FUJITSU'S MACHINE TRANSLATION SYSTEMS

Hiroshi Kushima
Applications Development Sec.
Scientific Systems Development Dept.
Fujitsu Ltd.

Hiroshi Uchida
Section Manager
Natural Language Processing Sec.
Software Laboratory
Fujitsu Laboratories, Ltd.

ATLAS-I

ATLAS-I is an English-to-Japanese machine translation system developed by Fujitsu Limited for use in the translation of technical documents, and is designed to run on Fujitsu's large-scale computers (M340R or larger models) and S3000 series of superminicomputers. This system was launched back in September, 1984, making it the first commercial E-J machine translation system to be marketed anywhere in the world. ATLAS-I is currently being used by some 70 organizations, some on a trial basis and others for practical applications, · and has recently been applied to tasks related to the provision of value-added network (VAN) services here in Japan.

TRANSLATION METHOD

The ATLAS-I system employs a direct syntactic approach to translation, supported by a small amount of semantic analysis.

The translation process is as follows. First, the system segments the source sentence into words, defining their individual parts of speech and meanings from built-in dictionaries. Then it analyzes the structure of the source sentence based on these word definitions, as well as on sets of grammar rules incorporated into its software components. Having done this, the system can determine the role played by each word in the source sentence, select the appropriate corresponding Japanese words and make the necessary conversion to a Japanese sentence structure (Japanese syntax). The system also inserts any auxiliary words required by the Japanese sentence structure prior to generating the target sentence.

Although the translation approach employed by the ATLAS-I system is syntax-oriented, there are times when syntactic analysis alone is not sufficient to determine the sentence structure and word meanings expressed in the source sentence. ATLAS-I therefore also makes use of techniques such as categorizing the semantic features and relationships of nouns and verbs.

For example, the English word "spring" has a number of different meanings. In its noun form, it can mean 1) a mechanical part that recovers its original shape when released after deformation; 2) a time or season of growth or development, especially the season of the year

between winter and summer; and 3) a source of supply, especially an issue of water from the ground. The corresponding Japanese translations of these meanings are expressed using three different words: "BANE," "HARU" and "IZUMI," respectively. The ATLAS-I system, therefore, categorizes these various translations of the noun "spring" according to their respective semantic features. "BANE" thus is categorized as an "object," "HARU" as a "time" and "IZUMI" as a "place." These semantic categorizations of the Japanese equivalents of the English noun "spring" are then registered in the system's bilingual English-Japanese dictionary together with the respective Japanese words. This technique enables the system to automatically select the optimum Japanese translation for the noun "spring" based on an analysis of that words relationship with other words in the source sentence.

Even though a certain amount of this type of semantic analysis is performed during the translation process, ATLAS-I is, first and foremost, a syntax-oriented system. It relies primarily on syntactic analysis techniques to translate a source language sentence into a target langage sentence; semantic analysis is simply a support function. And because it is syntax oriented, ATLAS-I is able to produce target language sentences that reflect the structural characteristics (syntax trace) of the original source language sentence.

MAN-MACHINE INTERFACE

ATLAS-I is an interactive machine translation system, which means it requires a human operator to perform the various rewriting and editing tasks necessary to produce a polished final translation. It is therefore equipped with a "translation editor" function.

The translation editor divides the terminal screen into two parts, right and left, displaying the source and target sentences simultaneously. This makes it quite easy for the human operator to compare the source and target sentences in preparation for performing the necessary rewriting, correction and dictionary display and updating work using the functions built into the translation editor. Since all this work can be performed on the same screen, it makes the process much more efficient.

APPLICATIONS

The ATLAS-I system was developed for the purpose of translating English technical documents, such as science and technology reports and product instruction manuals, into Japanese.

In the little over a year since it was first put on the market, ATLAS-I has been utilized by some 70 organizations, roughly half of which have been universities and research institutes. The second largest group of users have come from the manufacturing and related industries, and are using ATLAS-I to translate technical documents and manuals concerning new products, for the

most part. Although quite small in number, a few financial institutions have been using ATLAS-I to translate documents and reports concerning matters other than those of a technical nature.

