[From: *AFIPS Western Joint Computer Conference*, San Francisco, 1960]

THE OUTLOOK FOR MACHINE TRANSLATION

Franz L. Alt
National Bureau of Standards
Washington, D. C.

Introduction

The idea of using electronic digital computers for language translation seems to have arisen about 1946; it was first brought to widespread attention in a memorandum by Warren Weaver in 1949. Today there are about a dozen projects in the United States devoted to machine translation, mostly translation from Russian to English. Many additional projects are located in the Soviet Union, a few in other countries. Many of these projects have been in operation for a number of years. Yet today there is nobody in a position to feed some foreign text into a computing machine and produce automatically an acceptable translation. This is not to say that there have been no accomplishments. Some achievements may be noted, in some other areas there has been considerable progress; and it seems plausible that in a few years completely automatic translation of fair quality will be possible.

Most projects in the United States have been devoted to translation from Russian to English, although French and German have also been considered as source languages. In some projects, studies of the English language have been conducted in the hope of benefiting machine translation. The effort in the Soviet Union, from the scant information available, seems to be of a magnitude comparable to that in the United States, but spread over more language pairs. Although the two major projects in Great Britain have considered translation from French, German and Russian into English they have concentrated mainly on general questions. Not only do different projects work on different language pairs, but they frequently attack different portions of the translation problem. Many limit themselves to source texts in specific fields, such as chemistry or mathematics; some concentrate on the theories of linguistic structures; some on the compilation of a dictionary; some on methods of transcription of the original text onto punched cards or other machine media; some on so-called automatic programming for language translation, which in this case means the creation of general-purpose machine programs which can perform a variety of linguistic operations. When they deal with the central problems of machine rules for analyzing the source words and synthesizing target sentences, there are many differences in the methods which are used. In order to appreciate these differences, we shall first have to survey the main difficulties in the way *of* translation and the methods that have been proposed to overcome them.

Methods of Machine Translation

When the idea of automatic translation emerged, at first only word-for-word translation was considered. The inadequacies of this plan were obvious. They manifested themselves principally in the facts that some source words had more than one target word associated with them, and that the word order in the target language often has to be different from that in the source language. Speaking of multiple target words and of word order, however, is not the most useful approach to classification of problems. The change in word order can be understood by analyzing the grammar of a sentence. Multiple target words can be either different grammatical forms belonging to the same stem, or words having entirely different meanings. Thus, in the main, the problems encountered in translation are classified into syntactic and semantic problems.

The first attack on semantic problems was outlined in the memorandum by Warren Weaver[1] in 1949. Early ideas ran as follows: A word like "nucleus" might have one meaning in a context of physics, another meaning in a context of biology, etc. There are other words which occur only in contexts of physics, still others which occur in biology, and so on. In the dictionary, each word would be coded according to the fields in which it is used. If a word of multiple meaning is encountered, a number of neighboring words both before and after the questionable one would be searched in order to get a "majority opinion" concerning the field with which the present text is concerned.

This idea of word classification by context does not dispose of the problem entirely. The choice among multiple meanings may be determined not by context but by certain successions of words. For instance, the same Russian word may have to be translated into English as "prove" in the sentence "Prove the theorem of Pythagoras" or as "show" in the sentence "Show me how to construct a regular pentagon". Attempts are being made to establish large-scale statistical information about pairs of words occurring together. Another approach has been to attempt a more refined classification of word meanings, resulting in something like an oversized Thesaurus. Still other investigators maintain that the magnitude of the semantic problem can be greatly reduced by attempting whenever possible to find neutral translations which will cover as many meanings of the source word as possible.

The three approaches to the semantic problem which we have just outlined might be named the statistical, systematic and empirical approach. The same three methods of approach can be distinguished in dealing with the syntactic problem.

