

HUMAN v. MACHINE

TRANSLATION OF FOREIGN LANGUAGES

by

J. M. LUFKIN

This paper, which was presented at the 1965 International Convention of the Institute of Electrical and Electronics Engineers in New York, was first published in the June 1965 issue of that society's "Transactions on engineering writing and speech".

It is being reprinted by permission of the publication's editor as well as that of the author, who is an employee of Honeywell Inc., Minneapolis, Minn., an associate member of IEEE and a member of the American Translators Association.

What translation requires

THE translation of any kind of serious subject matter, whether it is done slowly and deliberately in an office, or simultaneously at a meeting or conference, demands an extensive knowledge of both the source language and the target language and a considerable acquaintance with the subject. These elementary truths need to be repeated from time to time because they are frequently forgotten. The whole world needs translators who have these qualifications, and it now needs them more urgently and in greater numbers than it has ever needed them before.

It is really astonishing that under these circumstances the most naïve notions about translation and, for that matter, about language in general, still prevail even among otherwise well-educated people. How often have we all heard that so-and-so speaks "five" or "six" or "seven" languages *fluently*! (That magic number *seven*; how often it turns up!) This story is almost invariably false. I say *almost* invariably false, because there really are a few people who can speak several languages fluently. But the person who tells the story is usually himself quite ignorant of any language other than his own, and the fellow who is said to perform this miracle usually turns out to be unable to converse, in most of his languages, at any level much above that of a small child. Winston Churchill is said to have replied, when somebody urged upon him, during the Battle of Britain, the services of a man who could speak "nine" languages fluently, "My! What a splendid head waiter he'd be!"* And a dozen years ago I had the unpleasant

task of checking up on the qualifications of three of our counter-intelligence agents, whose personnel records said they were "fluent" in the language of a certain country and who were turning in reports based on interviews they had presumably conducted in that language. None of them, it turned out, could answer a single one of a series of questions like, "Do you see a cat?" which were put to them orally. Since not one native American in a hundred can carry on an adult conversation in any foreign language, all you need, to pass for "fluent" is a little gibberish.

Standards for translators

Although surveys are published from time to time pretending to show how many scientists, for example, are fluent in a given foreign language, I have never seen such a report that I could believe because they were all based on the individuals' own opinions rather than upon an examination. But although the term *fluency* has been used so loosely that it no longer has a generally agreed upon meaning, another term has been quite well defined, and that is a "native knowledge" of a language. Our Civil Service standards for translators and interpreters define a "native knowledge" of a language as "the ability to speak or write a language so fluently that the expression of thought is structurally, grammatically, and idiomatically correct, and expresses the range of vocabulary in the language common to a person who has received his education through the high school level in a country of the language or its equivalent¹." That is an excellent, practical definition and I commend it to all who have to hire or make use of the services of translators or interpreters.

It is generally agreed that if the translator knows one language better than he knows the other—and he almost always does—he should be translating *from* the lesser known language *into* the better known one. This is not in order that his smoothness may cover up his mistakes. It is because it is easier to *grasp* a difficult concept than it is to *express* it in words; it is easier to *see* a thing than to *describe* it. Here again, our Civil Service standards are helpful². They insist upon this distinction, although it is frequently ignored elsewhere with unfortunate results.

* Mr. André Hadamard of the United Nations interpreting staff tells me that Churchill was paraphrasing a remark of Talleyrand's.

It is also generally agreed, among people who are experienced in technical translation, that a general acquaintance with the subject can be as essential as a knowledge of the language. A professor of French literature may be fully bilingual and he may have a huge vocabulary, but if he knows little or nothing about electronics he will not be of much help at a meeting between American and French engineers to discuss a new circuit theory. The fact that people are frequently asked to translate discussions they cannot understand is a result of the all-too-generally held notion that the translator deals with *words* rather than *ideas*.

Words are not ideas

But, "words are to serve and follow meaning, and not the meaning the words," as one of the greatest of translators³ expressed it 450 years ago. This has become a cliché among linguists, but it is not well enough known among others. As the Russian research linguist, O. S. Akhmanova has said, ". . . The most important form of translation in practice . . . presupposes . . . the replacement of whole statements in one language by equivalent statements in the other, i.e. equivalence in difference. This latter is the basic problem of language and the main subject of linguistics."⁴

The need for translations

The need for translations is in part suggested by the fact that Government agencies are now buying translations at the rate of more than three million dollars a year⁵. Furthermore, the National Defense Education Act has awarded more than three thousand scholarships for the study of foreign languages in the past five years⁶. The funding for this has been about three and one-half million dollars. As our military, economic, and cultural commitments throughout the rest of the world increase, so, surely, will the urgency of our need for people who can handle foreign languages. In fact, I expect this urgency to rise so steeply that within a very few years we will have what is unthinkable today: a requirement for a reading knowledge of a foreign language upon graduation from high school. To anyone who finds this suggestion preposterous, I can only answer that the absence of such a requirement is a vestige of a kind of isolation which we really have not enjoyed since the advent of the airplane.

