

A Introductory speech

By D. Panov (USSR)

Machine translation is one of the problems of modern science which interests very many people. The specialists are interested in it because they fully appreciate the vital part it can play in solving a whole series of other important tasks; and the general public are interested in it because they are intrigued by the idea of machine translation from one language into another on account of the very boldness of its conception and its apparent impossibility, and because the idea has the same romantic appeal as man's great triumphs over the forces of nature.

Just over ten years have elapsed since electronic computers were first constructed and the possibility was suggested of using them for translation purposes, if only for mechanizing dictionary search. At that time, the potentialities of the machine were treated with extreme caution, and pre- and post-editing of texts were considered essential. From that standpoint, the celebrated experiment conducted in 1954 by the IBM Corporation and Georgetown University was a revolutionary one and I hope the representatives of that esteemed firm will not be angry with me for using that term. Although it related to only the simplest of sentences, it demonstrated the possibility, in principle, of translating without any editing being required. I think I am not wrong in saying that it marked the beginning of the general popularity of the idea of machine translation. Together with these veterans, a number of new recruits began to interest themselves in MT, and our present meeting in this room is incontrovertible proof that the problem is now a very live one.

It is impossible, in outlining the history of machine translation, to omit mention of the extremely daring proposal made over 25 years ago in the Soviet Union. On the 5th of September 1933, the Soviet scientist P. P. Troyansky registered the invention of a "machine for selecting and printing words in the process of translation from one language into several languages simultaneously," and was issued with USSR Patent No. 40995. Troyansky's machine was never constructed, for the level of technique was then still inadequate; but many of his ideas anticipated present-day views on the solution of MT problems. The USSR Academy of Sciences has prepared a separate book on Troyansky's work, together with the necessary commentaries.

The problem of machine translation is interesting in the sense that it is one of those scientific problems where the means of practical execution were in advance of theoretical research. The old and universally esteemed science of linguistics proved itself backward in that connection. The linguists had been peacefully plodding away for years giving instruction on such respectable themes as verb forms in Old Icelandic, the special features of Chaucerian English or Rabelaisian French, the use of pronouns in Sanskrit, and so on; and when the bustling designers of electronic computers came along and told them: "We want to construct a translating machine. Tell us how to design it and what sort of operations it should carry out," the linguists tried to get out of it by answering that they knew nothing about machines. But the constructors replied "There is no need why you should. It is our job to draw up a programme of work for the machine. All you have to do is to tell us the rules we must follow to get an accurate translation." Then it transpired that our modern languages studies were far from satisfactory, or at any rate

were not sufficiently developed to enable us to frame an exact system of translation rules on the basis of which an algorithm could be worked out for solving the task mechanically.

It is hardly surprising, therefore, that most of the research on machine translation reduces to linguistics in one way or another. All the experts and scientific groups working in this field are in fact engaged in a single task, which can be defined as follows: "To provide an accurate description of the structure of an actual existing language with a view to translating it into another language." Defined in those terms, the problem sounds almost trivial, even though there are many possible ways of interpreting what it involves. Scholars employing mathematical methods in their work, understand it to mean that some mathematical description of the structure of a language has to be devised, and only when that is done (they consider) can the problem of formulating the translation algorithm be satisfactorily solved. My own view, however, is that this conception of the task is fraught with major, if not insuperable, difficulties. John von Neumann, one of the founders of modern computational mathematics and techniques, declared in his lectures on computing machines and the human brain that the language of the brain was not mathematical language. And since our normal language is intimately connected with that of the brain, it is very difficult, if not impossible, to describe it mathematically. Fortunately, a great deal can be done by approaching the task from a more practical angle and by studying language as it really is. As we all know, the physicists and chemists, by carefully observing nature and checking their observations experimentally, have succeeded in making an exact mathematical formulation of many natural laws, although a strictly logical base has still to be found in some cases. I feel that this approach to the problem offers the prospect of greater success achievable in a shorter space of time, and that, too, is a fairly important consideration.

There are of course a number of technical questions relating to the construction of special-purpose machines which have to be solved. We can divide MT problems into the following three categories:

- 1) The formulation of concrete algorithms of translation from individual languages into other languages. These may be subdivided into algorithms that do not use an intermediate language and those that do, whose object is to work out a whole series of algorithms for multi-lingual translation. This necessitates solving a number of linguistic questions such as the comparative analysis of the structure of languages for translation purposes, the compilation of specialized dictionaries, etc.
- 2) The formulation of new procedures for studying and describing languages by exact mathematical methods which will involve some reassessment of the results of traditional linguistics. Scholars working in this field consider that while some old results will take on a new form in the process and will assume a new and more objective significance, many of the established concepts in linguistics will prove to be useless in the new circumstances. In their view, great interest attaches to the use of symbolic languages for recording algorithms and to a number of other mathematical questions connected with study of the structure of translation algorithms, etc.

