[From: *Communication and Electronics* 28 (January 1957)]

The Technical Feasibility of Translating Languages by Machine

VICTOR H. YNGVE

NONMEMBER AIEE

RECENT DEVELOPMENTS in the design of electronic digital computers have suggested the possibility of translating languages by machine, but when one tries to write a good translation program for a computer, one finds that there is much to be learned about language. For this reason, no program has yet been written by which a computer could translate usefully; but a number of people have been challenged by the problem of filling in the gaps in present knowledge about language and translation.[1,2]

This increasing interest in the possibility of mechanical translation is not surprising. Translation on a large scale is becoming more and more urgent in a rapidly shrinking world still partitioned by many language barriers. There are at least 50 languages that possess written records worth translating, and at least 200 languages into which the scientific, technical, and cultural records of our civilization might be translated. The price that each language community pays for the luxury of expressing itself in its own language is the cost and the difficulty of communication with the rest of the world. Today, communication between language communities must be funneled through individuals who are to some extent bilingual. If special-purpose machines could be devised for translating written material, the cost of translation would be significantly reduced. Language communities could then retain their individuality and still gain the advantages of increased communication.

This paper will explore two types of mechanical translation. The first, which can be called word-for-word translation, is within immediate reach, but promises results that are crude, at best. The second type, sentence-for-sentence translation, which is the type being studied at the Research Laboratory of Electronics, Massachusetts Institute of Technology, promises translation of a higher quality, but a great deal of effort will be required to work out the linguistic details.

Word-for-Word Translation

Word-for-word translation is based on the use of a dictionary. The words of the input text are looked up in order, one by one, in a dictionary, and the meanings are written down in order, one by one. In determining the meanings, only one text word at a time is considered.

Word-for-word translations can be produced by machine in the following way: The material to be translated is typed on a keyboard device which transfers the characters to some input medium such as punched tape, punched cards, or magnetic tape for entry into the machine. The machine then compares the words one at a time with the dictionary entries that are stored in the memory. When each input word is located in the memory, the dictionary entry stored there provides a meaning that consists of an equivalent word or words in the other language. The equivalents found in the dictionary are then sent one at a time to an output printer that prints a word-for-word translation.

If a machine that can produce rough translations on a production basis is wanted as soon as possible, the word-for-word machine is the favored type because it is simple and can be implemented now. If translations better than word-for-word translations are wanted, work on an automatic dictionary should still be undertaken because any machine that translates will need a dictionary, and the technical problems arising in the design of an automatic dictionary are not trivial.

There are good reasons for taking the word, delimited by spaces or punctuation, as the item to be looked up, instead of some smaller unit. It has been suggested that if the roots of words and their inflectional endings were stored in the memory separately, much space would be saved in the dictionary. If a given word can be inflected in five ways, the number of entries in the dictionary could, perhaps, be reduced by nearly four-fifths, since the inflectional endings are common to many words. As another space-saving device, it has been suggested that compound words, such as *blackboard, counterelectromotive, Koordinatenanfangspunkt,* could be separated into their component parts, and each part looked up separately. Suggestions such as these for splitting up words have at least two disadvantages. One is that the separation of a word into parts before it is looked up requires a method for deciding where to make the separation. Part of the storage space saved has to be used for a more complicated program or look-up routine. Another disadvantage is that parts of words and inflectional endings have, in general, more different meanings than fully inflected words. Strict word-for-word translation completely avoids the segmentation problems and avoids some of the multiple meanings of parts of words—inflectional endings, in particular.

In the next sections, word-for-word translation will be considered in detail. The requirements of the task will be compared with the capabilities of present-day, general-purpose digital computers and of special-purpose machines that might be constructed for translation; then the usefulness of the product will be assessed.

Paper 56-928, recommended by the AIEE Communication Theory Committee and approved by the AIEE Committee on Technical Operations for Presentation at the AIEE Fall General Meeting, Chicago. Ill., October 1-5, 1956. Manuscript submitted June 11, 1956; made available for printing July 19, 1956.

VICTOR H. YNGVE is with Massachusetts Institute of Technology, Cambridge, Mass.

The work on which the paper is based was supported in part by the U. S. Army (Signal Corps), the Air Force (Office of Scientific Research, Air Research and Development Command), and the Navy (Office of Naval Research); and in part by the National Science Foundation.

