[*Impact of Science on Society* 10 (1960), 26-44.]

MACHINE TRANSLATION OF LANGUAGES: RESEARCH AND ORGANIZATIONAL PROBLEMS

by
E. DELAVENAY

E. Delavenay, Chief of the Documents and Publications Service of Unesco since 1950, was formerly Chief of the Official Records Division of the United Nations. He has been a teacher of English language and literature, and has taught French literature in English universities. His present functions as well as his personal interest have led him to take up the study of the applications of computer techniques to the problems of linguistics and international cultural relations, with which he has been concerned throughout his career.

INTRODUCTORY REMARKS

It was not so long ago that some of the most intelligent of men were still dubious about whether translation itself was really possible—a paradox that reminds us of Zeno of Elea denying the possibility of motion. Yet the perfectionists are given the lie by the completely bilingual 5-year-old child who can translate accurately into French what his kindergarten teacher tells him in English; and who may thus be said to have proved the reality of movement by walking!

We may therefore take it as axiomatic, or as a patent fact, that translation is possible, while provisionally granting the existence of a special preserve in which perfect translation is no more conceivable than an absolutely perfect performance of a piece of music or an ideal copy of a painting by a great master. This special field is that of literary translation, as against purely informative or 'indicative' translation or—to put the matter briefly and at the risk of rather over-simplifying a side-issue—as against scientific and technical translation.

Recently, the question has been mooted of the possibility of mechanical translation, that is, translation done entirely, or almost entirely, by a machine. This question, too, has been answered in the affirmative, and there are some optimistic spirits who have such faith in their analytical methods that they have even broken into the hallowed preserve described a moment ago. Miss Belskaya and Professor Panov vigorously maintain—and use excellent arguments to back their views—that machine translation of poetry will be feasible within the near future.

Let us assume, then, that machine translation is already possible for many types of work, and that further research and organization are all that is

needed. Let us also admit, provisionally, that it has certain limitations, just like any other human undertaking—limits to its scope of application and to its full automation.

FROM MAN TO ROBOT: PARALLELISM OF FUNCTIONS AND ORGANS

Professor Panov was, I think, the first to expound the principles of automation in translation work by drawing a very simple parallel between the work of the human translator and of the machine which can replace him. That parallel, in fact, cannot be bettered as an aid in trying to define the organs and operations of the machine designed to achieve the end in view: the production of a typewritten translation.

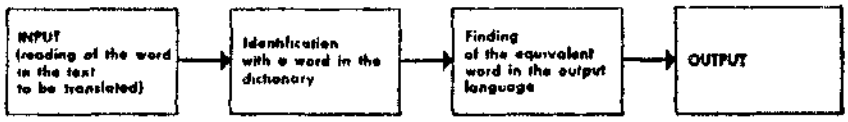
Setting a machine to do the work of a translator means automating a highly complex process, the main stages of which are as follows.

When I translate a written sentence from English into French:

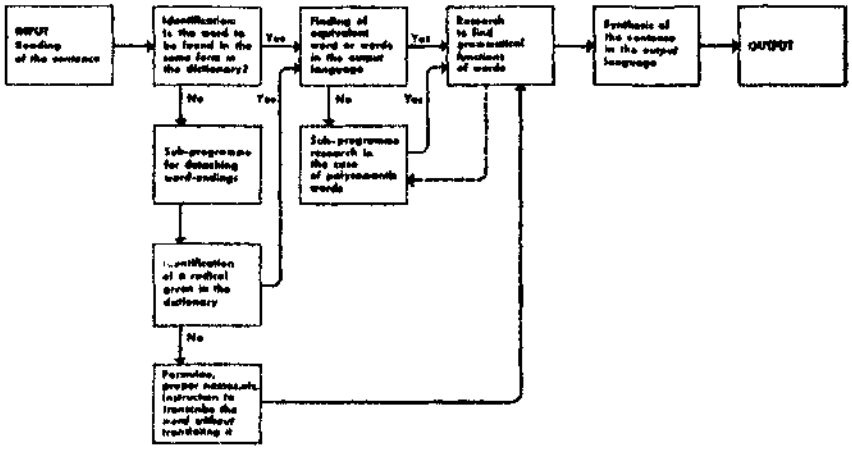
1. I read it—a function which may be termed 'reading' or 'input'. (If the sentence is spoken, I hear it, but for simplicity this function and the organ performing it may be described as 'reading' and 'the reader'.)
2. I perceive a meaning in the words and the sentence I am reading; this function may be described as 'identification' of the constituent elements of the sentence.
3. If the meaning of the words or word groups is not immediately apparent, this identification may involve a grammatical or logical analysis of the words or sentence.
4. Once that analysis has been completed, I search my memory or else look up a dictionary to find the equivalents of these words or their functions in French. A moment's thought makes it clear that the search for equivalents may relate either to words of equivalent meaning (semantics) or to words of equivalent grammatical value or function (morphology and syntax).
5. The next stage is the grouping of the French equivalents in an order in harmony with French usage and grammatical rules. This series of operations (establishment of word order and grammatical agreement) may be described as 'synthesis'.
6. Lastly comes the psycho-muscular operation of giving vocal expression to the translation of my sentence or inscribing it on a sheet of paper. This last operation may be termed 'writing' or 'output'.

