Mechanical Translation[1]

Victor H. Yngve

Translated by Sabura Muroga

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Ever since the automatic digital computer was developed, there has been much discussion of the possibility of using these machines to perform tasks different from those for which they were designed. One of the possibilities that has intrigued many people has been the idea that perhaps such general-purpose computers could be programmed to translate languages.

The translation of languages poses a task of large magnitude for the world. It has been estimated that over two-thirds of the technical and scientific literature is written in English and that over two-thirds of the world's engineers cannot read English. Such material must be translated into the other languages if it is to be used to its fullest. The situation is also difficult for the English-speaking scientist. Most scientists in America can read German and French with difficulty. Very few scientists in America today can read Russian and practically none can read Japanese, except those who learned Japanese as a child. Expert translators in these languages are difficult to find. Thus there is a great need for translation and when it was suggested that machines could possibly be used to assist with this very large task, it seemed worthwhile to investigate the possibility.

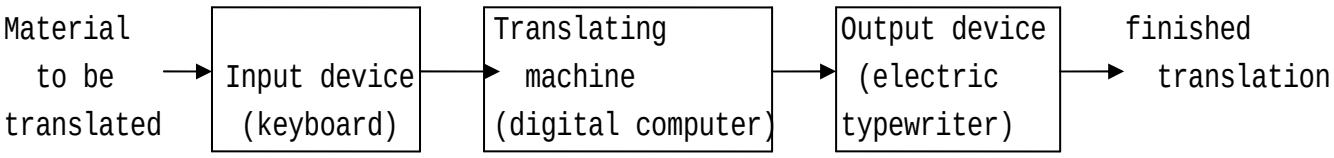
A great impetus was provided in this country in 1949 by a 12-page mimeo-graphed memorandum by Warren Weaver, then of the Rockefeller Foundation, which pointed out several possible avenues of research and discussed some of the problems.

THE DIGITAL COMPUTER

General-purpose digital computers have many properties which might make it possible to program them to translate. In the first place, they have a memory in which information can be stored and later looked up much as a word is looked up in a dictionary. They have the ability to manipulate information, to store it

[1] This work was supported in part by the Signal Corps, the Office of Scientific

here, move it there, compare it with information stored elsewhere. They also possess input end output devices which would be satisfactory for the purpose of Mechanical Translation. Nearly all general-purpose digital computers use an electric typewriter as input - either directly or through a punched paper tape or magnetic tape. Nearly all general-purpose digital computers have facilities for printing the output on a typewriter. Many of them already have provision for handling the roman alphabetic characters as well as decimal digits. Those that do not have these facilities could be easily modified to handle any kind of an alphabet. For the Japanese language the kana characters would probably be the most desirable for the input. Output in the Japanese language could be in the kana characters completely, or in some numerical representation of the Chinese characters also



Other input devices that ere under development and might be useful in the future are the photoelectric printed-character recognizer and the speech recognizer. If input came from a speech recognizer, the output could be to a speech synthesizer, However, since these devices are not yet fully developed, and may not be for many years, the present effort in mechanical translation concerns itself entirely with the written language, keyboard input, and electric typewriter output.

We have indicated in a rather general way that probably the mechanical prob-leme have been solved. If it is possible to translate languages by machine, it seems that machines already exist that could do the job. This is the opposite situation to the situation which we find in the development of the automatic digital computer. It has been known for a long time that complex differential and integral equations could be solved by numerical methods. In fact, numerical methods of solution were used quite extensively, and most of the techniques were already understood. The problem at that time was to build an electronic machine which could solve such equations rapidly by digital methods. The general-purpose automatic digital computer was the result. In the case of mechanical translation, machines that might be able to do the job already exist. The problem is to develop our scientific understanding of languages to the point where we know how to program such a machine to translate. The major problems to be solved are linguistic ones.