Only a little more than one half of the world's scientific and technical literature is now published in English, and as the output of scientific and technical literature increases, the proportion of it which is in English decreases. And although English is, to some extent, becoming a sort of standard international language for professional-level communication, there is little likelihood that it will ever become anything like the standard language that Latin once was. The generally received notion that "everybody" in Europe can speak English is an illusion. The professional man who reports this as a fact has not stopped to think out the truth of the matter; namely, that he has been

introduced, in all those conferences and laboratories, *only* to people who could understand English. I am by no means alone in this observation⁷.

A graduate student with the proper motivation can learn to *read* serious essays in a foreign language without prohibitive effort in a course consisting of ten two-hour class sessions and about two hundred hours of study. It might be assumed that anyone who invested that much effort in acquiring a skill would preserve it by reading a little every day so that he would always have it, but it is well known that this is not done. Nearly everybody in this country who has a Ph.D. is supposed to be able to read two foreign languages. I doubt if one in twenty can read a newspaper in either of those languages a year after he has finished his doctorate. Incidentally, the ability to read a newspaper is a fair test of a man's ability to deal with a language. It is frequently said, of an engineer or scientist, that he knows a certain language "well enough to understand technical articles in his field of specialization but not well enough for more general subject matter," or words to that effect. There may possibly be some truth to this if the articles contain practically nothing but symbolic expressions such as equations or formulac, but I am afraid that the whole idea is simply false. A man who cannot read a newspaper is on very shaky ground indeed if he pretends he can follow the complex statements that are common to practically all serious technical writing at the professional level. Furthermore, the ability to identify the subject matter of an article, or even to get the gist of it, is by no means the equivalent of an ability to understand its details. Finally, no matter how many pages of equations a technical discussion may contain, the expression of its essential ideas usually requires plain, everyday language.

The automatic translation programme

The first serious proposals for mechanical translation were apparently offered a little more than thirty years ago, independently, by a Russian, Smirnof-Troyanski, and a Frenchman, Artsouni. The idea was first taken up seriously both in this country and in Russia shortly after the war when electronic computers became available.

The cost of Government-sponsored research on automatic translation has now passed the ten-million-dollar mark and this does not include work on hardware, but only research in linguistics and machine programming.

A fifty-page survey of the work being done on machine translation was published in the National Science Foundation's *Current Research and Development in Scientific Documentation* No. 13 in November, 1964. This survey consists of short reports turned in by the project leaders themselves and therefore offers no comparative review or evaluation. However, this much is apparent: there are now at least fifty different active projects in machine translation, more than half of them being conducted outside of the United States. Nearly all of the fourteen new projects reported in answer to the National Science Foundation's 1964 inquiry are foreign. In America, several of these

projects have been abandoned in recent years, and a number of others have "backed up" into basic research on the structure of language.

In addition to at least thirty different centres for machine translation in Europe and Japan, the National Science Foundation report lists work being done at the Bunker-Ramo Corporation, Canoga Park, Calif., the University of California at Berkeley, the Computer Concepts Corporation, Los Angeles, Calif., Ohio State University, Columbia, Georgia Institute of Technology, Atlanta, Harvard University, Cambridge, Mass., the IBM Corporation, Yorktown Heights, N.Y., the Lockheed Missiles and Space Company, Palo Alto, Calif., M.I.T., Cambridge, Mass., the Summer Institute of Linguistics, Santa Ana, Calif., the University of Texas, Austin, Tufts University, Medford, Mass., and Wayne State University, Detroit, Mich. However, the centre at Georgetown University, Washington, D.C., where some impressive pioneer work has been done, is no longer active, and the projects at Washington State University, Seattle, and at the National Bureau of Standards have apparently been terminated.

At the request of a group of Government agencies, including the Department of Defense and the National Science Foundation, the National Academy of Sciences has created an Advisory Committee on Automatic Language Processing to advise the Government on mechanical translation and related subjects. This committee's executive secretary, who has his office in the Center for Applied Linguistics, is in Washington. The Center for Applied Linguistics is not a Government agency but a subsidiary organization of the Modern Language Association of America.

All of this effort, both in this country and abroad, is presumably based on the assumption that useful machine translations will be feasible sooner or later and that they will be cheaper than human translation or will cost no more and will have other advantages.

