Mechanical Translation and its Implications in Data Processing

THE mechanical translation of language has importance, not only in its own right, but also in commerce and industry, both as a means of improving communication and as a pointer to the use of new techniques in data handling and retrieval. This inter-relation between an essentially non-arithmetical process and a normal calculating machine exhibits an important idea: that many processes, thought to be typical of living organisms, can, upon analysis, be reduced to combinations of simple binary decisions and ordinary arithmetic.

In 1947 the present author first suggested that the storage organ of a digital calculator could be used to hold a dictionary of words and their foreign language equivalents if these were expressed in suitable numerical code, and also that the logical facilities which the machines contained might be suitable for grammatical processing.

EARLY IDEAS

Unfortunately no digital computing machine was available in 1947 and in the period between 1947 and 1951 when the first of the machines became available for normal use, the development of machine translation was restricted merely to ideas. The earliest proposals considered only the use of the computing machine as an automatic dictionary, but, in 1948, Booth and Richens produced several methods, all of which involved greater sophistication in the linguistic and data processing techniques. These ideas were tested in some experiments which used punched card machinery for the translation of several languages on a word-for-word basis and form the basis upon which most subsequent work has been modelled.

The Americans came comparatively late into the field, the two pioneers being Irwin Reifler at the University of Washington and Y Bar-Hillel at the Massachusetts Institute of Technology. Reifler suggested the pre-editor and the post-editor, the former to remove ambiguities in the foreign language text, for which purpose he need have no knowledge of the language into which translation is to be effected; the latter a technical expert in the subject of translation who would resolve ambiguities and attend to the correct technical phrasing of the results of the machine's operations. Generally the post-editor would be the person for whom the translation was intended. We will not dwell further upon either pre- or post-editor, since

Is the era of translation of languages

by computers just round the corner?

And if so, will the computer do the

translator's job efficiently and cheaply?

by ANDREW D BOOTH, DSc, PhD, FInstP

Reifler himself now believes that neither is necessary.

1953 CONFERENCE

In 1953 the Massachusetts Institute of Technology, in conjunction with the Rockefeller Foundation, held the first International Conference on Machine Translation, and the way in which the subject has grown can be seen from the fact that in 1953 about a dozen experts gathered together, whereas today some hundreds might be expected. The discussions of the Conference were inconclusive, but one positive result was the publication of a volume containing an account of all the work which had been done up to that time and a subsidiary outcome was a much publicised experiment which claimed to show translation from Russian into English by means of business computing machines.

In 1955, the Nuffield Foundation made a generous grant to Birkbeck College, University of London, of funds to pursue research in machine translation on a more extensive scale. The first entry of the Russians into the field was revealed in 1956 when I Mukhin of the University of Moscow surprised delegates at the Conference held by the Institution of Electrical Engineers in London by showing translations which he claimed had been produced on a Russian computing machine. Unfortunately there is internal evidence that these results were not produced by a machine, but the methods which were described were certainly applicable, and Russian work has since increased in tempo until at the present time in Leningrad and Moscow there is a team of some hundreds of linguists working on the resolution not only of English into Russian, but also of Hungarian and Chinese.

At present the major centres of research are: in the United States, the University of Washington, Seattle, Georgetown University, and the Massachusetts Institute of Technology; in England, at Birkbeck College and at Cambridge; and in Russia, at Leningrad and Moscow. New papers and reports are being produced almost daily and the subject is in a state of continuous expansion.

TRANSLATION TECHNIQUES

We discuss next some of the techniques which have been used in translation. It is worth dealing with these in some detail because of their implications in the more general field of data processing. The basis of machine translation is, of course, the dictionary. Unfortunately, even a cursory examination of the way in which ordinary dictionaries are constructed shows that they are quite unsuitable for translation by means of a machine. This is chiefly because they are based upon the assumption that a user will have considerable knowledge of the grammar of the foreign language, and that he can reduce inflected word forms to the particular base which is recorded in the dictionary. One elementary example is the Latin word '*amas,* another the French word '*cherchait.'* Neither will be found in any dictionary. When a human translator attempts to look up these words, he knows that' amas ' is derived from the Latin verb '*amo'* and that 'cherchait' is derived from the French verb '*chercher*'; a machine, however, does not have this prescience.

