Sync/Trans: Simultaneous Machine Interpretation between English and Japanese

Shigeki Matsubara†, Katsuhiko Toyama‡§ and Yasuyoshi Inagaki‡

†Faculty of Language and Culture, Nagoya University
‡Department of Computational Science and Engineeering, Nagoya University
§Center for Integrated Acoustic Information Research, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
matubara@lang.nagoya-u.ac.jp

Abstract. This paper describes Sync/Trans, an incremental spoken language translation system. The system has been being developed for efficiently translating a spontaneous speech dialogue between an English speaker and a Japanese speaker. Its purpose being to behave as a simultaneous interpreter, the system produces the target output synchronously with the source input. Sync/Trans has the following features: (1) the system consists of modules that work in a synchronous fashion, (2) the system translates the source language possibly word-by-word according to the appearance order, (3) the system utilizes grammatically ill-formed expressions for the speech output, and (4) the system corrects the grammatical ill-formedness of the speech input at a pretty early stage. An experimental system for translating English speech into Japanese speech has been implemented. A few experimental results have shown Sync/Trans to be a promising system for simultaneous interpretation.

1 Introduction

Immediate speech comprehension and production are essential to a smooth interaction in a spoken dialogue. A spontaneously spoken dialogue through interpretation systems thus demands that they should also participate in the dialogue without preventing the coherence.

Our intuitions suggest that efficient speech dialogue translation strongly requires a simultaneous interpretation which is one of the ambitious applications in artificial intelligence [10]. As an example, let us consider a dialogue between an English speaker and a Japanese speaker through an English-Japanese interpretation system and a Japanese-English interpretation system. Figure 1 shows a comparison with the dialogue using a conventional machine translation system. Since the conventional system cannot start the translation processing until the input of an entire Japanese/English sentence finishes, the waiting time [1] of the

[1] This means the time from the end of an utterance by the English/Japanese speaker to the start of the utterance by the Japanese-English/English-Japanese interpretation system.

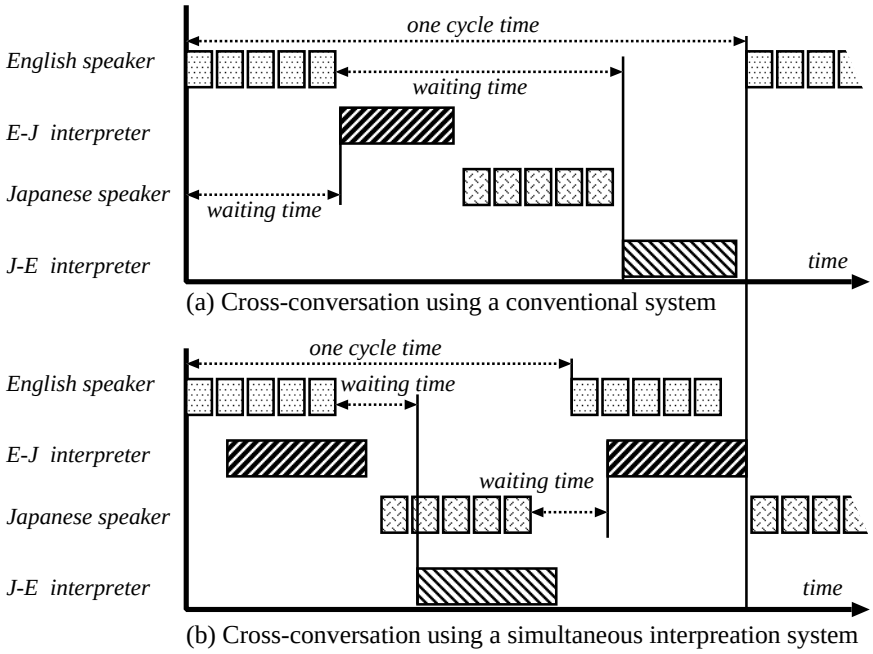

Fig. 1. Comparison between a dialogue through conventional systems and a dialogue through simultaneous interpretation systems

English/Japanese speaker is long. Therefore, the time required for one cycle [2] necessarily becomes long. On the other hand, since the simultaneous interpretation system can start translating it right after the start of the input, the waiting time is reduced. As a result, the time of one cycle also becomes shorter.

The following enumerates several problems that should be solved in the development of a simultaneous interpretation system.

- **Architecture** A machine translation system is usually composed of modules such as parsing, transfer and generation, which work sequentially in a compositional way [11]. That is to say, each module cannot start the processing until the previous module finishes processing an entire sentence. It is practically impossible for such the system to behave as a simultaneous interpreter.
- **Incrementality** Current most techniques for parsing, transfer and generation process a natural language on a sentence-by-sentence basis. Each module should be able to build up partial representations for incomplete

[2] This means the time from the start of an utterance to the start of the next utterance by the English/Japanese speaker.

input [4].