The need for an adequate grammar

The first serious attempts to get a computer to translate a foreign language look to us now, with the wisdom of ten years of hindsight, very foolish indeed. They seem to have been based upon the incredibly naïve assumption that all you need for translation is a dictionary and some simple transfer mechanism. There were, of course, trained professional linguists in this work; and they were not stupid. They knew that word-for-word translations would not do, but the early exhibits that were published of what the machines could do in the way of translations did make them look a little foolish. Examples of machine translation of serious subject matter, by the way, are exhibited much less often than they used to be. One reason is that most of the machines are not being used to "translate", for that crude early work resulted in an important discovery: We did not understand *any* language well enough to programme a machine to translate it. So our linguists have gone back to the drawing board, not to design new programmes

for new machines but to chart the structure of language as it had never been charted before. Needless to say, there was a good deal of agonizing reappraisal. Linguistic scientists had known, for thirty or forty years at least, that the traditional grammars—for example, the artificial Latin one forced upon English by 18th century grammarians—were hopelessly inadequate as a description of language. I do not know how many linguistic scientists suspected the existence of the enormous gulf between our best descriptive grammar and the description of language which we now know we must develop if we are ever to have machine translation. But a good many linguists here and in other countries set to work on this enormous problem. Imagine a little bird who flies once a year to the top of the highest peak in the Andes and sharpens his beak upon it. When he has ground that mountain down to sea level, so the pessimists tell us, these linguistic scientists will have succeeded in analysing language⁸. My own feeling is that this analogy is unfair because there is more than one bird; there is really quite a flock of them and they are all working very hard. But let us leave our beak sharpeners for a moment and turn to another group.

Cerebral logic

Another group said in effect, "Look, we do not *need* to understand the structure of any language in all of its details; we have only to imitate what goes on inside the brain of a competent translator."

Now there were several reasons to hope for success from this approach. For example, nerve impulses are known to be digital in a way. To be sure, they are chemical and mechanical as well as electrical, and these three aspects do merge at times, but they are basically digital, and like those in a computer they are reversible. Furthermore, these impulses can be measured, and they turn out to be quite slow compared to those in a present-day digital computer, for the nerve impulses seem to average about one millisecond. Besides, computers had been trained to do a number of operations which undeniably imitate one aspect or another of human thinking. Quite aside from mathematics, there is the spectacular example of computers programmed to play chess. No computer has yet been trained to defeat a Grand Master of chess, but some of them have turned in creditable performances. To do this they have to analyse fairly complex situations, and they have to examine long trees or chains of branching alternatives and make decisions after conducting searches⁹.

But the analogy between playing chess and translating a language turns out to have some serious limitations. First of all, the rules of the game of chess are so simple and so few in number that they can easily be stored entire in the memory of an electronic computer. Furthermore, the chess-playing computer, although it does look ahead (down various branches containing alternatives), does not look behind. In other words, it makes all of its decisions without regard for what has happened previously. The translator does almost exactly the opposite.

projects have been abandoned in recent years, and a number of others have "backed up" into basic research on the structure of language.

In addition to at least thirty different centres for machine translation in Europe and Japan, the National Science Foundation report lists work being done at the Bunker-Ramo Corporation, Canoga Park, Calif., the University of California at Berkeley, the Computer Concepts Corporation, Los Angeles, Calif., Ohio State University, Columbia, Georgia Institute of Technology, Atlanta, Harvard University, Cambridge, Mass., the IBM Corporation, Yorktown Heights, N.Y., the Lockheed Missiles and Space Company, Palo Alto, Calif., M.I.T., Cambridge, Mass., the Summer Institute of Linguistics, Santa Ana, Calif., the University of Texas, Austin, Tufts University, Medford, Mass., and Wayne State University, Detroit, Mich. However, the centre at Georgetown University, Washington, D.C., where some impressive pioneer work has been done, is no longer active, and the projects at Washington State University, Seattle, and at the National Bureau of Standards have apparently been terminated.

At the request of a group of Government agencies, including the Department of Defense and the National Science Foundation, the National Academy of Sciences has created an Advisory Committee on Automatic Language Processing to advise the Government on mechanical translation and related subjects. This committee's executive secretary, who has his office in the Center for Applied Linguistics, is in Washington. The Center for Applied Linguistics is not a Government agency but a subsidiary organization of the Modern Language Association of America.

All of this effort, both in this country and abroad, is presumably based on the assumption that useful machine translations will be feasible sooner or later and that they will be cheaper than human translation or will cost no more and will have other advantages.

