[From: *Aspects of translation* (Studies in Communication 2). London: Secker & Warburg, 1958]

The History and Recent Progress of Machine Translation

A. D. BOOTH
Reader in Computational Methods in the
University of London at Birkbeck College London

HISTORICAL

It is no doubt true to say that somewhere during the course of history the great philosophers must have considered the possibility of effecting translation from one language to another by means of a machine. There is no written evidence, however, that this suggestion was in fact made, and it is merely speculation on my part that, for example, Leibnitz would have been a likely originator.

The suggestion that modern computing machines could be used for the purposes of translation originated with the present author. It arose because, in 1946, various new uses for automatic digital calculating machines were being considered and these ranged from the more obvious applications to problems of mathematics and physics, to philosophical problems such as the mechanization of human thought processes, the playing of games and the translation of language.

By 1947 the problem had been put into quite formal language and was discussed in detail by Booth and Warren Weaver. Up to that time all that had been produced was a programme which would enable a computing machine to perform look-up operations which a human translator would perform with a dictionary.

Although in 1947 there were high hopes that an automatic digital computing machine would be available within the next few months, these hopes did not materialize and it was not until 1950 that the first working machine was produced. In the intervening period, Booth and Richens had considered in greater detail the structure of a dictionary which might be used with a computing machine and had given definite form to their ideas in a paper which was circulated to the people known to be interested in this aspect of the art.

American workers were not particularly active in the field and considerable opposition to a memorandum by Warren Weaver was forthcoming, not only from linguists, who might be expected to rebel against any suggestion that their art was a purely mechanical one, but also from such people as Norbert Weiner who, although well known for his progressive ideas in some other fields, did not appear unduly excited by the thought of machine translation. The Weaver memorandum was widely circulated in 1948 and was taken up with enthusiasm by Bar-Hilel at the Massachusetts Institute of Technology. Even so it was not until 1950 that Reifler began active work on machine translation, and this culminated in a set of memoranda which, in the absence of a suitable medium for ordinary publication, were circulated privately.

In 1951 the ideas of Booth and Richens were given a practical trial on standard punched card machinery. The results were exactly as was predicted, and the faults which were known to exist in this system of translation were quite accurately reproduced in the experiment. On the credit side it should be mentioned that the results were still considered worthy of further development.

Work in Great Britain was greatly hampered by the complete absence of any formal backing for research in the subject, and the two active workers in the field at this time were able to devote only such time to the project as could be spared from their normal University duties.

In 1953 the Rockfeller Foundation took a practical step by financing a Conference held at the Massachusetts Institute of Technology to which all active workers in the field were invited. The results of this Conference were in some ways interesting. There was the expected opposition to the new ideas by linguists. There was considerable discussion, particularly by Bar-Hilel, of the extreme difficulties attending any form of machine translation, and there was the hopeful attitude adopted by the British workers that the time for practical experiment had come. No formal conclusions were published by this Conference, but there was a measure of agreement that if further work was to be carried out, it should be devoted firstly to the production of microglossaries and secondly to statistical investigations directed at ascertaining facts about language which had hitherto not been produced by the linguists.

After the Conference, a spectacularly publicized experiment was conducted by the I.B.M. Corporation in conjunction with Georgetown University. In this, selected Russian sentences were translated, using a restricted vocabulary of about 250 words. This appears to be an isolated demonstration as the group concerned have not since engaged in active publication in the field.[1]

In 1955 the Nuffield Foundation made a generous grant to Birkbeck College, University of London, which enabled the project of machine translation to be taken up on a full-time basis. In particular, it was possible to obtain the services of young professional linguists whose ideas had not, through channelization, been so stultified as to make them incapable of appreciating the new techniques. Of the work of this group more will be said later, but it may be remarked here that the progress was spectacular and

[1] Since the above was written a number of reports have been circulated by the Georgetown University Institute of Languages and Linguistics.

the first language chosen for examination, French, was, to quote the words of one of the linguists, “resolved for all practical purposes” by the end of 1956. The work of the group is now being extended in a number of directions and particularly to a study of the more difficult problems of German.

In the United States, after the Conference, more work on a purely logical plane was done, much of this pointing out the difficulties which would beset the path of the would-be machine translator. The Rockefeller Foundation made possible the adoption, by the Massachusetts Institute of Technology, of full-time workers in machine translation, and this in turn led to the production of the first book on the subject, and to the periodical publication of a newsletter (‘M.T.’) devoted entirely to papers on machine translation.