Here, the statistical approach consists in searching through large amounts of source texts and enumerating the frequency of certain word sequences. For instance, how often does an adjective precede a noun, how often does it follow it? The systematic approach attempts to set up a system of rules—in other words, a machine program —which analyzes the syntactic structure of each source sentence. That is to say, it identifies subject, predicate, direct object, etc. of each sentence or clause. This frequently has to be preceded by a grammatical analysis of each word, just as conventional grammar is divided into morphology and syntax, the former dealing with the inflectional forms of each word, the latter with the function of each word in the sentence. Finally, the empirical approach starts by selecting a few very simple rules for translation, tries them on a body of text and notices where they fail, corrects the rules or introduces new ones to cope with the observed failures, tries the revised rules on a larger body of text, and so forth.

From a slightly different viewpoint we may distinguish between the use of conventional grammar and the design of new systems of grammar (or "linguistic structures", as they are called), which are intended to be better suited to mechanical analysis than is conventional grammar.

There are other differences among the various groups working on machine translation. For instance, in designing the dictionary or glossary, some propose to list in the glossary every inflectional form of every source word, while others propose to list only the stem, or equivalently some canonical form such as the infinitive of a verb and the nominative singular of a noun. Russian nouns have a dozen inflectional forms, adjectives and verbs many more. Thus the size of the required glossary is greatly affected by this decision.

Other differences are found in limiting the scope of a machine translation project. Some groups are satisfied to translate into a kind of pidgin English. Some are resigned to leaving certain semantic ambiguities unresolved and printing out multiple meanings. Some are willing to admit failure in a small percentage of all cases. Some will even admit undetected errors in the translation (The latter is a point of view which I consider dangerous). Some propose to use a man-machine partnership rather than letting the machine do the entire job. In these cases the machine prepares certain aids to translation—at best a kind of preliminary draft, and these are used by a "post editor" in producing a polished translation. Pre-editors are less frequently contemplated, but a certain amount of pre-editing may be combined with manual key punching of the text.

It is my feeling that, in dealing with the syntactic problem, the empirical approach is most likely to give an early appearance of success, by leading in a short period of time to a system for producing less accurate but serviceable translations. It will prove quite difficult to improve the quality of translations later on. The systematic approach may take a little longer in achieving its first concrete results, but these will be of higher quality and will lend themselves readily to further improvement by adding more and more refined rules of grammatical analysis. The statistical approach to syntax seems to me to be far from successful. Ob the prospect of success in the semantic problem I have no present opinion at all.

Example - the Approach of the National Bureau of Standards

The problems associated with mechanical translation may perhaps be better appreciated by looking in detail at one example. For convenience I propose to use the work now in progress at my own organization.

The project at NBS, under the direction of Mrs. Ida Rhodes, concentrates on problems of syntax and dictionary organization. In its approach to syntax it stays close to conventional grammar. We may visualize the proposed machine program as consisting of two parts, the first concerned with glossary look-up and morphological analysis, the second with syntactic analysis. We look upon constructing such a machine code as similar to setting up a mathematical theory of language. Some idiosyncrasies of the project arise from the fact that its staff consists mostly of mathematicians, and coincidentally we are using mathematical texts for our experiments.

The dictionary look-up proceeds on the assumption that, in any foreseeable machine scheme, memory capacity will be a bottleneck. At this point of our exposition this is a mere assumption, but we shall see soon that it is well founded. It will be impossible to store the entire dictionary in the internal memory of the machine. It will be stored on an external medium, for instance on magnetic tape or on a magnetic disc memory. Since reading from an external medium into the computer is relatively slow, every effort is made to store the dictionary as compactly as possible. This desire explains some of the unusual features of the machine program.

The program starts by reading a few hundred words of source text into the machine. Each word is decomposed into its inflectional ending, prefixes, suffixes, and the root. Endings, prefixes, and suffixes are identified by comparison with stored lists, and are replaced by numerical codes. The roots are put in alphabetical order and are then looked up in the dictionary by a single pass through the external medium stored in the dictionary. There, we find the source words stored in alphabetical order of the roots, listing under each root those combinations of prefixes and suffixes which occur in the source language in combination with this root, each followed by grammatical information about the source word and by one or more English equivalents. This information is read into the internal memory.