MACHINE REQUIREMENTS: SIZE

The size of the memory of a machine is of importance in word-for-word translations. Memory size depends upon the number of dictionary entries, the average length of each entry, and, in stored-program machines, on the storage requirements of the program.

The number of words in a language is difficult to estimate. If the entries in dictionaries are counted, numbers around a million are obtained. However, no dictionary contains all the words of a language; new words are being added every day. The vocabulary of a language is not a closed system; it is always open to additions. If translating is limited to one field of knowledge (electrical engineering, for example) most of the words belonging to other fields may never appear, and a glossary containing electrical engineering terms and certain general scientific terms might be adequate. But, like the vocabulary of an entire language, the vocabulary of a specific field is not closed; and an author of a paper in a special field often uses terms from other fields.

A glossary should contain the words most frequently used in a specific field. Since it is virtually impossible to list all the words of a language, or even of a specific field, new words will have to be added from time to time as they are needed. It is feasible to start with a rather small list of words and add new ones where they are first encountered or after they have appeared a given number of times. In this way, the glossary is always an approximation to a list of the words of highest frequency in the specific field.

The most important parameter to be considered in evaluating a glossary is not the number of words that it contains, but some parameter connected with the number of running words of a text that can be found in the glossary. The coverage of a glossary with respect to a certain test can be defined as the ratio of the number of running words of the text that can be found in the glossary to one more than the number of words that cannot be found. It is the average length of uninterrupted runs of text words that can be found in the glossary. The coverage of an English glossary with respect to a Russian article is zero. The coverage of a glossary containing the 50 or 60 most frequently used English words is about one, because these words make up about half of the running words in a text. The coverage of a glossary with respect to an article is numerically equal to the number of words in the article if the glossary contains them all. To have a reasonably high coverage with respect to a field such as electrical engineering, a glossary should contain at least 5,000 carefully chosen words.

In assessing the size of memory needed for an automatic dictionary, it is also important to form some estimate of the length of each entry. The distribution of lengths of the different words in a language is characteristic of the language. In English, the average word length is about 7 or 8 letters; in German it is somewhat longer. German has many words with more than 20 letters. Since word lengths are so variable, and since each entry may contain one or more words as equivalents, there will be considerable variation in entry size. Perhaps the average size will be something like 20 letters; the maximum size may run to more than 100.

The amount of storage needed for a minimum dictionary of 5,000 entries with an average entry size of 20 characters of

5 binary digits each is 500,000 binary digits.

Now that an estimate of the memory capacity requirements has been made, the adequacy of present-day machines can be examined. Many general-purpose digital computers have three kinds of memory: magnetic core, magnetic drum, and magnetic tape. A storage capacity of 500,000 binary digits is, by present standards, large for a magnetic core memory. Even many magnetic drum machines do not have this capacity. One is forced to the conclusion that magnetic-tape storage must be used with most present-day general-purpose computers if an adequate dictionary is to be stored.

MACHINE REQUIREMENTS: SPEED AND ECONOMY

While the ability of a machine to make word-for-word translations is determined primarily by the size of the glossary that it can store, its usefulness for producing translations economically depends on a number of factors. These include the cost of operating the machine, the access time of the memory, the register length, the type of commands available and their execution times, and the way in which the program is written.

In present-day machines the memories are organized by address. This means that the memory contains a number of pigeon holes or memory locations, each one designated by a number. The machine is arranged so that ready access is obtained to each memory location when it is referred to by its address. The average time that it takes to refer to a given address and extract the contents of the memory from that position is called the access time. For core memories the access times are very short, in the neighborhood of 10 to 20 microseconds. For drum and tape memories, the concept of access time is not quite as appropriate, since the actual time required to find something at a given address varies greatly, being dependent on the time it takes the drum or tape to move to that address. Average access times for drums are around 10 milliseconds.

In a dictionary look-up program, the access time is not the only factor in determining how quickly a given word can be found; although the word is known, its address is not. The binary representation of the word itself cannot be used as the address, because the spelling of words is highly redundant. For a 5,000-word memory, the address would need to be only 13 binary digits long; but the spelling of a word might require well over 100 binary digits.