Two alternatives are available to overcome the difficulty: the first to endow the machine with a knowledge of the structure of the language with which it is dealing, the second to modify the dictionary. Although the first alternative is not impossible, it is rather difficult and it is fortunate that one of the earliest ideas in this subject showed

that, by a simple re-orientation of the dictionary, the problem could be avoided altogether.

DICTIONARY OF STEMS

The basic idea is that instead of entering the infinitives of verbs, nominative singulars of nouns and so on in the dictionary, what should be recorded are the stems of the words. The stem is here defined as the longest segment of a word which is common to the majority of its parts. This qualification, ' the majority of its parts,' is necessary because in the case of irregularly formed words it may be necessary to store several sterns or even complete words themselves. Once the notion of storing stems is accepted, the process of looking up a word in a dictionary by means of a computing machine becomes considerably simplified. Numerous processes are possible but they usually involve first coding the foreign language word in terms of numbers, for example, A=01, B=02, C=03 ... Z=26, 'amo' = 011315, and then testing whether the given code number corresponds with successive dictionary word code numbers by seeing whether their difference is zero. With the stem-ending procedure, however, a zero difference is unusual and generally speaking the criterion of match is simply that the result of a subtraction between unknown word and dictionary entry changes sign at that point in the dictionary at which the stem defined in the manner mentioned above is to be found.

We have not time to discuss the minutiae of dictionary construction, but it may be well to remark that a certain art is needed in constructing and using such a dictionary if stems of one word are not to become confused with larger portions of other words. These problems, however, have been considered in detail and have been shown to be solvable in all cases so far encountered.

When the stem has been subtracted from the foreign language word to be translated, the remainder, called the ending, can be used to give linguistic information. In the early days this information consisted simply of a few grammatical notes, for example ' 2nd person singular present tense,' in the case of the 'as' ending of 'amas ' mentioned above; more recently, however, the ending has been used in conjunction with a dictionary of endings to give an appropriate suffix and affix to the word translated. For example, with 'amas' the stem dictionary would give '-lov-', the ending dictionary the prefix *'thou'* and the suffix *'-est.'*

METHOD OF CONSULTING

The actual way in which words are hunted out in the dictionary is of considerable importance. The earliest programmes stored the dictionary entries in consecutive locations and in ascending order of code number magnitude and compared the foreign language words starting at the beginning of the dictionary with all entries in sequence. On average it is clear that about half the dictionary must be scanned to obtain the meaning of a word selected at random. With quite small dictionaries this may involve considerable time, even on the fastest computers, and with large dictionaries the process is so unwieldy as to be useless. As an example, with a dictionary of 100,000 words some

Andrew D Booth was educated at Haberdasher's Aske's School, Hampstead. After some years of industrial experience and graduation with first class honours as an external student of London University, he went during the war to Birmingham University to work on crystallographic problems of explosives.

In 1946 he took up a Nuffield Fellowship at Birkbeck College, London, and in 1947 spent six months at Princeton, USA, as a Rockefeller Fellow and member of the Institute for Advanced Study. On returning to Britain he initiated the Birkbeck College Electronic Computer Project and in 1954 received the title of Director of the Computation Laboratory, Birkbeck College, and University Reader in Computational Methods, In 1957 Dr Booth was appointed Director of the newly-formed Department of Numerical Automation.

Ha was awarded the DSc, London, in 1951.

of the fastest modern computers might take ten seconds to find a word chosen at random.

The second approach which was developed was the so-called ' bracketing' method in which the unknown word is subtracted from an entry half way along the dictionary. If the result of this subtraction is positive, it will be seen that the dictionary equivalent lies in the lower half of the dictionary, if negative in the upper. Suppose the former case, then a comparison is made with a word one quarter of the way along the dictionary and the same criterion is applied.

