CULT

Chinese University Language Translator

Research into machine translation at the Chinese University takes a different approach than the others in that the Chinese University of Hong Kong places a heavy emphasis on pre-editing the source text instead of post-editing the target text. It is the only group taking this approach of computer-pre-editor partnership. All the other groups, who realized the FAHQT is not really attainable in the near future, have adopted a tendency to compromise in finding some computer-post-editor partnership.

A fixed set of pre-editing rules must be formulated to enable inexperienced and even mono-lingual people to transform quickly the input into machine-translatable form. With this arrangement, post-editing can be kept to a minimum, if not all together eliminated. Given time and better programming techniques, these pre-editing rules will gradually be reduced so that the computer will eventually take up this routine work. Pre-editing can therefore solve many of the present linguistic problems that are otherwise dependent on further research in natural language, computational linguistics, and transformation mathematics. In the present stage of development, very complex sentences can be translated with the aid of pre-editing. *

CULT (Chinese University Language Translator) was developed based on the principle mentioned above and has been rigorously examined and tested. Since the beginning of 1975, the CULT System has been used on a regular basis to translate two Chinese scientific journals, ACTA Mathematica Sinica

*An average of 5% of text is pre-edited by computer or editor.

and ACTA Physica Sinica, which are published by the Peking Academy of Science. This accomplishment indicates the correctness of our approach and the potential capability of CULT.

THE NEW LANGUAGE TRANSLATOR

Initially, CULT (Chinese University Language Translator) was designed as a special natural language translator with Chinese as the source language and English as the target language. Of course, a separate language translator will be required if English is to be used as the source language and Chinese as the target language.

The present translator consists of four modules, namely: 1) Dictionary look-up procedures employing the largest matching principle, 2) Syntatic analyzer, 3) Semantic analyzer, and 4) Output procedures including re-arrangement of word-order for the output sentences.

1. The Dictionary Look-Up Module

The basic dictionary look-up algorithm employs the "largest match" principle, designed for Chinese input (i.e., five digits numbers) and can readily be used for other non-alphabetic language input. However, an additional procedure for languages with alphabets (i.e., English, Malay, etc.) may be required to convert the alphabetic characters into numerical form by forming a "hash" before performing the look-up.

```
┌─────────────────┐
│   DICTIONARY    │
│     LOOKUP      │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│    SYNTACTIC    │
│    ANALYSER     │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│    SEMANTIC     │
│    ANALYSER     │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│     OUTPUT      │
│                 │
└─────────────────┘
```
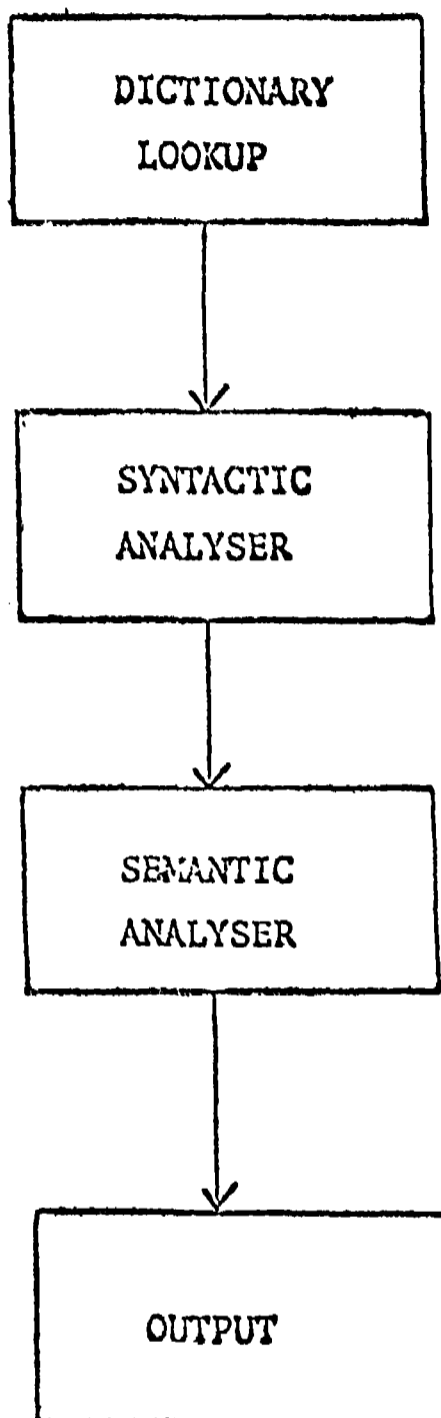
FIG. 1   TRANSLATION ALGORITHM

2. Syntactic Analyzer Module

The main function of the syntactic analyzer is to determine precisely the role that the individual words play in the sentence (i.e., to which parts of speech the words belong, whether noun, verb, etc.). The process is accomplished by means of a rather sophisticated true-false table.

While working on the machine translation of the Chinese scientific journals, a number of interesting linguistic difficulties experienced have been identified and defined. Previously, such structures would have to be pre-edited or post-edited in order to obtain the correct translation, but now they can be readily translated without any pre-editing.

3. Semantic Analyzer Module

At present, the semantic analyzer is able to offer only limited facilities, and the problem of semantic ambiguities is essentially resolved by: 1) a dictionary with specialized subject matter and 2) by pre-editing.

4. Output Module

The function of the output module is simply to rearrange word-order of the output sentence structure appropriate to the target language.

CONCLUSIONS

The successful translation of Chinese scientific journals, as well as non-scientific articles, by means of <u>CULT</u> has amply demonstrated the capability and the potential usefulness of the machine translation system in overcoming language barriers.

Though a number of   guistic problems are still to be defined and solved, the present machine translation system developed so far, if used with care and understanding, may contribute in some small measure in easing the desperate translation needs facing us today.

Automatic translation cannot be perfect. Whether it could even be high quality or not is dependent on how high the standards are set. The immediate goal is not to design a perfect automatic translation system or to achieve high quality machine translation, but to design a machine translation system that is better and more efficient than the ones we have today.

SHIU-CHANG LOH

Professor of Computer Science

Director, Machine Translation Project

United College   Chinese University

Shatin, New Territories   Hong Kong