The foregoing exposition is oversimplified in several respects. Sometimes the machine obtains, in place of a prefix, a group of letters which only looks like a prefix but which grammatically is not one. Such a group is called a pseudo-prefix. Similarly there are pseudo-suffixes. What is left after splitting them off *is* called the pseudo-root. For instance, if the method were applied to English, the word "conifer" would be split into the pseudo-prefix "con", the pseudo-suffix "er" and the pseudo-root "if". Another point that has been oversimplified is the arrangement of the dictionary. Although we called it "in alphabetical order of the roots" the program actually uses a more elaborate and more economical scheme.

The information obtained from the dictionary, together with the previously identified ending, results in a morphological description of the word in question, which is stored for later use. In the second part our code works on one sentence at a time and establishes the elements of each sentence, such as subject, predicate, direct object, etc. In this process we use a device which, to the best of my knowledge, is unique. As each word is identified, it is used to "predict" other sentence elements. For example, a transitive verb predicts a complement in the accusative case, etc. All these predictions are stored and each new word is compared with them. If no match is found, this very fact *is* stored in the expectation that it will be resolved by a subsequent word of the sentence. Thus, throughout this part of the program we have two pools of stored information which we call foresight and hindsight, and which assist in determining the syntactical function of each word in the sentence. This syntactical information, in turn, indicates how our translation should differ from word-for-word translation. For example, it indicates changes in word order, the insertion of English prepositions like of or by, etc.

This account of the method being developed at NBS is necessarily brief and omits many essential features, for instance, how we deal with clauses and phrases within a sentence, or with word forms which are morphologically ambiguous. What I have said will, however, suffice for our present purposes. To date the dictionary look-up and morphological analysis has been coded, and the syntactic analysis is in process of being programmed.

Survey of Projects

The most important ones among the mechanical translation projects in Western countries will be enumerated here, in geographical order from West to East, and each project briefly characterized. For a recent complete survey, see e.g. Bar-Hillel.

1. University of Washington, one of the oldest projects, concerned with translation from German and Russian to English. Extensive dictionary of inflected forms. Word-for-word translation, study of selected grammatical and semantic problems.

2. University of California, Berkeley, one of the youngest projects. Russian to English. Emphasis on economy in the use of the dictionary.

3. Rand Corporation, Russian to English, empirical and statistical approach, some work on semantic problems. Large corpus of transcribed Russian material in Physics and Mathematics.

4. Ramo-Wooldridge, Russian to English, empirical approach, emphasis on programming methods and on display of results in a form which facilitates continuing revisions. In its approach to grammar this project is close to one of the groups formerly at Georgetown University.

5. University of Texas, a newcomer. German to English, concentration on grammatical problems.

6. Wayne State University, a new group, Russian to English, statistical approach. Cooperates with Ramo-Wooldridge and with a group formerly at Georgetown University.

7. Georgetown University, one of the oldest groups, at one time comprised four separate projects. One *of* these is disbanded, its ideas being continued at Ramo-Wooldridge and Wayne. Another, Russian to English on an empirical basis, recently moved from Georgetown to a private corporation, CEIR.

A third project at Georgetown, called "General Analysis Technique", Russian to English, has made considerable progress with syntactic analysis, which is being developed in part empirically and in part systematically, staying fairly close to conventional grammar. A good-sized dictionary has been assembled and a large corpus of text examined. Heavy reliance is placed on post-editing.

The fourth of the Georgetown projects works on French to English, is strictly empirical, lays stress on advanced programming techniques.

There are also small-scale efforts devoted to Chinese and Arabic.

