[*Impact of Science on Society* 10 (1960), 16-25.]

MACHINE TRANSLATION AND THE HUMAN BEING

by

D. Y. PANOV

D. Y. Panov, Doctor of Science, Professor at the Moscow Institute of Physics and Technology, was the organizer and first Director of the All-Union Institute for Scientific and Technological Documentation of the Union of Soviet Socialist Republics. In 1954 he set up, in the institute, a working party for research into machine translation—the first of its kind in the USSR.

The research director of a large industrial corporation recently stated that, provided the cost of research into a particular subject was not prohibitive, he preferred to begin such research from the beginning, without looking up any results which might already have been achieved and published—because inquiries of that kind took up so much time and it ultimately cost more to solve problems on the basis of already-published results than to tackle them afresh. This statement may be an exaggeration; however, reference to the results of scientific research does constitute, in our day, a serious issue.

The number of scientific papers published throughout the world, and of publications relating to each of the separate branches of science, increases from year to year. In chemistry and the related fields, more than 50,000 books, articles, patent specifications, etc., are issued yearly. In other words, about 150 new publications dealing with chemistry make their appearance every day of the year, holidays included. Thus, to extract information required from this flood of reported results is obviously an overwhelming task. The difficulty is increased by the fact that the new papers are published in every language in the world, including a regular and rapidly increasing proportion in languages described as 'difficult of access'. The list of journals published in Russian, which many people still regard as a 'difficult' language, contains over 1,000 titles; about 700 journals are published in Japanese; and so on. Scientists would like to have access to all this material; but this is frequently impossible, even if they can secure the actual article, book or specification, because they do not understand the language in which it is written. And here the problem of translation becomes acute. As we know, translation is a costly and time-consuming business; and, what is even more important, there are not so very many competent translators in the world. A good scientific translation is extremely hard to obtain, because the translator is seldom familiar with the branch of science to which the article relates. When the article to be translated is a general one, the translator

often feels so sure of himself that he produces an entirely subjective rendering.

Moreover, scientific and technical material is not the only kind of matter that needs to be translated. General material is required as well—extracts from newspapers and general reviews, information given over the wireless, etc. Translations are needed in correspondence, whether by letter or telegraph; and, of course, they are essential to the successful functioning of international organizations. It is no exaggeration to say that the problem of translation is one of the most serious with which present-day mankind has to deal, since upon its successful solution depends our awareness of what is going on in the world, our understanding of other nations, the dissemination of knowledge and the possibility of establishing international co-operation in every sphere of intellectual activity.

MACHINE TRANSLATION

A very attractive way of dealing with the problem is presented by the use, for translation purposes, of electronic machines, similar to those employed for mathematical computation. At first sight the idea of translating from one language into another by mechanical means seems almost fantastic; but a minimum of reflection on the subject will show us that there is nothing incredible about it. A language is, after all, merely a specific system for giving material expression to ideas and concepts. The different meanings of Words, together with any variants of meaning, are given in dictionaries and grammars. There are no unknown words in any language, except those that are deliberately invented, and no concept that cannot be conveyed by the order in which the words are placed and by the connexion between the different words. Otherwise, a language would not fulfil its main function, which is to provide a means of human intercourse; it would be impossible to use it as a means of communicating one's thoughts to another person. But this, as we all know from experience, is being done all the time; in conversation and in reading we are able to discern extremely delicate shades of meaning, expressed in words. So obviously there must be some material means of establishing a precise, objective distinction between them. The essential thing, for this purpose, is that the distinction shall be drawn according to definitely established rules, and this is exactly what machine translation renders possible.

To mechanize the search for a required word in the dictionary is easy enough. All that is necessary is to write it out in some selected code (such as the Baudot telegraphic code) and then, with the help of the machine, to compare this word with all the other words in the dictionary. These other words, of course, must be written in the same code in the 'memory' mechanism of the machine. Difficulty arises when a dictionary does not give a

single, uniform meaning for the word—in the case of polysemantic words, for instance. But even here the question can be solved by properly analysing a few aspects of form.

A translator, when choosing between the possible meanings of any particular word he encounters, is usually guided by the context; that is, he considers the word together with whatever other words accompany it. For instance, the English adjective 'sweet' may have at least six different meanings in Russian. The choice between these meanings will be determined by the noun that this adjective qualifies. The Russian translation will use different adjectives to render 'sweet apple', 'sweet song', 'sweet words', 'sweet girl', and so on. But the right term can be selected by purely mechanical means, with the same formal symbols being used. It is entirely a matter of correct programming—of showing clearly and precisely what has to be done, and in what order. If this is established, the work can be carried out by a machine. Problems of this kind are more complicated than the calculation of definite integrals or the solution of differential equations with which machines frequently have to deal; but the machine is quite able to cope with them.

