[From: *International Social Science Bulletin* 10, 1958]

THE SOCIAL CONSEQUENCES OF AUTOMATION

TRANSLATING MACHINES

ANDREW D. BOOTH

Before the major subject matter of this paper is broached, there is one misapprehension which should be corrected: it is that, as far as is known to the author, no ‘translating machine’ as such exists at the present time. This does not mean that translation by machine has not been achieved but simply that up to the present it has not been found worth while to build a special machine for the purpose of translation. What then is the basis of the numerous reports which have been current to the effect that translating machines exist and have been used to perform translations from one language into another? To answer this question we may perhaps go back a few years into the history of our subject. The possibility that a machine could be made to translate first arose during discussions which took place in 1947 between the present author and Warren Weaver of the Rockefeller Foundation. The motivation for these discussions was the fact that high speed digital calculators were coming into existence about that time and that the mathematicians concerned with the design and use of these machines were trying to find applications wherein the new devices would show their superiority to the well-known business and accounting machines which were their predecessors. The earliest proposals for machine translation envisaged only that a language could be handled in a simple dictionary look-up manner. That is, a segment of text would be presented to the machine, and the machine would then see whether the individual words, considered as complete entities, were to be found in its storage organ. In this event the meaning associated with those words was to be typed out. Any words which could not be found directly in the storage organ would be output by the machine in the original language version. As an example of this consider the following collection of words:

L’étude des fonctions définies par une équation differentielle . . .

Should these words be presented to a computing machine, the first operation would be to look up the word *L’*. This would be contained in the machine’s dictionary and would be translated by some such output as *the.* In a like manner the word *étude* could be found as *study,* the word *des* as *of the,* but when the machine attempted to extract the word *fonctions* this would appear to be untranslatable since the word held in the dictionary would be the singular form *fonction,* that is without the *s*. In a like way the word *définies* could not be found. The resulting translation effected by the programme in 1947 would thus be:

The study of the *fonctions définies* by an equation differential . . .

It will be apparent that the method, though having a certain utility, left much to be desired.

On the return of the author to England, actual construction work was begun on a computing machine, and the interest engendered by these activities led to contacts with various other workers who were interested in the possibility of machine translation. In particular, Richens suggested that a way out of the difficulties of the direct looking up of words in the dictionary could be found by storing only the stems of the foreign language words in the machine. The stem in this context having its usual definition as the longest segment of a word common to all of its parts. Naturally, even at this early stage, cognizance was to be taken of the fact that for many foreign language words, for example the French irregular verbs *avoir* and *être,* it would be necessary to store several stems, depending on which particular part of the verb was involved. The suggestion of Richens was that, having stored the stem of a word, the ending was to a large extent irrelevant and even if this was not so it would be sufficient for the machine to make use of the foreign language ending to give some grammatical note regarding the precise function of the word in the sentence. Thus, for example, *fonctions* would appear as *function* (p) where the latter symbol indicates that the word is plural. In the same way a verb such as *cherche* would appear as *seek* (1 pps) which would mean first person, present tense, singular.

Unfortunately, in common with all the other high speed computing machines which were projected with so much hope in the middle 1940's, the machine at Birkbeck College did not become an operational reality until 1950, and even when this occurred the storage capacity of the machine was still such as to make it impossible to do useful machine translation. For this reason, work on machine translation in England proceeded slowly, and the only practical trials which were attempted made use of standard punched card machinery. These trials were so unspectacular, and gave results so much in accord with those predicted, that it was not thought worth while giving them any publicity.

In 1953 the first international conference on the machine translation of languages took place at the Massachusetts Institute of Technology. It was at this conference that workers in Great Britain discovered that there was considerable interest in machine translation on the other side of the Atlantic. So far as actual achievement was concerned, both sides were about even. In this country a small practical trial had been made but the methods of linguistic resolution which went into it were of a simple nature. On the other side of the Atlantic considerable linguistic investigation had been carried out which purported to show that machine translation was difficult if not impossible. To sum up the position at this stage, it is probably fair to say that the British workers considered that any experiment which could be carried out was worth while and might lead to an extension of the range of the subject, whereas American workers were concerned with the thought that machine translation would be difficult and that, furthermore, unless it could be done with extreme perfection it was hardly worth doing at all.