Important papers by Reifler, Oswald and Fletcher, and others, considered the theoretical aspects of machine translation, but workers in the United States do not, up to now, appear to have made practical use of a computing machine to do translations, although it must be said that the United States Air Force is at the present time constructing, at Rome Base near New York, a special dictionary storage machine which will be of incalculable value to the workers in the field.

In 1955 the Russians also became interested in the field and it was a surprise to most workers in the Western Hemisphere when, in 1956, a paper was read at a convention of the Institution of Electrical Engineers describing work on translation conducted at the University of Moscow. Certain examples, stated to have been produced on the Moscow machine B.E.S.M., were shown, but there appear to be good grounds for assuming that these cannot in fact have been directly produced by a machine. On the other hand, the description of the methods of producing a translation given by the Russian workers are of high quality, and parallel very closely the ideas of the British workers.

During 1956 numerous demonstrations of actual machine translation were given at Birkbeck College on the computing machine APEXC and although these do not satisfy all of the aesthetic requirements which may be desired of an ultimate experiment, they are nevertheless admitted by most people who see them to be impressive.

To sum up: at the present time it can be said that the British approach has been one of practical experiment taking the line that only by such means is progress to be attained; the American attitude has been to discuss the difficulties and to try to devise means of overcoming them before any practical experiment is made, which to us seems a method of doubtful virtue; and the Russian approach, so far as can be gathered from the scanty publication, is closely parallel to the British one. Apart from these groups of workers, there appears to be no really active work on this subject being carried on elsewhere.

THE MECHANICAL DICTIONARY.

The keystone of any translating process is the dictionary, and this was the first item to be considered by Booth in 1946, and later by Booth and Richens in 1948 and 1949. The early proposals considered only the storage of a dictionary of conventional type in the memory organ of a computing machine. It was realized of course that such a dictionary would be of rather limited utility in practice. This is because when a human linguist makes use of a dictionary, he is not merely utilizing the information contained in the dictionary itself, but also a considerable basic knowledge of the language from which translation is to be effected. To make this clear consider the following example: In a text the word *cherchais* is encountered. Every schoolboy is aware that this is a form of the imperfect tense, first person singular of the verb *chercher*. On the other hand it will be sought in vain in a dictionary of the French language. In such dictionaries all that will be found is *chercher* and meanings ‘to search’, ‘to seek’, ‘to look for’, ‘to be in quest of’ or ‘to mind’—to quote from one particular dictionary. The schoolboy, however, is aware that *chercher* is a regular verb and behaves in a certain manner, so that he can immediately associate the infinitive given in the dictionary with the inflected form given in the text and thereby produce the translation ‘I sought’. Not so, however, a machine, which could only find the exact equivalent of the unknown word among the store of words with which it was acquainted.

What, then, is to be done? The suggestion of Booth and Richens was that, instead of constructing dictionaries in the conventional manner, these dictionaries should be constructed on a new principle. The principle was that, instead of the infinitive of verbs, the singular forms of nouns and so on, there should be stored the stem of these words, the stem being defined as the longest segment of a given word which is common to all of its parts. It must be borne in mind that for irregular verbs like *avoir* and *être* there may be several stems but this poses no special difficulty for the machine, except to make necessary the storage of the same translation in those positions associated with various forms of the stem.

This simple suggestion led to a great improvement in the possibilities of machine translation. For example, it was shown that the stem-ending procedure made possible the considerable restriction of the number of words which it was required to store in the dictionary. Computing machines, even at the present time, have storage capacities which are limited to a few thousand words, and since a computer word is often much shorter than a word of real language, several computer words may have to be used up to contain information about one actual word. Thus even limited storage capacity is further reduced and if all of the parts of a verb such as *chercher* were stored, a very large number of positions would be required just to cope with a single word. Suppose that N words are to be comprehended by the machine, and that each of these words can take M common endings, it follows that if all of the parts of each of the words is written in the store, N x M different entries will have to be used. If, on the other hand, the words are split into stems and endings, then only (N + M) positions will be required, N for the stems and M for the endings.

In the early proposals, the ending was to be used to produce certain grammatical notes which would be attached to the translation of the stem, so that, to quote the earlier example, *cherchais* would appear as ‘Seek (1 P S I)’ where the letters within the bracket indicate that the part is first person, singular, imperfect. From these indications anyone who has a slight knowledge of English grammar could reconstruct the word in its proper form. It was also shown that by adopting the stem-ending procedure, it became possible to treat complicated compound words which occur in such languages as German. In this case, however, having looked for the first stem, considered as the longest dictionary structure contained in the word, it must not be assumed that what is left over is an ending, but a further test must be made to see if this can be found among the stems. It need hardly be remarked that occasional difficulties will arise due to the fact that certain words by themselves are incapable of exact translation.

