(6 pairs)

**lecturer:** *haikēi-desu-ga*  
**interpreter:** This slide shows the background of us.

The lecturer gives a lecture using presentation slides. By seeing this slide, the interpreter can predict and follow the contents which the lecturer speaks.

- Conjunction(11 pairs)

**lecturer:** *shitagatte*  
**interpreter:** Therefore

When a lecturer generates a conjunction, it can be interpreted

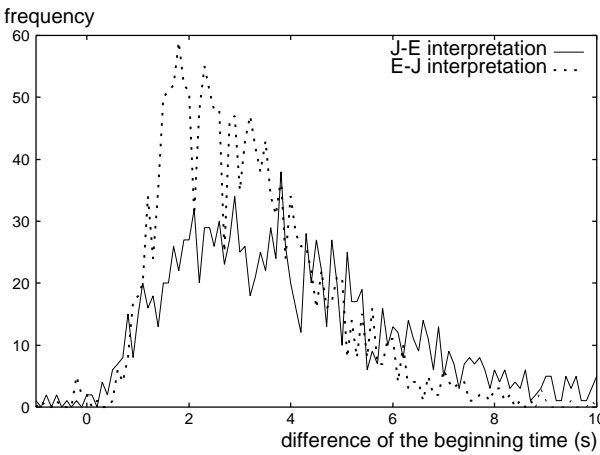


Figure 7: Distribution of difference between the beginning time of the interpreter's utterance and that of the lecturer's utterance

without waiting for the next utterance.

- Interpolating a subject(4 pairs)

**lecturer:** *bunrui-shi-ta-tokoro*

**interpreter:** we analyzed such patterns

A Japanese lecturer may omit a subject. At this time, an interpreter chooses the general subject "we" or "I" from the context, and generates it at an early stage.

- Interpreting a subject(3 pairs)

**lecturer:** *yuza-ga*

**interpreter:** the user,

A subject appears at the beginning of a sentence in Japanese and in English. So the subject can be interpreted immediately.

- Insertion of a filler(1 pair)

**lecturer:** *ijo-de owari-masu*

**interpreter:** Ah this completes my lecture.

While waiting for the next lecturer's utterance, the interpreter speaks a filler.

### 4.3 English-Japanese Simultaneous Interpretation

16 sets of the aligned English-Japanese simultaneous interpreting data are used for the investigation. Figure 6 shows the distribution of the number of morphemes in lecturer's utterances by the dotted line.

We have extracted the alignment pairs in which a lecturer's utterance consists of four or less morphemes. There exist 170 pairs equivalent to 9.90% of the whole. As a result, the same features of a Japanese-English interpretation were observed as follows:

- Lecturer's utterances consisting of a conjunction.(36pairs)

**lecturer:** For example,

**interpreter:** *tatoeba*

- Lecturer's utterances consisting of a subject.(32pairs)

**lecturer:** some of the machines

**interpreter:** *sono uchi-no kikai-no ikutsuka-ga*

- Lecturer's utterances of short clauses.(48pairs)

**lecturer:** Someone once said that

**interpreter:** *aru hito-ga ii-mashi-ta*

Next, we have examined the difference between the beginning time of the interpreter's utterance and that of lecturer's utterances, The dotted line of Figure 7 shows a distribution of the difference between the beginning time of the interpreter's utterance and of the lecturer's utterance. The average of difference of the beginning time is 3.2 seconds.

The following factors can be observed from the investigation of 30 alignment pairs with small difference of the beginning time:

- Interpreting a subject(7pairs)

**lecturer:** I arrived in Japan

**interpreter:** *e watashi nihon-ni tochakushi-mashi-ta-no-ga*

A subject appears at the beginning of a sentence in Japanese and in English. So the subject can be interpreted immediately.

- Insertion of a filler(4pairs)

**lecturer:** these are supporting countries.

**interpreter:** *aa koyu-yona kuni-ga shien-wo shi-te-iru wake-de ari-masu*

While waiting for the next lecturer's utterance, the interpreter speaks a filler.

