MACHINE TRANSLATION — PROSPECTS AND PROBLEMS

R. Sudarsan

Defence Scientific Information & Documentation Centre, Delhi—110054.

1. INTRODUCTION

About 25 years ago, Hugo Gernsback, editor-in-chief, *'Radio Electronics'* predicted that by the dawn of the coming century, language barriers will have vanished. His prediction has almost come true although we are still 13 years away from the next century. Electronic machines, particularly computers have made inroads into the translation activity. Will they replace the human translators? Such a situation may not arise in the near future, since fully automatic, high quality machine translation seems next to impossible. In spite of this, efforts towards achieving machine translation have not stopped. Significant progress has been made in the area of machine translation. Machine translation is a complex process of man-machine interaction.

2. METHODS OF TRANSLATION

For the purpose of understanding the concept of machine translation, it is desirable to examine various methods of translation. They are as follows :

   i) Human translation

   ii) Machine-aided translation (MAT)

   iii) Human-aided machine translation

   iv) Machine translation (MT)

2.1 Human Translation

Human translation is as old as mankind and the most widely used method. It is also the most reliable method. The other three are considered to be synonymous, and are labelled as MT. Some draw a distinction between MAT (also called computer assisted translation) and MT. In MAT, the role of the machine is to assist the human translator, as is obvious. Whereas, in the human-aided machine translation, the machine, besides assisting, does the translation as well. The MT, on the other hand, is the ideal case that falls in the domain of artificial intelligence.

2.2 Machine-aided Translation (MAT)

In the MAT, the machine used is the most commonly used tool of office automation — the word processing computer. It works as a personal *secretary* to the translator. Used as a factotum, it alleviates much of the drudgery of the translator's work. It can detect most of the typographical and spelling errors with the help of dictionaries and spelling checking programs. It edits and proof-reads before giving a camera-ready copy. The most important use of a word processor for a translator is in the form of a terminology base. Every professional translator, normally, has his own glossary of terms he comes across. But this glossary is maintained in a haphazard manner. A lot of time is wasted in culling out the meanings of the terms from such manual tools for various subjects. In MAT, a dictionary of definitions and possible choice of specific terms can be produced for use when translating one or more documents on the same topic. Glossaries can also be prepared on a computer simultaneously with the typing of a document.

In human-aided machine translation, the machine, while translating, should ensure that the contents of the texts in the source language and the target language are the same or roughly the same. To achieve this, the machine should have the qualities of the human translator built into it. Human translator has linguistic knowledge, common sense, capacity to perceive cause-effect relationship and other such human traits. So endowed, the human translator is able to perform semantic and contextual analysis with ease. So that it can simulate the human translation process, a machine has to be similarly 'endowed' by having these qualities, at least partly, built into it.

2.4 Machine Translation

2.4.1 Role of Machine

The process of MT involves mainly three phases viz. analysis, transfer and synthesis.

Source Language
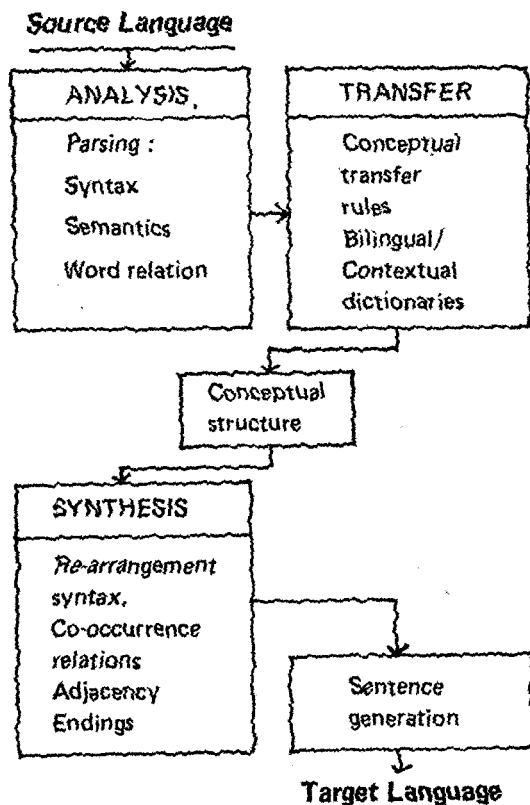
Process of machine translation

The analysis phase involves parsing or, other words, picking apart a sentence to determine what each word and ultimately the sentence means. Both syntactic and semantic parsing are done. The former concerns parts of speech, and word order, the latter the meaning or the sense. A parser uses grammatical rules as well as other sources of knowledge to determine the role each word plays in the sentence and how these words are related to one another.

