

ATLAS PROJECT

Susumu Sawai

Development Engineering Dept., Development Div.
Fujitsu Limited, Kawasaki, Japan

ATLAS project is a machine translation project in Fujitsu. The existing projects are ATLAS/I and ATLAS/U.

- (1) ATLAS/U is a long-range project to develop a multilingual machine translation system. ATLAS/U is designed to realize high-quality machine translation, using the meaning of a natural language and the knowledge of the objective world.
- (2) ATLAS/I is a practical project for realizing machine translation of a text with limited applications.

This paper explains ATLAS/I.

1. ATLAS/I MACHINE TRANSLATION SYSTEM

ATLAS/I is a machine translation system using a production system. ATLAS/I can gradually expand the sentential patterns that are translated by rewriting the production rule for pattern matching and updating dictionaries.

2. JAPANESE-ENGLISH MACHINE TRANSLATION

Machine translation involves three stages: input of the original Japanese text, translation, and editing of the translated English text. The postediting feature has the following functions for assisting users:

If machine translation from Japanese into English is not possible, it displays the original Japanese sentence.

It presents the translation so far completed and requests the user's assistance.

Figure 1 shows that ATLAS/I currently integrates three processes: analysis of an original Japanese sentence, structural translation

from Japanese into English, and synthesis of the English sentence. The "case" of noun phrases, which is the relation of noun phrases to verbs, is checked while the syntax is analyzed; case analysis is performed by referring to the production rules. When the matching rule is found, case and syntactic synthesis of English are performed, and the synthesized English sentence is printed out. These rules are described in knowledge representation called FKR-O, and synthesis of English sentence are performed by using the matching rule.

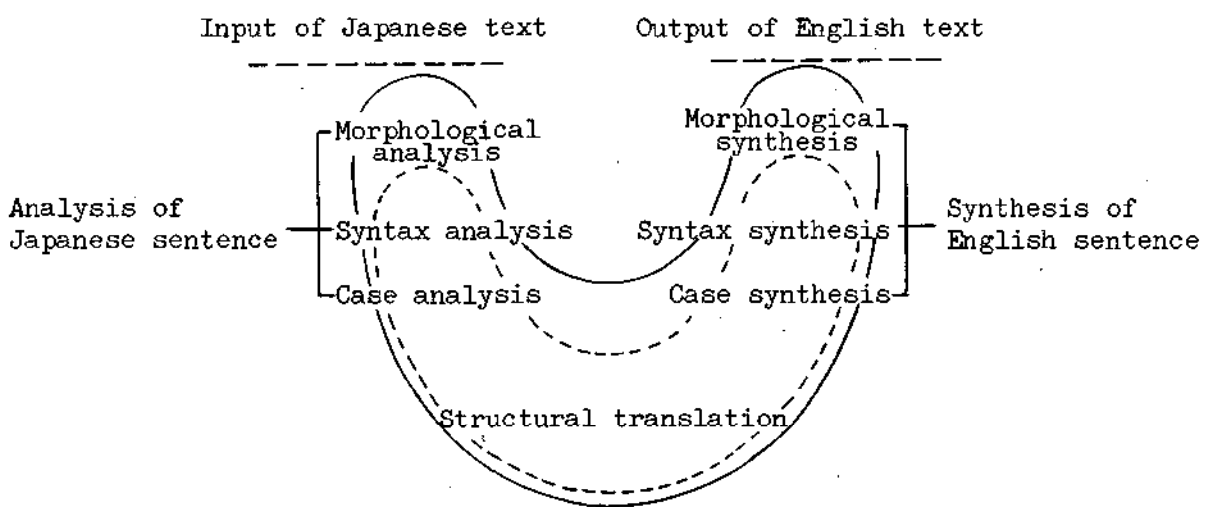


Fig. 1 Flow of ATLAS/I

Machine translation processes achieved by the use of FKR-O are delineated by dotted lines in Figure 1.

When the Japanese sentence in Figure 2 is typed in, morphological analysis is performed by dividing its strings into segments. This is followed by syntax analysis, case analysis, structural translation to English, synthesis of the English sentence, and output of the English sentence.

2.1 Morphological Analysis

Morphological analysis consists of preprocessing, phrase

dictionary reference, word dictionary reference, adjunct analysis, and proper noun processing.

- o Preprocessing

The Japanese sentence is segmented by the period (.), question mark (?), exclamation mark (!), commas (,), and parentheses which are converted to codes.

- o Phrase dictionary reference

Idioms, compound words, and word phrases written in hiragana (cursive form of the Japanese alphabet) and kanji (Chinese characters), are assigned attributes (codes) and an English equivalent. If any of these are not found in the phrase dictionary, they are segmented with a blank provided before each set of katakana (square form of the Japanese alphabet) and kanji.

- o Word dictionary reference

Each word beginning with kanji or katakana is assigned attributes and an English equivalent.

- o Adjunct analysis

After the phrase and word dictionaries are referenced, the character string is checked to extract adjuncts which are converted to codes. If there is any adjunct which cannot be extracted, the kana dictionary is referenced.

- o Proper noun processing

When any word or phrase is not found, even in the kana dictionary, it is given the semantic marker of a proper noun. Numbers are treated as proper nouns.

