[From: *Methodos* 8, 1956]

THE PRACTICAL REALISATION OF MACHINE TRANSLATION

ANDREW D. BOOTH

Birkbeck College Computer Laboratory,
21, Torrington Sq.
London, W. C. 1
England

[Received on April 28th, 1956]

It is nearly 10 years since the discussion between Booth and Warren Weaver which lead to the suggestion that mechanical translation might be a possibility. During this period much work has been done to reduce the naive ideas of the original discussions to a practical system whereby language can be translated. It is appropriate to look back now and see what has been achieved and to venture a prediction as to what is likely to be achieved in the next few years.

In the first place it is necessary to disentangle the subject from the mass of speculation with which it has been clothed since professional linguists became interested in it. In its original form machine translation was merely to be another application of a high speed computing machine. The principles upon which it was based were well understood and the philosophical problems, in so far as they existed, were not discussed.

The original proposition was simply that a computing machine, because it contains a large unit often known as the store or memory, would be able to hold in that store a dictionary, and since access to the store of a computer is extremely rapid, it might be that, by using a computing machine, the time wasted in looking up words when translating from a foreign language could be saved. This is as far as the original proposals went. There are no difficulties in principle in carrying out this scheme and the reason that no practical trial was made was simply that in 1947 no computing machine with an adequate store existed. By the time computing machines became more readily available, and certainly by the time that their stores became adequate, mechanical translation had become so invested with complexity that the original scheme was not deemed worthy of trial and some of the more exciting of the modern developments were still too difficult to achieve.

It has been said that the early proposals merely consisted in the looking up of words in a dictionary. Yet it is evident, on examining this proposition, that certain difficulties arise. Chief among these is the fact that when a human being examines a dictionary to find a foreign language word, he is using not only the dictionary itself, but also certain knowledge of the structure of the foreign language which he himself possesses. Without this knowledge it is almost impossible to translate a significant proportion of the words from a foreign language text using a dictionary alone, and the fact that it has been claimed that translation is possible by means of a dictionary is simply due to the fact that many of the foreign languages are similar in structure to the languages with which most translators are familiar.

This situation is brought clearly to the mind of the author of this paper by an example which occurred very early in his own school days when he attempted to look up the word Grosse in a German dictionary. Two difficulties arose, firstly that the dictionary was not written in the archaic script so that the letter ss could not be found in it, and secondly that in any case this word would not be given, but a simpler form, Gross.

The minimum requirements of a computing machine when applied as a mechanical dictionary, are that it should be at least as good as a real dictionary, and preferably that it should be as good as a real dictionary taken in conjunction with a user who is not a professional linguist, but who has a ‘smattering’ of the language.

The practical realisation of a plan of this sort is not difficult, and in the years between 1947 and 1950 R.H. Richens of the University of Cambridge, and the present author, evolved a system whereby the relatively small dictionary used in a high speed computer, together with certain rules of procedure would enable a significant number of words in a foreign language text to be looked up. The developments which made this possible were briefly as follows: first, the realisation that dictionary requirements could be greatly reduced if the dictionary itself were constructed on a new plan. This was simply that instead of one of the complete parts of the word which are normally given in a foreign language dictionary, there should instead be given merely the stem, or possibly stems, of the foreign language word. The definition of the stem, in this context, is simply that it is the longest segment of a word common to different forms of that word to which various endings can be added in order to produce real words in the foreign language.

At this early stage it was considered that the provision of stems would be sufficient. Endings could be dealt with either by looking up in a special ending dictionary and adding grammatical notes, or by ignoring them altogether. It will be seen that this division of words into steins and endings greatly reduces the size of the dictionary required to recognise a significant fraction of the words in a foreign language text.

Since the store of all existing high-speed computers is limited to a thousand words or so, each having about 10 letters, it is not possible to store anything like a complete dictionary of the language concerned. To overcome this difficulty it was proposed to store only those words which were relevant to the context of the passage being translated, and it was assumed that in most cases the person requiring the translation would be sufficiently familiar with the subject matter to know which types of word were going to be encountered in the given context. Sets of words peculiar to given subjects such as plant genetics, nuclear physics, X-ray crystallography, etc. were known as ‘microglossaries’. Apart from those words specific to the subject in hand, it was proposed to store also a number of words of general utility; these are sometimes called ‘cement’ words.