- Conjunction(4pairs)

**lecturer:** but after walking up and down

**interpreter:** *e shikashi ee michi-wo aruite iru uchi-ni*

When a lecturer generates a conjunction, it can be interpreted without waiting for the next utterance.

- An adverbial phrase and a prepositional phrase(2pairs)

**lecturer:** Here in Kansai

**interpreter:** *mm kansai-dewa*

The word order of Japanese is flexible in comparison with that of English. If the adverbial phrase and prepositional phrase appear at the beginning of the sentence, the phrase can be interpreted immediately.

#### 4.4 Comparison

The delay of the English-Japanese interpretations is smaller than that of the Japanese-English interpretation about 1.0 seconds. The following reasons can be considered as the factors: (1)The high simultaneity is attained by interpreting according to the order of an appearance of an English word because the word order of Japanese is flexible in comparison with that of English. (2)It is possible to predict the next utterance by an interpreter's background knowledge because the topic of an English lecture was common. (3)Since a verb appears at the end of sentences in Japanese, the sentence structure of the interpretation in a Japanese-English at an early stage.

#### 5 Conclusion

This paper has described an investigation of simultaneous interpreting corpus. The results of the investigation are as follows:

1. When a lecturer generates a conjunction, it can be interpreted immediately without waiting for the next lecturer's utterance.
2. Since a subject appears at the beginning of a sentence in both Japanese and English, the subject can be interpreted immediately.
3. While waiting for the next lecturer's utterance, the interpreter speaks a filler.
4. By controlling the speaking speed based on the quantity of the input utterance, the interpreter reduces the difference between the beginning time of the interpreter's utterance and, that of the lecturer's utterance.

These results will be available for the development of a simultaneous interpreting system.

**Acknowledgement:** The collection and transcription of the speech 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 Grand-in-Aid for COE Research of the Ministry of Education, Science, Sports and Culture, Japan.

#### References

- Y. Aizawa, S. Matsubara, N. Kawaguchi, K. Toyama, Y. Inagaki, "Spoken Language Corpus for Machine Interpretation Research", *Proceedings of ICSLP-2000*, Vol. III, pp. 398-401, (2000).
- J. Amtrup, "Incremental Speech Translation", *Lecture Notes in Artificial Intelligence*, 1735 (1999).
- N. Kawaguchi, S. Matsubara, K. Takeda, F. Itakura, "Construction of Speech Corpus in Moving Car Environment", *Proceedings of 7th European Conference on Speech Communication and Technology(Eurospeech-2001)*, pp. 2027-2030 (2001).
- K. Maekawa, T. Kagomiya, H. Koiso, H. Ogura, H. Kikuchi, "Design of the Corpus of Spontaneous Japanese", *Journal of the Phonetic Society of Japan*, 4-2, pp. 51-61 (2000). (in Japanese)
- S. Matsubara, K. Toyama, Y. Inagaki, "Sync/Trans: Simultaneous Machine Interpretation between English and Japanese.", In

- N. Foo (Ed) *Advanced Topics in Artificial Intelligence, Lecture Note in Artificial Intelligence*, Vol. 1747, pp. 134-143 (1999).
- S. Matsubara, Y. Aizawa, N. Kawaguchi, K. Toyama, Y. Inagaki, "Design and Construction of Simultaneous Interpreting Corpus", *Journal of the Japan Association for Interpretation Studies*, No. 1, pp. 85-102 (2001). (in Japanese)
- Y. Matsumoto, et al. "Morphological Analysis System ChaSen version 2.2.1 Manual", <http://chasen.aist-nara.ac.jp/>
- H. Mima, H.Iida, O. Furuse, "Simultaneous Interpretation Utilizing Example-based Incremental Transfer", *Proceedings of COLING-ACL '98*, pp. 855-861 (1998).
- A. Mizuno, "On Simultaneous Interpretation from Japanese into English", In *Journal of the Interpreting Research Association of Japan*, Vol. 5, No. 2, pp. 4-21 (1995). (in Japanese)