An example given by Richens was the word *desideremus* which could be decomposed into either *desid-eremus* or *desider-emus* with meanings ‘desired’ or ‘be idle’. At the time it was considered that such words would be so uncommon that they could be ignored. In fact, in the most recent work, means have been devised by which, in suitable cases, these difficulties can be resolved correctly. Having said so much about the dictionary as originally proposed by Booth and Richens, it is perhaps worth completing this part of the account by stating that Brandwood, in 1956, extended these ideas so that, instead of producing grammatical notes, the dictionary was made to produce the English output in its correct form. This is really an obvious extension of the original Ideas and it is surprising that it so long escaped discovery. What is done is the following: Opposite the foreign language stem in the dictionary, is contained the stem of the English language translation. When the stem has been detached from the foreign language word, the ending is now looked up in a separate dictionary of endings and in this will be contained any prefix and affix which must be added to the English stem in order to produce the correct text. For example to take the Latin word *amat*, this consists of the stem *am*, whose translation is ‘lov’, and the ending *at.* In the dictionary of endings the letters *at* would be accompanied by an indication that the English output is to be prefixed by the word ‘he’ and affixed by the letters ‘es’, so that the output would now appear as ‘he loves’. We ignore for the moment the difficulty that the prefix could be ‘he’, ‘she’ or ‘it’, but enough has probably been said to make clear the way in which a dictionary is used.

THE TECHNIQUE OF DICTIONARY SEARCH

The next important point concerns the actual mechanics of dictionary hunting on a computing machine. The earliest proposal for this was simply that each letter of the foreign language word should be given a number, so that, for example, A=1, B = 2, C = 3, and so on. In this way the foreign language word is converted into a number and these numbers could, in principle, be considered to indicate the position in which the word is stored in the machine memory. A moment's consideration, however, will show that this process is not workable, for even suppose that words were restricted to a maximum length of ten letters, which they are not, then each letter can take twenty-six different forms and the number of possible words, which is the same as the number of storage positions, is 26^{10}, which is approximately 1.4×10^{14}. This number is greater by a factor of about 10^7 than the size of any storage device which can be conceived, let alone built at the present time. It may be argued that there are not 1.4×10^{14} words in any language, particularly when their number is reduced by the use of the stem-ending procedure. This, however, does not affect the case, since, if the simple coding of foreign language words into positions represented by their numerical equivalents is adopted, all of this space must be provided.

To overcome this difficulty, the first suggestion was that each foreign language word should be compared by subtraction with the contents of a dictionary whose entries were arranged in ascending order of numerical magnitude. When this is done, it follows that, when the unknown word is first presented to the dictionary, it will encounter words of lesser numerical magnitude, so that the result of the subtraction will be negative. In the latter part of the dictionary, the reverse will be true and the result of the subtractions will be positive. When, however, the stem or the whole of the unknown word is encountered in the dictionary, the result of the subtraction will either be just about to change from negative to positive, or zero. Thus the hunting process simply consists in subtracting the foreign language word's numerical equivalent from the dictionary entries starting from the beginning and noting either that point at which a change from negative to positive occurs, or alternatively the point at which the result of the subtraction is zero. In the latter case, the zero subtraction gives the position of the word in the dictionary, and, in the former, the last negative subtraction gives the stem position. This process was, in fact, that used in early experiments. It is clear, however, that it requires, on the average, the examination of one half of the dictionary entries in order to define the position of the unknown word. The searching time on the Birkbeck College computing machine APEXC is about fifty comparisons per second so that for a thousand-word dictionary where five hundred examinations must, on the average, be made, the time for hunting would be of the order of ten seconds. This is quite long and could in fact be bettered by a human operator using even a rather large dictionary.

The first significant improvement in this situation was made by Booth in 1955, when the method of 'bracketing' was proposed. In this, use is made of an idea which is, in other contexts, of considerable antiquity. The unknown word is first subtracted from that dictionary entry which is half-way between the start and finish of the dictionary. The result of this subtraction may be positive or negative. If positive it is at once known that the unknown word lies in the first half of the dictionary, if negative, in the second. The process is now repeated but this time on the word at one quarter or three quarters of the dictionary, and so on, by successive partitioning into fractions of 2^{-n}, the unknown word is located unequivocally. It is easily shown that using this process an average of $\log_2 N$ hunting operations will be required, where N is the total number of entries in the dictionary.