The need for an adequate grammar

The first serious attempts to get a computer to translate a foreign language look to us now, with the wisdom of ten years of hindsight, very foolish indeed. They seem to have been based upon the incredibly naïve assumption that all you need for translation is a dictionary and some simple transfer mechanism. There were, of course, trained professional linguists in this work; and they were not stupid. They knew that word-for-word translations would not do, but the early exhibits that were published of what the machines could do in the way of translations did make them look a little foolish. Examples of machine translation of serious subject matter, by the way, are exhibited much less often than they used to be. One reason is that most of the machines are not being used to "translate", for that crude early work resulted in an important discovery: We did not understand *any* language well enough to programme a machine to translate it. So our linguists have gone back to the drawing board, not to design new programmes

for new machines but to chart the structure of language as it had never been charted before. Needless to say, there was a good deal of agonizing reappraisal. Linguistic scientists had known, for thirty or forty years at least, that the traditional grammars—for example, the artificial Latin one forced upon English by 18th century grammarians—were hopelessly inadequate as a description of language. I do not know how many linguistic scientists suspected the existence of the enormous gulf between our best descriptive grammar and the description of language which we now know we must develop if we are ever to have machine translation. But a good many linguists here and in other countries set to work on this enormous problem. Imagine a little bird who flies once a year to the top of the highest peak in the Andes and sharpens his beak upon it. When he has ground that mountain down to sea level, so the pessimists tell us, these linguistic scientists will have succeeded in analysing language⁸. My own feeling is that this analogy is unfair because there is more than one bird; there is really quite a flock of them and they are all working very hard. But let us leave our beak sharpeners for a moment and turn to another group.

Cerebral logic

Another group said in effect, "Look, we do not *need* to understand the structure of any language in all of its details; we have only to imitate what goes on inside the brain of a competent translator."

Now there were several reasons to hope for success from this approach. For example, nerve impulses are known to be digital in a way. To be sure, they are chemical and mechanical as well as electrical, and these three aspects do merge at times, but they are basically digital, and like those in a computer they are reversible. Furthermore, these impulses can be measured, and they turn out to be quite slow compared to those in a present-day digital computer, for the nerve impulses seem to average about one millisecond. Besides, computers had been trained to do a number of operations which undeniably imitate one aspect or another of human thinking. Quite aside from mathematics, there is the spectacular example of computers programmed to play chess. No computer has yet been trained to defeat a Grand Master of chess, but some of them have turned in creditable performances. To do this they have to analyse fairly complex situations, and they have to examine long trees or chains of branching alternatives and make decisions after conducting searches⁹.

But the analogy between playing chess and translating a language turns out to have some serious limitations. First of all, the rules of the game of chess are so simple and so few in number that they can easily be stored entire in the memory of an electronic computer. Furthermore, the chess-playing computer, although it does look ahead (down various branches containing alternatives), does not look behind. In other words, it makes all of its decisions without regard for what has happened previously. The translator does almost exactly the opposite.

Simultaneous translation

The simultaneous translator or interpreter, to take the most convenient example, develops at the very outset of a given piece of work a context-frame of reference which governs, either very strictly or very loosely depending upon how technical the subject matter is, his treatment of nearly every word or phrase. As he listens to the sentences coming from the speaker, he analyses each phrase for its meaning and recasts that meaning in the words, idioms, phrases, and sentences of the target language. He performs vocabulary lookups at a rate which has never been measured but which I suspect may run as high as several dozen of alternatives examined and rejected in one second. He considers the effect, upon the meaning of each word in the speaker's discourse, of all of the words which have gone before it, and of the few which come after it, in the lag of a few seconds which separates his words from those of the original speaker. Let us examine, as an example, a typical "subroutine" performed in the brain of a simultaneous translator.

Suppose the speaker is speaking in English and that the translator is translating him into German. Suppose that the speaker uses the word *take* in a sentence. This word can sometimes be translated by the German *nehmen*, but more often it cannot, and the German must use, instead, *annehmen*, *aufnehmen*, *benennen*, *einnehmen*, *entnehmen*, *festnehmen*, *hinnehmen*, *mitnehmen*, or *übernehmen*; or else, *bringen*, or *mitbringen*; or else, *führen*, *tragen*, *befördern*, *auffassen*, *vertragen*, *finden*, *fassen*, *erfordern*, *dauern*, or *brauchen*¹⁰. This example is quite typical. Suppose, for example, that the English speaker uses in a technical description the word *stock*. That word in English can be made quite exactly to stand for any one of thirty very different things or ideas. No other language uses one word for the same set of things and ideas, and it is quite likely that in one language or another there really *are* thirty different words for what we call *stock*. Or, to take an example from another language, the French word *porter* must be translated into English by one of thirty or forty different English words, the choice depending entirely upon the context in which the French word is used.

The translator's choice of terms in situations like these depends upon an analysis of the *context* in which the word is used by the speaker. A competent translator has most of these alternate terms in his head, and will examine some of them consciously, and others subconsciously. In fact, it is clear to me that most of the mental processes of translation—and especially those done at the speeds necessary for simultaneous translation—are subconsciously. This is one reason why we know so little about the *logic* of translation. If translation were a mere lexicographic operation, if it were merely the systematic substitution of one word for another, it would all be very simple indeed. But it is nothing of the kind; it is almost unimaginably more complex than that and *we do not* understand it. I suspect for example, that most of the logical transactions which take place in the translator's brain are conceptual rather than

verbal. There is apparently very little in print on this point, but the impression I have gained from experience has been confirmed in conversations with several members of the United Nations interpreting staff.