Various programming methods have been devised for searching a memory when the address is unknown. Clearly, it is not economical of time to start at the beginning and examine each entry in turn, through the whole dictionary, until the right entry is found. An improvement consists of using the thumb index device, whereby the search is divided into two steps. The initial one or two letters of the word are found in the index part of the dictionary. The index entry gives the position in the main list at which the search should begin. It is possible to extend the thumb index idea and divide the search into three or more steps. If this procedure is carried to the extreme, the bifurcation, or "20-questions," technique is reached. Each step of the search chooses one out of two approximately equal sections of the dictionary. If the division is exactly into two each time, one out of 2^n entries can be found in n trials. Another possibility is to try to obtain from the spelling of the word some function that specifies, as closely as possible, the address of the word. With this scheme, surprisingly good results can be obtained.

It is also possible to take advantage of the vocabulary statistics. The probability p_r that the rth word from the vocabulary will appear next in a text is given by the Zipf law[3] as $p_r = P\ r^{-1}$, where r is the rank of the word when words are ranked in order of decreasing frequency, and P is a constant with a value of approximately 1/10. A more accurate law has been derived by Mandelbrot.[4] It is $p_r = P\ (r+m)^{-B}$, where B is a constant in the neighborhood of one, and m is another constant. The constants P, m, and B are characteristic of different languages and, to some extent, of different authors and styles of discourse. According to these two laws, as few as 50 or 60 different words make up about half of the total words in a text. If these 50 or 60 words are placed in a core memory with effectively instantaneous access, the over-all access time can be reduced by a factor of 2.

For a memory such as magnetic tape, the access time to a given entry depends on its position along the tape in relation to the reading heads. The most effective way to use a magnetic-tape memory is to look up the words in batches, not one at a time. Each batch of words is alphabetized in the fast memory of the machine. All of the words in a batch can be looked up in one pass of the tape, if the words on the tape are also alphabetized. The average rate of access can be greatly increased by this method at the expense of some delay in translation and some storage and manipulation requirements in a fast memory. The concept of random access time is almost meaningless when words are looked up in batches The appropriate parameters of the system are the average rate of information flow through the machine, and the average delay.

Present-day digital computers have been designed to handle arithmetic at very fast rates. A dictionary search program for one of these machines usually involves a great deal of arithmetic, not because it is required by the problem but because the machine does it so well. A modern fast computer would not be as economical as a machine specifically designed for translation. Such a machine should be capable of handling entries of variable length and designed specifically for searching the memory when the address is not known. It should be specially designed for matching sequences of binary digits, and determining whether the numerical value of one sequence greater or less than that of another.

Memory research is progressing rapidly. In another few years, high-capacity magnetic-disk memories, special magnetic tape or drum arrangements, or various types of photographic memories will probably be developed to the point where they will be available for translating machines. It may not be necessary to resort to word-splitting techniques in order to store a high-coverage glossary a machine that is economical to use.

The simple word-for-word translation schemes that have been discussed have the advantage that they can be implemented now. To decide whether or not their implementation is worth while, the potential usefulness of the translations should be considered.

USEFULNESS OF OUTPUT

If translation, from German to English, for example, is considered as the selection of one out of all possible sequences of English words for each German sentence, a word-for-word translation is a good first approximation. English sentences average about 20 words. Assuming a vocabulary of 10^5 words, there are 10^{100} possible sequences of this length. A word-for-word translation may supply an average of about two meanings for each German word. If the word-for-word translation is good, the correct English translation can be found by choosing the correct meaning for each word from those supplied, and then choosing the correct word order. The initial choice of one out of 10^{100} possible sequences has been reduced to one out of 220 x 20!, or about one out of 10^{24}. For each sequence left to choose from, approximately 10^{76} have been ruled out. It is thus seen that the word-for-word translation has done a Herculean job, but there is still much work for the reader before he can read the translation. He must still choose one out of 10^{24} possibilities. This he presumably does on the basis of his knowledge of English sentence structure and on the basis of his knowledge of the subject matter. The problems of multiple meaning and of word order are the two difficulties that he must face.