Most projects at Georgetown are concerned with source texts in the field of chemistry.

8. National Bureau of Standards, a relatively recent project, Russian to English, systematic approach to grammatical (morphological and syntactic) analysis, staying close to conventional grammar. Emphasis on economical use of dictionary. Uses source texts in mathematics.

9. University of Pennsylvania, systematic approach to grammatical structure of languages, esp. English.

10. IBM Corporation, concerned mostly with hardware, with some supplementary systems studies.

11. Massachusetts Institute of Technology, one of the oldest projects, has pioneered in developing new theories of the grammatical structure of languages. Also works on syntactic problems of German-to-English translation and on general-purpose programming systems for machine translation in general.

12. Harvard University, Russian to English, has compiled a large dictionary and worked systematically on methods and machine codes for dictionary compilation and updating, and on word-for-word translation. More recently concerned with grammatical analysis on lines similar to National Bureau of Standards, and with general theory of language structures. A related project is in operation at the Arthur D. Little Company.

13. Birkbeck College, University of London, apparently the oldest group in the field. German, French and some Russian to English. Morphological analysis of source words, some syntax, largely empirical.

14. Cambridge University, England. General linguistic theory, semantic problems.

15. University of Milan, Italy. A highly theoretical, long-range approach.

There are a few other projects, either too small or too new to be characterized separately. They include one at the National Physical Laboratory in England and one in France.

Economic Considerations

We now turn to the discussion of the cost of using electronic computers for language translation. Cost depends on the system used, i.e. on equipment and methods. Although we shall mostly deal with cost in dollars, there are elements of cost not readily expressible in terms of money, such as the damage caused by time delays or by erroneous translation.

There are so many uncertain and unpredictable factors in the situation that we can discuss only orders of magnitude. In some cases we can estimate within a factor of 2 or better; in other cases we are lucky to come to within a factor of 10 of true cost. Sometimes, rather than assign a cost to a process, we shall state the conditions under which this process becomes competitive with others.

Human Translation

The cost of translating "by hand", i.e. by human translators, seems to range from 0.25 to 4 cents or more per word. If quoted per page, one hears figures ranging from $12 to $25 per page. The cost depends on the qualifications of the translator, on the technical difficulty of the text, and on the degree of perfection, smoothness of style and appearance desired. It is cheaper in foreign countries than in the United States. As an indication of order of magnitude we may use the figure of 1 cent per word. In this connection "word" means the usual average of five characters followed by one space. To avoid confusion with the "machine word", the unit of memory space in computers, some authors use the word "grubit" for the latter. It happens that on many machines a word corresponds to about six characters.

Apart from the monetary cost, human translation suffers from its slowness and from the scarcity of experienced translators, especially for technical subjects. The quality of translation is variable, and from time to time gross errors appear which could be avoided by translators with greater technical competence or by the use of very detailed technical dictionaries. By and large, however, the quality of human translation is far superior to anything in sight of mechanical translation.

The cost of mechanical translation may be divided into the two elements of initial cost and current cost.

Initial Cost

Under this heading we shall discuss the cost of developing programs and codes for automatic translation.

As always in cost accounting questions, there is some ambiguity about where some specific items should be charged. In the present problem, it could be argued that the initial cost of developing the hardware should be treated as an element of the initial cost of translation. It is, however, my contention that mechanical translation will, if at all, be performed by general-purpose equipment which would be developed in any case, regardless of its application to automatic translation.

Current spending in the United States for the development of machine programs for translation probably amounts to between 2 and 3 million dollars per year. The effort maintained in the U.S.S.R. is of the same magnitude, though perhaps at lower cost. Including projects in other countries, an amount between 5 and 10 million dollars per year is being invested in this work. One may speculate that at the present rate it will take at least five years before the system is really perfected, notwithstanding certain glowing newspaper accounts about systems supposedly now in operation. It is true that certain simple-minded things can be done right now, certainly the word-for-word translation and probably something a little better. It is also very probable that some concrete and respectable progress will be visible two or three years from now but I should estimate at least five years before a really satisfactory system is reached. Even this may be an underestimate. Before we are through, the world-wide investment in these programs may be $50 million.

