Incremental Spoken Language Translation based on A Normal-Form Conversion of CFG

Shigeki Matsubara , **Hiroshi Ogawa** , **Katsuhiko Toyama** and **Yasuyoshi Inagaki**‡
Faculty of Language and Culture, Nagoya University †
Department of Information Engineering, 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

An incremental translation system for spoken language, which can produce the target language synchronously with a source language input, has been proposed so far. Producing so much source language structures in the CFG-based parsing stage, however, the system has two fatal problems: 1) difficult to work in real-time, and 2) difficult to generate correct results. This paper introduces a special form of CFG, called M normal-form, and describes a method of converting any CFG to M normal-form. Since it produces less structures in the incremental parsing, the system using M normal-from of CFG can generate translation results with higher quality. Experiments with a prototype system have shown the effectiveness of the conversion for fine-grained incremental machine interpreting.

1 Introduction

Immediate speech comprehension and production by a spoken dialogue system are essential to a smooth human-computer interaction. A cross-lingual spontaneous speech conversation through a machine interpreting system demands that the system should also engage in the conversation without preventing its coherence. Our intuitions suggest that the system should behave as an simultaneous interpreter.

In order to develop a simultaneous interpreting system, working synchronously with a spoken language input, a technique for translating natural language incrementally is strongly required (Iida et al., 1996; Menzel, 1997). Towards such an ambitious application, some attempts have been recently made so far. Kitano has proposed a technique of incremental sentence production for modeling simultaneous machine interpretation (Kitano, 1990; Kitano, 1993). As well, Amtrup has introduced chart-based incremental transfer framework for processing head switching in German-English machine translation (Amtrup, 1995). Furuse and Iida has developed a technique of incremental translation utilizing transfer patterns for language constituents (Furuse and Iida, 1996). Furthermore, Matsubara and Inagaki has proposed an incremental English-Japanese translation system (Matsubara and Inagaki, 1997b; Matsubara and Inagaki, 1997a). In these studies, Matsubara and Inagaki has achieved *fine-grained incremental machine interpreting* [1], in spite of the difference in word-order between source and target language.

Generally speaking, the more incrementally machine interpreting system translates, the more smoothly and efficiently the cross-lingual spoken conversation proceeds. We would like to emphasize that an essential key to success of the fine-grained incremental spoken language translation is to allow the system to produce grammatically ill-formed sentences. Above all, utilizing repair expressions for the target language production is indispensable to simultaneous interpretation. Because the case that the system revises its understanding state after it produces the target language for an initial source fragment once is pretty common. However, we have to pay attention to that utilizing repair expressions so frequently brings the rather opposite effect in the sense that the target spoken language becomes unnatural for the users. As a result of an incremental translation experiment, there were a lot of cases that the system corrects itself so frequently that the user could not understand it exactly, in spite that the result is not incorrect (Matsubara et al., 1997a).

This paper considers a technique for reducing repair expressions which a fine-grained incremental translation system produces. A repair usually occurs at the time when the system finds it to be incorrect later though choosing one from some possible source trees before in a deterministic way. This means that the frequency in generation of repair expressions may become lower if the possibilities of

[1] *Fine-Grained Incremental Interpretation* means that the processor analyses each small part of a sentence, such as each word or morpheme, immediately it is encountered. In contrast, *Coarse-Grained Incremental Interpretation* occurs if the processor waits until larger chunks of a sentence are encountered (Charter et al., 1995).

choosing incorrect ones are getting fewer. There are two ways of solving the problem. One is to cut down the opportunities of choosing the only result from the candidates. This is nothing but to make the translation unit larger. The other is to reduce in advance the candidates of the trees extracted from an initial fragment. In these solutions, being against a fine-grained incremental interpreting, an approach of making the processing unit larger cannot be adopted in this study.