To make the reader's task easier, various suggestions have been made for the arrangement of the multiple choices on the page. They could be strung out in linear order, perhaps with parentheses; or the most likely meaning for each word could be put first or written in bold type; or the various meanings could be listed in columns. Another question concerns the number of meanings that should be provided for each word. Perhaps one broad meaning would be less confusing than two or three narrower meanings. Another suggestion for making the reader's task easier is called partial translation. Partial translation is based on the fact that only a few hundred articles, prepositions, conjunctions, and adverbs make up a large majority of the running words in a text and that these words, together with a few inflectional endings, have especially long multiple-meaning lists. The suggestion is that these few hundred items be given in their original language, not to confuse the reader with long lists. The reader, after a brief introduction to the grammar and sentence structure of the foreign language, would be able to understand these words and endings and the constructions in which they are used. All of the technical terminology would be translated for him.

The reader of the translation plays a vital role. In a sense it is he who makes the translation, because he alone understands the meaning. The machine only recodes the foreign language so that it can be more easily translated by the reader. Experiments have shown that word-for-word translations can be deciphered more easily if the reader is familiar with the topic that is being discussed. If he knows what the author is trying to say, he can recreate the work with very few dues. Thus, word-for-word translation could be useful for people who want to keep abreast of large amounts of foreign technical literature in their field, but it would probably not be adequate for a careful reading of any one article.

It has been suggested that rough word-for-word translations should be revised and edited by an expert. For certain purposes, revision may be advisable, but perhaps the best course would be to let the final reader make his own revision, since he is an expert in the field and can use his background knowledge in deciding what the author probably meant.

It is very important that the final reader be considered a link in the translation chain. Any improvement in the translation which speeds up the reader's comprehension is worth while, since the reader's time is expensive.

CONCLUSIONS

Word-for-word translation can be implemented now by using present-day general-purpose digital computers or, more economically, special-purpose machines. The output might be useful, but would definitely be of poor quality, probably so poor that the reader would have to be highly motivated before he would want to use it.

Sentence-for-Sentence Translation

Attention will now be given to the question: How can a machine produce something better than a word-for-word translation? The deficiencies of word-for-word translations will be examined, not from the point of view of how a reader may cope with them, but from the point of view of how a machine may do so. A diagnosis of the difficulties will be attempted, and a remedy will be proposed. This will necessitate some examination of the way in which words are combined to make sentences in a language.

ANALYSIS OF DIFFICULTIES

The major shortcomings of word-for-word translations have been grouped under multiple-meaning difficulties and word-order difficulties. It is now advantageous to reclassify them as lexical difficulties and structural difficulties.

Lexical difficulties include the multiple meanings of words like *power, force, integrate, potential,* and so on. It has been suggested that the use of a field glossary would restrict these words to one meaning, the technical one. A startling exception to the observation that most nouns, verbs, and adjectives have few meanings is the word *run* which, according to Webster, has 54. (The average number of meanings per word in *Webster's Collegiate Dictionary* is about 2.5.)

Structural difficulties include word order, inflectional endings, multiple meanings of most articles, prepositions, conjunctions, and adverbs. Most of the multiple-meaning difficulties in a text are structural because a majority of the running words in a text are structural words, with more than the average number of meanings. The 54 meanings of *run* are divided into four lists that are structurally different: transitive verb, intransitive verb, noun, and adjective. It is clear that a machine method of handling structural problems is highly desirable.

Diagnosis of the difficulties is easy. After all, a word is usually not ambiguous when its context is considered. In word-for-word translation, no attempt is made to translate a word in its contest, but a meaning or meanings are assigned to it on the basis of its form alone. From this point of view, a word-for-word translation might be called a literal translation, whereas a translation which takes into account the context might be called a free translation. Whether a translation is literal or free really depends upon the point of view. A word-for-word translation translates words literally, but translates letters and syllables freely. Words may be translated freely if larger blocks are translated literally. In general, a literal translation will be better if it is based on larger blocks of text.

A way is needed to use the context in translating the words. Investigation has shown that using the context of a complete sentence is a great deal better than using smaller parts of a sentence, say a phrase, but is not significantly worse than using the context of two or more sentences. A sentence seems to be a natural language unit. To see how the information in a sentence context can be used for translation, more knowledge of language structure is needed.

STRUCTURE OF LANGUAGE

The most significant thing which may be said about a sentence is that it has a structure. Some hint of this is given in the grammar taught in grammar school. where words and groups of words are classified in terms of nouns, verbs, subjects, objects, and so on, and simple sentences are parsed in order to show the interrelation of their parts.