Following the conference, a practical demonstration of machine translation was carried out by the International Business Machine (IBM) Corporation with the assistance of Dr. Leon Dostert of the University of Georgetown. This effected a translation of some Russian sentences making use of a restricted dictionary of about 250 words. In 1955, work in this country was greatly stimulated by the generous grant, by the Nuffield Foundation, of an award which enabled Birkbeck College, University of London, to pursue machine translation on a more ambitious scale than had hitherto been possible. In particular, it made it possible for the computational laboratory to enlist the services of skilled professional linguists. With the advent of the linguists, and with the computing machine APEXC in a working state, progress on machine translation became very rapid. A number of machine programmes taking account of all the suggestions which had previously come forward were constructed by the present author and by Dr. K. H. V. Booth, and these formed a useful basis for the work. At an early stage it was necessary to decide what language or languages were to receive the full force of the linguistic attack. The decision for this selection rested with the present author. In view of the fact that the Americans were believed to be actively engaged in the study of Russian, and that the converse was true of the Russians themselves, it was decided to work on the French language. The reason for this decision lay simply in the fact that both the present author and his mathematical assistants were acquainted with this language and could therefore talk on the grounds of familiarity, if not of equality, with the professional linguists. In the course of some six months, Mr. L. Brandwood, Nuffield Fellow in machine translation in this laboratory, produced his first version of a set of grammatical and syntactical rules for the resolution of the French language. These were subsequently developed until now it is not unfair to say that the problems attending the resolution of French scientific texts have been solved. The further extension of this work depends not so much on the discovery of some new principle but merely on the demonstration by linguists of deficiencies in our existing programme and on the drafting of the rules on which the rectification of these errors is to be based.

Before going further we may give here a short example of the actual work produced by the machine when applied to a French text. The following example is taken from the first chapter of the third volume of Goursat's *Cours d'analyse mathématique* and the translation given below is that produced by the machine in accordance with the rules provided by Brandwood.

> Nous démontrons dans ce chapitre le théorème fondamental de M. Poincaré, après avoir étudié les intégrales d'un système d'équations differentielles, considérées comme fonctions des valeurs initiales. Cette étude a déjà été faite, dans le cas où les seconds membres sont des fonctions analytiques. Nous la reprenons dans le cas général, au moyen de la méthode des approximations successives de M. Picard.

> We demonstrate in this chapter the fundamental theorem of M. Poincaré, after having studied the integrals of a system of differential equations, considered as functions of initial values. This study has already been made, in the case where the second members are of analytical functions. We adopt it in the general case, by means of the method of successive approximations of M. Picard.

It will probably be admitted from the examination of this small example that the machine translation is indistinguishable from that produced by many human operators. Particularly is this true if the human operator is a linguist and not a professional mathematician, who would in fact re-write the passage in standard mathematical English. A number of persons to whom this translation has been shown without explanation gave it as their opinion that the translation was done by a human being, and expressed great surprise on hearing that it was in fact done by a machine.

At this point it may be worth while spending a small amount of time describing the machine as it exists at present. APEXC is a typical example of a medium speed digital calculator. Its storage organ takes the form of a magnetic drum which has a capacity of 8,192 words. These words are composed of 32 binary digits and this means, in effect, that each computer word can store six alphabetical characters coded in standard international teletype code. For this reason it is necessary, in practical translation, to take at least two computer words for each foreign language word and the same number for the translation. This in turn restricts the possible number of foreign language words which can be held in the present computer store to something of the order of 1,800. In practice far fewer than this number has been put on the drum since there has been no attempt to use the machine for what might be described as practical commercial translation. Any number of words, from the number contained in the single sentence under examination upwards, is adequate to demonstrate principles. The arithmetical portion of the machine consists of units to perform the operations of addition, subtraction and multiplication, together with certain shift operations and discriminations which are necessary to make the machine obey a set of instructions in a manner which is not known initially to the person who is setting the problem. The input to the machine is via a teletype tape read by a mechanism capable of absorbing 25 alphabetical characters per second and the output is handled by a punch which is capable of the same speed. The whole unit consists of some 450 thermionic valves, occupies about 4 square feet of floor space, and has a power consumption of between 1.5 and 2 kilowatts.

Mention has been made of the inadequate storage capacity of the machine and also of the word length. Inadequacy of storage capacity was a problem which was present in the minds of experimenters from the earliest conception of machine translation. The method proposed for overcoming the difficulty was that of the 'idio glossary'. This is simply a collection of the most frequent words which occur in the specific subject of the text under translation. Certain early estimates suggested that about 1,000 words specific to the subject in hand, and about the same number of words of general linguistic usage, would make possible the translation of over 90 per cent of the given text.

