Simultaneous Machine Interpretation using Finite State Transducer

Koichiro Ryu[1] Shigeki Matsubara[2] Yasuyoshi Inagaki[3]

[1,2]Information Technology Center, Nagoya University
[3]Faculty of Management and Information Science, Aichi Institute of Technology
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan
Tel. +81-52-789-1532 Fax. +81-52-789-4384
[1]ryu@el.itc.nagoya-u.ac.jp [2]matubara@itc.nagoya-u.ac.jp

Abstract

In this paper, we present a method for simultaneous machine interpretation. It is based on the pattern-based machine translation and uses translation patterns acquired from a parallel corpus. By using word-for-word translations as data, our method makes it possible to generate translations simultaneously. To translate an input utterance at high speed, our method converts translation patterns into a finite state transducer (FST) on ahead and uses the FST to translate it. In order to confirm the effectiveness of our method, we have experimented using spoken dialogue sentences.

1 Introduction

To support cross-language communication, several speech-to-speech translation systems have been developed so far (Vrranz, Comelles and Farwell, 2005; Shimizu, Ashikari, Sumita, Kashioka and Nakamura, 2006). The previous systems, because of their sentence-by-sentence translation, can not start to translate an input utterance before it is completed. The following problems may arise by using such systems in cross-language communication:

- The conversation time becomes long since it takes much time to translate an input
- The listener has to wait until the system begins to translate an input, and therefore the waiting time becomes long

One of the effective methods for solving these problems is that a translation system begins to translate an input utterance without waiting for the end of the speaker's utterance like a simultaneous interpreter. This has been verified as possible by a study on comparing simultaneous interpretation with consecutive interpretation from the viewpoint of efficiency and smoothness of cross-language communication (Ohara, Matsubara, Ryu, Kawaguchi and Inagaki, 2003).

So far, some studies have presented methods to realize simultaneous interpretation systems (Ryu, Matsubara and Inagaki, 2006; Mima, Iida, Furuse, 1998). Some of these methods are transfer-based machine translation and have been evaluated on its realization by experiments using prototype systems. However, such methods have issues that maintenance is a complex task and customization is expensive and time-consuming.

Corpus-based machine translation is one of the methods to solve the above problems. Recently corpus-based machine translation has been addressed, its usefulness has been confirmed by some studies. Moreover, spoken language parallel corpora have been constructed for the purpose of studying on speech processing and speech-to-speech translation (Takezawa, Sumita, Sugaya, Yamamoto and Yamamoto, 2002; Tohyama, Matsusbara, Kawaguchi and Inagaki, 2005).

However, compared with the previous studies on corpus-based machine translation, there exist some new problems for simultaneous machine interpretation. The previous studies on corpus-based machine translation have proposed a method for improving the quality of translations. On the other hand, simultaneous machine interpretation has the restriction on speaking time (when-to-say) and how to translate an input utterance (how-to-say). Our method addresses these tasks.

In this paper, we present a method for simultaneous machine interpretation. It is based on the pattern-based machine translation, and uses translation patterns acquired from a parallel corpus. By using word-for-word translations as data, our method makes it possible to generate translations simultaneously. To translate an input utterance at high speed, our method converts translation patterns into a finite state transducer (FST) on ahead and uses it to translate an input utterance. In order to confirm our method to be effective, we have experimented using spoken dialogue sentences.

This paper is organized as follows: Section 2 discusses simultaneous machine interpretation using pattern-based approach. Section 3 introduces our method for simultaneous machine interpretation, and Section 4 reports the experimental results.

2 Pattern-based Simultaneous Machine Interpretation

Machine translation is one of the most important research topics in natural language processing. However, the longstanding major issues are how to acquire knowledge for translation, such as grammar rules and lexical translation dictionaries. Corpus-based machine translation is considered as an effective method to solve the above problem. Statistical machine translation, example-based machine translation and pattern-based machine translation are typical corpus-based approaches.