It is rather illuminating to consider some of the quantitative aspects of sentence structure. Of the 10^{100} arbitrary sequences of 20 words, perhaps only 10^{30} to 10^{50}, are actually English sentences. Practically all of the remainder are mere incoherent jumbles of words. Sentence structure thus places very strong constraints on the sequences of words. Another way of appreciating the degree of constraint imposed by sentence structure is to take a 20-word sentence from a newspaper and rearrange the words. Of the 20! different arrangements, only a few are English sentences. That there must be some structure or system underlying these constraints is apparent. A man in his lifetime would not be physically able to hear or speak more than 10^9 sentences, yet he can instantly recognize whether any given sequence of 20 words is included among the 10^{30} to 10^{50} English sentences or is left in the remaining 10^{100} ungrammatical jumbles of words. He can do this by recognizing whether or not the sequence conforms to one of the allowable sentence patterns. Recognition of sentence pattern is based primarily on formal features of the sequence and not on the meaning; for the meaningless lines by Lewis Carroll:

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.

are instantly recognized as conforming to English sentence patterns.

Constraints powerful enough to restrict the sequences of words to only those that conform to the sentence structures of a language can be represented by the following scheme. A set of structural choices is encoded into a sequence of words. The first step is to choose one of several possible sentence types. For English these might be:

sentence = 'I' + predicate *a*

or

sentence = 'you' + predicate *b*

or

sentence = subject *c* + predicate *c*

etc., or

sentence = 'it' + predicate p + 'that' clause

etc.

In this "sentence" stands for all sentences in English; "predicate *a*" stands for all predicates that can follow the first person, singular subject "I"; and so on. Take the example:

sentence = 'it' + predicate *p* + 'that' clause

This means that the sentence will start with the word 'it,' followed by a particular type of predicate, predicate *p,* and followed in turn by a clause introduced by the word 'that.' The particular predicate of type *p* and the particular 'that' clause will emerge as a result of further choices. For the predicate, the choices might be:

predicate *p* = verb 'to be' + adjective *b*

or

predicate *p* = verb 'to be' + prepositional phrase

or

predicate p = verb 'to be' + past participle *c*

etc.

The first of these might be chosen. Further choices would decide the tense of the verb, the present tense, for example. For the adjective, the choices might be:

adjective *b* = 'clear'

or

adjective *b* = 'true'

or

adjective *b* = 'ikely'

etc.

If 'clear' is chosen, the sentence so far is: "It is clear + 'that' clause." Note that adjective *b* would not include adjectives like hot, red, small, and so on.

An encoding scheme of this sort can allow for certain types of recursiveness in sentence structure. For example:

It is clear that it was likely that it was true that...

This type of recursiveness is indicated by

'that' clause = 'that' + sentence

which leads back again to the beginning of the set of choices.

An encoding scheme of this sort, starting as it does with a set of choices or directions for producing a certain sentence and ending with the words of that sentence, can be thought of as a grammar of a language from the point of view of the speaker rather than of the listener. Schemes from the speaker's point of view have the advantage that each set of choices leads unambiguously to a unique sequence of words. On the other hand, a given sequence of words may be ambiguously represented by two or more different sets of choices. As an example of this, consider that

predicate *a* = 'am here now'

or

predicate *a* = 'see it now'

etc., and

predicate *b* = 'are here now'

or

predicate b = 'see it now'

etc.

Then 'see it now' can be produced ambiguously as a first-person singular or a second-person predicate. A grammar from the listener's point of view could resolve this ambiguity only by consideration of the context 'I' or 'you.' Similarly, words can have multiple meanings from a literal point of view, but be unambiguous if their role in the sentence structure is known. As an example, consider this English sentence:

The (Der) man showed the (dem) boy the (das) picture.

and this part of a German sentence:

Der (the) Mann, der (who) der (of the) Tat fähig ist,. . .

In this way, considerations of sentence structure can resolve many of the multiple-meaning problems.