THE HISTORY OF MACHINE TRANSLATION

The first attempt to mechanize translation appears to have been made in 1933 by a Soviet citizen, P. P. Troyansky. He proposed the construction of a machine which would select and print words in the process of translation from one language into another or into several different languages simultaneously.[1] A certificate of authorship was issued to P. P. Troyansky for this invention, but it was not actually followed by the construction of the machine. This is understandable, since the necessary mechanisms did not exist at the time. Troyansky suggested that the translation should be made with the assistance of two editors. One would be acquainted only with the language from which the translation was to be made (the 'source language'), and would supply the text with the aid of 'logical analysis symbols' indicating the sentence-construction, the parts of speech and their form. The other, knowing only the language into which the translation was to be made (the 'target language'), would take the text delivered by the machine and convert it into current phraseology in that language, with the help of the accompanying logical analysis symbols. It is interesting to note, however, that even at this early stage P. P. Troyansky wrote that '. . . the process of logical analysis could itself be mechanized, by means of a machine specially constructed for the purpose . . .'.

American and English scientists began to concern themselves with the problem of machine translation in 1946. Initially they, too, considered that

1. *P.P. Troyansky's Translating Machine*, published by the USSR Academy of Sciences, Moscow, 1959.

the process would involve editing, at least at the final stage. In 1954 the International Business Machines Corporation, in co-operation with the University of Georgia, carried out a celebrated experiment in translation from Russian into English, using the IBM-701 electronic computer. Although only very simple Russian phrases were chosen for translation (see examples in Table 1) and the mechanized vocabulary was very small (250 Russian words), this experiment proved that machine translation, without the need of editing, was in principle feasible.

TABLE 1

Качество угля определяется калорийностью.	The quality of coal is determined by calory content.
Крахмал вырабатывается механическим путем из картофеля.	Starch is produced by mechanical methods from potatoes.
Обработка повышает качество нефти.	Processing improves the quality of crude oil.

At the end of 1955 the USSR Academy of Sciences, using a BESM electronic computer, made the first experiments in the translation of entire passages of a scientific text from English into Russian. For these experiments a vocabulary of approximately 1,000 English words was used, and here again no editing was required (see examples in Table 2).

TABLE 2

In problems of this type numerical methods become a necessity due to absence of other methods for getting the requisite information out of the differential equations.	В задачах этого типа численные методы становятся необходимостью, обусловленной отсутствием Дргих методов для получения необходимого сведения из дифференциальных уравнений.
Even in cases where explicit or implicit solutions are known, it is sometimes easier to obtain a numerical solution than attempt to calculate numerical values from the known solution. values from the known solution.	Даже в случаях, где явные или неявные решения известны, иногда более легко получить численное Решение, чем пытаться вычислить численные значения из известного решения.

Research in the field of automatic translation is at present proceeding in the USSR, England, France, Japan, China, Sweden and other countries. Algorisms of translation from English into Russian and Japanese, from Russian into English, from French and German into Russian and English, from Chinese, Japanese, Magyar and a number of other languages into Russian, and so on, are being studied.

TWO ASPECTS OF THE PROBLEM

It has now been demonstrated that electronic computers can be used to obtain translations, from one language into another, which do not require editing. But the first translations, carried out with universal computers, were clearly unsatisfactory. A computer used for translation purposes is working much below capacity, for translation enlists only a few of its resources. For this reason, if mechanical translation is to become worth while in practice, it will be necessary to construct special electronic translation machines, of the same type as the computers but of rather simpler structure in the sections performing arithmetical operations. And here there arises a dilemma, on the solution of which the further development of work on machine translation will depend.

The answer to the problem of machine translation depends on two factors —the actual machine, and the algorism of translation, which is indispensable for preparing the machine's work programme. Opinions may differ as to the relative importance of these two aspects of the matter, some people being inclined to give priority to the machine and others to the algorism. To put the question more simply, we may say that holders of the former view would be in favour of constructing a more powerful machine and using a simpler algorism, while holders of the latter view would prefer to work out a more elaborate and flexible algorism and to use a less powerful machine. The matter must be considered, to a certain extent, in relation to the problem of the vocabulary for automatic translation. There may be one large dictionary, aiming at universality and suitable for the translation of texts dealing with a variety of subjects; or there may be a series of restricted dictionaries, known as 'micro-dictionaries', each of which is intended for the translation of texts dealing with only one field (physics, radio-technology, mathematics, etc.); when a text is to be translated, the appropriate micro-dictionary must be inserted in the machine. A translation machine with an extensive vocabulary must have a very capacious memorizing apparatus, and the time it takes to select words ('access time') must be brief. Such machines would have to be specially designed, for apparatus as adaptable and reliable as this does not yet exist. On the other hand, the existing types of memorizing mechanism are quite sufficient for dealing with micro-dictionaries, and the method of inserting them, when required, presents no difficulty. In short, if the trend of development is to be in the direction of powerful, highly-specialized machines, working with a comparatively simple algorism, they will have to be specially designed; if the trend is to be towards a more complex algorism, it will be possible to construct machines on the basis already established in the field of computers. We consider the latter course to be the better, since it would yield practical results more rapidly. The possibility of working out the necessary algorisms in this event is

clearly demonstrated by the example of the English-Russian algorism of automatic translation prepared in the USSR and tested with satisfactory results.

