Stochastic Methods of Mechanical Translation

Gilbert W. King, International Telemeter Corp., Los Angeles, California

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, a machine translator has a much easier problem - 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 generalized 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.

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 inflections (paradigmatic forms) of each stem are 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 are the factors which make 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.

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.

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.

The simplest type of problem is "the unconscious pun" which a human may face in seeing a headline in a newspaper in his own language. 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 the words belonging to more than one field.)

A second type of multiple-meaning problem where the probability of correct selection can be increased substantially and can also be mechanized is the situation where a word has different meanings when it is in different grammatical forms, e.g. the two common and annoying French words: pas (adverb) "not", (noun) "step, pass, passage, way, strait, thread, pitch, precedence", and est (present 3rd sin-

gular 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', etc." Experiment shows this rule (and a similar one for est) has a confidence coefficient of .99 of giving the correct translation.

A more complicated type arises when a word has several meanings as the same part of speech. Here we can only look forward to an approach such as that suggested by Yngve, using the syntax rather than grammar. This type, of course, has by far the largest frequency of occurrence.

The formulas above use grammar (and we hope someday syntactical context) to increase the probability. The human mind uses in addition other types of clue. A fairly simple type, and hence one easily mechanized, is the association of groups or pairs of words (without regard to 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 machine 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. Available capacity is the only criterion for acceptance of a word group for entry in the dictionary. The probability that certain word groups are idiomatic is so high that one can afford to enter them in the dictionary.

In principle, the same solution applies to word pairs. 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, and not for others. The exceptions lie in the same ground that we cannot approach with grammatical clues, but which may be solvable with the syntactical approach, although at the moment the amount of information which would have to be stored seems to be much too large.

The choice of multiple meaning like "dream/consider" (Fr. songe) 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 meaning 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, not only because of the large number of meanings for these words but also because of their frequent occurrence. Booth and Richens proposed printing only the symbol "z" to indicate an unspecified word; others have proposed leaving the word untranslated, and others have proposed 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 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 in the store, and hence correct; in 6% of the cases it would have been associated as "de 1'", "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 2% would have been obtained by a more elaborate rule: "If followed by adverbs and a verb, then 'to'." The single example of de le + verb probably would not have been programmed or stored.

There remain then 8-10% of the cases where "in, on, from" should not be translated at all. In some of the 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 this incorrect percentage.

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

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