A fine-grained incremental translation system, which is based on very orthodox compositional machine translation consisting of incremental parsing, transfer and generation, has been developed so far (Matsubara et al., 1997b). Our interest is to reduce the syntax trees produced by incremental parsing without using except syntactic information. To accomplish this, in this paper, we introduce a normal-form of a set of context free grammar (CFG) rules. It is called M normal-form. The incremental parsing can produce less syntax trees by converting any set of CFG rules into M normal-form in advance. As well, transfer rules conversion in advance enables the incremental transfer processing to choose one from less trees. In the result, the incremental generation can produce a target language including less repair expressions. Experimental results have shown the effectiveness of the M normal-form conversion.

This paper is organized as follows: Section 2 explains the incremental spoken language translation system in brief. Section 3 describes M normal-form conversion of CFG. Section 4 reports on experimental results.

2 Incremental Speech Translation

This section explains an overview of the system for fine-grained incremental English-Japanese machine interpreting. For further details, the reader should refer to (Matsubara and Inagaki, 1997a; Matsubara and Inagaki, 1997b).

The system is composed of three stages: incremental parsing, incremental transfer and incremental generation. These stages work in an orthodox fashion as being shown below, except executing the procedures on a possibly word-by-word basis.

- **(1) Parsing stage:** produces syntax trees for the initial fragment using context free grammar rules whenever a word is inputed.
- **(2) Transfer stage:** brings target expressions by applying transfer rules to one tree whenever syntax trees are revised.
- **(3) Generation stage:** produces the target language whenever the transfer processing is executed.

Executing step-by-step for every additional word, as it turned out, each module comes to work synchronously with speech input. Both incremental parsing and transfer are based on a chart processing technique (Kay, 1980). The chart is used as the data for representing source-structures for initial fragment. The below in Section 2 illustrates incremental parsing.

$s \rightarrow np\ vp$	$det \rightarrow$ the
$np \rightarrow det\ adj\ n$	$adj \rightarrow$ red
$np \rightarrow det\ n$	$n \rightarrow$ table

Figure 1: CFG rules

2.1 Incremental Chart Parsing

The chart has been introduced into representing parsing results. Since the chart can represent a partial sentence structure as an active edge, it is effective to utilize charts for incremental parsing. Incremental chart parsing (Matsubara et al., 1997b) is carried out applying to edges two main rules: bottom-up rule and fundamental rule. The edge formed by a fundamental rule could be used for the transfer processing and the one formed by a bottom-up rule is not. The method is a kind of the so-called mixed mode left-corner chart parsing method based on reachability relations. However, the method differs from the others in the point 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 of another active edge are introduced.

3 A Normal-Form Conversion of CFG

This section provides an informal explanation of M normal-form.

3.1 M Normal-Form

First of all, let us consider incremental chart parsing of the initial fragment "The red table" (Haddock, 1989), in a sentence, for instance, "The red table looks very nice, isn't it." with a simple grammar shown in Figure 1. For the first word "the", analyses (A) $[[[the]_{det}[?]_{n}]_{np}[?]_{vp}]_{s}$ and (B) $[[[the]_{det}[?]_{adj}[?]_{n}]_{np}[?]_{vp}]_{s}$ are realized. It should be noted that two phrase-structure trees are produced in spite that "the" is not ambiguous syntactically. This means no more than that there exist two possibilities with regard to the categories of the coming words. One is "n" and the other is "adj, n". When the next word comes, it will be decided. To a greater or less extent, such phenomena happen to common left-corner parsing utilizing accessibility relation. What needs to be emphasized is, however, that more trees are produced in incremental parsing because the system have to estimate the more coming words.

$$
\begin{array}{ll}
s \rightarrow np\ vp & det \rightarrow \text{the} \\
np \rightarrow det\ new & adj \rightarrow \text{red} \\
new \rightarrow adj\ n & n \rightarrow \text{table} \\
new \rightarrow n &
\end{array}
$$

Figure 2: Converted CFG rules

Let us thus consider making a certain tree represent both (A) and (B), for instance, a tree (C) $[[[the]_{det}[?]_{new}]_{np}[?]_{vp}]_s$ where a category "*new*" can be rewrote to both "*n*" and "*adj n*". This is nothing but to rewrite phrase-structure grammar. So, we propose converting CFG rules to a special form in advance. For example, in the above case, the grammar is converted to rules such as Figure 2 by introducing a category "*new*". Using the grammar such as Figure 2, the parsing produces the tree (C).