2.1 Pattern-based Approach

A pattern-based machine translation system acquires translation patterns from parallel corpora and translates an input utterance by combining their translation patterns. A translation pattern consists of a source pattern and a target pattern. The system translates an input utterance using translation patterns whose source pattern matches an input utterance. For example, we consider to acquire translation patterns from the source sentence, "I'd like to reserve a room at a service counter" and its Japanese translation "Sabisu-kaunta-de heya-wo yoyaku-shi-tai-no-desu-ga". We can acquire the following translation patterns:

S : I'd like to reserve NP1 at NP2 → NP2 de NP1 wo yoyaku-shi-tai-no-desu-ga (1)
NP : a room → heya (2)
NP : a N1 counter → N1 kaunta (3)
NN : service → sabisu (4)

In translation pattern (1), the source pattern is "I'd like to reserve NP1 at NP2" and the target pattern is "NP2 de NP1 wo yoyaku-shi-tai-no-desu-ga". When a system translates the input utterance, "I'd like to reserve a seat at a ticket counter", it uses the above translation patterns (1) , (3) and the following translation patterns (5), (6) to translate it.

NP : a seat → seki (5)
NN : ticket → chiketto (6)

Then the system can generate the Japanese translation, "Chiketto-kaunta-de seki-wo yoyaku-shitai-no-desu-ga".

2.2 Problems

To realize simultaneous machine interpretation using pattern-based approach, we should expand how to make translation patterns and how to use translation patterns. Acquiring translation patterns from a parallel corpus, our method has the following features:

1. **Parallel corpus**
   In simultaneous machine interpretation, it is important to translate an input utterance simultaneously. However, it is generally difficult to translate an English utterance into its Japanese translation simultaneously, because of the difference of word-order between English and Japanese. If a system can train techniques utilized by professional simultaneous interpreters, the system translates an English utterance into its Japanese translation simultaneously . Our method utilizes the techniques using a parallel corpus consisted source utterances and these word-by-word translations.

2. **translation timing**
   In simultaneous machine interpretation, it is necessary to decide the output timings of translation words. Our method attaches a translation pattern with the output timing of each translation word in its target pattern. Output timings are decided by the correspondences between the source pattern and the target pattern.

Utilizing translation patterns acquired from a parallel corpus, our method has the following features:

Table 1: Examples of word-by-word translation

S1 : I would like to reserve a room with bath. T1 : watashi-wa furotsuki-no heya-wo yoyaku-shi-tai-no-desu-ga. (I) (with bath) (room) (would like to reserve)
S2 : Your room number is 1001. T2 : Okyakusama-no heya-no bango-wa 1001 desu. (Your) (room) (number) (1001) (is).
S3 : I want to reserve a single room on May eighteenth T3 : Watashi-wa shinguru-no heya-wo gogatsu juhachinichi-ni yoyaku-shi-tai-no-desu-ga. (I) (single) (room) (May) (eighteenth) (want to reserve).

1. **Incremental transfer**
   The conventional pattern-based machine translation begins to translate an input utterance after the utterance is completed. On the other hand, it is necessary to translate an input utterance before the utterance is completed. To translate an input utterance simultaneously, our method has a function of incrementally applying translation patterns to the input.

2. **Time processing**
   In simultaneous machine interpretation, it is important to translate an input utterance at high speed. In usual pattern-based machine translation, the more the number of translation patterns increases, the longer the processing time is. Our method solves this problem using a finite state transducer (FST) converted from translation patterns.

3 Simultanous Interpretation using Finite State Transducer

In this section, we describe our method for simultaneous machine interpretation using an FST converted from translation patterns.

3.1 Word-for-Word Translation

In our method, we make translation patterns from a parallel corpus consisted source utterances and their word-by-word translations. The style of word-by-word translation is that the word order of translations becomes as similar word order of its source utterances as possible. We show the examples of word-by-word translation in Table 1. As word-by-word translations are created by professional translators, we can utilize the techniques of professional translators by using them.