Reading or input, identification of the input material by comparison with words stored in the memory, analysis when the words are not simple or do not coincide exactly with those stored in the memory, the search for equivalent words in the second language and their registration in a memory cell, combination or synthesis of these words in order to form a French sentence, writing or output—these, more or less, are the stages of my work as a translator which the machine will also have to reproduce.

When I am faced with an extremely simple translation problem, my work is in accordance with the following pattern.



If I wish to translate a single word, such as 'paper', it is easy to find the simple equivalent, *papier,* assuming that this word has no other meaning. The same applies to simple groups of words, such as 'the little horse', *le petit cheval;* 'the cat eats the mouse', *le chat mange la souris.* But problems are involved even with such words as 'le', 'petit', 'cat' and particularly with the inflected word 'eats', and the programme to be carried out is then more complex (see figure below).



Once the meanings and values of words and their combinations have been identified it is relatively easy, even for a machine, to translate the sentence.

But how is this work to be done?' And what organs will be used?

Human beings read with their eyes, the visual impressions received being transmitted by the nervous system.

Identification is effected by means of nerve centres, assisted by the eyes where the dictionary has to be consulted.

Analysis is carried out externally, with pencil and paper, and internally, in the nerve or brain centres, by processes which have not yet been subjected to searching investigation.

Equivalent words are found either in the human memory or in the dictionary (which may be regarded as an auxiliary memory).

Synthesis is carried out either in the memory in the case of a simple sentence,
or on paper in the case of more complicated sentences.
The final result (output) is recorded by hand, in writing or in typescript

The 'organs' used during this process are also to be found in a translating machine.

The reading is done by means of punched cards or tapes or, preferably, by magnetic tapes; an even better method, which will be feasible once certain technical problems have been solved, will be the use of photo-electric cells or some other photoscopic process.

The identification is done by comparing signals representing a word fed into the machine with signals stored in the memory-dictionary. This is a simple subtraction operation if the signals are binary digits. Other methods apart from arithmetical subtraction are conceivable. In every case, the operations are carried out in circuits which are connected up in much the same way as the neurones of the nervous system.

The analysis is conducted by means of what are termed sub-programmes. In this case, the parallel with performances by human organs can still be drawn, though with difficulty, provided we consider the matter in terms of actions rather than static conditions.

Let us take an example. The word 'loved' is fed into the machine. The dictionary does not show this word, which must therefore be looked up in a different form. A sub-programme for detaching word endings now comes into play. At a given stage, the machine will be asked the question 'Are the last two letters of the word ed?' and will give an affirmative reply. It will then be immediately instructed to look up the rest of the word (i.e., lov-) in the dictionary, where the French translation will be found. Once the word-ending ed has been identified as indicating the past tense in English, an appropriate symbol will be entered in a memory or register (the equivalent of the sheet of paper used by a schoolboy for making his translation notes). The sub-programme for detaching endings will thus make it possible to identify simultaneously, or almost so: (a) the ending of the word to be translated; (b) its radical; (c) the ordinary meaning of this radical in French; (d) its grammatical function (e.g., verb); (e) its tense (in the case of a verb).

There are of course other features relating to the word which still have to be ascertained (such as its subject, and whether the latter is singular or plural, and what type of complement the word governs, etc.). Other sub-programmes for dealing with the rest of the words in the sentence will make it possible to store in the machine's memory all the data necessary for sentence construction in French.

Let us now revert to the other operations involved in an over-all programme of machine translation. The last one mentioned was identification, which in this case implies a series of analyses. (As we have already seen, identification is often accompanied by analysis and a search for equivalent words.)

These operations are carried out in the electronic circuits of the machine through the agency of successive sub-programmes. They may be described briefly as the putting of a series of questions to which the answer will be either yes or no, each reply leading either to the registration in a memory organ of a result identified by the mechanism, or to a further search, which may culminate in turn in either a definite result or a renewed search. The sub-programmes may thus be compared to a succession of switches on a railway line, each switch either bringing the train to a standstill (recording of a result) or directing it to a further destination (continuation of the search). The same applies to synthesis, of which we shall now try to obtain a clear idea. Let us revert to our simple sentence, adding another word so that it becomes: 'The cat eats the white mouse'. Let us assume that the machine has identified each of the words in the sentence as follows: (a) definite article (gender to be determined); (b) noun rendered in French by the word *chat,* masc. sg.; (c) verb meaning *manger,* third person, present indicative; (d) definite article (gender to be determined); (e) adjective meaning *blanc* which has to agree with the following noun; (f) noun meaning *souris,* fem. sg.; (g) noun and adjective to be placed in correct order.