The rules can be produced by introducing a new category for a set of rules whose left hand side (LHS) and the leftmost of right hand side (RHS) are identical respectively. In this paper, we call such special form of context free grammar rules *M normal-form*, and such rules, simply, *M rules*.

3.2 *M* Normal-Form Conversion

An algorithm for M normal-form conversion is shown in Figure 3. Let α, β, ω be categories and Σ, Δ, Z be categories sequences. There is no difference in acceptable sentences between a grammar and the converted grammar. The incremental parsing can produce less syntax trees by converting any set of CFG rules to M rules in advance according to Figure 3. As well, since they correspond to the grammar rules in a one-to-one way, transfer rules can be also converted in advance. The incremental transfer processing have only to apply the converted rules to a syntax tree.

Input: a set of CFG rules G
Output: a set of converted CFG rules G'

$G' := \{ \}$;
while $G \neq \{ \}$ **do**

step 1 Dividing rules:
divide G into sets consisting of the following rules:
1) LHS is identical.
2) The leftmost of RHS is identical.

Step 2 Excluding rules:
if a set consists of only one rule
then add the rule to G'

Step 3 Converting rules:
for each set consisting of rules:
$\alpha \rightarrow \beta\ \Gamma$
$\alpha \rightarrow \beta\ \Delta$
$\vdots$
$\alpha \rightarrow \beta\ Z$
1) introduce ω:
a category rewritable to $\Gamma, \Delta, \cdots, Z$
2) convert them to rules:
$\alpha \rightarrow \beta\ \omega$
$\omega \rightarrow \Gamma$
$\omega \rightarrow \Delta$
$\vdots$
$\omega \rightarrow Z$
3) add the rules whose RHS is α to G
4) add the rules whose RHS is ω to G'

Figure 3: An algorithm of M normal-form conversion

Table 1: Comparison of translation result between before M normal-form conversion and after that

item	before	after
success rate(%)	59.7	70.5
repair occurrence per sentence	3.25	2.25
translation time per word (sec.)	2.65	0.046
parse trees per word	522.7	43.6

4 Evaluation

An experimental system has been developed in GNU Common Lisp 2.2.2 on a workstation. We have made an experiment on the system to evaluate the effectiveness of converting CFG grammar rules to M normal-form. The conversations, whose task is travel application, consisting of 4 dialogues and 278 spoken English sentences in ATR Dialogue Database have been used. The average length is 7.2 words. The dictionary and context free grammar have made in the scale of 570 words and 185 rules respectively.

The set of CFG rules has been converted according to the M normal-form conversion algorithm. 159 rules became the target of the conversion, as a result, the number of the rules has increased by 77 to 262. The success rate, number of trees and translation time have been measured on the grammar before the conversion and after that. Experimental results are shown in Table 1.

4.1 Experimental on success rate

The success rate was examined. As Table 2 shows, we classified the sentences according to their translation results. In the case of using the grammar before the conversion, 166 of sentences, classified into A) or B), are correct, providing a success rate of 59.7%. On the other hand, the experiment with M normal-form grammar translated 196 sentences correctly. The success rate increased by 10.5% to 70.5%. In addition, the number of repair expressions occurrence

Table 2: Success rate of 278 sentences

type	sentences before	(rate) after
A)correct (no repair)	0 (0.0 %)	35 (12.6 %)
B)correct (repair)	166 (59.7 %)	161 (57.9 %)
C)unnatural	76 (27.3 %)	38 (13.6 %)
D)incorrect	35 (12.6 %)	43 (15.5 %)
E)parsing failure	1 (0.4 %)	1 (0.4 %)

per sentence on the average decreased from 3.25 to 2.25. M normal-form conversion resulted in that 35 sentences in B) and 30 sentences in C) are classified into A) and B) respectively, because of the reduction of repairs. The result shows the effectiveness of M normal-form conversion for fine-grained incremental spoken language interpreting. The reason why 8 sentences in 3) are classified into 4) is that the heuristic about the order of the grammar rules was destroyed a little by the conversion.

