# STOCHASTIC METHODS OF MACHINE TRANSLATION

Gilbert W. King and Irving L. Wieselman

International Telemeter Corporation

September 21, 1956

This work is being done with the support of the Rome Air Development Center, United States Air Force.

# STOCHASTIC METHODS OF MECHANICAL TRANSLATION

## Redundancy of Language

It is well known that Western languages are 50% redundant. Experiment shows that if an average person guesses the successive words in a completely unknown sentence he has to be told only half of them. Experiment shows that this also applies to guessing the successive word-ideas in a foreign language, How can this fact be used in machine translation?

It is clear that the success of the human in achieving a probability of . 50 in anticipating the words in a sentence is largely due to his experience and the real meanings of the words already discovered. One cannot yet profitably discuss a machine with these capabilities. However, the problem facing the machine translator is easier in one respect. It does not have to make a choice from the wide field of all possible words, but is given in fact the word in the foreign language, and only has to select one from a few possible meanings.

In machine translation the procedure has to be generalised from guessing merely the next word. The machine may start anywhere in the sentence and skip around looking for clues. The procedure for estimating the probabilities and selecting the highest may be classified into several types, depending on the type of hardware in the particular machine-translating system to be used.

## Equipment

It is appropriate to describe briefly the system currently planned and under construction. The central feature is a high density store. This ultimately will have a capacity of one billion bits and a random access time of 20 milliseconds. Information from the store is delivered to a high-speed data processor. A text reader supplies the input and a high-speed printer delivers the output. The store serves as a dictionary, which is quite different from an ordinary manual type. Basically, of course, the store contains the foreign words and their equivalents. The capacity is so large, however, that all inflexions (paradigmatic forms) of each stem is entered separately, with appropriate equivalents. In addition, in each entry identification symbols are to be found, telling which part of speech the word is, and in which field of knowledge it occurs. Needless to say many words have several meanings, may be several parts of speech, and may occur with specialized meanings in different disciplines, and it is trite to remark that these factors are what makes mechanical translation hard.

Further, in each entry there is, if necessary, a computing program which is to instruct the data processor to carry out certain searches and logical operations on the sentence. These rules will be formulated from a relatively brief study of the language. By assembling statistics on the occurrence of these rules, and new situations, the rules will be modified and extended, and possibly deleted.

Operation

In operation, each sentence is considered as a semantic unit. All the words in the sentence are looked up in the dictionary, and all the material in each entry is delivered to the high-speed, relatively low-capacity, store of the data processor. This information includes target equivalent, grammar and programs. The data processor now works out the instructions given to it by the programs, on all the other material, - equivalents, grammar and syntax belonging to the sentence, all in its own temporary store.

At this point some of the stochastic selection of meaning has been done in the memory, by identifying an inflected form which limits the meaning, or a word pair or idiom. It is important to realize these translations are only the most probable.

The data processor now proceeds to make its best guess at multiple meanings, by making use of the programs it now has in its memory, delivered from the various entries. At present we visualize only two kinds - grammatical and graphical. As stated, above, no machine at present can use clues from meaning. Each of these kinds of programs instructs the data processor to look for further clues in the particular sentence itself, which is essentially infinitely variable, in contrast with the memory, which identifies only fixed patterns of words. These searches are testing of logical equations, typical of the logical circuits of present-day computers.

With these facilities in mind, we may now examine some of the procedures that can be mechanized to allow the machine to guess at a sequence of words which constitute its best estimate of the meaning of the sentence in the foreign language.

Field of Knowledge

The simplest type of problem is "the unconscious pun" a human faces is seeing a headline in a newspaper in his own language, such as the now classical "Harvard Prexy fights erection in Square". He has to scan the text to find the topic discussed, and then go back to select the appropriate meaning. This can be mechanized by having the machine scan the text (in this case more than one sentence is involved), pick out the words with only one meaning and make a statistical count of the symbols indicating field of knowledge, and thus guess at the field under discussion. (The calculations may be elaborated to weight in words belonging to more than one field.)

Grammar

A second type of multiple-meaning problem where the probability of correct selection can be increased substantially and can also be mechanised is the situation where a word has different meanings when it is in different grammatical forms. Here the machine seeks clues to guess the part of speech the word is, and makes its guess on this basis. The clues are the parts of speech, given by the memory of neighboring words. Examples are the two common and annoying French words: pas, adverb "not"; noun, "step", and est, present 3rd singular verb, "is"; noun, "east". The probability of selecting the correct meaning can be increased by programming such as the following for pas: "If preceded by a verb or adverb, then choose not: if preceded by an article or adjective choose step." Experiment shows this rule (and a similar one for est) has a confidence coefficient of . 99 of giving the correct translation, as it is almost always an adverb. It is too difficult by hand to collect a sufficiently large sample of pas as a noun to get an estimate of how often the above rule makes an incorrect translation in these cases.