To write the sentence *Le chat mange la souris blanche*—that is to make a synthesis of the words in the sentence (or in school terms) to produce a French translation—the machine will have to arrange the French words found in the dictionary in the correct order, with nouns and adjectives agreeing in number and gender and verbs in the correct tense. To do this it will consult tables. There will be tables of definite articles giving the word *le* before *chat* because the latter is masculine, and *la* before the feminine word *souris;* a table of word order, which will make it say *souris blanche* and not *blanche souris,* and so on. This, very briefly, is the process of synthesis. To avoid elaboration, all that need be said here is that this process of consulting tables is just as feasible for synthesis as it is for analysis, and can be rendered entirely automatic, as it is in computation. The real problem is to build up valid tables which are not unduly complicated.

So far as output goes, suffice it to say that the signals registered in some kind of electronic memory will be able, when required, to work a typewriter, just as my hand writes or my mouth utters the sentence I have translated from English whenever I please.

For the sake of clarity, we have deliberately simplified our initial exposition. Simplicity of operation, however, obviously does not come about automatically, but is the fruit of prolonged and extensive effort. In the sequence of operations described above, the obvious aim of each of these operations is to reduce work to the simplest possible pattern, the identification of each separate word in the source language being immediately followed by its translation. But the twists and turns of human language, the social and cultural differences between peoples, the growth of languages

and the existence of linguistic structures that are at once fossilized and living stand in the way of achieving such simplification, and our efforts will depend upon an initial act of faith in the possibility of doing so.

BRIEF HISTORY OF MACHINE TRANSLATION RESEARCH

While the idea of making the consultation of bilingual dictionaries completely automatic may not have struck fifth-form boys obliged to look up the meaning of words in a Latin dictionary running to 800 or 1,000 pages as a practical proposition, it was bound to occur to present-day scientists similarly obliged to read the reports of their colleague in an increasing number of unfamiliar languages. The pioneers in this field were Smirnov-Troyansky and Booth, who, in 1933 and 1946 respectively, attempted to mechanize the consultation of dictionaries with a view to the production of more or less acceptable translations in telegraphic style.

Appetites were whetted by these initial endeavours, and as early as 1952, when the first conference was convened by the Massachusetts Institute of Technology (MIT), with the help of the Rockefeller Foundation, the idea took shape of going much further and rendering the translation process fully mechanized, or almost so. We have reached a point today where theoretical discussion as to how far translation can be mechanized is almost academic: between the simple mechanical consultation of the electronic dictionary and the 99.5 or even 100 per cent automation of translation lies a whole range of possibilities, all of which can lay claim to theoretical and practical value.

The following dates are landmarks in the history of machine translation.

1946: First discussion of the problem of mechanical dictionaries between Booth and Warren Weaver.

1946-1948: Work by Booth and Britten at Princeton on the automation of dictionaries.

1948: Richens suggests the automatic stripping of word-endings.

1949: Warren Weaver's historic memorandum gives a new impetus to research by drawing the attention of American scientists to certain aspects of language study capable of close analysis. In particular, Weaver induces the American sinologist, Reifler, to explore the system of signals either contained in the alphabet or in extra-alphabetical signs, visible or invisible but perceptible by the human reader and recognizable by a machine (e.g., spacing, position, punctuation, etc.).

1950: Reifler's first report. Work by Oswald and Fletcher on German.

1952: Definition by the first MIT conference of two short-range research targets.

1954: Experiment in machine translation by Dostert, and Garvin and Sheridan on an IBM-701. First issue of the periodical *Mechanical Translation* edited by William Locke and Victor Yngve.

1955: Publication of the first collection of essays on machine translation, edited by Locke and Booth.

1955: Beginning of Soviet research by the Panov and Lyapunov teams.

The scope of research has been greatly expanded since 1955, as witness the many centres now researching in the subject in the Soviet Union and the United States of America, to say nothing of others in the United Kingdom, Japan and Italy.

The spade-work done between 1946 and 1955 clearly showed that research by any team looking for quick, practical results was bound to follow a definite cycle or rather 'spiral' of development, and we have every reason to hope that by keeping to this course we shall progressively approach the fuller automation of the translation process.

As we have seen, then, that Booth starts from the idea of the mechanical dictionary look-up. The machine is conceived of as a means of helping the English or American scientist to read papers by his Russian or Chinese colleague by expeditiously providing him without any effort whatever on his part, with the meaning of each word in the order in which they occur in the original. The scientist is expected to make sense of this word-for-word translation because of his familiarity with the subject.

The obvious objection to this procedure is that the same word can have many different meanings, and it is met by giving every one of the meanings leaving the scientist to make his own choice among them.

We linguists and professional translators obsessed with the niceties of language and the difficulties of our art, greeted Booth's proposal with superior smiles. But we were wrong, and Booth and Richens were right, as the results obtained by Professor Oettinger's team at Harvard were to prove.