3.2 Parallel Syntactic Structure

Our method needs a parallel corpus annotated with syntactic structures and alignments. We show the example of word-by-word translation with syntactic structures in Figure 1. The left side of Figure 1 shows the syntactic structure of the English utterance “I would like to reserve a room with bath on next Sunday.”, and the right side shows that of its Japanese translation “watashi-wa furotsuki-no heya-wo tsugi-no nichiyobi-ni yoyaku-shi-tai-no-desu-ga.” To express syntactic structures there are nine kinds of syntactic categories that are shown in Table 2. Each node in a parse tree is annotated with a label (e.g. NN1) indicated a syntactic category (e.g. NN) and ID (e.g. 1). Each label expresses the correspondences between a source utterance and its translation. For example, label “NN1” in Figure 1 indicates that the English word “I” is correspondent to the Japanese word “watashi” and label NP1 in Figure 1 indicates that the English phrase “a room with bath” is correspondent to the Japanese phrase “furotsuki-no heya”.

3.3 Acquisition of Translation Pattern

We describe how to acquire translation patterns from a parallel corpus with syntactic structures. One translation pattern is acquired from one node in an English parse tree. For example, from the node attached syntactic label “NP1” in Figure 1, we acquire its child nodes, “a NN2 ADJP1”, as the source pattern of a translation pattern. The child nodes of the node attached with “NP1” label in a Japanese parse tree, “ADJP1 NN2”, is acquired as the target pattern of the translation pattern. Then

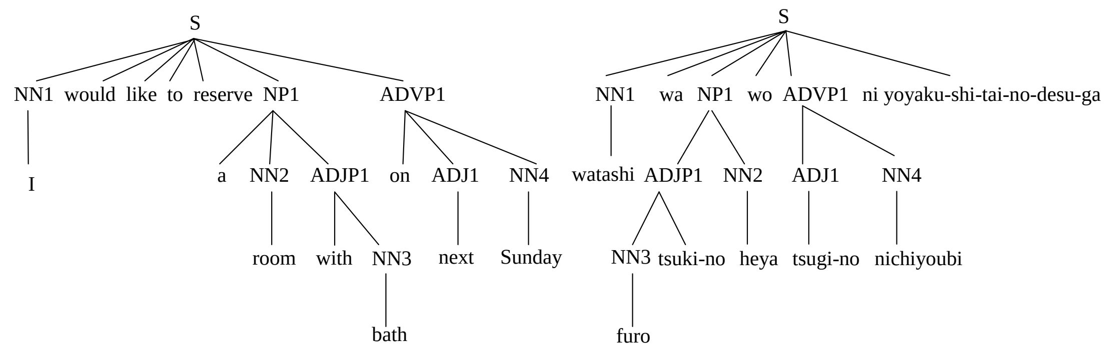

Figure 1: Example of a source sentence and its translation with word alignments

Table 2: Syntactic labels

Unit	Explanation	Kinds of translation pattern
Word	Consisting of one word	Noun (NN), Adjective (ADJ), Adverb (ADV)
Phrase	Consisting of two or more words and not including a predicate	Noun phrases (NP), Adjective phrase (ADJP), Adverb phrase (ADVP)
Clause	Consisting of two or more words and including a predicate	Noun clause (NC), Adjective clause (ADJC) Adverb clause (ADVC)

we can acquire translation pattern "NP: a NN2 ADJP1 → ADJP1 NN2" from syntactic label NP1. Figure 2 shows the example of translation patterns acquired from the parallel parse tree in Figure 1.

3.4 Decision of Output Timing

We explain how to decide the output timing of components consisted of translation words or syntactic labels in a target pattern. Our method decides output timing by the correspondences between a source pattern and its target pattern. Our method assumes that a syntactic label in a target pattern can be outputted after the corresponding component in its source pattern is input. Under this assumption, we can acquire translation patterns with output timings in Figure 3 from translation patterns in Figure 2. In "NP" category of Figure 3, the components "ADJP1" and "NN2" can be outputted after the corresponding component "ADJP1" is input.

3.5 Conversion from Translation Patterns into FST

Translation patterns of each category with output timing are converted into the network of each category. Figure 4 shows the networks converted from translation patterns in Figure 3. One transition of each network corresponds a pair of a source component and a target component. Each path from the initial state to the final state corresponds one translation pattern.

We show how to convert networks into an FST in Figure 5. First, the network of S category is as an initial FST. Next, the transitions of the initial FST whose input label include a syntactic category are substituted by the network of the syntactic category. For example, the transition attached NN1 is substituted by the NN network in Figure 4. We substitute transitions attached syntactic labels in an FST recursively.

