MECHANICAL TRANSLATION

DEVOTED TO THE TRANSLATION OF LANGUAGES WITH THE AID OF MACHINES

VOLUME THREE, NUMBER TWO NOVEMBER, NINETEEN FIFTY SIX

News

UNIVERSITY OF WASHINGTON
(from the Univ. of Washington News Service)

Research to prepare a machine for the translation of Russian into English is being conducted at the University of Washington under a contract with the International Telemeter Corporation of Los Angeles. The device is being developed by the corporation under a contract with the Rome Air Development Center, Griffis Air Force Base, Rome, N.Y.

A team of University experts under the direction of Dr. Erwin Reifler, professor of Chinese, is preparing representative samples of all language problems that may be encountered in Russian-English translation. These samples will test the ability of the machine to translate scientific and other material and to handle idioms, syntax, words with multiple meaning and other translation problems. Other principal investigators working with Dr. Reifler are Prof. W. Ryland Hill of the electrical engineering department, and Dr. Lew R. Micklesen, assistant professor of Russian.

Research by Dr. Reifler, a pioneer in mechanical translation research, and other linguists working with engineers in various parts of the world has shown that mechanical translation is possible. "Studies have now reached the stage where a machine, designed specifically for translating, can be constructed," Dr. Reifler said. "The same machine, with different bilingual memories, could be used for any set of two languages."

This will be the first practical test of mechanical translation using a machine specifically designed for the purpose. Previous research has been done with computers designed for mathematical calculations. The distinguishing feature of the Telemeter machine is that its "memory" unit stores information photographically on a rotating transparent disc. This makes possible the storage of large amounts of information in a limited space.

The "photoscopic" device stores information in the form of 30 million minute black and white areas arranged in concentric tracks on the rotating transparent disc. Information is read by shining a light through the disc and converting the resulting alternations of light and dark to electric signals by the use of a photocell. These signals are then processed by means of conventional computer techniques.

BABEL

Babel, an international journal of translation, published by the International Federation of Translators with the assistance of UNESCO, devoted its October, 1956, issue entirely to mechanical translation. Abstracts of the articles can be found in the bibliography section. The journal, a quarterly now in its second year, is devoted entirely to matters of interest to translators. Articles are included on literary translation, scientific translation in various countries, and the role of the translator in improving scientific and technical terminology. It may be obtained from Babel-Verlag, Bonn, Hausdorffstrasse 23, Germany.

INTERNATIONAL CONFERENCE ON MECHANICAL TRANSLATION

MIT PLAYED HOST on October 20 to an International Conference on Mechanical Translation. This issue of MT contains the proceedings of the conference in the form of eight reports from the groups represented at the conference.

The conference was attended by some thirty workers in the field from England, Canada, and the United States. Papers were received from the group in the Soviet Union. Prof. Panov of the Academy of Sciences of the USSR had visited several of the projects a month earlier, but was unable to stay over for the conference.

The conference was preceded by four days of informal discussion at MIT's Endicott House in Dedham, a suburb of Boston. Informal papers from each of the participants were circulated prior to the discussions, which were organized around a series of panels covering a dozen or so topics. Some of the material from the informal discussions will be published in subsequent issues of MT.

RAND Project

The Rand Corporation of Santa Monica, Cal., has announced the establishment of a project in mechanical translation. The following letter was received from Mr. David G. Hays:

"RAND has now established a small project in the area of mechanical translation in which I will be involved for a substantial part of my time during the next year. There will also be some programming and computer time available, and we hope to begin developing a file of Russian language scientific text on punched cards. Perhaps later in the year we will be able to offer duplicates of these punched cards to others who are working in the field.

The main purpose of the project at RAND is to try out some of the notions that I brought into the discussion at the conference last fall. I still believe that a great deal of statistical analysis of linguistic behavior is prerequisite to program machine translation. We do not expect to complete this kind of analysis at RAND, but we hope to make progress in development of techniques for the analysis."

GEORGETOWN UNIVERSITY

Professor Leon Dostert announces that a series of papers dealing with various problems of mechanical translation is being prepared in mimeographed form by the staff of the Georgetown Mechanical Translation Project and will soon be available on request to others who are working in the field. Requests should be addressed to: Professor Leon Dostert
Institute of Languages and Linguistics
Georgetown University
Washington, D. C.

UNIVERSITY OF MICHIGAN

An Associated Press dispatch in the newspapers of December 11 and 12 reported the development of an electric computer which automatically translates Soviet scientific papers into English at the University of Michigan. Several statements concerning the feasibility of mechanical translation and the results achieved so far were included in the dispatch. Later, the following letter was received from Mr. Andreas Koutsoudas, one of the staff members of the Engineering Research Institute of the University of Michigan, with the request that the letter be published in Mechanical Translation.

To the Editor: Dec. 12, '56

We wish to disclaim to our professional colleagues any responsibility for the publicity release on our mechanical translation project which was printed in today's newspapers. The implication that we are very far along toward practical translation was due to an overenthusiastic misinterpretation by a journalist.

Andreas Koutsoudas

[It is true that research in our field has been misinterpreted because of overenthusiastic claims in some press releases. Below is an excerpt from a letter received by the Mechanical Translation Project at MIT:

"I am extremely anxious to see the machine translator at MIT: I do not know whether this machine is running regularly or whether you allow visitors, but if it would be possible I would be most grateful to see it."

No, the machine is not "running regularly", but, as always, we are happy to welcome visitors.]