- 3) The construction of physical systems for machine translation, such as specialized translating machines, special input and output devices (including reading devices), etc.

One point that must be stressed, however, is that premature differentiation between MT problems is undesirable; it is advisable, rather, to establish the closest possible contact between the various lines of research mentioned, between specialists in different disciplines—linguists, engineers and mathematicians—interested in MT, and between representatives of the various groups working on the subject. Only by a broad front of united endeavor will it be possible to achieve significant successes in such a new and difficult field as machine translation.

It is encouraging to find that productive work on this subject is going ahead in so many countries. During the past few years dozens of papers and specialized works have been published on this subject. Our own Conference certainly does not nearly cover everything that is being done in it, although all the various groups and lines of work are more or less represented.

The paper by Dr. Giuliano and Dr. Oettinger describes the fine work done at the Computation Laboratory of Harvard University on compiling an automatic Russian-English dictionary which permits word-by-word translations. Examples of experimental translations obtained with this dictionary are given. The extensive and detailed research which led to its compilation will be of undoubted value to all those working on the subject and this paper should obviously be regarded as relating to the first set of problems that I mentioned above.

Dr. Yngve's paper relates to the second set of problems, and gives details of the special system designed at the Massachusetts Institute of Technology to facilitate the recording and programming of translation algorithms. The system has been called the COMIT system and in conjunction with the special compiler and interpretative routines worked out at MIT, gives an automatic programming scheme for translation algorithms. It is, therefore, of considerable methodological importance for MT work. I should mention that a number of scientific centres in the Soviet Union are also dealing with the questions of standardizing the recording of translation algorithms and are devising an automatic programming system on this basis. Our literature describes proposals by various authors concerning the forms which the recording might take. The paper by Dr. Harper and Mr. Hays is interesting since it describes the use of machines not only as an aid in the construction of translation algorithms but also to construct a grammar and to establish a programme for structural analysis for MT purposes. The paper suggests a method for making an objective and accurate classification of the words in the language studied and for finding a table expressing the relationship between one word class and another. This extremely solid and interesting piece of work relates partly to the first and partly to the second set of problems.

The paper by Messrs. Takahashi, Wada, Tadenuma and Watanabe on the special-purpose translating machine constructed in Japan relates to the third set of problems. For

this machine, an algorithm for translation from English into Japanese was prepared which so far has only been used for translating texts of a limited nature, but its scope no doubt will be enlarged in the course of time. This "Yamato" machine is the first special-purpose translating machine to be built. The Japanese scientists who have only lately begun to concern themselves with MT, have scored notable successes on which they deserve congratulation, and I am very happy to note, in this connection, that their approach to the formulation of an MT algorithm has a great deal in common with that of some of our own experts in the Soviet Union.

Miss Belskaya's report gives an account of the algorithm she has constructed for English-Russian machine translation. The results of research on a number of genuine texts, together with a general description of the grammatical structures of English sentences on which the algorithm is based are given. Many detailed examples of the results achieved are also supplied.

I have indicated the scientific groupings in which the papers submitted to this conference can be classified; but other types of classification are possible apart from the purely scientific one, for example, that based on the four characteristic stages of computer development traced by Mr. MacWilliams, one of the leading specialists in this field. The history of the achievement of machine translation can probably also be divided into the same four stages:

- 1) "Talking," when everything lies ahead and we say "How wonderful our translation algorithm is going to be!"
- 2) "Complacency", when the algorithm has finally been constructed and we say how good it would be if only it worked.
- 3) "Enthusiasm," when the algorithm is really in operation and we say: "It works, although we still have a lot to learn."
- 4) The final stage—again "talking," when the comment runs: "It is working well at last, and we are happy we managed to produce it, but we would be happier still if we had made it in some other way. In any case we shall make the change in a new and simpler variant."

But to which stage each of the papers submitted here belongs I leave it to their authors to decide. One thing is certain, however; they throw light on the subject from a number of angles. They differ as regards the methods that have been adopted, the languages studied and the nature of the results obtained; and these differences, I feel, are not fortuitous. They reflect the situation now obtaining everywhere in the field of machine translation. This is true, at any rate, of my own country, where many people are fascinated by this novel and interesting problem and are exploring new avenues and testing their strength in various directions. Work is proceeding in Moscow and elsewhere, and it is still not clear who has chosen the best road and will get the best results. My own view is that in order to hasten success it would be extremely valuable to arrange for broad discussion of these problems, and for international co-operation by experts in the matter. This would also help to hasten the solution of the noble task of promoting understanding between nations.