Graphical

A different type of clue, which is basically very simple, and hence one easily mechanized, is the recognition of groups or pairs of words (without regard to their individual meaning). These are the well-known idioms and word pairs. In the system proposed the probability of correct translation of words in an idiom is increased almost to unity by actually storing the whole idiom (in all its inflected forms) in the store. The search logic of the memory is peculiar in that words, or word groups, are arranged in decreasing order on each "page", so that the longest semantic units are examined first. Hence no time is lost in the search procedure, and available capacity is the only criterion for acceptance of a word group for entry in the dictionary. The probability of certain word groups being idiomatic is so high that one can afford to enter them in the dictionary.

In principle, the same solution applies to all word pairs where a particular meaning of one or both of them is selected on the basis of the association. For example état has several meanings, but usually état gazeux means gaseous state. Can one afford to put this word pair in the dictionary? Only experiment, with a machine, can determine the probabilities of occurrence of technical word pairs. Naturally, there will be room for some, but the capacity of any conceivable memory will not accommodate them all.

In any case it would be impossible to store a large number of word pairs, because they are often separated e.g. ne.... que. Hence such graphical clues must be sought, not in the memory, but in the data processor. Here the que entry would impart the program "If preceded by ne within three words, translate as unless."

## Unspecific Words

The choice of multiple meaning like state/account (Fr. état) is not of first importance - the ultimate reader can make his own choice easily. The multiple meaning merely clutters the output text.

The choice of multiple meanings of the so-called unspecified words like de (12 meanings), que (33 meanings) is much more important for understanding a sentence. The amount of cluttering of the output text by printing all the multiple meanings is very great, owing to the large number of meanings of these words and their frequent occurrence.

Booth and Richens proposed printing only the symbol "z" to indicate an unspecified word; others have proposed leaving the untranslated word and others always giving the most common translation.

These seriously detract from the understandability. At the other extreme, one could give all the meanings. In the case of unspecified words, the reader can rarely choose the correct one, so he is given very little additional information and this at the expense of reducing the ease of reading.

The stochastic approach of printing only the most probable permits the best effort in making sense and prints only one word, so it is easy to read. What is the probability of successful translation?

Let us look at a few unspecified French words. Large samples of de have been examined. In 68% of the cases of would be correct; in 10% of the cases de would have been part of a common idiom is the store, and hence correct; in 6% of the cases it would have been associated as de l', de la which are treated as common word pairs, and hence in the store. In another 6% of the cases it would have been correctly translated by the rule sent to the data processor from the store: "If followed by an infinitive verb, translate as to." Another 8% would have obtained by a more elaborate rule: "If followed by adverbs end a verb, then to." The single example of de le verb probably would not have been programmed or stored.

There remains then 8-10% of the cases where the correct translation was in, on, from or vacuous. In some of these cases of could have been understandable, just as in the title of this paper "Stochastic Methods of Mechanical Translation" and "Stochastic Methods in Mechanical Translation" are equivalent. Further study, of course, may reveal some other rules to reduce the incorrect percentage.

Not all unspecified words can be guessed with as high a probability but the bad cases seem more subject to programming.

<u>Meaning Clues</u>

The large majority of choices made by a human are based on the actual meanings of the other words in the sentence. No current machine can duplicate this in a truly intellectual manner, but can be made to carry out the process as its probably actually comes about in the human mind, namely by associations.

Yngve has proposed a scheme along these lines, in which words are assigned to, and associated by, classes. This will take a considerable amount of preliminary study, and may require more storage capacity than is currently available. In addition this scheme requires subsidiary reference to a large store. The additional time of access to out store is prohibitive.

It is possible to embark on a similar approach, in which words are assigned to numbered classes according as they aid in selecting a multiple meaning of another word. For example, <u>gazeux</u> belongs to class 1 because it selected "state" for <u>état</u>. The entry for <u>état</u> would have the rule: "If followed by adjective of class 1, translate as "state". The insertion of this type of clue would not require an ordinate amount of storage space. Of course, the accumulation of such empirical class labels will take some time, but can be started immediately and can be used without the whole job being completed. These investigations can be made quite efficient by the use of our mechanical translating system. The output of the system is designed to print in red any selected semantic unit. Thus large bodies of text can be put through the machine and only these sentences with <u>état</u> printed, and the English equivalent, chosen by the rule (if already introduced), printed in red. In this example the text is examined by an expert in the field to assemble the adjectival classes associated with <u>état</u>.

## <u>The Stochastic Method</u>

It will be possible in the translating system under construction to attack the multiple-meaning problem by selecting only one meaning in all cases, with a high probability of being correct. We believe such output will have a much greater chance of being read than these cluttered with multiple meanings. Even though at times the translation will be rejected as meaningless, this is better than a universal apathetic rejection because of poor readability.

In summary, we believe that this type of attack can be quite successful, but only after a large scale study <u>with the aid of the mechanical translation machine itself</u>.