A mechanical dictionary could be self-sufficient provided words and ideas fitted exactly. But this is not the case; words represent concepts which on the whole are arbitrary and ill-defined. We only have to consider such French words as *et, ou, à, par, de, être, avoir, oser, penser, estimer, paraître, aimer, donner, façonner, amicalement, seulement, annonceur, pain, charcuterie, jardin, outil,* and word sequences such as *il n'est pas venu, de plus en plus, à qui mieux mieux, se faire du souci, mettre les bouchées doubles* to realize that neither simple words nor groups of words are units which are strictly comparable with each other as regards their meaning, grammatical or logical role in a sentence, structure or possible relationship with other words. Just try to translate these words into English or Russian and you will be faced with many problems in finding exact equivalents. The only thing to do is to study each word in relation to its immediate context and the meaning of the sequence or sentence as a whole.

In view of the impossibility of making individual words the basis of machine translation, the first thing that had to be done was to try to break up words which are not simple so as to mechanize the analysis of their

constituent elements, (e.g., features denoting gender, number, tense, person, etc.). Richens showed that tables of paradigms, such as we find in grammars, could be built into the storage part of the machine. It was thus possible to confine the dictionary to the stems or 'bases' of words, with references back and forth to the tables of endings.

This fairly easily disposed of the problem involved in the changing forms of a word: declensions in Latin, German or Russian, different forms to denote gender and number in French, and conjugations in all languages where the verb takes different forms according to person, tense and mood.

The attempt to deal with morphological inflections, incidentally, shed practical light on a phenomenon with which we were thoroughly familiar in theory. Inflections provide valuable information about the grammatical function of words in the sentence and hence, about the relationships between the verb, the subject, the object and the various complements of the verb. In so far as the machine has to record in a memory a full description of the special features of each word in the sentence, the analysis of word-endings affords a means of condensing and simplifying this description and the preliminary research.

But some languages are almost completely devoid of inflections. English has preserved only a few endings such as -s, -ing, -ed, -er, est and -th from the highly inflected Anglo-Saxon system. Yet English, with its rich vocabulary and exceptional flexibility, is capable of expressing every possible shade of thought. Hence in mechanizing translation, a study had to be made of the rules of syntax and language structures in order to derive from the positions of words in a sentence information about their functions which could not be ascertained from the words themselves or from their endings. It was a case of eliciting information about the function and even the nature of a word by studying its context rather that of analysing the word and its constituent elements.

The point is that the same word in English may, without any change in form be a noun or a verb or an adjective, and its grammatical role may well be impossible to determine properly except from its position in relation to other words. The same, I believe, is true of Chinese.

This means that research in language structure—with particular reference to Jespersen's work on analytic syntax and Charles Fries's studies of the structure of English—has now been put to practical use; and although we should avoid trying to explain everything by structural linguistics, this line of development has yielded good results. When I read an English sentence such as 'He gave the child a swim-suit', I can often identify the verb and its behaviour in relation to its complements and even without fully understanding the sentence, deduce that the first noun (child) is in the dative, the second (suit) in the accusative and that 'swim' has an attributive or adjectival function in relation to 'suit'.

The machine itself is capable, with the help of appropriate sub-programmes, of examining the order or words and their interrelationship and of recording the results in the appropriate cell of a register or memory until it has deciphered all the grammatical constituents of the sentence.

The idea of the automatic dictionary has thus led naturally to that of similarly automatic grammars to supplement it. Nor is there any contradiction between the two ideas: analysis has simply resulted in an attempt to endow the machine with all the tools used by a translator.

And so we must be wary of suggesting misleading alternatives and becoming ensnared by preconceived theories of translation or language. There are in fact three complementary and inseparable stages in the analysis of texts: analysis of the meaning of words culminating in the automatic dictionary; analysis of the form of words, culminating in tables of inflections; analysis of the relationships between words, culminating in tables of linguistic structures for a given language, and perhaps in tables of equivalent structures for another language. But this third type of analysis will bear on aspects of word relationships which have hitherto been given scant attention.

These three stages, corresponding to the three basic aspects of any grammatical analysis of language are, I repeat, inseparable, for it is impossible to obtain full information about the meaning of a sentence or word sequence from a glossary or from morphological tables, or from tables of syntax or language structure taken singly. Where the human translator relies on what he calls intuition ('I think this sentence must mean such-and-such'), the machine can only follow its pre-established programmes blindly and try by means of the three stages of analysis to work out a word-for-word or sentence-for-sentence equivalent of the original text fed into it. Not only is it impossible for the machine to dispense with any stage of analysis; it must come back to each in turn to complete its investigations. To attempt an abstract definition of the order in which the machine has to cope with vocabulary, morphology and syntax, would be to approach one of the most important research problems from the wrong angle. The approach will obviously depend on the type of language to be translated, while the relative importance of the three elements in 'deciphering' the text will depend on the language's special features. Whereas, for Latin and Russian, the starting point might be grammatical analysis with particular stress on morphology, for English the most important thing might be to find the verb and see how it fits into its context. Structural analysis is therefore the proper approach for deciphering English. However, such analysis is impossible—for any sentence—without lexical information, which means that we come back to the inevitable and primary task of consulting the dictionary and the word-endings.