Now there are good reasons for saying that this investment need not be amortized. Present research in automatic translation is of great scientific interest, will result in deep insights into the working of the human mind and in great improvements in our use of computers in general, and is worth being carried on for its own sake. But let us take the narrow viewpoint and insist on amortizing it. The question is, over how long a period of time or what volume of work? Let us assume that, after a number of systems have been perfected and the best of them selected, we may wish to translate, over a number of years, a total of 40,000 volumes averaging 500,000 words each, or a total of 20 billion words, so that the amortization cost per word is 1/4 cent. (We may disregard interest cost for our purpose.) Forty thousand volumes is the size of a medium-sized technical library; in a large technical library probably the new accessions alone total 10,000 volumes in one or two years. Remember, also, that translation will be made from and into several languages. Thus, even if amortization is extended over a small number of years, it will cost only a small fraction of a cent per word. In other words, work on automatic translation systems is an excellent investment.

Equipment Costs

The current cost of machine translation will depend on the hourly cost of the machines used and the amount of machine time for translation.

These factors, of course, vary greatly from one machine to another. As a starting point, I propose to discuss the most advanced machines on which complete performance and cost specifications are available, machines like the Stretch computer of IBM or the LARC of Sperry-Rand. These computers come with a variety of optional equipment, and depending on this we may estimate their hourly cost at between $1,000 and $2,500. (This is intended to be first shift rental, without overhead or other charges.) A little later we shall be more specific about the kinds of optional components we may wish to include in a system designed optimally for language translation.

These computers are scheduled to perform operations like addition, subtraction, transfers, logical operations, etc. at the rate of about one microsecond per operation. Multiplications and divisions take longer, but these operations are hardly used in a machine program for language translation.

Apart from the computing machine itself, we have to consider terminal equipment. There seems to be no problem at the output end. Conventional tape-controlled printers can print a thousand words in something less than a minute, at a cost of a fraction of one dollar. More advanced equipment will undoubtedly be even more economical.

The input operations are indeed critical for the success of machine translation. If a Russian text had to be punched into cards, say, by hand, the cost of this operation and of checked and revised punching to the required degree of accuracy would be almost comparable to the entire cost of human translation. Fortunately there seems to be reason to expect that reading machines, capable of scanning printed pages of Russian texts and recording them on some mechanical medium suitable for machine input, will be available in the next few years. Some rudimentary machines are already in existence. We have no specifications, nor reliable cost figures, for an entirely satisfactory system, but it seems likely that the cost of operating this equipment will be considerably less than the cost of key punching.

Machine Time Requirements

I do not yet know how many instructions will be required for the entire translation program. It will certainly be several thousand, it may be several tens of thousands. Not all of these will be executed for every source word which is to be translated. Most of them will come into play only occasionally, while some will be executed repeatedly for every word. Let us estimate that 10,000 instructions have to be executed for each word to be translated. On next-generation computers the execution time of most instructions is about one microsecond, so roughly speaking we can translate at the rate of 100 words per second. This translation process has to be overlapped by reading from the dictionary. It has been estimated that the dictionary will have to contain 50,000 Russian stems, and that even in a highly condensed version of storing in the dictionary such as that of NBS, about 20 machine words will be required for each Russian stem. This is a total of one million machine words, and is probably also on the high side. There is no difficulty about storing one million machine words on a magnetic disc memory. To read through this dictionary in alphabetical order requires about one second, and during that second the machine can simultaneously translate about 100 words. Thus, the optimal procedure will appear to be to read the source text into the machine in batches of about 100 words, alphabetize them (or rather their roots), read through the dictionary and extract the information pertaining to these 100 words, and then proceed to translate while the next hundred words are being read in and looked up. Internal memory required *is* only 2,000 words for the dictionary information pertaining to 100 source words, storage space for all instructions required by the program, and some temporary working storage. A total of 32,000 words of internal memory is probably adequate. The hourly cost of a machine with 32,000 words internal memory and a magnetic disc memory is on the order of $1,000. The machine translates at the rate of 100 words per second, or at a cost of 1/3 of a cent per word. This, as we said before, compares with one cent per word by human translator. Since all our estimates are quite crude, all we can say is that the cost of machine translation by this scheme is probably not higher than by human translator.