When it comes to English-language technical documents, many times the person or persons requiring the information contained in such documents can obtain that information simply by reading the document. They needn't go to all the trouble and expense of having it translated, especially by a computer. But in those cases where the documents or reports in question comprise hundreds of pages, the people in need of the information contained therein are numerous and the time available to extract that information is short, then the ATLAS-I machine translation system is the most effective means of accomplishing the task. It can translate large volumes of source text extremely fast, and produces the target text in the form of easy-to-read hardcopy printouts. Another major advantage of the ATLAS-I system is its ability to translate the difficult technical terms stored in its dictionaries correctly and accurately every time.

ATLAS-II

ATLAS-II is a Japanese-to-English machine translation system marketed by Fujitsu in June, 1985. The technology incorporated into this system, however, is designed to enable it to be upgraded to a multilingual machine translation system in future.

The ATLAS-II system runs on any of Fujitsu's FACOM M series OS IV/F4MSP general-purpose large-scale computers, and is equipped with a 50,000-word basic Japanese dictionary, a 50,000-word basic English dictionary, plus two science and technology dictionaries, one Japanese and one English, containing 250,000 technical terms each. When run on Fujitsu's M380 computer, ATLAS-II can churn out 60,000 words of translation every hour.

In the three months since its commercialization, ATLAS-II has been installed at 30 user locations, and is finding increased application in the VAN service business.

TRANSLATION METHOD

The ATLAS-II system employs a semantic transfer approach to machine translation. The semantic transfer method is based on the construction of intermediate linguistic representations that represent (express) the sentence structure and meaning inherent in the source language text, which in this case is Japanese. The translation process itself consists of three major steps. In the first step of the process, the system analyzes the source language sentence and convertes that sentence into an intermediate representation. It then transfers (converts) this source language intermediate representation into a target language intermediate representation. The third and final step in the process involves generating the target language

sentence from the target language intermediate representation.

The semantic transfer method·employed in the ATLAS-II system makes use of a semantic network representation as its intermediate representation. This semantic network representation is called the conceptual structure. The conceptual structure expresses the interrelationship between the various concepts expressed by the words that comprise the source language sentence. For example, if we have a sentence like "WATASHI-WA MIZU-WO NONDA" (a literal English translation of which might read "I drank water," where "WATASHI" means "I," "WA" is the postpositional particle that designates "WATASHI" as the subject of the sentence, "MIZU" means "water" and "WO" is the postpositional particle that designates it as the object of the sentence, and "NONDA" is the past tense of the verb "NOMU," which in this case means "to drink,"), the subject of the verb "NOMU" is "WATASHI," the object of the verb "NOMU" is "MIZU" and the actual act of drinking was done in the past, as expressed by the past tense form "NONDA." So this one very simple sentence contains all these meanings and semantic relations.

Now, with a very simple sentence such as this, you can see how the source language intermediate representation will match up perfectly with the target language intermediate representation, thus eliminating the need to transfer (convert) from one to the other. However, with more difficult sentences, ones where cultural differences have a

much greater influence on sentence structure and wording, the semantic concepts and their relations inherent in the two languages do not necessarily match up. When this is the case, it is necessary to transfer the semantic concepts inherent in the source language sentence into concepts that exist in the target language at the intermediate representation (semantic concept) level.

The semantic transfer method does not use Japanese-to-English or English-to-Japanese dictionaries. Instead, it makes use of dictionaries that describe the meanings of words found in Japanese and English dictionaries. Thus, during the Japanese-to-English translation process, the ATLAS-II system "grasps" or "understands" the meaning of the source language (Japanese) sentence by referring to its Japanese language dictionaries, and then generates a target language (English) sentence with an equivalent meaning based on that understanding, plus the results of references to its English language dictionaries.

As a result of its ability to perform translations based on an understanding of the meaning of source language sentences, the ATLAS-II system builds up a store of conceptual information related to words. For example, the subject of a sentence containing the verb "to drink" has to be a living creature, and examples of living creatures found in our known world include "humans," "dogs" and "birds." And words such as "man," "woman," "father," "brother" and "I" all refer to humans. Based on this type of "knowledge," ATLAS-II is

capable of inferring that the word "I" can serve as the subject of the verb "to drink."

The ATLAS-II system might, therefore, be described as a machine translation system capable of understanding the meaning of source language sentences, and then, based on that understanding, re-expressing that meaning in target language sentences, using a process very similar to the one that goes on inside the minds of human translators.