It has already been pointed out that words are divided, from the dictionary point of view, into a stem defined in the manner described above, and an ending. Since the early days of machine translation a more sophisticated approach than the original one of meaning plus a few grammatical notes has been evolved. In this method the machine contains two dictionaries, one of stems and one of endings. If these are stored separately, it is clear that $m \times n$ words become comprehensible to the machine for a storage of $m + n$ stems and endings so that if, for example, n and m were each 100, some 10,000 words would be comprehensible for the storage capacity of only 200 stems and endings.

The way in which the machine deals with a foreign language word may be of interest. When the word is presented to the machine it occurs, of course, as groups of binary digits and these in effect constitute a number. The dictionary words in the machine are stored in consecutive positions in ascending order of numerical magnitude. The earliest proposals for looking up an unknown word in the dictionary required that the word (expressed as a number) be subtracted in turn from the words listed in the dictionary, starting at that word having the least numerical significance. It is clear that at the beginning of the dictionary the result of the subtraction will be negative, that when, the exact equivalent is found the result will be zero, and that for words of higher numerical magnitude than the given word the result of the subtraction will be positive. The unknown word finds its match at the point when the result of subtraction turns from negative to positive. This procedure will require on the average half as many look-up operations as there are words in the dictionary, so that if the dictionary contained say 10,000 words, and the look-up rate was 50 words per second as is the case with APEXC, the actual time required to locate a given word would on the average be 100 seconds.

A modification of this procedure, which saves a considerable amount of time, occurred to the present author. This method makes use of the well-known principle of bracketing. The unknown word is first subtracted from that word which is halfway through the dictionary. If the result is negative it follows that the unknown word lies in the second half of the dictionary; if the result is positive it is known that it lies in the first half. The operation is repeated at that word either one-quarter or three-quarters of the way through the dictionary and this in turn leads to a further partition into one-eighth, three-eighths, five-eighths or seven-eighths. It is easy to show that after $\log_2 n$ look-up operations the unknown word will have been located exactly. Thus, for example, in the dictionary of 10,000 words previously mentioned, some 14 look-up operations would be adequate; for a more extensive dictionary of one million words the number of look-ups is still only about 20. It follows that the time taken to locate any given word in the dictionary is reduced to about half a second. The subtraction process, of course, will give zero result only if the foreign language word finds its exact counterpart in the dictionary. When it turns out that the stem only occurs, things are so arranged that a transition from negative to positive occurs at that point in the dictionary when the true stem is being tested. At this juncture of the process, the equivalent of the foreign language word is to be found in the same storage location as the stem and could, in principle, be typed out. In practice it is retained stored away and the ending, which is the portion of the foreign language word which remains after the subtraction of the stem, is tested by comparison with the ending dictionary. The ending dictionary gives: first an indication of the ending which should be affixed to the English stem, and second an indication of any prefatory words which should be added. Thus, for example, in the Latin word *amas* the stem would be *am* and the ending *as:* These would lead to an English stem *lov,* and an English prefix and affix respectively of *thou* and *est.* The machine operation would thus result in an output of the phrase *thou lovest.* The foregoing treatment has assumed that words can always be translated as they stand; in practice this is not so. The most elementary examples in the French language occur in the inversion of the order of nouns and adjectives which are often present; furthermore complicated pronoun structures such as

Je le lui donne
Je ne le lui donne pas
Je ne vous le donne pas.

mean that the machine has to perform extensive rearrangement of the English text if it is to make sense in that language. All these things are achieved by having the machine absorb the foreign language text one sentence at a time. The machine operations then result either directly in an output, if this is known to the machine to be possible, or alternatively in the accumulation and rearrangement of the English words in the machine storage organ prior to output in sensible English order.

A word may be appropriate at this point concerning the treatment of idioms, which are of two kinds: grammatical and verbal. Both can be treated by a similar process, in which when the computing machine sees a word which may form the key word to an idiom it suspends any translation operation until it has made an analysis of succeeding words to see if these are in fact idiomatic. For example, with the French idiom *boîte de nuit,* the machine would first see the word *boîte.* This might be translated as *box* or alternatively may form the basis of an idiom. Machine instructions say quite simply: examine the next two words, if these are *de nuit* then ignore the normal translation and produce the translation *night-club.* If on the other hand, for example, the words were *de chocolat,* then the output would be *box of chocolate.*