It will be clear on a little reflection and analysis that for a dictionary of D entries about $\log_2 D$ comparisons are needed. To put this into perspective we may remark that the number of comparisons required for a dictionary of about 100,000 words is 17 and this would take even slow computers only ¼ of a second and fast machines a time of the order of 1/1000 of a second. Before leaving the dictionary, it is worth mentioning that the process of sequential hunting through a dictionary which was used in the early experiments is now returning to favour. The reason for this lies in the fact that for very large texts an efficient method of search is first to sort the text into alphabetical word order, numbering the words in accord with their position in the sentences being translated, then to compare all of the words with the dictionary at one pass. Having obtained the meanings and any grammatical notes which are appropriate by this process, the output is then sorted into original text order again and processed as is appropriate. This distinction between methods of search which are appropriate when few words are to be processed and when whole strings of input data are available at the same time is quite typical of two classes of problem which occur in sorting applications in commerce.

WORDS IN CONTEXT

Anyone who has attempted to translate even simple French will be familiar with the fact that a word-for-word translation is generally ludicrous and frequently meaningless, and that only by considering words in conjunction with their environment among other words, that is of syntax, can an adequate translation be made. The problem of using context has been solved by associating with each dictionary word a number of category numbers. One of these category numbers indicates that a word is a noun, pronoun, adjective, verb, and so on. Another category number shows that a word belongs to certain fields of human endeavour. For example, ' scalpel' has a meaning which is restricted to the biological field, whereas ' nucleus ' spreads over biology, physics, sociology and mathematics. By means of these category numbers and an associated dictionary of structural forms which indicates, for example, that a well formed sentence must have a subject, verb and object, that qualifying words must be correctly associated with the words upon which they act, and so on, it is possible to produce a passably correct rearrangement of text between the source language and its translation.

LEARNING FROM EXPERIENCE

An essential technique in translation is that of learning from experience. Much has been written on the subject of learning machines mostly of a vague and unsatisfactory nature. In the present context, however, it is possible to give a precise meaning to this term, and it is as follows: in processing any text it will be found that many words have multiple meaning, for example the words 'bear,' 'revolution,' 'nucleus,' and so on. As the machine processes the words of a new text, it counts up the number of occurrences of words in the different subject categories.

Ambiguous words will contribute to several categories, but the unambiguous ones which define the subject matter will contribute predominantly to a single category. After some portion of the text has been processed, the machine will learn by the fact of maximum occurrence in one category that all of the alternatives outside this sphere are to be rejected. By this means the quality of translation and the reduction of ambiguity will increase as the text continues to be processed. If this is considered undesirable, however, a blank run can be made in which no translation is attempted but in which the machine fixes upon the subject matter of the paper concerned. This process, however, is dangerous, since if the author changes his field in the middle of the text, the machine will make mistakes.

For this reason we favour firstly an initial indication to the machine of the general subject matter, and secondly the use by the machine of category information to vary its own consciousness of categories as the translation proceeds, so that, for example, if during a paper on nuclear physics the subject changes to the social consequences of strontium 90, the machine after a small period of ambiguity during re-education would change its

internal category indications, so that the sociological words were correctly translated.

EFFECT ON DICTIONARY MAKING

It is worth reiterating that these translation techniques have considerable impact on other fields. For example, it is the opinion of the author, and of many of the more enlightened linguists with whom he has conversed, that the method of dictionary layout adopted for translating machines could well be adopted in the future for dictionaries which are intended for human use, so that once conventional dictionaries are replaced by these new ones, a person relatively unversed in the structure of an unknown language could still make effective use of the dictionary for translation.

COMMERCIAL USES

The detailed techniques of dictionary search have applications in industry and commerce. Where large amounts of information can be collected together and processed, say, at the end of the week, the method of successive comparison of pre-sorted information with an ordered store of facts may well be the appropriate one, but, where information is required randomly and instantaneously, for example in giving a client in a bank the state of his account on demand, then the partitioning method will be more appropriate. Again, the construction of dictionaries from words gives pointers to the way in which files might be constructed. For example, a dictionary or file which is to be partitioned must be arranged strictly in ascending order of numerical or alphabetical magnitude. This is inconvenient if new data are added at any considerable rate and in the latter case straightforward sorting techniques involving comparison with the whole dictionary may be better since in this case it is possible to place new information at the end of that already stored.