- **Difference in word-order** To a greater or less degree, the word-order of a source language is different from that of the target language. The difference might cause the system to have a loss to the simultaneity of the output with the input. Most of simultaneous interpretation systems which have been proposed so far (e.g., [6, 7, 1, 2]), have not solved this problem.
- **Grammatical Ill-formedness** Grammatically ill-formed expressions appear very frequently in spontaneous speech. It is necessary to investigate a method of incrementally translating the ill-formed input.

This paper describes Sync/Trans, a system for incremental spoken language translation between English and Japanese. The aim of this study is to pursue the possibility of a simultaneous interpretation and to realize a speech dialogue on it as Figure 1(b) shows. The characteristic features of Sync/Trans corresponding to the above four problems respectively are as follows:

- *Synchronous performance:* The system is composed of mainly two modules: parsing and transfer, which work almost synchronously with the input. In other words, each module starts the processing right after the start of the processing in the previous module.
- *Incremental speech translation:* The system makes a translation result of the spoken source language on a possibly word-by-word basis according to the appearance order [3]. Each module can predict the precessing result halfway through the input. The next module performs the processing using the result at any time.
- *Utilization of grammatically ill-formed expressions:* The system utilizes grammatically ill-formed expressions characterizing spoken language for the translation results [8, 9]. This is a key to the success of translating between English and Japanese, which are different in word-order, in an exceedingly incremental way.
- *Correction of grammatical ill-formedness at an early stage:* Finding out grammatical ill-formedness in an input sentence, the system corrects it immediately. This enables the system to robustly proceed with the translation processing.

This paper reports on the evaluation of Sync/Trans through a few translation experiments. However, we will concentrate on the evaluation of the English-to-Japanese translation part of Sync/Trans, because the current implementation is restricted on the translation of English into Japanese.

2 English-Japanese Translation

2.1 Overview

We have developed a system for translating English speech into Japanese speech incrementally. Figure 2 shows the configuration of the system. The system is composed of eight components: speech recognition, speech synthesis, incremental parsing, incremental transfer, parsing rules, transfer rules, a lexicon

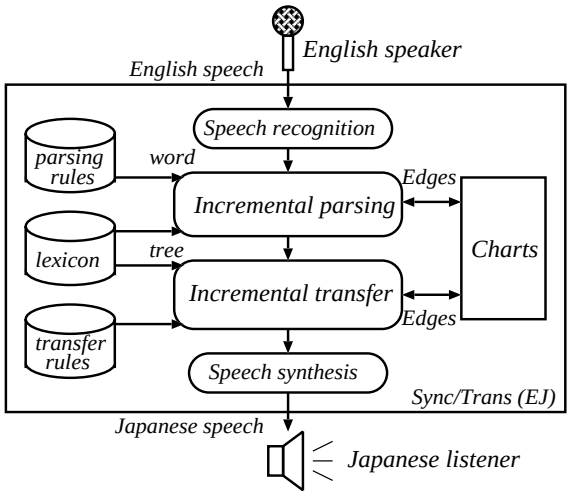

Fig. 2. Configuration of the English-Japanese part of Sync/Trans

and a chart. The chart component is the data structure which represents the possible phrase structures of the source language halfway through the input. Executing incremental parsing and transfer sequentially for each word, as a consequence, the system can translate a spoken English sentence synchronously with the appearance.

2.2 Production of Grammatically Ill-formed Sentences

The expressions such as repetitions, inversions, ellipses, repairs and hesitations are grammatically ill-formed but natural in Japanese daily conversations. In order to incrementally and synchronously translate between English and Japanese which are different in word-order, the system utilizes these expressions in an effective way.

For example, although the standard Japanese translation of an simple English sentence (2.1) is (2.2), the system generates (2.3) synchronously with the input of (2.1).

(2.1) Ken met her in the park yesterday.

(2.2) *ken-wa* (Ken) *kinoo* (yesterday) *koen-de* (in the park) *kanojo-ni* (her) *atta* (met).

(2.3) *ken-wa* (Ken) *atta* (met) *kanojo-no, anoo, kanojo-ni* (her) *koen-de* (in the park) *kinoo* (yesterday) *atta* (met).

We can say that Japanese people can understand (2.3) easily in spite of the grammatical ill-formedness.

Input (2.1)	Output (2.2)	Output (2.3)
Ken	*ken*	*ken*
met	*wa*	*wa atta*
her		*kanojo-no*
in		*anoo*
the		*kanojo-ni*
park		*koen-de*
yesterday	*kinoo*	*kinoo*
	koen-de	*atta*
	kanojo-ni	
	atta	

Fig. 3. Timing of the output of (2.2) and (2.3)

Figure 3 shows the timing of the output of (2.2) and (2.3). It is obvious that the system can output (2.3) synchronously with (2.1).