3.6 Translation Method

We describe a method to simultaneously translate an input utterance using an FST. In an FST used by our method every path from the initial state to the final state represents a mapping of source words to target words. We input source words into an FST word by word. It repeatedly transits from a state to other state by selecting a transition that is in agreement with an input word. As the FST is nondeterministic, one transition is randomly selected when there are two or more transitions that is agreement with an input word.

4 Experiment

4.1 Overview

To evaluate our method, we conducted a translation experiment. A source utterance is input word by word as text, the translation toward

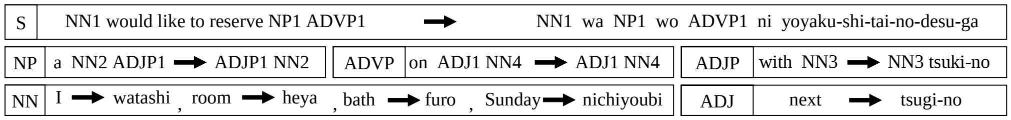

Figure 2: Translation patterns

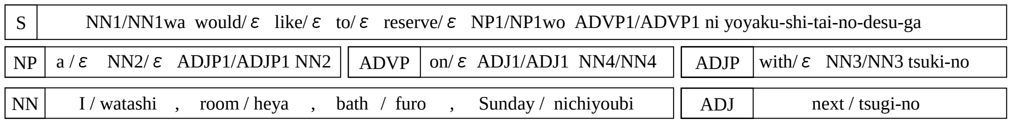

Figure 3: Translation patterns with output timing

an input utterance is target words that are outputted by an FST. Training data for acquiring translation patterns is also used as test data, so the experiment is closed. The data is 52 utterances and its word-for-word translations in CIAIR simultaneous interpreting corpus (Tohyama, Matsusbara, Kawaguchi and Inagaki, 2005). Table 3 show the statistical data of the source utterances. Syntactic structures are annotated on each utterance and its translation. The substitutions of transitions cause a problem that exists a huge number of transitions not to be used for translation. To solve the problem we divided noun phrase categories into the six (numeral, price, room, date, time and the other). The number of substitutions is four times. Table 4 shows the number of states and transitions. We implemented the system in C language on a 2.0-GHz Pentium4 with 2 GB of RAM. The OS was Linux.

Table 3: Statistics of source utterances

Item	Number
Utterances	52
Words	389
Translation patterns	164
Average number of words	7.48

Table 4: Number of states and transitions

Substitutions	States	Transition relations
0	200	253
1	936	1673
2	9264	22176
3	86179	222981
4	794586	2066456

4.2 Results

Table 5 shows the performance of each substitution times. 47 translations in 53 utterances were correct when the number of substitutions is four times. 17 utterances in source utterances of the correct translations were translated before the input utterances are completed. Moreover, the Table 5 indicates that the more number of substitutions is increased, the more number of correct is increased. The translation time was about 0.01 seconds per sentence. This result indicated that the utterances can be translated at high speed by our method. We examined the translations estimated as wrong translation. According to the examination, the reason of wrong translation was that FST could not transit before the input utterance was completed.

Table 5: Result (No. of Substitutions: 0-4)

Substitutions	Correct (Simultaneous)	Wrong
0	7	45
1	24	28
2	32	20
3	41	11
4	47(17)	5

5 Conclusion

In this paper, we proposed a method for simultaneous translation using an FST. Our method translates an input utterance by transiting from a state to any other state and outputting the translation words labeled to the transition. We introduced how to acquire translation patterns from a parallel corpus and how to convert these translation patterns into an FST. We developed a prototype simultaneous translation system. We evaluated our system with 55 English utterances in CIAIR simultaneous interpreting corpus (Tohyama, Matsusbara, Kawaguchi and Inagaki, 2005). The ex-