AUTOMATIC AND MACHINE TRANSLATION

The terms automatic translation and machine translation are often employed as though they were synonymous. But this is not quite the case. The term 'automatic translation' has a wider meaning than the term 'machine translation', since every machine translation is automatic, whereas an automatic translation is not necessarily made by a machine. It may be made by human beings, but in obedience to special rules leaving no initiative to the translator, who is thus obliged to work in a completely automatic manner. In both instances there has to be an algorism of translation; but for automatic translations, carried out by human agency, the algorism may be written out in ordinary language, whereas for machine translations it must be adjusted to the machine's work programmes written in the appropriate code. This algorism, without which there could be no machine translation, produces a further result: it establishes simple and objective rules for translation which, if scrupulously followed, ensure that the translations will be accurate and linguistically correct. It establishes its own translation norms, and produces standardized translation.

This seems to us to be of considerable significance. We have no doubt that, within a few years, machine translation will have great successes to its credit and that the machines will to a large extent relieve human beings of the task of translation. But it would perhaps be over-optimistic to suggest that even 15 or 20 years hence the need for 'hand-made' translations will have ceased to exist—particularly in view of the spread of scientific knowledge and education among the populations of a number of countries which, until recently, had no possibility of acquiring them. The population of those countries runs into hundreds of millions, and even on the most optimistic basis it is difficult to suppose that all the translation problems arising in this connexion could be solved by machines alone. But if human translators have to be brought in, the existence of rules of automatic translation will do much to ease their task and deliver them from the worst calamity that can befall a translator—the necessity of editing and correcting bad translations, which reduce the qualified translator to despair and take up more of his time than if he had done the work himself from the start. The algorism of an automatic translation, once it has been prepared, can be either employed for the preparation of a work programme for a translation machine, or published in book form, as a guide for inexperienced translators or even for those who do not know the language concerned, and as a work of reference for experienced translators.

THE ALGORISM OF AUTOMATIC TRANSLATION

The algorism which has so far been most thoroughly elaborated seems to be that used for translations from English into Russian. It consists of the following parts: (a) a scheme for looking up words and determining the meanings of words with more than one significance; (b) a scheme for analysing English sentences; (c) a scheme for constructing Russian sentences.

The last part of the algorism—the scheme for constructing Russian sentences—is the least interesting, since it works without difficulty in accordance with the established rules of grammar. For the purpose of automatic translation without a machine it is quite superfluous, since anybody acquainted with the Russian language can rapidly put together the required Russian sentence without the help of a system, from the data worked out by parts (a) and (b) of the algorism.

The dictionary work is covered by part (a) of the algorism. As already stated, we consider that the correct method to adopt is that of the micro-dictionaries. Experience has shown that a vocabulary of about 2,500 English words is sufficient for the translation of any text dealing with applied mathematics. This gives grounds for assuming that a dictionary comprising between 2,500 and 3,000 words is amply sufficient, and that quite a wide range of subjects can be covered by dictionaries of that size. The use of micro-dictionaries not only reduces the demands on the machine's memorizing mechanism; it also simplifies the analysis of the sentence to be translated, because in micro-dictionaries synonyms are cut down to a minimum, as also are the number of words with more than one meaning. In automatic translation, as has already been explained, the precise significance of a word with more than one meaning is ascertained by analysing the context. This process deserves to be looked into, for it explains the resemblance between the methods of solving problems of automatic translation and those applying to a number of entirely different logical problems. It supplies a typical example of dichotomy, very similar to the process adopted when determining the species of a flower or butterfly in botanical or entomological classification.

Table 3 illustrates the simplified system of selecting a meaning, applied to the polysemantic words 'many' and 'much'.

The symbols used in this scheme are as follows: the formula A (B, C) signifies that if the question asked at A brings an affirmative reply, we should proceed to check point B; whereas if the reply to A is negative, we proceed to check point C. The symbol A (0) indicates a final answer, which renders further checking superfluous. We will now see how the system works in specific instances.

1. 'The subject would have been *much* better standardized.'