2.3 Chart-based Framework

To represent incomplete structures gained on incremental parsing, we have adopted a chart-based framework in both parsing and transfer [9].

Incremental chart parsing produces edges labeled the term whose category is a sentence. The edges represent the possible structures of the entire source language inputed up to the point of time. The altering point for the orthodox bottom-up chart parsing method [5] is that the operations of applying a parsing rule to an active edge and replacing the leftmost undecided term in an active edge with the term labeling another active edge are introduced. On the other hands, the incremental transfer produces the Japanese expressions by applying transfer rules to an edge in a top-down fashion. How the system utilizes grammatically ill-formed expressions is described as the transfer rules.

2.4 Translation of Grammatically Ill-formed Sentences

The system can translate grammatically ill-formed source sentences incrementally by correcting the error immediately after the parsing fails. Correctly constructing the structure for the well-formed part in the ill-formed sentence, the system can reproduce it as the translation result to some extent.

For example, an English sentence (2.4) can be considered as one in which a word "going" of the part "going by train" is omitted.

(2.4) I think by train is best.

The system inserts a category, e.g. gerund, immediately after"by" is inputed. The Japanese people can understand correctly the semantic contents of (2.5) which is the translation result of (2.4).

Table 1. Translation result of 278 sentences

type	sentences	rate(%)
A) correct (no repair)	96	34.5
B) correct (repairs)	132	47.5
C) unnatural	33	11.9
D) incorrect	16	5.7
E) failed	1	0.4

(2.5) *watashi-wa* (I) *omoimasu* (think) *densya-de-ga* (by train) *ichiban-ii* (is best) *to-omoimasu* (think).

3 Evaluation

An experimental system has been implemented in GNU Common Lisp 2.2 on a workstation. We have made a few experiments with the dialogues in ATR Dialogue Databse, whose task is the application of travels.

3.1 Basic Experiment

To evaluate the effectiveness of the system, to begin with, we have made a translation experiment using 4 dialogues. The dialogues consist of 278 spoken English sentences, the average length of which is 6.8 words. The system has been implemented in the scale of 476 English words and 204 grammar rules. In order to enhance only the real-time processing of the system, the English input was restricted to grammatically well-formed sentences. To satisfy this requirement, we have excluded extra-grammatical phenomena such as hesitations and errors from the source sentences in advance.

The success rate was examined. As Table 1 shows, we have classified the source sentences according to the translation results. 228 sentences classified into (A) or (B) are translated correctly, providing a success rate of 82.0%. The result shows the system to be available for spoken language translation. Although the successful Japanese sentences gained on the system are different from those on a conventional system in the sense that they include much ill-formedness, they represent the semantic contents of the source sentences correctly.

3.2 Translation Processing Unit

Many causes of the translation failure in the above experiment is that many repairs appear too frequently in the translation results (33 sentences classified into (C), accounting for 11.9%). In particular, the longer the source sentence becomes, the more repairs appear in the translation. In order to solve the problem, we can consider to relax the restriction of the word-by-word basis.

Table 2. Translation results on a delay of one word

type	sentences	rate (%)
A) correct (no repair)	122	43.9
B) correct (repairs)	113	40.6
C) unnatural	27	9.7
D) incorrect	15	5.4
E) failed	1	0.4

Table 3. Translation results of 278 sentences without error correcting

type	sentences	rate (%)
A) correct (no repair)	21	7.6
B) correct (repairs)	106	38.1
C) unnatural	9	3.2
D) incorrect	45	16.2
E) failed	97	34.9

We have tried to make an experiment on a system translating with a delay of one word. We have used the same 4 dialogues. Table 2 shows the success rate. 235 sentences (accounting for 84.5%) are translated correctly. Although the frequency of the repairs averaged 1.07 times a sentence in the above experiment, the one word delay reduces the frequency to 0.74 times. In general, there exists a trade-off between the translation unit and the translation accuracy. This result shows that it is important to pursue an effective translation unit for simultaneous interpretation.

3.3 Translation of Grammatically Ill-formed Sentences

In Section 3.1, we have made the experiment on the assumption that all input is well-formed. However, such an assumption is not realistic for spontaneously spoken language translation.

We have made an experiment with the same 4 dialogues consisting well-formed 181 sentences and ill-formed 97 sentences on the system having 391 English words and 94 grammar rules. Figure 3 and 4 show the results of the translation without error correcting and with error correcting. As a result of introducing error correcting, 27 sentences providing a rate of 9.7% are newly added to the correct translation results. This result shows the error correcting method to be effective for the spontaneous speech translation.