SOME PROBLEMS OF METHOD

Our historical survey has led us directly to the problem of the methods to be used in applied research. One question we were able to consider was whether priority should be given to any one of the recognizable stages in language analysis for the purpose of machine translation.

We must now turn our attention for a while to the importance of general theories as compared with applied research. In particular, we have to consider whether research is as yet sufficiently advanced for the formulation of a general theory of language suitable for practical application or whether it would not be preferable to put all our energies into developing our knowledge and understanding of the phenomena of language and translation, with particular reference to the formal definition of the mutual behaviour of certain categories of words which are ill-defined or not defined at all in the traditional grammars.

All science moves forward in stages: hypotheses are advanced and must be subjected to repeated factual checks. Applied linguistics has now reached the point where the most urgent task is the preparation of a comprehensive and systematic inventory of scientifically observed facts.

But scientific observation implies an attempt to classify material in accordance with criteria based on observed characteristics. Nothing is more dangerous than to begin with a classification accepted *a priori* or based on minor or secondary characteristics. One of the scientific necessities in applied linguistics today is a reappraisal of classifications in general (classification of parts of speech, for example) which delves beneath the old terminology based on Latin and Greek grammatical usage and adduces new word classes based on the behaviour of words in their immediate context.

What I have in mind is not so much the methods and theories of the advocates of the structural approach as the modified form of structural linguistics flowing more or less spontaneously from studies of context or word sequence with a view to machine translation.

We should take good care not to waste our time in pointless discussion of general theories of language, or thought, or any other aspect of the problems before us. Professor Panov, rightly anxious to see research yielding quick and concrete results, warned his readers against placing undue confidence in the virtue of more or less general theories, and the attitude he specified is one of scientific empiricism.

Machine translation, as a new subject, has naturally inclined people to rely on general theories or would-be universal explanations whereby the problems can be strikingly presented to the general or newspaper public in a nutshell. It is tempting, of course, to use a simple formula for explaining everything, especially in the absence of any obligation to see that a concrete programme is successfully carried through under practical conditions.

The theory of information (or rather the mathematical theory of communication), cybernetics, structural linguistics and mathematical linguistics are very general terms often cited in connexion with the problems of machine translation. Far be it from me to belittle their importance. Each of these disciplines has its part to play, spotlighting some particular aspect of the innumerable problems involved in translation from one language to another.

In the same way, while the linguist's strong sense of the individuality of language may often set him at loggerheads with the logician, that is not to say that their viewpoints are irreconcilable. The linguist's analysis is perforce closer to dynamic reality, while the logician's amounts rather to a systematization of thought emphasizing the underlying unity of our mental processes. The linguist stands in need of the logician, but he will look askance at the suggestion that language—that means of self-expression built up of the collective memories of a community and reflecting the often highly personal means of analysis of the speaker—can be reduced to a strictly logical formal system. Such a reduction might indeed hold good from the purely logical point of view but be totally useless for elaborating a practical translation programme.

The same is true when we come to the systematic application of mathematics to language. For the time being at least, whether we like it or not, electronic computers are designed to handle figures and figures only. It has been found possible to encode letters of the alphabet in digital form so that they can be fed into the machine, and this system has its advantages. The question is, will the situation still be the same in 10 years' time, or will we have found some means of registering graphemes in visual form and phonemes in the form of modulations? This already seems to be within the bounds of possibility, and the further question therefore arises: what part will mathematics then have to play? The answer will doubtless depend on two factors—on the success of mathematical linguistics, particularly of the current work on reducing at least some of the aspect of linguistic modes of expression to formal mathematical systems, and on the future importance of computation in the strict sense of the term (digital computation and logical computation) in the auto-programming and actual programmes of machine translation. Lastly, it is quite clear that the present-day algorismic systems whereby a French sentence can be converted into a succession of signals suitable for feeding into a machine offer obvious practical advantages and thus ensure mathematics a leading place in machine translation.

It is also clear that the linguist nowadays must become a mathematician and make use of the theory of sets as well as of graphs and Boolean logic, and even more, of course, of statistics, for language, because of its complexity as a social phenomenon, is not reducible to purely formal systems, but is amenable as such to statistical analysis.

To sum up, no general theory will help us for the time being to solve the

problems of translation. But every one of the scientific methods of analysis will be useful, and the research worker should be in a position to exploit them all. His special field, however, will be the detailed and accurate observation of a mass of facts with the aid of all the resources of mechanography and electronic computing.

His main attention will thus be focused on two objects—on the operation and principles of automatic digital computers, and on language-analysis methods designed to make it possible to 'feed' sentences into a computer with a view to obtaining equivalent sentences in another language. It is in this field, in fact, that studies and research already abound, if only because of the complexity of electronic computers, and because of the host of problems raised by their use for language work and the multiplicity of 'bilateral' language programmes (from English into Russian, say, or from Chinese into French). If we multiply the number of problems raised by a bilateral programme of machine translation by the number of possible bilateral programmes, the product is, in fact, enormous. The bibliography of articles on particular aspects of bilateral programmes is being endlessly enlarged, and provides a most fruitful and rewarding field for research.