Statistical investigations seemed to indicate that with about 1,000 words of general utility and about 1,000 specific to the subject under translation something like 90 % of the words in most scientific texts could be translated, and it was, and is, maintained that this proportion is sufficient to justify the operation of translation on a computing machine.

It may perhaps be argued that the cement words are unnecessary. Many scientists ‘read’ foreign language texts by the simple process of looking up or recognising the nouns and some special verbs and also the symbols and diagrams which occur in the passage being translated.

Translation made by the stem ending procedure together with micro-glossaries would fall short of the sort of thing which the skilled human translator could produce.

These ideas of Booth and Richens received a small practical trial on a punched card installation. The results were not regarded as sufficiently important to justify extensive publicity and publication, but the experiments were reported to the conference on mechanical translation which was held at Massachusetts Institute of Technology in 1953.

During the period under consideration, work had been carried out on the other side of the Atlantic, notably by Bar-Hillel, Reifler, Oswald and Fletcher and Perry. This work paralleled that which had been done in England but to a large extent by-passed the simple procedures, which were advocated here and had preceded to far more sophisticated considerations which it was hoped might lead to the translation of language in a more literary fashion.

Among the ideas which were produced during this period it is perhaps worth mentioning those of Reifler on pre - and post - editing. The first stumbling block encountered, when any translation but the simplest is envisaged, lies in the difficulties of ambiguity which are inherent in the original text. It may perhaps be argued that a good author would not commit ambiguities. This, however, is not true; it is quite easy for an author to write passages which are to such an extent peculiar to the language in which he is writing that it perhaps is not immediately recognised that they are ambiguous. Examples from English are the expressions ‘She cannot bear children’ and ‘These men are revolting’. These sentences as they stand have two completely distinct meanings, and it is quite impossible to decide which of them is intended without reference to the context. At the same time the sentences concerned are neither badly written nor are they particularly unfortunate examples of their kind.

Reifler’s pre-editor was to be a person versed only in the foreign language from which it was required to effect translation. His function would be to remove all known ambiguities which the language itself contained. Since he would not have to be expert in any language but his own it should not be difficult to obtain a sufficient number of such people to do the work concerned, and it was considered that editors might perform the function of preediting before papers appeared in their journals.

The post-editor was to be a native of the language into which translation was to be effected. He would not necessarily know the foreign language although it would be an advantage that he should, and his function would simply be to render the output of a translating machine into acceptable prose in the target language. Both Reifler and most of his colleagues now appear to have abandoned the ideas of pre- and post-editing.

In the original Richens-Booth proposals the post-editor was to be the scientist for whom the translation was made, and it was assumed that if he had sufficient knowledge of the subject of translation he could resolve ambiguities in the machine output, either on his own, or possibly with certain auxiliary reference to the machine.

Up to the 1953 conference, when linguists had deigned to be interested in machine translation at all they had tended to be extremely sceptical of its possibility. It is still true that many workers in the field appear from their publications and utterences to be more interested in demonstrating the difficulties of the subject rather than in putting the matter to the practical test. This attitude perhaps marks the difference between the English and the American schools. In England it was proposed to conduct limited tests to see what could be done, and thereby obtain encouragement for further efforts; whereas in America, it appeared to be thought not worthwhile to conduct any tests at all unless they could produce a completely literate form of output.

We find ourselves completely in disagreement with the American point of view. It seems to us that unless the experimental method, as it is known and understood in the exact sciences, can be applied to this new science of mechanical translation then it is unlikely that translation by machine can become a really practical operation. It is easy to object that it is known *a priori* that the output will be imperfect and therefore that nothing at all should be done. This is likely, eventually, to bring the whole subject into disrepute.

After the 1953 conference in America one practical experiment on translation from Russian was conducted. It made use of a limited vocabulary of 250 words and had the desirable feature that certain grammatical rules were imposed upon the output from the machine so that the resulting prose was acceptable in form. Apart from this experiment little practical work seems to have been done in the United States.