To see how these economic factors might depend on our choice of equipment, let us suppose that we wish to use the same translation system on one of today's machines, like an IBM 704 or a Univac. Here we have no disc memory, so that the dictionary has to be stored on magnetic tape. This will take something like half an hour for reading into the machine. The internal memory of these machines is at most 32,000 words, of which we can assign at most 20,000 for copying the pertinent portions of the dictionary, reserving the remainder of the memory for instructions and working space. Twenty thousand machine words of memory will hold the dictionary information corresponding to 1,000 source words. Thus, the procedure would be to read the text in batches of 1,000 words, alphabetize them, copy from the dictionary and translate. The translation time for 1,000 words, at 10,000 instructions each, is about five minutes. Thus the computer is poorly used, most of its time is spent in reading through the dictionary over and over again, and the total time required to translate 1,000 words is over half an hour, giving a cost of about 15 cents a word. Here again, since all our estimates were quite crude, we can only say that machine translation by means of today's computers is probably more expensive than by human translator, perhaps by an order of magnitude.

There are other factors to be considered. Many instructions are used quite infrequently. For example, the Russian word for "and" requires a whole subroutine all by itself. Probably many other words require special treatment which has to be incorporated into special subroutines, and these might in the end account for a large portion of the instructions used. There is no reason why such subroutines could not be stored in the dictionary with the word to which they pertain. Thus, we can to some extent trade dictionary space for internal memory capacity. The same can be done in reverse order by selecting a fairly large number of frequently occurring words and storing all dictionary information about them in the internal memory, reserving the external dictionary for the less frequently used words.

It may turn out that our extreme insistence on conserving space in the dictionary has been unnecessary. Perhaps we can loosen up a bit, waste a little space in the dictionary and store information in less concentrated form. This may result in a dictionary several times larger, but on future machines all reading from the external dictionary will be overlapped by computing, and the computing time would actually be reduced by this more generous allotment of external storage space.

Our present crude estimates of the requirements for different translation systems do not enable us to choose definitely a *beet* among them.

translation.

Apart from the computing machine itself, we have to consider terminal equipment. There seems to be no problem at the output end. Conventional tape-controlled printers can print a thousand words in something less than a minute, at a cost of a fraction of one dollar. More advanced equipment will undoubtedly be even more economical.

The input operations are indeed critical for the success of machine translation. If a Russian text had to be punched into cards, say, by hand, the cost of this operation and of checked and revised punching to the required degree of accuracy would be almost comparable to the entire cost of human translation. Fortunately there seems to be reason to expect that reading machines, capable of scanning printed pages of Russian texts and recording them on some mechanical medium suitable for machine input, will be available in the next few years. Some rudimentary machines are already in existence. We have no specifications, nor reliable cost figures, for an entirely satisfactory system, but it seems likely that the cost of operating this equipment will be considerably less than the cost of key punching.