We now turn to another aspect of translating machines. It is this: would it be worth setting up a central installation for the translation of a single language? The answer is unequivocally 'no,' although, of course, work on precisely this basis has been carried out in the United States and in Russia for the particular language pair English-Russian. In a country such as England, however, the answer given above seems to be the only possible one, but when the problem of translating between many languages is considered, several other interesting logical factors become important. Suppose that there are n languages, a, b, c, d ... n, and that it is required to translate bi-directionally between them.

For a long time it was thought that such bi-directional translation would involve the construction of n(n-l) sets of dictionaries and grammatical procedures. The present author in a moment of aberration then suggested that a way of simplifying this requirement would be to invent a meta-language, that is, some intermediate language into which all other languages can be translated and from which translation can be made back to these languages. It can be seen that inserting such a meta-language M, the total number of dictionary and grammatical pairs required for the n language system described above is 2n. Further reflection, however, shows that this is untrue since, by using one of the languages of the original set as a meta-language, the number of language pairs required to be processed becomes reduced to 2n-2. This simple numerical fact, although perhaps a slight over-simplification of the problem, seems to be completely overlooked by a number of workers in this field.

RESULTS ACHIEVED

After so much preliminary talk, the reader will probably be wondering what sort of results have been produced by the researches which we have described. In particular, does a translating machine exist? To which the answer is ' no. ' Has language translation been carried out on a machine ? To which the answer is ' yes.' Is translation on a machine at present an economical proposition? To which the answer is again ' no.'

Many examples of machine translation could be quoted, but in the present instance we shall give only two which show both the strength and the weakness of the method. The first example was derived from a scientific text for which mechanical translation processes are particularly suited and it is given below:

'L'étude des fonctions definies par une équation différentielle dans tout leur domaine d'existence, eat un problème dont la solution compléte, dans ce cas général, depasse actuellement la puissance de I'analyse. On a cependent obtenue des resultats du plus haut intérêt en se limitant á I'étude des integrales inftniment voisines d'une integrale connue.'

' The study of the functions defined by a differential equation in all their sphere of existence is a problem of which the complete solution, in the general case, surpasses at present the power of analysis. One has nevertheless obtained results of

highest interest while limiting oneself to the study of infinitely adjacent integrals of a known integral.'

The second example of a more literary style is the following:

' *L'homme qui m'a parlé hier est très malade ; il est tombé de son cheval devant sa maison.*'

'The man which did talk to me yesterday is very ill; he did fall from his horse before his house.'

This shows the process up in a less favourable light.

SPEED OF TRANSLATION

The speed of translation for French at the present time is about 3,000 words per hour, but there is no reason why, in the very near future, it cannot be increased by a factor of between five and ten, even without the construction of special purpose translating machines. Even so it turns out that the price of translation by machine is perhaps £10 per 1,000 words, and this compares unfavourably with the rate for human translation which is about £2 10s. for the same volume of work.

To summarize the discussion and the remarks which we have made, we can conclude first that machine translation has had a considerable impact on linguists by making them consider with more precision the basic structure of their art. Second, that it has contributed ideas on the processing of large collections of information. Sorting and handling techniques developed for machine translation have proved of considerable value for the more general application and, for example, such improbable things as the construction of a concordance or list of distinct words in a foreign language text has suggested methods by which information derived from a business letter can be collated with information contained elsewhere in files. Third, that it has illustrated how machines can, in a rudimentary sense, be made to learn, and finally that it shows how contextual information can be used to produce an improvement in the results of a given process. We have considered this in the context of language translation, but more generally the methods can be applied to the process of deciding the particular meaning of a statement or occurrence in the context of other operations in the same field: for example, the location of a mutilated order or invoice from a file containing all invoices.

The future of machine translation itself is perhaps less rosy. Despite the prestige value of work in this field, it appears to the author unlikely that translating machines will be of practical utility in the commercial and industrial fields for a number of years to come, and although the United States and Russia may seek to use the somewhat inefficient process to ease their manpower problems in the translation of technical material, it does not by any means follow that this would be a rational approach in a more leisurely society such as our own.