The memory problem

To continue our analogy between brain function and computer operation, the memory problem can be said to have two aspects: size or capacity, and access time. The mathematician Von Neumann estimated the capacity of the human memory to be about 2.8 times 10^{20} binary digits of information¹¹. For readers who do not live on familiar terms with numbers of that magnitude I can only say that 2.8 times 10^{20} centimetres is the distance light travels in 280 years. However, more recent theorists have suggested that much of the information in the brain is received and presumably stored not by discrete-point or digital processes but by continuous-line or analogue mechanisms which will ultimately prove even more difficult to measure and describe. In any case, and whatever the actual capacity of the human memory may be, it is certainly enough to enable a translator or interpreter who has access to neither dictionaries nor grammars nor other reference books to do perfectly creditable translations of extremely wide-ranging technical discussions, on his feet, at the rate of 150 or even 200 words a minute. In other words, the translator carries in his brain a very large part of both the vocabularies and the grammars of both languages, a great deal of information about the subject itself, and finally, if he is to do a smooth, exact, and complete job, a great deal of information about the world in general outside of the subject matter and the two languages involved. How all of this information is stored there, and by what fantastically subtle and complex associative logic the operating centre of the brain gets access to it, *we do not know*. I think we need to know for a great many reasons and not merely so that we can design machine translation systems. I do not claim that merely because most of it is unconscious we do not understand the process of storage and retention at all, but only that we do not yet know enough about the cerebral logic involved in *any* intellectual process to construct a convincing model of it. This subject is of very great importance. Von Neumann, for example, said "I suspect that a deeper mathematical study of the nervous system . . . may alter the way in which we look on mathematics and logic proper."¹² And so we have a number of groups working on this problem, another flock of little birds at the top of another mountain in the Andes diligently sharpening their beaks.

Learning machines

But there is still another approach. A group of physiological psychologists now tells us "Human thinking can be explained in information-processing terms without waiting for a theory of the underlying neurological mechanisms."¹³ This is a comparatively recent development. It is a form of the celebrated Black Box Theory: you don't

know what's in the box; you can lift it and shake it and rattle it a little, but you can't open it and study its insides. But you notice that it has input apparatus and output apparatus and so you monkey with the controls until the input-output relationship is about what you want. These people, or others closely related to them, are developing what are called *learning machines*. These are devices which can be made to learn from their own experiences. They are radically different from digital computers, which have information inserted in them, although digital computers may include devices which have learning components and a learning machine may include a computer. But the essential feature of the learning machine is that it makes use of experience to modify its own internal structure. A classical adaptive mechanism such as an adaptive autopilot is not really a learning machine. It does modify its response, that is, its output, so that that response bears a specified relation to the input or experience it receives, but for the moment or for the instant only. It does not acquire a "set" of permanent characteristics. It does not, in other words, acquire a new behaviour pattern as the result of past experiences. Now a learning machine does, and it may be possible that we shall some day have a learning machine to which we could feed a complete textbook of physics, for example, in Russian, together with a complete translation of that textbook into English. The machine's experience would consist in comparing the Russian with the English, and the result would be an enormous set of very complex equivalents stored in its memory. The hope is roughly this: If we could give such a machine a big enough capacity, and a big enough self-adaptive mechanism, we could then feed it a hundred such textbooks, together with translations of all of them. Then, after some such experience the machine could take in a Russian text alone, and produce its own translation. This whole idea is fantastic, but not quite as fantastic as it may seem to those who are not aware of what has been done recently in the field of artificial intelligence. Some machines have now been built which are actually capable of generalizing from experience¹⁴. So there is another mountain in the Andes with another flock of birds working on this problem.

"Synonyms" do not exist

As every translator and every editor knows, one of the great obstacles to the public understanding of the complexity of language is the generally accepted notion that synonyms are two or more words with the same meaning. Synonyms do not exist in this sense between any two languages, and not, as far as I know, within any one language. The reason for this is that *pure meaning* does not exist, partly because a word has meaning only within a phrase or a statement. This has become a cliché among linguists, but it is not well enough known among others. The hopelessly crude word-for-word translation theory was based upon a faith in synonyms of this kind. It is sometimes argued that word-for-word translation is more