Machine Time Requirements

I do not yet know how many instructions will be required for the entire translation program. It will certainly be several thousand, it may be several tens of thousands. Not all of these will be executed for every source word which is to be translated. Most of them will come into play only occasionally, while some will be executed repeatedly for every word. Let us estimate that 10,000 instructions have to be executed for each word to be translated. On next-generation computers the execution time of most instructions is about one microsecond, so roughly speaking we can translate at the rate of 100 words per second. This translation process has to be overlapped by reading from the dictionary. It has been estimated that the dictionary will have to contain 50,000 Russian stems, and that even in a highly condensed version of storing in the dictionary such as that of NBS, about 20 machine words will be required for each Russian stem. This is a total of one million machine words, and is probably also on the high side. There is no difficulty about storing one million machine words on a magnetic disc memory. To read through this dictionary in alphabetical order requires about one second, and during that second the machine can simultaneously translate about 100 words. Thus, the optimal procedure will appear to be to read the source text into the machine in batches of about 100 words, alphabetize them (or rather their roots), read through the dictionary and extract the information pertaining to these 100 words, and then proceed to translate while the next hundred words are being read in and looked up. Internal memory required is only 2,000 words for the dictionary information pertaining to 100 source words, storage space for all instructions required by the program, and some temporary working storage. A total of 32,000 words of internal memory is probably adequate. The hourly cost of a machine with 32,000 words internal memory and a magnetic disc memory is on the order of $1,000. The machine translates at the rate of 100 words per second, or at a cost of 1/3 of a cent per word. This, as we said before, compares with one cent per word by human translator. Since all our estimates are quite crude, all we can say is that the cost of machine translation by this scheme is probably not higher than by human translator.

To see how these economic factors might depend on our choice of equipment, let us suppose that we wish to use the same translation system on one of today's machines, like an IBM 704 or a Univac. Here we have no disc memory, so that the dictionary has to be stored on magnetic tape. This will take something like half an hour for reading into the machine. The internal memory of these machines is at most 32,000 words, of which we can assign at most 20,000 for copying the pertinent portions of the dictionary, reserving the remainder of the memory for instructions and working space. Twenty thousand machine words of memory will hold the dictionary information corresponding to 1,000 source words. Thus, the procedure would be to read the text in batches of 1,000 words, alphabetize them, copy from the dictionary and translate. The translation time for 1,000 words, at 10,000 instructions each, is about five minutes. Thus the computer is poorly used, most of its time is spent in reading through the dictionary over and over again, and the total time required to translate 1,000 words is over half an hour, giving a cost of about 15 cents a word. Here again, since all our estimates were quite crude, we can only say that machine translation by means of today's computers is probably more expensive than by human translator, perhaps by an order of magnitude.

There are other factors to *be* considered. Many instructions are used quite infrequently. For example, the Russian word for "and" requires a whole subroutine all by itself. Probably many other words require special treatment which has to be incorporated into special subroutines, and these might in the end account for a large portion of the instructions used. There is no reason why such subroutines could not be stored in the dictionary with the word to which they pertain. Thus, we can to some extent trade dictionary space for internal memory capacity. The same can be done in reverse order by selecting a fairly large number of frequently occurring words and storing all dictionary information about them in the internal memory, reserving the external dictionary for the less frequently used words.

It may turn out that our extreme insistence on conserving space in the dictionary has been unnecessary. Perhaps we can loosen up a bit, waste a little space in the dictionary and store information in less concentrated form. This may result in a dictionary several times larger, but on future machines all reading from the external dictionary will be overlapped by computing, and the computing time would actually be reduced by this more generous allotment of external storage space.

Our present crude estimates of the requirements for different translation systems do not enable us to choose definitely a best among them.

Several of them appear to yield cost figures of the same order of magnitude. Whichever will be the final choice, it does seem likely that some system will be found whereby machine translation will be considerably more economical than human translation.

References

1. Weaver, W., Translation. Machine Translation of Languages (W. N. Locke and A. D. Booth, eds.), Wiley, New York, 1955.

2. Bar-Hillel, Y., Ihe present status of automatic translation of languages. Advances in Computers *1,* Academic Press, New York, 1960.

3. See, for instance, a series of papers on this subject in Proceedings of the Eastern Joint Computer Conference 1959, in press.