In the transfer phase, transfer dictionaries are used to find the corresponding words in the target language. Contextual dictionaries are used to remove ambiguities. Finally, the target words chosen are fed into the synthetic routines. In the synthesis phase, appropriate endings are added and words are rearranged according to the rules of syntax of the target language.

2.4.2 Role of Human Translators

The role of human translator in MT is changed to that of a pre-editor and a post-editor. He works as a linguist. Since it is not possible to incorporate all human knowledge and vocabulary into the machine, it is necessary to modify the source text into a machine understandable form. The pre-editor controls the input by rewriting the text to be translated in conformity with the contents of the dictionary stored in the computer and according to the rules established especially for it. Pre-editing not only improves the source-text but enables better translation. It is also a step used to update the dictionary. Although the computer output is a translation that is half-done, it has performed the unpleasant part of the translation job—the tedious and monotonous job of translating the same words and expressions several times in the same text. The translator, playing the role of the post-editor, has only to see how those words are connected and make the output grammatically correct and technically precise. It is on syntax, word order and semantic congruity that the computer stumbles most often. The job of the translator is enriched because he has to analyze the text, and identify the areas that need frequent correction. He can thus improve the system linguistically. This reduces the frustration

4

of the post-editor and augments the quality and productivity of future computer outputs.

In between the pre-editor and the post-editor there is the dictionary coder who builds the terminology base, which includes parts of speech, homographs and conflicting entries in the dictionary. Based on the subject knowledge and the work done by the pre-editor, decisions will have to be taken as to what expressions should be entered into the terminology base. There may be two coders, one for the source language and the other for the target language. Both must have an excellent command over the languages, mainly at three levels : morphology, syntax and semantics.

3. MACHINE TRANSLATION SYSTEMS

The technology of machine translation has made rapid progress during the last three decades. Several machine translation systems were developed and used. Some of the most significant machine translation systems in use and being developed are described below :

3.1 Logos

Developed by Dr. Peter Toma in the late 1960s, it was operationalised by the European Commission in Luxembourg. It is a word-processor based system which translates between English and German. It has a speed of 1500 words/hour.

3.2 Systran

It is a comprehensive translation system which can be grouped under the human-assisted machine translation. It takes care of multiple meanings and makes a choice during the translation process. Initially, it was designed to translate between English and French because the target country was Canada. Now, the Systran is available for translation from English to German. The German language has an infinitely large capacity to produce compound nouns, making it difficult to create a comprehensive German to English system. Systran has an amazing speed of 3,00,000 words/hour.

3.3 Meteo

The English to French Meteo system developed by the Montreal University, Canada, has been highly successful. Its success is obvious from its application in meteorology. In this field, there is no ambiguity of words, sentences are short, verbs and nouns are repetitive. Moreover, rules of grammar and syntax are very simple. With a vocabulary of 2000 words and expressions, the METEO system has been providing forecasts in translation for the public and mariners in Canada four times daily since 1977. It translates over nine million words annually but not without revision by the human translator.

3.4 Eurotra

Eurotra is another ambitious project being developed by the European Commission. It will be a multilingual system, translating between all European Community languages. It will be portable from one make of computer to another. It will also permit addition of new languages and subjects. Its estimated cost is $22 million.

4. OTHER AIDS IN TRANSLATION

The development of online terminology is an area which promises to be very useful to the professional translator. Glossaries and dictionaries become out-dated because of the delays in publishing and the rapid advances made by science and technology. A clearing-house for sharing terminology files at nominal cost can be developed.

5. COST FACTORS

According to a study, a workload of translation of a minimum of one million words per year is required to justify the acquisition of a MAT system. Private translation agencies and publicity departments of multinational companies can use these systems. But, what about a non-profit government organization like INSDOC and DESIDOC? The articles translated, for example, in DESIDOC are diverse in nature. Articles in subjects like aeronautics, medical science, chemistry, ballistics, to mention only a few, are translated. These are translated from four languages into English. It is difficult to contemplate a system which will cater to such multifarious fields.

REFERENCES

1. Cancel, A.C. Over Byte of the translator. Paper presented at American Translators Association Conference, 1984.

2. Caron, A. The Meteo system and the translation. Paper presented at American Translators Association Conference, 1984.

3. Cohen, P.M. Information technology and the translator. Aslib Information, 13(4), 1985.

4. Mauro, P.D. and Russo, M.J. Computer assisted translation at Xerox Corporation. Paper presented at American Translators Association Conference, 1984.