

## **Machine Translation of Natural Languages\***

J. McDaniel†

The possibility of substituting a machine for a human in the process of translating printed literature in one natural language into its meaning equivalent in another natural language, was first mooted when the first electronic digital computers were being constructed in the late 1940's. It was realised that these machines, having the ability to perform sequences of coded instructions on numerical data at great speed and with high reliability, should, if properly instructed, be able to copy human translation processes on numerically coded literature. At first the problem of how to instruct digital computers to translate was attractive only as an academic exercise, but it has since then gained much more practical and economic importance with the realisation that the speed of technological and scientific progress in the world would be much enhanced if the international exchange of published information was made more comprehensive and more immediate. This could be possible with only a few electronic computers per country to keep pace with the translation of information published in languages foreign to that country, and they would be a hundred times cheaper than the alternative, a host of human translators.

### **Research into machine translation**

The USA was the first country to devote a large research effort to the study of machine translation, and work was concentrated on translation from Russian to English. The USSR followed soon after with studies of many more languages than those being studied in the USA. Now, outside of the USSR, there are some 20 or so research projects in machine translation in six countries, with Russian still receiving most attention as the source language. All projects so far have restricted their studies to translation between one pair of languages. Only theoretical consideration has been given to the problem of universal translation between any pair of languages.

\*Abridged and updated version of a paper presented at a meeting of the Institute on 16 February 1966.

†Autonomics Division, National Physical Laboratory, Teddington.

Research on machine translation (MT) was started at the National Physical Laboratory (NPL) in 1959. From the outset, studies have been restricted to the translation of Russian scientific and technical literature into English. The electronic digital computer which is being used in the translation experiments is the ACE computer, designed and constructed by laboratory staff. It has six magnetic tape units on which to store the masses of linguistic data essential to the translation programmes. It must be emphasised that research on MT so far has revealed no requirement on the machine being used that cannot be met by most current large electronic digital computers. Such computers need no special adaptation for doing MT and can immediately switch to mathematical computations or business data-processing problems when not engaged in MT. Assuming MT will be proven to be feasible, then any large digital computer establishment could produce programmes to do MT on its machine.

### **Machine dictionary**

The first requisite in any MT programme is an automatic machine dictionary. This is directly analogous to the printed dictionary familiar to all language students, in that it is an ordered list of words of the source language (assumed to be Russian for the rest of this article) with, for each word, a grammatical description, and the equivalent word or words in the target language (English in the NPL project). Naturally, the MT dictionary has to be "printed" in a form that the computer can understand, and this is usually on magnetic tape, patterns of magnetisation on the tape representing the numerically coded linguistic data of the dictionary. The computer can "read" data on tape at very high speed and select portions to be transferred into the calculating unit of the computer for linguistic processing. The dictionary at NPL which is now in operation has about 20,000 entries and is stored on two magnetic tape reels. The vocabulary of the dictionary has been selected particularly to cover the literature of mathematics, physics and electrical engineering. Russian being a highly inflected language, most words have many different inflected forms depending on their particular function in given sentences. Considerable economy in the size of the NPL dictionary has been achieved by having, in most cases, all variant forms of a word referred to one dictionary entry, which entry represents the morphological stem of that word. This is in contrast to another form of dictionary organisation in which all forms of words have separate, distinct entries representing them in the dictionary.

The NPL dictionary will be used to "look-up" batches of several thousands of Russian text words entered into the computer *via* punched cards. Comparison of individual words with the dictionary, in the order in which they occur in text, would involve much wasted time with the computer skipping over large sections of tape to get to the matching entries. Putting many text words into dictionary order before they are looked up means that they can all be matched in one continuous scan through the dictionary. The words are then sorted back into original text order.

## **Multiple meaning of words**

There is a third major problem in MT, one on which, as yet, little useful progress has been made. The problem is that of multiple meaning of source language words, i.e. in our case, where Russian words need several English words in their dictionary entries to fully describe their range of meanings. A theoretically simple but practically clumsy way of avoiding this problem is, simply, to list all such meanings in the translation and let the reader choose the one most appropriate to the context. Aside from the possibility of the reader choosing wrongly, this solution would prove very onerous to the reader. At NPL, lacking a practicable means of semantic classification and analysis that might make the automatic choice of meanings possible, we are planning to reduce the problem by limiting the field of applicability of our dictionary, as described above, and by reducing the specificity of English equivalents so that, on average, fewer are required.

