

L. BRANDWOOD

Previous Experiments in Mechanical Translation

Disregarding experiments in simulated machine translation, which in place of a computer employs a person with no knowledge of the language to be translated acting in accordance with a set of given rules, only three actual attempts have been made, to my knowledge, to translate in this way. These were by the Institute of Languages and Linguistics of Georgetown University in association with the International Business Machines Corporation (1954, Russian to English); by the Computation Laboratory of Birkbeck College, London, under the direction of Dr. A. D. Booth (1955, French to English); and by the Institutes of Scientific Information and of Precise Mechanics and Computing Technique at the U. S. S. R. Academy of Sciences, Moscow (1955/6, English to Russian).

Little can be said about the IBM experiment, because little definite is known. An account of it is given by Dr. L. E. Dostert in *Machine translation of Languages*, chap. 8. pp. 124 ff., but he has unfortunately allowed himself to be sidetracked into discussing problems arising out of the experiment rather than the experiment itself, and the reader is consequently left to guess at the exact details of the latter. From what the author does say, however, it may be gathered that the experiment was designed to translate a number of specified Russian sentences into English.

The lexicon required consisted of 250 terms, each with one or, at most, two English equivalents, but whether this figure included the different (inflected) forms of nouns, verbs etc. or not, is not stated. Nor, in fact, is anything said on the subject of morphology, except that with a "complex" Russian word (i. e. one divisible into a stem and ending) the two parts were coded and stored separately.

For dealing with syntax and the choice of the correct English equivalent, where more than one was given, a set of operations "limited in number but widely significant in scope" was provided.

The actual number was six, and they read as follows:

1. The machine finds only one English equivalent for the Russian word — no problem.
2. The machine finds 2 equivalents and makes a choice between them based on a subsequent word in the sentence.
3. The same as no. 2, only the choice is based on a preceding word.
4. A Russian word is superfluous in translation and has to be omitted, (cp. in French, *la France*, *sortir de la maison*).
5. The English translation requires the addition of a word not to be found explicit in the Russian. (c. p. French, *je vais partir* = *I am going to leave*.)
6. The order of two words has to be reversed.

I

Evidently these are only the bare outlines of the program. What would have been of the greatest interest to all, and especially to those engaged in research on mechanical translation, namely an explanation of exactly how the difficulties involved, for example those of morphology and lexicographical choice, were overcome, is not forthcoming. If it had been an account of a scheme proposed, instead of one actually carried out then the absence of these details and, more important still, of

examples of texts translated by the machine¹⁾ might have been pardonable: as it is, the omission of such valuable information can only be looked upon with the deepest regret.

The Birkbeck experiment was of a popular, and therefore elementary nature, being intended as a demonstration to the public of the practicability of mechanical translation through the medium of television. An account of it appeared in the B.B.C.'s *European Review* but for the convenience of the reader it may be said briefly that the material consisted of a sentence (2) involving no translation difficulties apart from the correct recognition and location of each French word in the dictionary and the output of its English equivalent. In short, the solution demonstrated was of mechanical rather than linguistic problems involved in translation by machine.

The Russian experiment is described (3) in much greater detail than the American, over which it shows a considerable improvement. The main facts about it are as follows. The dictionary, consisting of 952 words, was divided into an English and a Russian section, the former containing not only each English word and an indication of the location of its Russian equivalent in the Russian section, but also the grammatical notes relevant to this equivalent (e.g. gender, declension, conjugation etc.), and the latter the Russian words in the order required by the references in the English section. Alongside those English words with more than one meaning, however, was given instead of a reference to the Russian section, one to a special program called "the vocabulary of polysemantic words", which enabled the machine to determine the correct Russian equivalent for the particular meaning by analysing the context of the word in question. No further information is given about this "vocabulary of polysemantic words", which clearly may take one of two forms — ideal or practical. The "ideal" one would permit the solution of ambiguity by a compilation of rules accounting for every possible context in which each of the polysemantic words (in this experiment 121) could occur. The "practical" one aims only to provide a solution by accounting for those contexts which actually do occur in the passage or passages of text under consideration. Since the realization of the former would involve a storage capacity not to be found in any existing machine, there can be little doubt that the latter was the method employed in the experiment.

The look up routine is the same as that described by Dr. Booth in his article, but the manner of identification differs in that it does not utilise a stem and an ending dictionary. Instead, if the machine fails to identify an incoming word with any entry in its dictionary, it goes through a process of discarding and rectifying the inflectional ending or endings, then after each of these alterations re-comparing the word with the entries in its dictionary. One such process for identifying the comparative of adjectives ending in -y (e. g. "sillier") runs as follows:

1. (sillier) Test two final letters for er Yes.
2. Discard -er; develop indication for comparative degree.
3. (silli-) Vocabulary test for remainder: No.
4. Discard one more final letter.
5. (sill-) Vocabulary test; No.
6. Add -y to remainder.
7. (silly) Vocabulary test; Yes.

¹⁾ One sentence with its translation is reproduced from a newspaper by Dr. A. D. BOOTH in an article in *Discovery XV*, p.280 (1954) but this is insufficient material from which to draw conclusions.

²⁾ C'est un exemple d'une traduction faite par la machine a calculer installée au laboratoire de calcul de Birkbeck College, Londres, the translation of which, as printed out by the machine was: This is an example of a translation made by the machine for calculation installed at the laboratory of computation of Birkbeck College, London.

³⁾ Proceedings of Inst. of Electrical Engineers, in press.

This procedure, involving as it does four dictionary searches (as compared with the one stem-dictionary search plus one ending-dictionary search of our own method) is most uneconomical in terms of machine operating time, and this disadvantage does not appear to be outweighed by any particular merit in another direction.

Once the correct Russian equivalent for each English word has been determined, it is given its correct inflection by a combination of the grammatical information on it included in the English section and an analysis of its role in the English sentence. Finally the words of the translated sentence are rearranged in accordance with the rules of Russian grammar.

In conclusion it may be said that the account appears a genuine endeavour to acquaint the reader with the workings of the experiment, and that the experiment itself, despite some shortcomings, which probably exist to a greater degree in print than in fact, represents the most concrete contribution to the progress of mechanical translation made to date.