[*Machine Translation Summit*, September 17-19, 1987, Hakone Prince Hotel, Japan]

MACHINE TRANSLATION USING CONTEXTUAL INFORMATION

Electrotechnical Laboratory, MITI
Shun ISHIZAKI

1.INTRODUCTION

Japanese and English are so different from each other that shallow analysis is fairly insufficient for translation between them. The transfer approaches using correspondence between parsing trees have not obtained sufficient quality of translation because English has digit sentence grammar while Japanese does not. It is said that Japanese has a discourse grammar rather than a sentence grammar.

Useful machine translation systems would be able to utilize not only sentence level correspondences between two languages, but also paragraph level information. Japanese paragraph structure is so different from English that contextual level information is necessary for "the useful system".

Our group at ETL has been developing CONTRAST(CONtext TRAnslation SysTem) which uses contextual information for analysis and generation.

2.UNDERSTANDING AND GENERATION

The CONTRAST utilizes grammatical, lexical and contextual information for analyzing and generating text. It understands input texts using background knowledge stored in the concept dictionary as well as grammatical and lexical information.

Each intermediate representation has a coherent tree structure about a topic included in the text. The generator decides a paragraph structure, sentence structures in the paragraphs, and sentence styles in the sentence structures.

3.GRANULARITY OF CONCEPT REPRESENTATION

The system doesn't use primitives such as those described by Schank. The problem is complexity included in the analysis procedure. It requires an enormous amount of inferences, which would prevent us from attaining practical MT systems.

The granularity of concept representation adopted by CONTRAST is not so coarse that each language has a "lexical-entry to concept" dictionary. One lexical entry has only one concept, but a concept may have a few lexical expressions.

It adopts a concept conversion method, which is driven when the system cannot find any lexical expression from a concept in the intermediate expression. The concept is converted into more precise expressions including concepts of finer granularity or more general concepts until a correspondence between the concept and a lexical expression is obtained. This procedure seems efficient especially for multilingual translation systems such as the ODA Project.

4.PERSPECTIVE

What is a practically useful MT system in the future? What functions does it require? It depends on the user levels of translation techniques. The system must have 100% quality of translation when users have no knowledge on the target language. The practicality of MT systems should be measured by reduction of cost at which translation is done by computer-assisted systems compared with the human translation.

The MT method using contextual information will provide necessary functions to future MT systems. It will enable MT systems to come into general use in the near future. The ODA MT Project will give a chance to us to obtain such high level quality of machine translation.

1: CONTRAST (CONtext TRAnSlaTor)

2: Machine Inference Section, Information Sciences Division
Electrotechnical Laboratory, AIST, MITI.
1-1-4 Umezono, Sakura-mura, Niihari-gun, Ibaraki, Japan 305
Tel: 0298-54-5423

3: Research system in the developmental stage.

4: An interlingua method for multilingual machine translation.
A machine translation system using contextual information.
It adopts a language independent concept structure named CRS (contextual representation structure) as the intermediate representation.
Topic area is of Newspaper articles.

5: Japanese and English.

6: Text understanding:

Syntactic analyzer (Augmented Context Free Grammar: 100 rules)

↓

Semantic analyzer (100 procedures)

↓

Contextual analyzer

Text generation: 8 major rules embedded as procedures

Paragraph-level generation

↓

Sentence-level generation

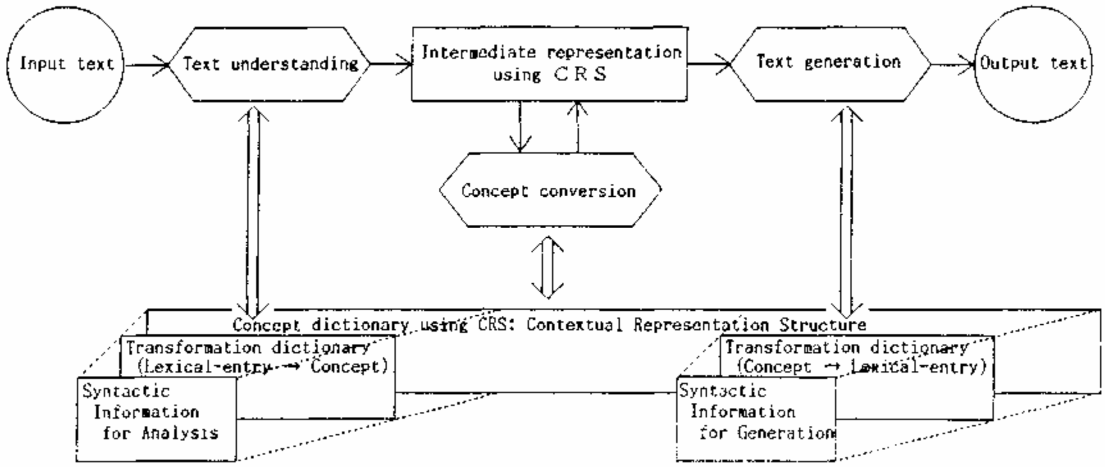
↓

Word-level generation

Concept conversion:

It converts input concepts into more precise concept representations using conversion rules named CONSTRAINT when it doesn't find any corresponding concepts to the output.

7: Concept dictionary (250 concepts)
Transformation dictionary
("lexical-entry in Japanese -> concept" : 200 entries)
("concept -> lexical-entry in English" : 100 entries)
Syntactic dictionary (500 words in Japanese and 100 in English)


Input text
Text understanding
Intermediate representation using CRS
Text generation
Output text
Concept conversion
Concept dictionary using CRS: Contextual Representation Structure
Transformation dictionary (Lexical-entry → Concept)
Syntactic Information for Analysis
Transformation dictionary (Concept → Lexical-entry)
Syntactic Information for Generation