Table 4. Translation results of 278 sentences with error correcting

type	sentences	rate (%)
A) correct (no repair)	21	7.6
B) correct (repairs)	133	47.8
C) unnatural	15	5.4
D) incorrect	100	36.0
E) failed	9	3.2

4 Japanese-English Translation

This section describes the Japanese-English translation part of Sync/Trans briefly. There exist some difficulties which should be overcome in developing a system of incremental Japanese-English translation. One of the difficulties is as follows:

- In spite that the system should output the English verb at an early stage, the verb usually appears in the end of a Japanese sentence.

As an example, let us consider translating the following Japanese sentence into English incrementally.

(4.1) *kinoo* (yesterday) *ken-wa* (ken) *mado-wo* (the window) *hanma-de* (with an hammer) *watta* (broke).

The standard translation of (4.1) is (4.2).

(4.2) Ken broke the window with a hammer yesterday.

It is impossible in principle to output the English verb phrase "broke the window with a hammer" before the Japanese verb "*watta*" appears.

To overcome the difficulty, we are investigating to predict a English verb at an early stage. In the dialogue task restricted to some extent, it might be possible to predict the appearing verb from the other noun phrases in the sentence. If the verb "broke" can be predicted from "Ken" and "the window", the system can product an English sentence (4.3) synchronously with (4.1).

(4.3) Yesterday, Ken broke the window with a hammer.

Figure 4 shows a comparison between the timing of the output of (4.2) and (4.3).

5 Concluding Remarks

This paper has described Sync/Trans, an speech-to-speech translation system which we have been studying. Sync/Trans has the features: synchronous architecture, incremental translation, utilization of grammatical ill-formedness and early error correcting.

Input (4.1)	Output (4.2)	Output (4.3)
kinoo		yesterday
ken-wa	Ken	Ken
mado-wo		broke
hanma-de		the window
watta	broke	with a hammer
	the window	
	with a hammer	
	yesterday	

Fig. 4. Timing of the output of (4.2) and (4.3)

This paper has provided a few experimental results on the English-Japanese part which we have implemented as a first step towards simultaneous interpretation. From the results, we have found that incrementally translating the source language in an appropriate unit and correcting the errors in the source sentence at an early stage, are effective for speech-to-speech translation. In the near future, as soon as the Japanese-English translation part is implemented, we are planning to evaluate Sync/Trans on a speech dialogue between an English speaker and a Japanese speaker. We might be able to confirm the effectiveness of Sync/Trans as a spontaneous dialogue interpreter through the evaluation. To this end, high-accuracy speech recognition and real-time language processing are essential.

Acknowledgements. The authors would like to thank Satoru Asai for his valuable contributions on an earlier version of this study. The authors are also deeply indebted to Yoshihide Kato and Yoshiyuki Watanabe for their considerable assistance with the experiments. This research has been supported in part by The Hori Information Science Promotion Foundation.

References

1. Amtrup J.W.: Chart-based Incremental Transfer in Machine Translation, *Proceedings of the 6th International Conference of Theoretical and Methodological Issues in Machine Translation*, pp. 188-195 (1995).
2. Furuse, O. and Iida, H.: Incremental Translation Utilizing Constituent Boundary Patterns, *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 412–417 (1996).
3. Iida, H., and Sumita, E. and Furuse, O.: Spoken-Language Translation Method Using Examples, *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 1074–1077 (1996).
4. Inagaki Y. and Matsubara S.: Models for Incremental Interpretation of Natural Language, *Proceedings of the 2nd Symposium on Natural Language Processing*, pp. 51–60 (1995).

5. Kay M.: Algorithm Schemata and Data Structures in Syntactic Processing, *Technical Report CSL-80-12*, Xerox PARC (1980).
6. Kitano, H.: Incremental Sentence Production with a Parallel Marker-Passing Algorithm, *Proceedings of the 13th International Conference on Computational Linguistics*, pp. 217–222 (1990).
7. Kitano, H.: *Speech-to-Speech Translation: A Massively Parallel Memory-based Approach*, Kluwer Academic Publishers (1993).
8. Matsubara S. and Inagaki Y.: Utilizing Extra-Grammatical Phenomena in Incremental English-Japanese Machine Translation, *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation*, pp. 138–145 (1997).
9. Matsubara S., Asai S., Toyama K. and Inagaki Y.: Chart-based Parsing and Transfer in Incremental Spoken Language Translation, *Proceedings of the 4th Natural Language Processing Pacific Rim Symposium*, pp. 521–524 (1997).
10. Menzel, W.: Parsing of Spoken Language under Time Constraints, *Proceedings of the 11th European Conference of Artificial Intellingence*, pp. 560–564 (1994).
11. Landsbergen J. and Jong F.: *Compositional Translation*, Rosseta M. (Ed.), pp. 57–84 (1994).