In England, the steady progress of simple experiments has culminated quite recently in a public demonstration of machine translation from the French. The French language was chosen for two reasons; firstly that it is structurally similar to English, and secondly the purely utilitarian one that it is a language in which most scientists, and in particular the author of this paper, are reasonably well versed. The latter condition is an important one since, at the moment, there appears to exist in this country at least no linguist sufficiently familiar with the operations of high speed computing machines to be able to prepare his material for machine use. If the intermediary of a mathematician or physicist has to be used then it is rather important that both the linguist and the machine programmer should understand one another. Otherwise it is extremely unlikely that an efficient programme will be produced.

For the limited experiment in French translation a number of words were selected from a standard French literary text. No particular effort was made to construct a micro-glossary of a particular scientific subject since it was considered that any production of a given author would be sufficiently homogeneous to make a sample taken from the text into an adequate micro-glossary.

The first experiment to be conducted simply put the stem ending dictionary in the store of the computing machine and produced an output from any input supplied to the machine in terms of this stored information. At this stage the effect of the ending was simply to add a set of grammatical notes following the translation of the stem. The second phase in the operation was to make the ending which has been detached from the stem modify the English output so that instead of grammatical notes the English output resembled that which would be produced by a human translator. This means, for example, that the word ‘ayant’ which originally would have been translated ‘have’ (p.p) would now appear as ‘having’. It is apparent from this simple example that the machine has not only to obtain a single word equivalent translation of the original foreign language word but also to add to the English text certain words which are required to represent the meaning contained in the French text. From the machine point of view this implies that a word occurring at the present time may in fact modify the translation which has been already produced, and this means in turn that it is no longer possible to contemplate anything in the nature of word by word translation. The very least that can be done is to prefix each of the dictionary words with a note which tells the machine that the present word cannot be translated at all as it stands and must await succeeding words before the essential ambiguities presented can be resolved. This idea, which had originally been evolved in connection with the translation of idioms, for which it supplies a complete solution, now finds itself easily applied to the problem of adding or subtracting words from the original text before this is produced in translation.

This particular operation of the machine has been shown to work in a satisfactory manner and from this point onwards progress has been made towards simplifying the actual way in which the machine is to make the translation itself. It was mentioned that early attempts were directed towards splitting foreign words into stems and ending where these are appropriate. Now whilst it is true that a machine can very conveniently detach the longest stem from a word and look this up in a dictionary by purely arithmetic processes, and can also perform the same service for the ending, it is, nevertheless, not immediately apparent that this is the most satisfactory method of procedure. The dictionary, which has to contain all of the endings which are required, may be of considerable extent, and unfortunately it is still true to say that the store of a computing machine is far too limited in size for the requirements of machine translation. It follows that anything which can be done to reduce the storage requirements for the translation problem will be welcome.

Progress towards the solution of this problem was made by L. Brandwood in this laboratory, who observed that it is not necessary in the French language to consider the whole of an ending, in order to determine what modification must be made to the English stem to produce the correct output. He has shown that a simple set of rules which examine, not the whole ending but simply selected letters from the ending, leads to a simplified machine programme which, nevertheless, can produce under suitable circumstances an adequate translation. As an example of the sort of thing which is meant the following refers to verbs:

1) A stem which ends in *r* or *re* implies the infinitive and 'to' should prefix the root ending.

2) The ending '*ant'* implies the present participle and 'ing' should be added to the root meaning.

3) The fact than an ending contains the letter 'r' is sufficient to indicate that it is either future indicative or conditional.

The work carried on in this laboratory on machine translation is a part of a continuous process of development, it is only now that we feel ready to put the matter to the practical test of our linguistic colleagues. It is felt that, with all its imperfections, what has been achieved is significant. Although it *is* possible by indulging in the more obscure forms of expression in French, to arrange that the machine will produce non-sensical output, it is, nevertheless, true that the principles whereby these ambiguities and difficulties could be resolved are now well understood, and it is only shortage of time and machine storage space which prevents any situation conceived by the linguist from being provided for by the programmers for the machine. This does not mean that it is now possible to effect literary translation from the French; it means that within the terms of reference which we have set ourselves limited experiments can now be conducted on small segments of French text, and these will produce an adequate translation. We think, furthermore, as new situations are presented the present methods will enable difficulties to be overcome, at least in principle, providing that the new programme required can be held in the store of the machine.