THE LANGUAGE PROBLEM

A great deal is already known about the language problem. Scholars for many generations have written grammars and dictionaries of the various languages that are of interest. Most of these books, however, were written for a different

purpose; they were written either for the purpose of describing the standard language to native speakers who speak a non-standard dialect and wish to learn the standard dialect, or for the purpose of instructing non-native speakers in *the* language. More recently, some contemporary linguists have been attempting to provide instead a scientific description of the languages as they are observed. A stimulus for this in America has come from the interest in describing the many native American Indian languages which are in danger of dying out.

It seems, however, that mechanical translation will require still a different scientific understanding of language, an understanding of the type that we do not have at present. The providing of such an understanding of language poses a great challenge to the linguist and to the scientist.

It is a very interesting fact that most languages which are of importance for mechanical translation have developed a unit of discourse of about the size of the English or the Japanese word. Interesting theoretical light has been shed on the reason for this by B. Mandelbrot, who considers language *as a* code and applies ideas from information theory.

WORD-FOR-WORD TRANSLATION

It has already been indicated that the machines that exist have a memory in which information can be stored and referred to as in a dictionary. It seems clear that any translation scheme must include a dictionary. Word-for-word substitution is certainly mechanically feasible. The question that remains is whether merely a word-for-word substitution is of any use as a translation. It hae been shown that for related languages such as English, German, French, Buseian, a word-for-word substitution is of some use, but there is a great difficulty in reading it. The same is true between English and Japanese, perhaps more so. There are two great difficulties in word-for-word substitution as translation. The first one may be called the problem of multiple meanings, the second one may be called the word-order problem. They are actually aspects of the same problem.

The problem of multiple meanings arises even with respect to one lan-guage alone, because a given word of any language is used in several ways. Words are usually almost unambiguous when used is a context of a sentence or a paragraph, but when they are removed from the context, we can only look them up in the dictionary and we find that there are several meanings.

Examples

hand	14 meanings	ne	18 meanings
to	20 mean ings	agaru	19 meanings
good	20 meanings	kakaru	20 meanings
beat	28 meanings	ki.	21 meanings
touch	41 meanings	yoi	24 meanings
run	54 meanings	deru.	25 meanings

The selection from among the various listed meanings is made on the basis of the context. Now when translating .into another language, the various different meanings assigned to one word in the input language may be assigned to different words in the output language. and. these different output words: which contain as part of their meaning, part of the meaning of the input word, will in turn have other meanings not implied at all by the given input word.

Moncii	speech	1, manner of speaking:speech 2, the power or ability to speak 3. a public oration or discourse 4. a. particular language or dialect
	dispute	1. to argue pro and con,discuss,debate 2.. to deny the truth or validity of

Thus there is no single word that can be chosen to substitute for the input word in all situations; and if a list of words is provided, they will in general. carry additional meanings not intended at all in the original, so that there is the danger of misinterpretation.

The extent to which the meanings of corresponding words in different languages are the same depends partly on the degree of similarity of the cultures. Language is a human behavior pattern, As it is a social behavior pattern . there there usually being at least two persons involved in a communication by language. The details of this behavior pattern are influenced to large extent by the society and culture to which the individuals belong.. Since the various countries of the world have developed historically different cultures, there is a difficult problem in translating between the languages. However, there is one aspect of culture, at least, which is largely international in that it has been contributed to by workers of all major countries. This is science and mod-ern technology. It seems that, here, the difficulties of translation will be fewest, since all scientists are talking about essentially the same thing. Much of scientific vocabulary is already international. The major difficulties that remain are that many of the words that are used by the scientist are not drawn

from the store of internationally defined and understood terminology but from the store of words of the scientist's own national language. Many of these words are associated with the grammatical and syntactic systems that are historically traditional in the various different languages. These grammatical and syntactic systems differ in a great many respects both because of historical cultural differences and because of other reasons.