To make clear how this effects an improvement, it may be pointed out that with the dictionary of one thousand words previously mentioned, about ten look-up operations will be required, and this would take, on the APEXC, one fifth of a second, which is fast compared with any human process. The bracketing method has the great advantage that if the dictionary size is increased, the look-up time increases only in logarithmic proportion. Thus, for a dictionary of a million words, under the ordinary successive comparison method, it would be necessary to examine about half a million words in order to locate the unknown one, and this would take a time of the order of ten thousand seconds on APEXC. Using the bracketing method, however, only about twenty look-up operations are necessary and the time for these would still be of the order of only half a second. In fact the look-up procedure is not quite as simple as that described. The discrimination at any stage tests not only whether the result of a subtraction is positive or negative, but also whether it is zero and under these circumstances it is easily shown that the number of look-up operations will, on the average, be only $\log_2 N - 1$, so that again an improvement is effected. It has been suggested by Tocher that it might be worthwhile dividing any large dictionary into two parts, the first containing only the words known to occur most frequently, and the second containing special and infrequent words. In this way the most frequent words would be hunted in a comparatively short time and the result might be a more efficient process. A detailed examination of this method, making use of the Zipf-Estoup law, has shown that the optimum partitioning would in fact produce a saving of at most one look-up operation, so that it appears hardly worth while in view of the complication of the machine programme which would result.

GRAMMATICAL ASPECTS OF MACHINE TRANSLATION

The means so far described will produce only a comparatively inferior translation in most cases, even with a language so structurally resembling English as French. It would produce results of no value at all with languages of different structure, such as Latin and German. The means by which such languages can be treated however, is not difficult to conceive, and at the present time programmes for handling this type of work are in course of production.

The principles involved can be seen quite clearly even with the French language, and we shall take this as the first example. The chief defect of word-for-word hunting in a dictionary lies in the fact that many foreign languages have a different word order from that encountered in English. In French, for example, the order of nouns and adjectives is frequently the exact reverse of normal English usage. An improvement would thus be produced if the machine, whenever it encounters a noun in the French language, does not produce an immediate output, but examines to see if this is followed by a qualifying adjective and, in the latter event, produces first the translation of the adjective and then the translation of the noun; for example, in the phrase *une equation differentielle,* it would be required to produce the output ‘a differential equation’ instead of ‘an equation differential’. Rearrangement of this sort is only a particular case of a much more general and complicated process, such as the rearrangement of verb-pronoun structures in *je ne vous le donne pas.*

The way by which such processes are performed is quite simple. In the dictionary, accompanying each foreign language word and its translation, is a storage space in which can be written an indication of the part of speech or other structural function of the word concerned. When a word is encountered, this structural position is examined, and its contents are compared with entries in a dictionary devoted to structures. In this way, if an unequivocal output is possible, the machine will immediately translate the given word. If this is not the case, however, the process of absorbing words from the input of the machine continues until sufficient structural symbols have been associated together to define an output configuration and, when this has been done, the translations of the input words are reproduced at the output in that order which is dictated by the dictionary of structures. There is no technical or conceptual difficulty in programming this process for a computing machine, and, for a language such as French, the actual dictionary of structures is quite limited.

When German is considered, however, a more difficult situation presents itself, and here the recent work of Brandwood has shown that, for existing computing machines, it is very likely that only a limited number of the different patterns encountered could be handled by the machines which are likely to be available in the near future. This leads naturally to the thought that certain incomplete translations may be acceptable. Unfortunately, however, existing linguistic data give very little guidance on the point, and at Birkbeck College considerable work is being carried out to clear up the question. To do this it is necessary to make analyses of the frequency with which different sentence structures occur. Such analyses do not appear to have been made, at least on the scale required, in the past and considering the time consumed by human linguists in studies of this type, it was thought necessary to have available some speedier means of analysis. For this reason, a recent activity of the group at Birkbeck College has been to devise methods by which a computing machine can analyse tracts of foreign language text presented to it and to produce statistical data of the type required. Quite apart from this statistical work, the machine will produce at the same time either a concordance or alternatively a dictionary of the words peculiar to the text under examination. Furthermore it will produce these dictionaries in the exact form required by the machine for subsequent use in translation. It is worth noticing that the early idea of using a microglossary, that is a list of words peculiar to a given subject, is likely to be necessary for some considerable time to come unless a great revolution in the design for storage organs for computing machines occurs.