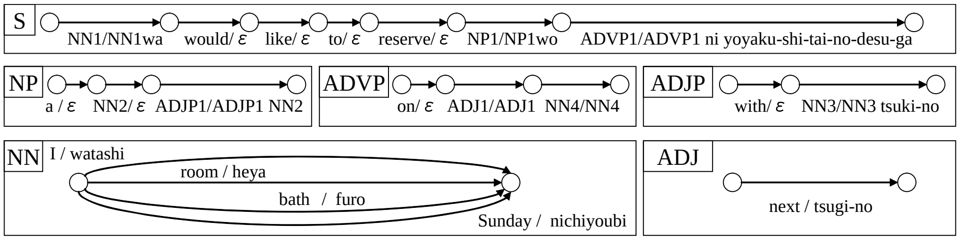

Figure 4: Network of each category

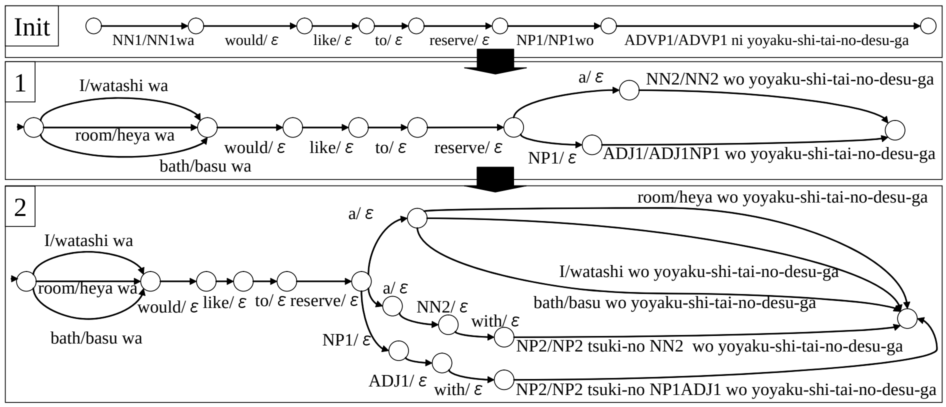

Figure 5: Substitutions of transition

perimental results showed the effectiveness of the proposed method.

The FST which we made from translation patterns has an ambiguity problem. To solve the problem we need to train transition probability from training data. Since our method simply outputs translations labeled to the transition of FST, it might penerate Japanese translations with strange expressions. Therefore we will propose a method of connecting fragments of translations well.

Acknoledgments

The authors would like to thank Prof. Dr. Toshiki Sakabe at Nagoya University for his precious advices.

References

V. Arranz, E. Comelles and D. Farwell. 2005. The FAME Speech-to-Speech Translation System for Catalan, English and Spanish *Proceedings of MT-SUMMIT X*, pages 195-202.

H. Mima, H. Iida and O. Furuse. 1998. Simultaneous Interpretation Utilizing Example-based Incremental Transfer, *Proceedings of COLING/ACL-1998*, pages 855-861.

M. Ohara, S. Matsubara, K. Ryu, N. Kawaguchi and Y. Inagaki, 2003. Temporal Features of Cross-Lingual Communication Mediated by Simultaneous Interpreting: An Analysis of Parallel Translation Corpus in Comparison to Consecutive Interpreting, *The Journal of the Japan Association for Interpretation Studies*, pages 35-53. (In Japanese)

K. Ryu, S. Matsubara and Y. Inagaki. 2006. Simultaneous English-Japanese Spoken Language Translation Based on Incremental Dependency Parsing and Transfer, *Proceedings of COLING/ACL-2006 Main Conference Poster Sessions*, pages. 683-690

T. Shimizu, Y. Ashikari, E. Sumita, H. Kashioka and Satoshi Nakamura, 2006. Development to of Client-Server Speech Translation System on a Multi-lingual Speech Communication Platform, *Proceedings of the IWSLT-2006*, pages 213-216.

T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto and S. Yamamoto, 2002. Toward a Broad-Coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World, *Proceedings of LREC-2002*, Vol. 1, pages 147-152.

H. Tohyama, S. Matsubara, N. Kawaguchi and Y. Inagaki, 2005. Construction and Utilization of Bilingual Speech Corpus for Simultaneous Machine Interpretation Research, *Proceedings of Eurospeech-2005*, pages 1585-1588.