4.2 Experimental on translation speed

Translation time per word was examined on a SUN UltraSPARC-II workstation (Main memory: 512MB, CPU: 248MHz). The time on the average decreased from 2.65sec to 0.046sec. Almost of the translation time is spent on parsing. Because the parsing have to produces all of possible trees though both transfer and generation have only to deal with one tree. The result leads to the conclusion that M normal-form converting is effective for speech-to-speech translation with high synchronization.

5 Concluding Remarks

This paper has proposed a technique for reducing the repair expressions produced in an incremental spoken language translation system. A special form of a set of CFG rules, called M rules, and a method of converting any set to such form have been introduced. Since incremental parsing can produce less trees by using M rules, incremental transfer might be able to avoid choosing incorrect trees as the target. Therefore, the system can decrease the frequency of repair expression occurrence in the target speech. Experiments with a prototype system have shown the technique to be effective for fine-grained incremental machine interpreting.

Acknowledgements: The authors would like to thank Satoru Asai for his valuable contributions on an earlier version of this study. This paper owes much to the valuable discussions with Yoshihide Kato. This research has been supported in part by The Hori Information Science Promotion Foundation.

References

J.W. Amtrup. 1995. Chart-based Incremental Transfer in Machine Translation. In *Proc. of 6th International Conference of Theoretical and Methodological Issues in Machine Translation*, pages 131–138.

N. Charter, M. Pickering, and D. Milward. 1995. What is incremental interpretation. In D. Milward and P. Sturt, editors, *Incremental Interpretation*, pages 1–22. Edinburgh Working Papers in Cognitive Science, Vol. 11.

O. Furuse and H. Iida. 1996. Incremental Translation Utilizing Constituent Boundary Patterns. In *Proc. of 16th International Conference on Computational Linguistics*, pages 412–417.

N.J. Haddock. 1989. Computational Models of Incremental Semantic Interpretation. *Language and Cognitive Precesses*, 4(3/4):337–368.

H. Iida, E. Sumita, and O. Furuse. 1996. Spoken-Language Translation Method Using Examples. In *Proc. of 16th International Conference on Computational Linguistics*, pages 1074–1077.

M Kay. 1980. Algorithm Schemata and Data Structures in Syntactic Processing. In *Technical Report CSL-80-12*.

H. Kitano. 1990. Incremental Sentence Production with a Parallel Marker Passing Algorithm. In *Proc. of 13th International Conference on Computational Linguistics*, pages 217–222.

H. Kitano. 1993. *Speech-to-Speech Translation: A Massively Parallel Memory-based Approach*. Kluwer Academic Publishers.

S. Matsubara and Y. Inagaki. 1997a. Incremental transfer in english-japanease machine translation. *IEICE Transactions on Information and Systems*, E80-D(11):1122–1129.

S. Matsubara and Y. Inagaki. 1997b. Utilizing Extra-Grammatical Phenomena in Incremental English-Japanese Machine Translation. In *Proc. of 7th International Conference of Theoretical and Methodological Issues in Machine Translation*, pages 131–138.

S. Matsubara, S. Asai, Y. Inagaki, and K. Toyama. 1997a. Incremental Spoken Language Translation Utilizing Grammatically Ill-formed Expressions. In *Proc. of 3rd Conference of Pacific Association for Computational Linguistics*, pages 188–194.

S. Matsubara, S. Asai, K. Toyama, and Y. Inagaki. 1997b. Chart-based Parsing and Transfer in Incremental Spoken Language Translation. In *Proc. of 4th Natural Language Processing Pacific Rim Symposium*, pages 188–194.

W. Menzel. 1997. Parsing of Spoken Language under Time Constraints. In *Proc. of 11th European Conference of Artificial Intellingence*, pages 560–564.