useful for certain pairs of languages than for others, but I suspect that these pairs are probably, like Swedish and Norwegian, so similar as to be not worth the expense of translating. In any case, to a linguistic scientist the picture looks something like this: The words for the colours of the rainbow in two different languages appear at first to be more or less interchangeable, but upon closer examination it turns out that the word in one language and the corresponding word in the other really have two very different meanings. They do not represent the same colour; they represent two different overlapping bands of the spectrum. This is not a minor point; it is tremendously important—not because colours are important but because the same sort of spectral shift is common to many, many thousands of pairs of words in any two languages. Let us take a familiar example from French and German. The German word *du* and the French word *tu* are sometimes translated by the archaic English word *thou*, because there is no modern English equivalent of *du* or *tu*. But the matter is not that simple. The German *du* and the French *tu* are by no means interchangeable. The French expression is much more intimate than the German one, as anyone who speaks both languages and has lived in those countries can testify. These examples represent a whole class of semantic difficulties which beset the machine translators.

Word-for-word translation

The word-for-word method, even with its most ingenious refinements, results in a more-or-less useless translation. Professor Yngve of M.I.T. says that this method can be made to solve "80 per cent" of the problems, but that "the remaining 20 per cent . . . make all the difference between an acceptable and an unacceptable translation." This has not been said often enough, and the outsider looking into this field for the first time must be warned that most of the exhibits of machine translation that he will see in the literature on this subject have been either pre-edited or post-edited by human beings. The best example of machine translation that I have seen appeared last year in a publication of the Select Committee on Government Research of the United States House of Representatives¹⁵. Here is a paragraph from that exhibit, described as a "raw" translation from the Russian. The term "raw" means at least that the text has had no post-editing. I assume that the Russian text had no pre-editing before it went into the machine.

XXII congress of Communist party of Soviet Union—congress of builders of communist society—was most important stage in movement/motion of our country to Communism. Congress lead! up/disappointed results to grandiose achievements of Soviet people, great revolutionaries conversions, carried out in our country posse XX congress of CPSU under leadership of Communist party and her/its Lenin Central Committee headed by N. S. Khrushchev. Congress took received new, third Programme of our party, correctly called Communist manifesto of contemporary epoch, giving concrete, scientifically proved plan of construction of communist society in our country.

Why an original with so much hot air in it was chosen for this exhibit I do not know. However, the result brings

to mind Dr. Samuel Johnson's famous observation about women in the pulpit: "Sir, a woman's preaching is like a dog walking on his hinder legs. It is not done well; but you are surprised to find it done at all."

The analysis of syntax and context

The problems we have just considered are solved by the human translator with the help of the context or environment in which the terms appear. Again, our understanding of how the human translator does this is not complete enough to enable us to programme machines to do it or even to programme machines to teach themselves to do it. Much the same can be said of the problem of the effect of syntax upon meaning. By syntax I mean phrase and word arrangement within sentences together with inflexional endings and their effect on the meanings of specific words. In a highly inflected language like Latin, the inflexional or case endings of the words tell a great deal about their meaning, although English, for example, has lost almost all of its inflexion. But even in Latin, this is not the only key to meaning; there are many other syntactic devices which affect it. Once again, these have never been codified for any language. In fact, the two great problems of computational linguistics are the analysis and symbolic representation of the effects of syntax and of context upon meaning¹⁶.

Even some of the most sophisticated attempts to make exact descriptions of the effect of syntax upon meaning have resulted in disappointment. For example, one American researcher said "it was felt originally that some two dozen rules might suffice to describe the noun phrase in either language. German was expected to require more than English, due to the greater variety of possible inflexional combinations. In actuality, about three hundred rules interpreting the noun phrase in German have been accumulated."¹⁷ The Russians have also had their problems. I. A. Mel'chuk describes how two linguists derived a set of rules for determining the syntactic function of a certain class of Russian adjectives from a study of 700 examples, only to be refuted by another linguist who was able to show contradictions for nearly all of the carefully derived rules¹⁸. Mel'chuk goes on to point out that general rules of this kind must be based upon the analysis of whole sentences—a task far greater than any that has been undertaken. "No grammar," one American researcher has reported, "which is even remotely complete has ever been compiled for any language, not even for Latin or Sanskrit, which has been under investigation for more than two millennia. It is probable that a reasonably complete grammar would be so complex that its compilation could not be feasible without the aid of a data processing system designed for that purpose."¹⁹ This is an understatement. A number of computers which were expected to be used for machine translation have been turned aside to the problem of investigating the structure of language itself. This, in fact, is what is usually meant today by the term *computational linguistics*.