## **Evaluation experiment**

The essential goal of the MT project at NPL is to provide English-speaking scientists with accurate and easily readable translations of Russian texts in their specialist fields. We aim to accept the minimum translation quality commensurate with this goal.

In order to test how far towards this goal we have reached, we are currently running an evaluation experiment. We are providing practising scientists with machine translations of Russian articles, chosen by themselves as pertinent to their interest, and asking for their evaluation of them from the information-transfer point of view (not from the linguistic point of view). We expect to process about 60,000 words of text. This experiment should give us not only a measure of performance to date, but also show more clearly what are the deficiencies in our machine translations that affect most strongly the reader's understanding of them.

## **Discussion**

Mr. J. C. Rippon (Metal Box Co. Ltd.) stated that for £5,000 a year one can employ two translators, covering 16 languages, who will do about 500 translations (varying from single-page letters to 60-page manuals) and also literature searches and selective abstracts. He asked what

was the extra efficiency of machine translation, except for cover-to-cover translation on behalf of a publishing firm, and commented that people who can type Russian faultlessly enough to feed a computer are very rare, perhaps as rare as technical translators, who will still be required because they can do so much more than the machine.

Mr. J. McDaniel replied that the advantage is for a large central translating organisation. For a perfect translation into a foreign language, the machine will probably not compete. Typists of Russian can be trained in six weeks up to the speeds they have attained in English. The cost of a new machine dictionary could be £15,000 for a restricted technical field.

Mr. Rippon observed that in any article there is much redundant material, and that a translator can ignore this or give the gist of it orally in a few minutes whereas a machine cannot. The speaker said that the cost for machine translation will be farthings per word, so the overall cost may not be so very different.

Mr. J. E. L. Farradane (Northampton College of Advanced Technology) mentioned that you can go back to a translator about obscure passages, but you cannot do this with a machine. In some languages there are many words equivalent to only one English word, or *vice versa*; for example, Scandinavian languages have many words for snow whereas English has only one. He asked how the machine dealt with this difficulty. In reply the speaker said that the terminology in technical literature is fairly international.

Mr. Farradane, quoting examples, said that "take" can be translated by "*porter*" or "*mener*," "*Le portier ferme la porte*" can mean either "the porter shuts the door" or "the strong porter carries her." Mr. McDaniel replied that the previous text will give guidance and enable a decision to be made, and mentioned that this sort of thing does not happen much in technical texts.

Dr. D. J. Campbell (Pressed Steel Co. Ltd.) asked how in machine translation one dealt with the present tense of "to be," which is rarely used in Russian. The speaker replied that if there is no other main verb "is" or "are" are added, but that this is not without difficulty where there is another main verb or where the present tense of "to be" is implied, e.g. by an adjective in predicate form.

Mr. J. Hetheron (Metal Box Co. Ltd.) said that Mr. McDaniel had presented a very useful and valuable paper and asked whether he had done any market research on the demand for machine translation. He asked the speaker whether the initial impulse towards it was from the computer manufacturers, and whether it would die, like the phlogiston theory.

As to market research, the speaker welcomed articles on electrical engineering subjects for translation, saying he would like to receive such material within a month or six weeks. It would be helpful if those who forwarded the originals would comment on the quality of the translations when they were sent out. He stated that the impulse towards machine translation did not come from computer manufacturers, and that

*Continued on p. 21*

## **Machine Translation of Natural Languages**

*Concluded from p. 8*

it would not die like the phlogiston theory: there was a future for it if the necessary technical effort was put into it. He questioned the effectiveness of use of material already available in libraries and whether enough use is made of knowledge already available in English, and mentioned that questions of cost and economics are largely unsolved even for those areas.

The Chairman (Mr. F. Liebesny, British Aluminium Co. Ltd.), in thanking Mr. McDaniel for his paper, said that he had shown that there are no limits to the effective use of machine translation, which will make possible a considerable increase in the amount of material translated.