As in the natural sciences, it is such special studies that count most, for the final and complete realization of machine translation depends essentially on listing, systematically studying and classifying a vast amount of linguistic data. The actual system of research may not yet have been perfected, but it can already be entirely visualized and its general structure defined independently of any general theory, just as the structure of chemical research was being defined in Lavoisier's day.

It follows that whatever progress is made towards machine translation, it will depend on the rational organization of work in two directions: the scientific training of staff for machine translation research, and the preparation of programmes for the machine.

The main purpose of the work done in these two directions is the creation of new social units, of a virtual symbiosis between man and a new tool capable of increasing his productive capacity manyfold.

But like any other creation of human society this tool cannot be improvised. A machine translation research team is not created overnight, even where a great number of machines are available for experimental work or for the actual production of translations.

THE FUTURE ORGANIZATION OF A MACHINE TRANSLATION RESEARCH TEAM

As indicated, we have urged the priority of the applied linguistics research workshop or laboratory over theoretical studies; for while the latter are of the utmost academic interest and are capable of illuminating our subject when required, the issues at stake are practical ones, and all research, not to

speak of all practical achievement, must be based on a study of language as a living reality.

How can machine translation research workshop be set up, and what will be the general principles governing team-work?

Wherever headway has already been made in this field, linguists and specialists in electronic digital computers have had to co-operate and therefore to learn to work together.

The first problem is to arrive at a mutual understanding and to 'speak the same language'. Three parties are involved—the linguist, the mathematician and the machine. Each uses a special language. The machine has its signals, its electronic impulses starting with input and resulting in output; language for it means an order, a precise and unambiguous instruction. The mathematician is trained in strict analysis, in the painstaking verification of his tentative assertions. The linguist is deeply conscious of the niceties and refinements of language and is perhaps better able to apprehend intuitively than to reduce his knowledge to precise formulae. If he is a translator, he often has had to give more thought to the irregularities and anomalies encountered in his work than to the vast body of repetitions and linguistic rules which he applies unconsciously. The linguist has made a practice of recognizing the exception, whereas the mathematician is trained to deduce the general rule from the individual cases.

The machine imposes a language, a choice, a standardization of the phenomena fed into it, and the mathematician or 'programmer' is its interpreter. The linguist has to learn to be selective, and to see a fact in the light of a single interpretation, and not of two; he must look for similarities rather than differences. Each has to familiarize himself with the other's methods, at least sufficiently to understand and respect them. The linguist contributes inventories of facts, while the mathematician tries to make them fit a system in which practically all phenomena can be reduced to formal rules. The former is the champion and interpreter of the recalcitrant stuff of language, the latter is the organizer who seeks to create rules from the diversity of facts.

The second problem is the choice of the machine to use. Should we wait for the ideal machine, the one best equipped to cope with complexities of language, or should we go ahead with what we already have?

The existing research teams have given us our answer. What Booth has done with the APEXC, and Panov, Belskaya and their colleagues have done with the BESM, Giuliano and Oettinger are doing at Harvard with the Univac I and Reifler, Micklesen, Wall and Hill at Seattle with an IBM-650, may be described as clearing the ground by demonstration and classification. Their work may be likened to pilot projects pointing the way to the machines and programmes of tomorrow. Why should we wait? It would mean condemning ourselves to waiting for ever, spurning the bronze axe and keeping to flint on the plea that the iron age will soon be here.

Besides, the team must be trained, and it will be more useful for it to use slow machines not too far in advance of the psychological reactions and knowledge of a group of persons who are acquiring their collective skills and forging their instruments. While linguists and mathematicians are thus training each other, the problems must be tackled one at a time, and there is no point in rushing ahead too fast or reaching too far.

The third problem is the choice of a source or input language and of a target or output language. Here, again, we need to have a clear view of the problem at the outset and avoid a dispersal of effort. The choice, for example, may be dictated by the more pressing need for translations from one given foreign language rather than another. It would probably be advisable for any team to confine itself initially to a single input language, and certainly to a single output language until it has been able to split up and form a second team whose members will already have been thoroughly trained in linguistic analysis for the purpose of machine-programming.

However, there is one need perhaps which has not been sufficiently stressed, owing to the fact that most of the work has been concentrated on translation into English. This is the need for synthesis, or the preparation of programmes enabling the machine to produce correct sentences in the output language. In the case of a language such as French, which is still relatively rich in stylistic turns of phrase, forms, and traditions, a degree of selection and simplification is necessary in order to compose the output language. From the means of expression available in the language, those have to be chosen which are essential, and those that are not discarded from the outset. The synthesis of a sentence by the machine means the reconstruction, from a sentence in a foreign language, of a French text that is intelligible and acceptable though it may lack polish. One of the very first tasks will be to determine just how rich or poor a glossary the machine will require for ordinary needs: for example, how many expressions should be stored in its memory in order to render the word 'no' in the sentence 'I have no bread' (e.g., *je n'ai* pas, point, aucun, miette, etc.). In many respects this process has much in common with the very interesting work being done on Basic French at the Ecole Nonnale Supérieure at Saint-Cloud.