Some hardware considerations

It is clear that the major problems are ones of linguistic analysis, not of computer programming or even of computer design. However, the problem of ready-access storage for both vocabulary and for analytical instructions is a very real one. Even so, it will surely be solved long before the linguistic problems are solved. For example, "associative" or content-addressable memories are now being developed which are said to promise as much as a thousand-fold increase in the speed of information retrieval. Such memories will no doubt be more urgently needed at first for purposes other than language translation, but they are definitely on the horizon and they offer a promising solution to the problem posed by external storage which, although its capacity may be fully adequate, takes altogether too much time for the computer to consult because it must be read all the way through for every search. Internal, quick-access storage equipment has not, during the past decade, been big enough for the requirements of language translation. But here also the solution seems to lie well within the foreseeable future. At least one computer manufacturer now claims to have produced a core memory that will store eight million characters any one of which is accessible in eight microseconds²⁰.

The machine amuck

A machine equipped with a fairly complete vocabulary but without the sorting mechanism that enables the human translator to eliminate the implausible meanings, would generate huge strings of alternate meanings. For example, an English sentence with two words like *open*, and *stock*, which can have about forty different meanings each, could, in machine translation, result in 1,600 different sentences. Most of these, to be sure, would be meaningless. But a considerable number of them would make sense, with the *wrong ideas*. In fact, since principles ought to take account of the worst possible examples, it is quite possible for a sentence in English to contain twenty words of which ten have multiple meanings, not perhaps as multifarious as those of *open* and *stock*, but with, say, ten variant meanings each. A translating routine that was not closely governed by syntactic and contextual controls could generate *ten milliard* sentences. Translators need have no fear of losing their jobs to machines under these conditions. This arrangement could provide long job-life-security for an army of post-editors (the term *post-editor* is a euphemism used in machine translation circles to mean *human translator*). If, on the other hand, that expenditure of manpower should be considered extravagant, the machine could be instructed to generate only one of the possible alternatives. However, in the case just cited, the odds against that single sentence being the right one would be about 9,999,999,999 to 1, a ratio which will not be readily accepted by people who have been caught up in the preoccupation with *reliability* which has lately become so fashionable.

The outlook for the future

What is the outlook for the future? The first few years of research into machine translation were more or less dominated by the optimists. You have to have optimists. As far as I know, practically nothing worthwhile gets done without them. But in 1962, Yehoshua Bar-Hillel, a giant in this field who had devoted eleven years of his life to it, said that machine translation of natural languages had "ended in failure."²¹ It is instructive to compare this remark with a much earlier statement by the late Norbert Wiener who said to Warren Weaver in 1946, "I am afraid that the boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any mechanical translation scheme very hopeful."²² On the other hand, the optimists are fond of a five-year period. In answer to the question, "When shall we see a machine translate?" Professor Locke said in the 'Scientific American' in January, 1956, "My best guess is within five years. By that time there should be in operation one or more models turning out a good deal better than a word-by-word translation."²³ In 1960, Franz Alt of the National Bureau of Standards said that he thought that it would be "at least five years before a really satisfactory system is reached."²⁴

We used to make fun of the Russians for their repeated reinstatements of their "five-year" plans. But we underestimated what courage, and talent, and a tremendous lot of hard work could do. Let us not underestimate our linguistic scientists. The road ahead of them is going to be very hard and very long. But if they persist, they will soon know more about language and about the nature of information and its transmission than their predecessors have learned in several hundred years—whether they succeed in machine translation or not. And, finally, no one need fear that machines will replace translators any more than that computers will replace mathematicians; the demand for good machines and qualified people in both fields will go on increasing throughout the foreseeable future.

BIBLIOGRAPHY

Although there are many books on translation in general, the best survey of the problems of scientific and technical translation is "Scientific and technical translating and other aspects of the language problem", a Unesco publication in the Documentation and Terminology of Science series, Paris, 1957, with a second edition in 1958. This planet's most impressive journal for translators and interpreters is "The Incorporated Linguist". Another periodical devoted to translation in general is "BABEL", the official organ of the International Federation of Translators, published by Langenscheidt in Berlin with the assistance of Unesco. The American Translators Association (ATA) publishes "ATA Notes", devoted to subjects of general interest to translators, and distributes "The Translation Inquirer—A Monthly Service Publication for Professional Translators and Translation-Minded Laymen". "The Translation Inquirer" is published by the Interlingua Division of Science Service with the co-operation of the New York Chapter of the American Medical Writers' Association and the ATA. The ATA's mail address is P.O. Box 489, Madison Square Station, New York, N.Y. 10010. The ATA seeks to regularize the translating profession and to attract the best talents to it. The American Association of Language Specialists, an association of professionally qualified conference interpreters, is presided over by Vladimir Pojidaeff, of 550 Fifth Avenue, New York. On the profession