Usually each possibility of choice of structure in a language has a meaning assigned to it. A distinction has been made between the "structural" meaning and the "lexical" meaning. The two sentences "The chair is on the rug." and "The rug is on the chair." have different meanings because the words appear at different positions in the structure. On the other hand, the difference in meaning between "The chair is on the rug." and "The chair is on the floor." is a lexical difference. Lexical meaning tells what sentences talk about; structural meaning tells what they say about it.

Just as different languages are different lexically, they can also be different structurally. Languages that are historically related are usually similar in structure. Even if two languages have certain structures in common, it may be that they assign different meanings to these structures. Languages may also differ in what aspects of the meaning are carried structurally and what aspects are carried lexically. The distinction between subject and object in a sentence is indicated structurally by word order in English, but in German it may be indicated by the lexical difference between *der* and *den*. Similarly the meaning that is assigned lexically to the word *if* in English, is sometimes indicated structurally by word order in German.

TRANSLATION OF STRUCTURE

If translation by machine is to be done on a sentence-by-sentence basis, grammars are needed for each language. The use of sentence structure to resolve word order and multiple-meaning problems and to translate the structural meanings correctly must be based on adequate grammars. These grammars should consist of descriptions of each language which are complete enough to make the distinctions in meaning necessary for translation. A description of this sort for even one language has not yet been provided, because the concepts involved are just now beginning to be understood and because languages are extremely complex in structure. Once two of these descriptions are available, they can be set side by side, and the equivalences stated. It must be decided which sentences in language *A* and language *B* are equivalent. The underlying assumption is that it is possible to find such correspondences. Of course, this assumption is not strictly valid, for no two languages are completely translatable. It is probably very nearly valid for related languages such as German, English, French, and Russian, and for scientific discourse, or other types of careful exposition.

If a description of the structure of two languages and a statement of the equivalences in meaning between them are given, only one more thing is needed for a mechanized translation procedure: a mechanical method of recognizing the structure of a sentence from the sequence of words composing the sentence. The precise form of this recognition routine must await the description of the structures to be recognized. However, the process must operate from the point of view of the person who reads or hears a sentence. It must tentatively recognize small groups of words, place them in a tentative structure, and so proceed, each time eliminating certain possibilities. Once the structure of the sentence to be translated has been recognized, it can be translated into an equivalent structure in the other language, and the correct meanings can be chosen from the multiple meanings of the words.

Results Expected

The translations produced by a routine that translates on a sentence-for-sentence basis will be vastly better than the output of a word-for-word translating machine. Most output sentences will be grammatically correct. There will be very few word-order problems remaining. Many of the multiple-meaning difficulties will be solved, particularly those involving the most frequent words of the language and the very frequent inflectional endings. There will be a partial solution to multiple-meaning problems such as the different meanings of *can* which are distinguished on the basis that one meaning is connected with a verb, one with an auxiliary verb, and one with a noun. Some multiple-meaning problems will remain. Of these, some can be resolved by judicious use of a field glossary or by related techniques. The rest must be resolved by the good sense and intelligence of the reader of the translation.

Conclusions

The state of the art of mechanical translation may be summed up in the following manner:

1. Word-for-word translations can be made now on high-speed, general-purpose, digital computers.
2. Word-for-word translations could be made more economically by means of special-purpose machines built with existing technology.
3. Word-for-word translations promise to be considerably cheaper than man-made translations.
4. Word-for-word translations are very crude, but may be useful when more accurate translations are not worth the additional cost.
5. If something better than a word-for-word translation is desired, the best procedure would be to take into consideration the sentence structure in designing a translation routine.
6. Providing translation routines on a sentence-for-sentence or structural basis requires a considerable amount of detailed linguistic work.

References

1. Machine Translation of Languages (book), W. N. Locke, A. D. Booth, editors. Technology Press, Massachusetts Institute of Technology. Cambridge, Mass.; John Wiley & Sons, Inc., New York, N. Y.; Chapman and Hall, Ltd., London, England, 1955.
2. Mechanical Translation, Massachusetts Institute of Technology, Cambridge, Mass , vols. 1, 2. 1954 and 1955.
3. Human Behavior and the Principle of Least Effort (book), G. K. Zipf. Addison-Wesley Press, Cambridge, Mass., 1949.
4. Simple Games of Strategy Occurring in Communication Through Natural Languages, Benoit Mandelbrot. *Transactions,* Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., Mar. 1954, pp. 124-37.