[From: *Information theory*: papers read at a symposium on 'Information theory' held at the Royal Institution, London, September 12th to 16th, 1955.; edited by Colin Cherry (London: Butterworths, 1956).]

19

INFLUENCE OF CONTEXT ON TRANSLATION

A. D. BOOTH
Birkbeck College, London

IN CONSIDERATION of the application of high speed computing devices to the translation of language two problems immediately come to the fore. First, that the storage capacity of all existing machines is far too small to allow an adequate treatment of a complete real language and, secondly that a straightforward word-for-word translation gives rise to considerable ambiguity in the output.

There is no real difficulty in treating idioms, so long as adequate storage is available. Much confused thinking in the past has made this subject into a mountain where one does not exist.

Whilst it is easy, in principle, to use a computing machine so as to include contextual evidence in making translations, the real problem, at the moment, is to find a situation or language which is sufficiently restricted to make possible its running on an existing machine. With this in view, and also because of its humanitarian aspects, it was considered a worthwhile project to record the conversion of Standard English into Braille.

When this problem was first brought to our notice it appeared to be a trivial one, which possessed no points of interest to the mechanical translator. On further consideration, however, it appeared that the operations carried out were not completely trivial, and that many of the problems which occur in translation to contracted Braille occur also in the translation of real language: the essential simplification being that the word in real language is replaced by the letter in Braille, and, similarly, that the word-group or idiom in real language goes over into the letter group of Braille.

It is not here worth entering into a philosophical discussion as to the distinction between translation and transcription. In the paper by CLEAVE (see page 184) the title 'A Programme for Braille Transcription' has been used so that, by taking the more restricted position, we do not render ourselves open to criticism.

It appears to be of great importance to achieve some results in the mechanical translation of language in the near future, otherwise the whole subject is likely to fall into disrepute and, for this reason, it seems appropriate at Birkbeck College to continue our researches on simple systems of language which will, nevertheless, be of practical utility as an example of the sort of thing involved. It may be mentioned that, at the present moment, a micro-glossary and micro-grammar in meteorological French are being constructed as a joint project of the College and of a group working under Professor Firth at the Institute for African and Oriental Studies.

A word may not be out of place regarding the character of possible

demonstrations of mechanical translation. What is needed is not the giving out of acceptable texts to the audience who may present them to the machine for translation, because this does not require, if dishonestly performed, any function on the part of the machine other than storage; and even this limited facility can be side-tracked by the use of a previously prepared set of output tapes.

It seems that a far more searching test is to supply a micro-glossary and restricted grammar from which any passage may be constructed and presented to the machine. It is only by such a means that the strengths and weaknesses of a translation system can be ascertained; and this situation will be seen merely to be the experimental method which is applied in scientific research.

Another activity of our laboratory has been to study coding processes which are appropriate to mechanical translation, and it is worth mentioning one milestone of recent progress. The elementary operation in mechanical translation is the looking up of a presented word in what amounts to a 'mechanical dictionary'. The earliest method suggested to perform this operation was to give each letter a number, such as a = 1, b = 2, c = 3 *etc* and then to store the translation, or other information concerning the unknown word, in a storage location the number of which was that given by the aggregate number of the word concerned, thus, 'et' would appear in position 520 and 'elle' in 512125. This system, although simple in principle, involves the use of a store with a capacity of the order of 10^{15} words in order to store an actual number of words which does not exceed 10^{7}. This is quite impracticable.

The second system to be considered, and the one which until very recently appeared to be the best available, stored the code-number corresponding to the given foreign language word, followed by the coded translation, or other information, in a single storage position. The machine store was then filled in such a manner that successive storage positions contained code numbers in ascending order of numerical magnitude. To search for the information concerning the unknown word, the code number of that word was subtracted in turn from the contents of each storage location, starting from that having least numerical magnitude. The results of this subtraction are negative until the nearest equivalent to the unknown word is reached, at which point the result of subtraction becomes positive. The remainder of the storage position then contains the required information. It will be seen that this process requires, on the average, an examination of half of the words in the dictionary.

A new system is based on a numerical process which can be used for such things as the extraction of *n*-th roots of numbers, the solution of high order algebraic equations and the conversion of numbers to binary form. The process can be simply described in the following way.

With the dictionary arranged in the same manner as that just described, the unknown word is first subtracted from the entry, which is half-way between the start and finish of the dictionary. If the result is positive the subtraction is again performed, but this time on a word a quarter of the way along the dictionary: while if the result is negative the subtraction is performed on that word which is three-quarters of the way along the

dictionary. The process is repeated so that the dictionary is sub-divided successively into lengths $1/2^n$. It is easily shown that for the *Shorter Oxford Dictionary* a maximum number of trials will be 15, and for a dictionary containing one million entries about 21. This process can be made the basis of an amusing lecture demonstration, and can also be applied to the current parlour game of Twenty Questions.

Discussion is recorded after the next paper by J. P. Cleave.