The fourth problem is the choice of a set of texts or 'corpus' to be used for the analysis of the input language, and choice of methods for the analysis of this corpus. Every team, as soon as it is set up, will need to go into this problem and its choice will be guided by certain important considerations. Obviously, it will be preferable for the texts to be scientific, so as to limit as far as possible the initial extent of the vocabulary problems to be solved and to keep grammatical and structural problems within the narrowest bounds. A corpus based on one of the humanistic fields of learning or on general literature would rake too many problems and would set the team chasing too many hares. A series of texts on mathematics, astronomy, chem-

istry or any other exact science will be the most immediately useful and suitable for circumscribing discussions of minor aspects of linguistic work.

The fifth problem will be study of the work done by other teams. In the field of machine translation, no research team of any kind can operate at present without keeping in very close touch with what is being done by others —not in any competitive spirit but with a view to close collaboration and the mutual exchange of information. Bibliography and the scanning of journals and documents on all aspects of related work will be an absolute necessity and will cover a host of subjects such as the organs and working methods of computing machines, general and applied linguistics, and all branches of work relating to translation and language problems, together with the extensive documentation on bilateral programmes of linguistic analysis for machine translation purposes.

This list of tasks and problems, though far from exhaustive, enables certain conclusions to be drawn.

In the first place, the research must be on a collective and not on an individual footing. It involves a close acquaintance with different techniques and branches of knowledge which no single person can now possibly command as a basis for practical action. The only way it can be profitably organized is on a sufficiently large scale to make use of all the resources of a large body of specialists, most of whom would be working full time; although many marginal aspects of the over-all field of research could be entrusted to experts working as temporary advisers.

The second point is that this research is international, first because it is multilingual, and secondly because national frontiers are not coterminous with linguistic frontiers.

In the third place, the number of bilateral programmes is enormous. In theory, a multilingual country like the Soviet Union, which has 70 official languages would require $70 \times 69 = 4{,}830$ bilateral programmes. It will therefore be necessary to explore the possibilities of simplifying and extending the general coverage of programmes. In addition to immediate bilateral studies, this will involve a work of synthesis which, by bringing out the essential principles on which bilateral programming is based, will chart possible methods of preparing the blueprint for a prototype universal programme.

The fourth point is the emergence of a twofold task, outlining an all-round approach; there is need first for the rapid development of simple programmes (e.g., an English-French, or a Russian-French programme for scientific translations) so as to build up the stock of lexicographical and grammatical knowledge which those programmes yield as a valuable by-product; next, for the investigation of the theoretical bases of future programmes, with a view both to the formulation of multilateral programmes, and to the extension of programmes along more ambitious lines to cover such fields as humanistic studies, literature and poetry.

THE PRIMARY NEEDS: THE TRAINING OF RESEARCH WORKERS AND THE DEVELOPMENT OF EXCHANGES BETWEEN THEM

On the threshold of what may be the most marvellous scientific adventure ever conceived by man, one of the paramount requirements of any research —the need for research workers—is also clamant in the field of applied linguistics that now concerns us. The fact that we have to deal with disciplines which are wholly beyond the individual research worker's capacity and require the constitution of research teams in which each worker plays his part means that we must construct a vast block-diagram of research showing the place of each in the common task. One of the first areas charted is that for the training of the men who will then demonstrate their skill by rapidly proceeding to occupy the other squares in the diagram and finish the job.

A new programme of studies will need to be devised by an Institute of Applied Linguistics or some other appropriate institution operating at the higher education or post-graduate research level, and will involve the patient guidance of the members of this new profession towards practical achievements, the formulation of methods best suited to this type of work as distinct from traditional linguistic teaching, the combination of theoretical teaching and laboratory work, systematic corpus analysis and experimental studies in reducing sentences to their prime elements. The work of such universities as Georgetown, Seattle and Harvard, of the Massachusetts Institute of Technology and the Rand Corporation, and of Birkbeck College in London, the Steklov Institute of Mathematics in Moscow, the Institute of Precision Engineering of the USSR Academy of Sciences, and the University of Leningrad points the way forward and serves as a stimulus to linguists in other countries where university teaching follows more traditional lines, so that it is perhaps more difficult to merge the various disciplines. The need is for theoretical and practical studies, coupled with the constant and persevering checking of working hypotheses by laboratory experiments involving mechanical equipment and all the human and mechanical skills to serve it. The work of British, American and Russian research teams must be followed up and additions made to the constantly growing body of practical results which enlarge our theoretical knowledge thanks to the repeated testing they undergo on the machine.