AMBIGUITIES AND IDIOMS

Some consideration has been given to the question of the resolution of ambiguities by machine. Ambiguities are of several types. There are the simple ones in which a word has different meanings according to the subject in which it is used, there are more complicated ones in which a word associated with other words has a different meaning—what may be described as a structural idiom—and there are cases in which a word can have completely different meanings according to the context in which it occurs.

The resolution of ambiguities of the first sort had been clearly envisaged by Booth and Richens. They were to be treated either by outputting all of the meanings so that the appropriate one could be selected by a reader, or alternatively by the use of the microglossary. In this the particular word is given only that meaning which is known to be appropriate in the subject matter which forms the material of translation. When this subject matter is unknown, it is proposed that a preliminary analysis should be made to find out what it is about, and this process is facilitated by the use of what we now call context numbers. These are simply numbers which accompany any given word which may occur in different subjects and which reduce to a single number when the word is unambiguous. Thus the word *noyau* can have the meanings ‘nut’ (1), ‘nucleus’ (2), ‘centre’ (3), ‘kernel’ (4) in different contexts. If each of these meanings is given a particular number, then, for example, the meaning ‘centre’ (3) would quite often occur in a sociological context and if the machine runs through the given passage and counts up the number of occurrences of each particular context number in that passage, it is likely that a maximum count will occur for context number (3), and under these circumstances all other types of context number would be deleted. The machine would thus output unequivocally the meaning under context number (3).

This involves an initial run through the passage concerned, but an alternative procedure accumulates occurrences of context numbers progressively as the text is processed. In this way, at the start, a number of alternatives will be given for each word, but as translation proceeds, these will become more and more channelled into a particular line and eventually unambiguous translations will be given. Yet again, as it will almost invariably happen that the subject of translation will be known beforehand, if it is desired to use a dictionary in which each word is accompanied by a number of meanings, the context number of the subject for translation can be fed to the machine before the unknown passage so that the machine can make an appropriate selection.

The second type of ambiguity, the idiomatic, can also be dealt with by the structure numbers mentioned previously. The phrase *boîte de nuit* serves quite admirably as an example. Literally translated it means ‘box of night’; in its idiomatic context, however, it means a night club. The machine instructions for dealing with an expression of this sort are quite simply as follows: whenever the word *boîte* is encountered, no translation must be effected until the next word is examined. If this is *de* still no translation is possible and the third word must be looked at. If this is *nuit,* then there is the unequivocal output ‘night club’. If however it were, for example, *chocolat* then the output would be ‘box of chocolate’—a perfectly normal translation of the words. Whilst this method of approach is quite simple in principle, it is unfortunate that the number of such idioms in any real language is rather large, and it does not appear at all likely that translating machines of the near future will have sufficient storage capacity to handle any extensive number of idioms of this type. Since, however, we, at least, envisage the use of the machine only for translating technical material, it is unlikely that the lack of idiomatic expression will be of much importance since good technical writers would not normally make use of idioms.

The third type of ambiguity has already been illustrated in the word *desideremus,* which is, in fact, two completely different words according to context. The situation is even more clearly revealed in the two expressions:

She cannot bear children

and

These men are revolting.

In these two sentences the words *bear* and *revolting* have two completely different meanings, and, even on the basis of the whole sentence in which they occur, it is impossible to ascribe a unique translation. The resolution of difficulties of this sort is very much more complicated, not in principle, but in the actual amount of programme material which is required. In principle, the difficulty can be resolved by the use of some form of context count. For the words illustrated above, these context counts might be, in the first case, gynaecology, giving a specific connotation to the word 'bear', and in the second case, revolutionary activity, again defining closely the meaning. Since, however, a good author would be unlikely to use sentences of this type, it is thought that no particular attention need be given to them. This is perhaps an appropriate point to mention the original idea of Reifler who suggested a pre- and post-editor for text. The pre-editor was to remove known ambiguities from the original text, for which purpose he need not have any knowledge of the language into which it was eventually to be rendered, and the post-editor was to clear up various ambiguities in the machine output. Reifler himself and most other workers now regard neither of these persons as necessary, but in the case of the two examples given above, a pre-editor would recognize the sentences as being ambiguous in his own language and replace the offending words by ones more closely defined.