.2 Syntactic Analysis, Case Analysis, Structural Translation to English, and Synthesis of English Sentence

Case analysis consists of surface case analysis and deep case analysis.

- o Syntactic analysis

Noun relations, noun phrase relations, and complex sentence relations are rearranged according to the production rules of the FKR-O descriptions.

- o Surface case analysis

The relation between a noun and a postposition is analyzed to check for matching with the production rules of the FKR-O descriptions. If it matches with any of the production rules, the surface case is determined.

- o Deep case analysis

The relations between a verb and surface cases are analyzed. This pattern matching is intended to analyze the surface cases to extract their deep cases. The deep case extracted by pattern matching means "when," "where," and "who does what how." It is possible to determine (when), place (where), actor (who), and object (what) by referencing the grammatical dictionary of FKR-O.

- o Structural translation to English

The production rules of the FKR-O descriptions contain a Japanese-English pattern translation table so that pattern matching can be followed by selecting a structural translation pattern. The English synthesizing program is invoked on the basis of this selected pattern.

- o Synthesis of English sentence

The English sentence is synthesized by providing lacking words in the Japanese, such as prepositions (at, in), definite article (the), and so on. Plural form processing is also performed.

The final constituent of the English sentence synthesizing process, the morphological synthesis, is still undeveloped.

Japanese sentence = コンピュータライゼーションの発展に伴い、コンピュータの利用分野は著しく拡大した。

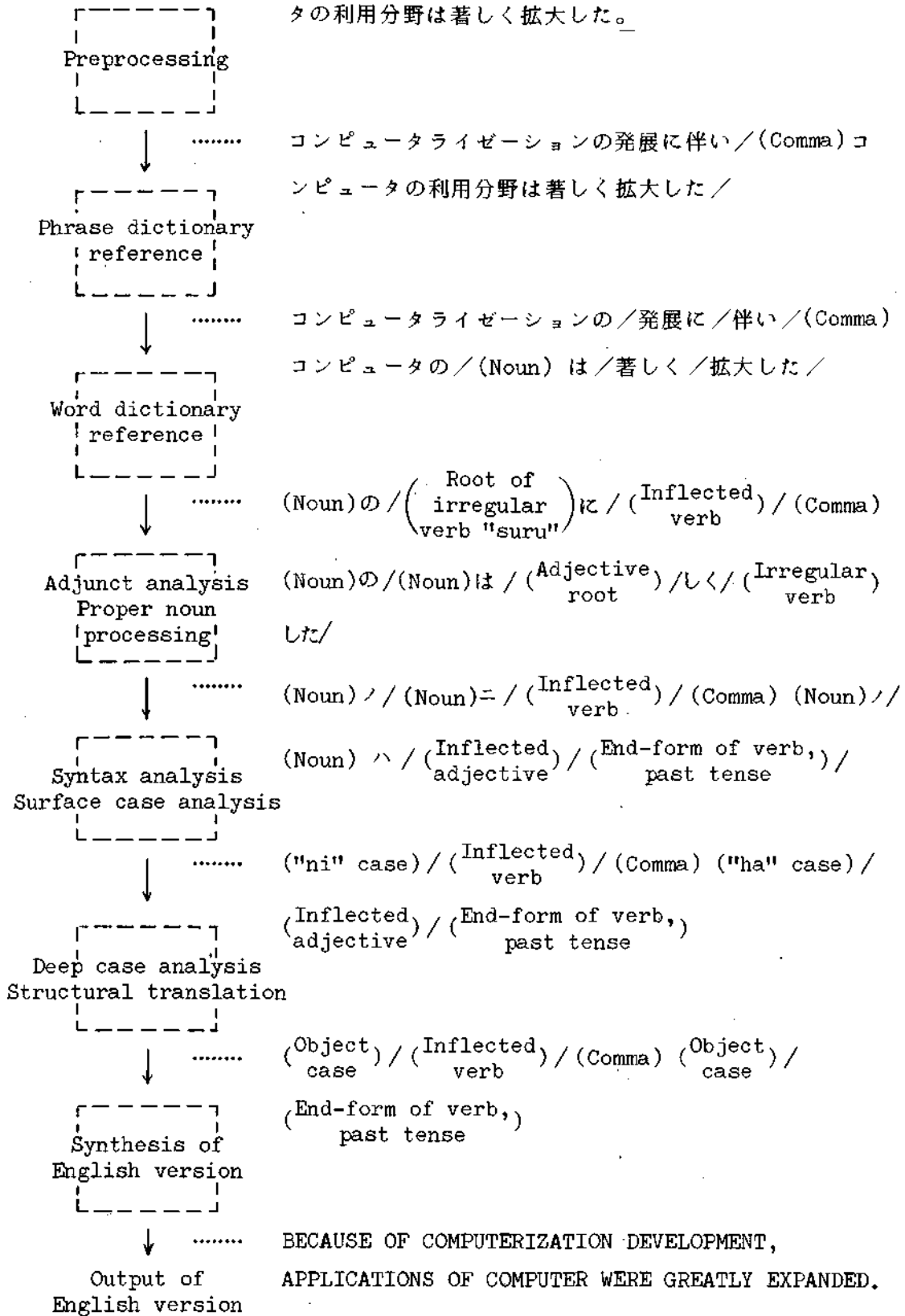


Fig. 2 Example of Japanese-English machine translation