Experiments on word-for-word translation have been carried out between German and English and between Russian and English. Some experiments on a *more* limited scale have also been carried out from a large number of different languages into English. Those who have been working in the field of word-for-word translation have attempted to solve the multiple-meaning problems by submitting several alternative translations of each word. The reader is then asked to choose the best meaning from among the choices presented. The experiments have shown that it is, in general, possible but difficult for a person confronted with such lists of meanings to choose the best one, using the context as a guide. The reason that it is difficult for him to do this is that the context information is expressed in different ways in different languages. Context information is expressed through the choice of words and through their order in the sentence. The words in the sentence must come in some order. In most if not all languages, the order of the words has become traditional and has been given tasks for carrying the meaning. For example in Japanese:

1. Inu ga hito o kanda. Kite o inu ga kanda.
2. Hito ga inu o kanda. Ino o hito ga kanda.

The difference in meaning between 1 and 2 depends entirely on the order of the words. Some languages such as Chinese and English have standardized the word order even more:

1. The dog bit the man.
2. The man bit the dog.

Any strict word-for-word translation will have difficulty when the traditional word orders of the two languages involved are different. The relative importance of word choice and word order is different in different languages. For example, Chinese and English rely heavily on word order: Latin, Russian, and Japanese rely heavily on word choice. These latter are highly inflected languages. It seems that translation from highly inflected languages into less highly inflected ones would be easier on a word-for-word basis. Much of the work on word-for-word translation has been concerned with the separation of inflectional endings from the remainder of the word, the stem. The possibility

of separating stems and endings provides a saving in the amount of storage space necessary in the memory of the machine. Each stem needs to be stored only once, the ending being looked up in a small auxiliary list.

kakimas u
kakimas hita
kskimas ho
kakimas hitaro
kakimas hitara
kakimas hite

Even more experiments, carefully carried out, will be necessary if word-for-word translation is ever to be of any use.

TRANSLATING THE SENTENCE

It is a further very interesting fact that most languages of importance to mechanical translation have developed a unit of discourse of about the size of the English or the Japanese sentence. In a sentence-by-sentence translation, very few difficulties of the multiple—meaning type remain. If an English scientific paper is cut up into pieces, each containing only one sentence, these can be placed at random in arbitrary order and translated rather easily into Japanese with little loss of the meaning. When the sentences are restored to their original order, a fairly good translation appears. It would seem, therefore, that sentence-for-sentence translation would be acceptable. The question remains: Would it be feasible mechanically? It is certain that sentence-for-sentence translation cannot be done on the same basis as word-for-word translation - storing each sentence in a dictionary and looking up its translation when it is needed. There are certainly too many sentences. It has been estimated that the number of possible sentencee in a language may be greater than 10^{80}.

If sentence-for—sentence translation is to be done at all, then it seems that perhaps the only way is to take the word-for-word translation as a first approximation and improve on it by taking into account the context by a series of grammatical and syntactic rules. Since the machine will never know the meaning of what is translated, it will be in a worse position than a person knowing only Japanese trying to translate from Korean into Italian, using only rules that were given to him in mathematical or symbolic form. The development of such a system of translation is a great challenge to the linguist and the persons who wish to have a scientific understanding of language. Research along this line is being carried forward between various related languages in the Indo-European family, such as Russian and German into English. It is anticipated that it will be easier to solve the problem between these related languages, and it is hoped that such solutions will assist in the eventual solution of the problem of mechanical translation between unrelated languages, for example,

Japanese and English.

THE FUTURE

At this time very little can be said about the economics of mechanical translation. It seems that if mechanical translation becomes possible, special-purpose machines will be designed which will reduce the cost. There is no question but that there is great need for more translation in the world today, particularly in the fields of science, technology, government, and business. Whether mechanical translation will ever be able to help in these areas remains to be seen. There will always remain certain ambiguities and difficulties in translation. Even human translators have difficulties.

The use to which mechanical translation is to be put is not a scientific or engineering problem, but depends upon many other important things. Mechanical translation could be used to encourage the increased use of the various national languages by allowing all authors to write and read in their own national language and have all materiel translated into other languages for foreign readers and all foreign literature translated into the national language for the native readers. On the other hand, mechanical translation could be used to promote some international language. All written material could be translated into this one international language which all scholars would learn as a second language. Eventually some might prefer to write directly in the international language. However, it is certain that no matter in what way mechanical translation is used, it will result in increased communication, and it is to be hoped that with increased communication will come greater understanding and co-operation.