If the training of research workers is necessity No. 1, the second is the organization of exchanges between them. The International Conference which was recently organized by Unesco was a confirmation of previous efforts in this direction rather than a beginning. International exchanges have been proceeding ever since 1952 when the first International Conference on Machine Translation was convened at the Massachusetts Institute of Technology with the help of the Rockefeller Foundation, and the scientists in the various countries are attentive readers of everything published on the work

of their colleagues. On all sides, a sense of fellowship in research is developing, and is clamouring for stronger expression—a sense of the need for closer and more regular exchanges which every new achievement intensifies.

Let us reflect for a moment on what is at stake. This is international research, in the real sense of the word, with the aim of facilitating and expediting scientific exchanges by surmounting the barriers that languages have erected between scientists and between men in every walk of life. The workers who are striving to penetrate the secrets of language with a view to machine translation are every one of them linguists, and specialists in the foreign languages on which they are working: their aim is to pave the way for exchanges, their aspiration is that nation shall speak peace unto nation.

This task, which, as we have seen, is a collective one on the national or regional scale, exceeding the capacity of single individuals, is essentially and inherently one for international team-work. The scope of requirements and of the research to be done makes it a fit subject for international co-operation, which alone will enable quicker results to be obtained through the division of labour.

How is this division to be achieved?

There are two well-defined phases in any general programme of translation from one language to another—analysis of the first language and synthesis of the second. We can already foresee how work will be divided up so far as synthesis is concerned. Each language study group (English, Russian, French, German, Spanish, Chinese and Japanese) can have only one synthesis programme at the outset since its aim is to convert algorisms into intelligible sentences in the national or regional language.

But the work of preparing this synthesis calls for the establishment of specific relationships between algorisms and linguistic means of expression: in other words, it involves analysis of the output language. This means that the German laboratory working on translation into German could compare and exchange its findings with French or English laboratories analysing German as an input language. The problems, while not wholly identical, are partly so.

This gives an idea of how exchanges covering a fairly wide field could be organized at various levels, and the same will apply to families of languages. The Romance languages will in all probability be able to use programmes with certain common features for translation from English or from Russian, and the same will apply to the Scandinavian or Slavonic groups.

Can such co-operation be aided or accelerated by organizing research internationally? This is a question we are bound to put even if it is impossible to give an authoritative or final answer.

An international organization responsible for promoting scientific and cultural exchanges could hardly fail to meet the wishes of scientists, if only by encouraging (even without organizing) co-operation between them. But

it is for the scientists to come forward in the first place and specify their needs and possibilities, with details of the assistance they desire in order to extend their work.

A start could be made with a clearing house for the exchange of information and programmes on a larger scale than is possible through the publication of material even in specialist reviews, and with a clearing house for the microfilming and exchange of those duplicated documents with which we are familiar but which are never circulated in an adequate number of copies. Another useful activity might be the organization of exchanges of students at the post-graduate level, study tours and visits by research workers and teachers to the various research centres. Lastly, in so far as work for machine translation impinges on neighbouring fields of research—and it does so to a very large degree, particularly in the fields of documentation, bibliography and abstracting—an international body, even on a small scale, should keep in touch with all those specializing in the mechanization of exchanges of documentation, scientific abstracts and bibliographies and keep its members abreast of any work in those fields likely to be of interest to them.

A minimum programme would be to publicize, in all appropriate cases, the work now being done, to communicate the findings of the most advanced teams to newly constituted ones, and to promote research by disseminating the results of previous work and saving newcomers from wasting their energies on research that has proved unproductive. It could be developed fairly rapidly in the form of technical assistance to countries needing to use up-to-date methods and international equipment for studies in applied linguistics for purposes of internal communication.

I am alluding here to the multilingual countries of Asia—to India with its 14 official languages and its sixty or so local languages; to Ceylon, where two or three languages are used side by side, and to the countries in which a European language of convenience exists alongside a vernacular language likely to develop as new sectors of society obtain access to science and education and need to cultivate and develop their language. International action in the world emerging from the second world war has taken different forms, of which technical assistance to the so-called underdeveloped countries is one of the most vital and practical. An international body of specialists in machine translation would soon have a part to play in this special field of international co-operation.

There is one last aspect of the foregoing problems on which a few concluding words should be said. International organizations have a tremendous need for translations, which are a heavy drain on their budgets—to say nothing at this early stage about the oral interpretation of speeches. The United Nations, Unesco, the International Labour Organisation, the World Health Organization and their regional bodies each spend an increasing

amount on the translation of speeches, documents and reports into a number **of** languages which are rather arbitrarily, and with difficulty, kept down to four or five. Hitherto, these organizations have paid little attention to the promises of the scientists who have heralded the coming automation of translation. Their attitude will change, for the promises are becoming more explicit. In conclusion, I must express my pleasure at seeing my translator colleagues and friends—international civil servants and members of the International Federation of Translators—taking an interest in the problems of automation and preparing themselves for the part they are to play in that respect.