From this point it is proposed to do two things: firstly to turn over the programmes so far developed for the French language to linguists for experiment, so that defects of the methods can be ascertained. Secondly to study the application of methods which have been applied to the French language to certain other languages, notably German, Latin and possibly Greek.

Some other points have been occupying the attention of our linguists and mathematicians in recent months, and these concern first of all the possibility of multiple translation, from a number of languages into a number of different ones, and secondly the general proposition of the economics of translation. Consider first the question of translation from anyone of n languages to any other. If a set of dictionaries is constructed to enable translation from any language A into any language B or vice-versa to be achieved, then it is clear that for n languages $(n — 1)^2$ of these dictionaries will be required.

It has been suggested that a way out of this large scale production of dictionaries would be to construct some intermediate or meta-language into which any of the languages could be translated and from which translation could be effected to any other language. This would involve the production of only 2n dictionaries. The difficulties of producing such a meta-language are, however, considerable. In the first place it seems unlikely that any existing language would be adequate. The language has, first of all, to be well adapted for coding on a machine, and this cannot be said of any real language known to the author of this paper, since the word structure of real languages involves great redundancy.

The construction of a new artificial language which resembles a real language has occupied linguists and scholars for a millenium at least and it seems very unlikely that agreement could be reached on a universal language to serve the purposes of machine translation. This being so it is necessary to enquire whether or not there exists some coding system which will enable the functions of an intermediate language to be performed without the necessity for the construction of the language itself. Work has only just started on this particular aspect of the subject and it rather appears that something which resembles the universal decimal classification or one of its variants may form the solution to this particular problem. The idea will be that a segment of foreign language text will first be scanned by the machine to find the ideas which it contains. For example, ‘the dog ran towards the man’ from which it can be seen that there are nouns which express the idea of subject and object, a verb which expresses the idea of motion, and a sentence structure which gives an indication of the way in which subject object and verb are tied together. Now this aggregate of ideas could possibly be related to a basic dictionary in which such ideas are classified. From this point it may be possible to express the same ideas in any of the languages into which it is required to produce an output. The difficulties are great but nevertheless there seems cause for hope when it is considered that a number of coding systems are at present in use for roughly the same purpose, that is, the reduction of ideas common to a large number of fields to a basic minimum which again can be expanded into the fields in which they are required.

The second of the points mentioned, that of economic feasibility, can be answered with a fair degree of precision at the present time, but the answer which is given cannot be regarded as at all satisfactory. In operation as a translating machine a computing machine is hopelessly inefficient for a number of reasons. In the first place the storage organ is too small, in the second place, paradoxically, it is usually too fast. The smallness manifests itself not only in the numbers of words which can be accomodated, but also in their size. There seems no doubt that these disadvantages could be overcome, by means of suitable techniques involving either magnetic drums, or photographic imprints on micro-film or micro-disc. Both of these systems give the flexibility which is required and will also give an access time which is such that any word can be located, and dealt with from the syntactic point of view, in a time which is of the order of 1 second.

On the output side printing devices are now available which could certainly cope with the printing of any word which is required in a time of the order of one second. It is, therefore, not unreasonable to assert that a machine which could be produced at the present time could conveniently translate at the rate of about one word per second, that is at about the rate of 3,600 words per hour. This rate of translation, even if the product were of such a quality as to compete with human operators, is not impressive since it must be remembered that computing machines cost, at the moment, about L. 30 per hour in operating time. It is dangerous to predict the cost of human translation but it is quite safe to say that it is very considerably less than this figure, probably about l/10th. Naturally with developing technology the price of machine time will be likely to decrease. When this happens translation will become economically more feasible. However, it is not in this aspect of commercial utility that the strength of mechanical translation is likely to lie. The really important thing is that once provided with a translating machine and with the basic programme and dictionaries, translation can be effected from and to any one of a wide selection of languages. This means that a central organisation, provided with such a machine, could supply all of the translations which are needed. These would be available at any time and would be independent of human operators. Human translators are difficult to obtain, particularly when it is necessary that they have some knowledge of a scientific field. It is perhaps not too much to say that for the more exotic languages translators might be almost impossible to come by, and it is in this field particularly that the machine will come into its own.