Enough has probably been said to show the way in which the machine operates, at least in the non-technical manner, which is all that can be done in an article of this sort. There are two other points which should receive mention, however, and these concern the resolution of ambiguities, and translation between several languages. It was mentioned earlier that the dictionary to be held in the machine was restricted by inserting only those words which were known to occur in texts dealing with the subject of the material under translation. This idea, too, can be extended to enable ambiguities to be removed so that for example the French word *noyau* in nuclear physics would usually be translated *nucleus* whereas for other subjects it might be translated as *nut, kernel,* or *centre.* This type of ambiguity resolution is sufficient in most cases where the subject of translation is known. When, however, the subject of translation is unknown, as in a literary text, then another procedure can be adopted. In this, each noun in the passage to be translated has associated with it a so-called 'category number'. In the case of ambiguities, several category numbers will accompany each noun and these category numbers will be associated with each of the possible different meanings. When an unknown text is presented to the machine, it first runs through this text and accumulates the number of times the category numbers associated with each noun occur. In the case of those nouns which are unambiguous, and there should be many, large category count numbers of a particular type will arise. The ambiguous nouns are then translated by those variants which have the same category number as that which occurs the maximum number of times unambiguously. From some trials it appears that this method of ambiguity resolution has a wide application and will lead to an adequate clarification in most cases. Yet another means of helping with the resolution of ambiguity, although not of removing it, is to replace all the possible translations of the ambiguous word by one translation, which has that meaning which is nearest to all the possible meanings. Again, in the case of words which have been tested, this procedure appears to offer promise. When translation between several languages is considered, the major new problem is that of volume. If, for example, bi-directional translation is required between each pair of languages selected from a group of *n;* it is clear that $n\ (n—1)$ sets of dictionaries and machine instructions will be required. To reduce this somewhat prohibitive volume of linguistic work to more practicable proportions, it has been suggested that an intermediate language, M say, might be used, In this event the number of sets of data is reduced to $2n$ for the situation defined above. The proposal to construct, artificially, such a meta-language can be rejected on two grounds: numerical and logical. In the first place, if any one of the original *n* languages is taken as the meta-language, the number of sets of dictionaries and instructions at once reduces to $2n—2$ which is certainly less than $2n$. And secondly assuming that M is such that it makes possible—as it must—accurate translation between all language pairs, it reduces itself, in fact, to a process which is at least as complex as that which uses one of the original languages as the basis for all translations between other language pairs, and is more complex for translations in which the basic language is involved. It is not possible, at the present stage of development of machine translation, to decide on the particular language which is to be chosen as the basis, although the volume of printed work would strongly suggest that it should be English.

At the present time work is proceeding with the resolution of the German language. The basic syntactical operations have been studied and codified by Oswald and Fletcher. Brandwood, in this laboratory, has extended their treatment to cover grammatical analysis and, so far as can be seen, with a machine which has adequate storage facilities and computing speed, the problems of German will yield in much the same way as those of French, at least so far as the scientific literature is concerned.

What now of the future? It has recently been announced that in the United States a machine specially for the purpose of translation is to be constructed at great cost. According to the rather meagre press reports so far published, it appears that this machine will simply take the form of a dictionary and will not make use of any of the extensive work mentioned above. Our feeling is that for a relatively modest sum of the order of £250,000, it would now be possible to produce a translating machine which would be potentially applicable to any language as soon as the linguistic work is done, and which would be immediately applicable to French and German and probably to Russian. This machine would have a storage capacity of about 100,000 words and would have an output speed of about 50 words per second. It could be constructed in a period of about two years assuming that the sum of money mentioned were available. It would differ from our present high-speed digital computing machines in only a few particulars, these being an extended storage capacity, the provision of high-speed alphabetical printing, the removal of the multiplying unit which forms a complicated part of most computing machines, the extension of the word length to something between 100 and 200 binary digits and the provision of one or two auxiliary operations which would enable the machine to detect parts of words and parts of sentences more readily than is the case at present. One other question which may be worth consideration is the provision of adequate input facilities. In the present machines and in that just envisaged, input would be via a standard teletype tape, which, of course, would have to be prepared for any new text. Work is at present proceeding on the development of devices for the direct reading of printed characters and for the direct recognition of spoken words. Our particular approach makes use of the extensive logical facilities implicit in the structure of an electronic digital computing machine and it is hoped that within the next two years a practical working character-recognising device will be available. Spoken word recognition is still at an early stage, although even with the present techniques the digits 0 to 9 can be recognised with almost complete accuracy so long as they are spoken by one person. Probably in the long run it is by the development of such techniques that the translating machine of the future will become not only a reality but also a very practical addition to the armament of such organizations as the United Nations. The prospect appears interesting, particularly when it is realized that a single machine can deal with the whole gamut of languages once the initial coding has been carried out. The storage space required on paper or magnetic tapes for the programmes for all languages would be quite modest, and would greatly reduce the difficulties involved in assembling large numbers of human translators for various languages at short notice.