of translating in the United States, there is a comprehensive survey by Henry Fischbach in "BABEL", 8 (3), 119-123, 1961. On simultaneous interpreting (or, as it is sometimes called, simultaneous translation) as a profession, there is an excellent article by Robert Ostermann in "The National Observer" for April 5, 1965, p. 20. The classic general introduction to the science of linguistics is Leonard Bloomfield's "Language", New York, Holt, Reinhard and Winston, 1933 and 1963. An excellent alternative is H. A. Gleason, Jr., "An Introduction to Descriptive Linguistics" by the same publisher, 1961. There is no survey of the state of the art of machine translation. The best general survey of this subject intended for the reader who is neither an engineer nor a linguistics scientist is Emile Delavenay, "La machine à traduire," published as vol. 834 in the "Que-sais-je" series, Paris, Presses Universitaires de France, 2nd edn., 1963. Reports issued by the various groups working in this field are more or less unreadable to the outsider because of technical jargon. The leading periodical in the field of machine translation is "Mechanical Translation", a quarterly issued by the American Association for Machine Translation and Computational Linguistics at the Massachusetts Institute of Technology. This association also issues a newsletter called "The Finite String".

REFERENCES

- ¹"Position Classification Standards", U.S. Civil Service Commission. There are two of these, one for translators, Rept GS-031-0, and one for interpreters, Rept GS-033-0. Both were issued in Dec., 1957, and slightly amended in June 1958.
- ²"Position Classification Standards" *ibid.*
- ³Luther, M., Defence of the translation of the psalms.
- ⁴Akhmanova, O. S., I. A. Mel'chuk, R. M. Frumkina, and E. V. Paducheva, "Exact Methods in Linguistic Research", translated from the Russian by D. G. Hays and D. V. Mohr, Rept R-397-PR, Rand Corp., Santa Monica, Calif., 1963, p. 20.
- ⁵"Study Number IV, Documentation and Dissemination of Research and Development Results", Report of the Select Committee on Government Research of the House of Representatives, Rept 39-993, Washington, D. C.: U. S. Government Printing Office, 1964, p. 69.
- ⁶"Study Number V, Federal Student Assistance in Higher Education", Report of the Select Committee on Government Research of the House of Representatives. Washington, D.C., U. S. Government Printing Office, 1964, p. 35.
- ⁷Curtis, R. W., "International conferences", in *Internat'l Science and Technology*, Nov., 1962, pp. 38-41.
- ⁸This analogy of the beak-sharpening bird comes from something I read as a child; it may have been by Hendrick Van Loon. (Since writing this article, I have been told by Mr. N. G. Neuweiler of "La Revue Polytechnique", Geneva, that the originator of this image was Buddha.)
- ⁹There have been many descriptions of automatic chess playing, but an excellent and convenient one is in Simon, H. A., and A. Newell, Information processing in computer and man, "Am. Scientist", 52, 281-300, Sep., 1964.
- ¹⁰For a discussion of the differences between these terms, see Farrell, R. B., "Dictionary of German Synonyms." Cambridge, Cambridge University Press, 1961, pp. 351-357.
- ¹¹Von Neumann, J., "The Computer and the Brain". New Haven, Conn., Yale University Press, 1958, p. 64.
- ¹²Von Neumann, *ibid.*, p. 2.
- ¹³Simon and Newell⁹, p. 282.
- ¹⁴Bloch, H. D., The perceptron: A model for brain functioning. I. *Rev. Mod. Phys.*, 34, 123-135, Jan., 1962, and Bloch, H. D., B. W. Knight, Jr., and F. Rosenblatt, Analysis of a four-layer, series coupled perceptron, II. *ibid.*, pp. 135-142. Also, Rosenblatt, F., "Principles of neurodynamics: perceptrons and the theory of brain mechanisms". Washington, D.C., Spartan, 1961.
- ¹⁵"Study Number IV, Documentation and Dissemination of Research and Development Results"⁵, p. 71.
- ¹⁶For an idea of the scope of these linguistic problems, see "Natural Language and the Computer", P. L. Garvin, ed., New York, McGraw-Hill, 1963.
- ¹⁷Pendergraft, E. D., A generalized computer system for language translation, The Linguistics Research Center, University of Texas, Austin, 1964. (This report describes work done in 1960).
- ¹⁸Mel'chuk, I. A., in Akhmanova, *et al.*,⁴, p. 49.
- ¹⁹Pendergraft¹⁷, p. 28.
- ²⁰Dullberger, L. H., and J. A. Strasser, "Electronics markets 1965", *Electronics*, Jan., 1965, pp. 87-110.
- ²¹Bar-Hillel, Y., Machine translation: The end of an illusion. "Information Processing 1962", (Proc. Internat'l Federation of Information Processing Congress, Munich, 1962), C. M. Poplewell, ed. Amsterdam, 1963, pp. 331-332.
- ²²Quoted in Locke, W. N., "Translation by machine", *Sci. Am.*, 194, 29-33, Jan., 1956.
- ²³Locke, *ibid.*
- ²⁴Alt, F. L., The outlook for machine translation, "1960, Process. Western Joint Computer Conference," pp. 203-208.