In line 1 of the scheme in Table 3, we have to check whether the word

TABLE 3

1(2, 3)	Check the immediately preceding word for 'how'.
2(0)	Skol'ko (numeral, invariable).
3(4, 5)	Check the immediately preceding word for 'as'.
4(0)	Stol'ko ze (numeral, variable).
5(7, 9)	Check given word for 'much'.
6(0)	Not to be translated (adverb).
7(1,11)	Check the immediately preceding word for 'very'.
8(0)	Mnogij (adjective, hard stem, with a sibilant).
9(8, 12)	Check preceding word for preposition, and following word for noun.
10(0)	Mnogo (adverb).
11(12, 10)	Check following word for noun.
12(0)	Mnogo (numeral, variable).

'how' does not precede the word 'many' or 'much'. That word does not appear in the sentence, so we have a negative reply and must go on to line 3. Here we have to check whether the preceding word is 'as'. Again the answer is in the negative, so we proceed to line 5. Here we check whether the word is 'much'. The answer is in the affirmative, and we go on to line 7. We look at the preceding word and find that it is not the word 'very'. We proceed to line 11 and look at the following word. In this instance that word is 'better', i.e., is not a noun; so we must move on to line 12, which gives us the final translation, 'mnogo' (adverb).

2. 'This is most useful, and for *many* reasons.'
   We proceed in the following successive stages: 1—the answer is 'no', so we proceed to 3; 3—the answer is 'no' so we proceed to 5; 5—the answer is 'no', so we proceed to 9; 9—the answer is 'yes', so we proceed to 8; 8—gives the translation, 'mnogij' (adjective).
3. 'There should be as *many* equations as there are unknown quantities.'
   1—the answer is 'no', so we proceed to 3; 3—the answer is 'yes', so we proceed to 4; 4—gives the translation, 'stol'ko ze' (numerative).

These examples show clearly how the meaning of a polysemantic word is determined in the process of automatic translation, and also reveal the restrictions that the procedure naturally entails. If a word has too лапу meanings, the scheme becomes extremely complex, and its use as a working method would cause waste of time owing to the necessity of analysing a great many variants, only one of which would actually be needed. For this very reason it is undesirable to try to set up large dictionaries embracing a wide range of subjects. It is preferable to have highly specialized micro-dictionaries with a strictly limited number of polysemantic words.

The most important function of an algorism of automatic translation is the analysis of the sentence which is to be translated. For it is this analysis which yields all the grammatical features of the individual words covered by it, and makes a correct translation possible. And from the linguistic

standpoint the most important and interesting result achieved in the sphere of automatic translation is I. K. Belskaya's discovery that this part of the algorism can be applied universally, irrespective of the texts for which the algorism is actually intended. Belskaya demonstrated that the systems of analysing English, based on a study of texts in applied mathematics, can also be successfully applied to literary translation as well. For example, the scheme of analysis based on E. Milne's thesis 'Numerical solution of differential equations', Hardy's 'Divergent series', and so forth is perfectly suitable for the analysis of writings by Dickens, Galsworthy or Aldridge.1 It will only provide a standardized translation, with no pretensions to 'artistic' merit; but this fact shows very convincingly that, in solving the problem of working out an algorism of automatic translation, the linguist should combine in a clear and accurate system the chief elements in the language which form the basis of all writing, whether scientific or literary.

AUTOMATIC TRANSLATION AND LOGICAL PROPOSITIONS

Among the problems now being solved by computers, increasing importance is attributed to what are known as logical propositions. In these propositions it is sought to discover one solution or another depending on the fulfilment or nonfulfilment of certain conditions. The planning of production, the problem of making the best use of transport, and so forth, come under this heading. The problem of automatic translation falls within the same category. Earlier in this article we gave an example of one of the schemes used for the solution of that problem. And that scheme is a typical example of the handling of logical propositions. In addition to the linguistic aspect, which we have been considering up to now, the problem of automatic translation also has a mathematical aspect, connected with the preparation of a work programme for the machine, based on an algorism which has to be evolved. This question can be settled without necessarily considering the specific linguistic content of the various sections of the algorism, and we are then confronted with the need to devise means of solving extremely complex propositions in logic. Each successful construction of a programme for automatic translation will undoubtedly be equivalent to a success in the development of programmes for dealing with other logical problems.

CONCLUSION

Machine translation from one language into another is one of the most interesting scientific and practical problems confronting mankind at the

1. I. K. Belskaya, 'Machine translation methods and their application to (the) Anglo-Russian scheme', paper read at the International Conference on Information Processing. Unesco, 1959.

present time. Its solution is of interest to all nations and to all international organizations, particularly those like Unesco. International co-operation between scientists and manufacturers would be of great value in dealing with this matter, and it might be worth considering the possibility of setting up a special international scientific organization for that purpose.