Editors

Machine Translation Development at the University of Washington

Erwin Reifler, Far Eastern Department, University of Washington, Seattle

MACHINE TRANSLATION development at the University of Washington is a joint enterprise of the Department of Far Eastern & Slavic Languages & Literature and the Electrical Engineering Department.

MT research at our University began in November 1949. We realized very early the importance of a close cooperation between linguist and engineer and the advantages of working jointly for a definite project with well defined linguistic and engineering conditions and limitations. The result was the planning of an MT Pilot Model by Dr. Thomas M. Stout, then of our Electrical Engineering Department, and its construction under the supervision of Prof. Hill.

During my research, I developed linguistic solutions for the identification by machine of grammatical categories, of both predictable and unpredictable compound words whose constituents occur in the machine memory, and for the automatic recognition and transfer to the output of words which, both graphically and in meaning, are shared by the two languages concerned in the machine translation process. It is for the purpose of testing the fundamental engineering feasibility of these linguistic solutions that the pilot model was planned.

Along with these researches went a steady development of an adequate terminology by the linguists and engineers of our group working in close cooperation.

At present, I am continuing research in all categories of words which can be omitted from the machine memory without any loss in the intelligibility and accuracy of the output text. I am also studying the problem of how to deal with proper and geographical names, which are also members of the general vocabulary of a language but should be left untranslated.

My research has been supported by two grants from the Rockefeller Foundation.

While my research, though primarily based on German language material, took into consideration the identical or analogous phenomena of a variety of languages, Dr. Micklesen directed his investigation primarily toward the Russian language and particularly toward the application of my results to Russian.

Supported by two grants from the Graduate School of our University, Dr. Micklesen carried out two studies. In one he investigated the process of compounding in the Russian language and elaborated proposals for the economical dissection of compounds by machine. The other developed into an exhaustive analysis of MT form classes of the Russian language, the prerequisite for the mechanical determination of intended grammatical and non-grammatical meaning. He also worked out a complete tabulation of all subclasses of Russian paradigmatic form classes and determined the number of distinctive forms in each paradigmatic set. These classes are purely formal, representing the most economical (structural) breakdown into Stems and endings.

Dr. Micklesen has also been very much interested in the theoretical aspects of the linguistic problems of MT. As a structural linguist, he has been especially concerned with fitting the results of MT research into the general framework of present-day linguistic thought. He recently contributed a chapter entitled FORM CLASSES—STRUCTURAL LINGUISTICS AND MECHANICAL TRANSLATION to "For Roman Jacobson" (Mouton & Co, The Hague, 1956).

Professor Hill has given much of his time to the study of the engineering aspects of a program for machine translation using a high capacity store. The recent development of large-capacity, rapid-access storage systems permits adopting a point of view different from that previously employed. It is no longer necessary to reduce the number of entries by dissection of stems and endings or by the use of "ideoglossaries". In fact, the vocabulary can be expanded to include idiomatic sequences as well as single words.

From the machine standpoint even a whole string of words which for reasons of source-target semantics has to be handled as an entity can be entered in the store and given an idioma tic translation. Such strings of words are the longest representatives of what we call "semantic units". Furthermore, punctuation marks and even the graphically very distinctive space

Continued on page 41

Mechanical Translation Work at the University of Michigan

A. Koutsoudas and R. Machol, Willow Run Laboratories, University of Michigan

THE PRINCIPAL differences between the work at The University of Michigan and other work in machine translation is in the emphasis placed on the problem of multiple meaning and the approach to that problem. Our approach consists in translating small groups of words, listing in the dictionary multiple meanings under each word in the group, and finding algorithms which make it possible to choose the proper set of meanings for the group. Some 9f the dictionary meanings under each multiple-meaning word will be vacuous and some will be redundant. The algorithms are based on the pattern of vacuous translations in the dictionary for the group of words under consideration. For example, for a particular idiomatic three-word sequence, the fourth meaning under the first and third words might be vacuous, and the entire idiom will be translated under the second word. The algorithm will be such as to lead the machine to pick the fourth meaning for each word in this case. These algorithms are discussed in more detail in the article on page

Since the problem of multiple meaning cannot be solved apart from the entire problem of translation, rules are also being prepared for the syntactical and grammatical aspects of translating Russian into English, and a large corpus of Russian is being processed. At the present time 64,000 running words (128 pages) of material from the Journal of Experimental and Theoretical Physics is being coded onto punched cards, and experiments are being carried on in which technicians simulate a computer in translating according to the stated rules. Theoretical frequency studies are also underway. These studies will use the results of the punched-card analysis. The theoretical aspects are based on equations comparable to those of Zipf's law. It is hoped to be able to predict answers to such questions as: How many different words will be found in a million running words? How many new words will be found in a second sample equally large? How many words must there be in a dictionary to ensure having 99% of the words in a sample randomly chosen from a certain field?

The University of Michigan also presented to the meeting a recent idea for a Universal Font of type for technical periodical literature. It is assumed that within a generation machine translation will be a fait accompli, as will machine reading (i.e., the scanning of printed matter with the production of signals suitable for feeding a computer). All of the great mass of technical periodical literature will then be routinely translated into many languages. At that time a number of trivial problems will arise, involving differences in type faces (fonts), diacritical marks, displayed matter (e.g. equations), underlining, the use of italics or boldface to convey special meaning, etc.

When mechanical reading and translation are routine, these trivial problems will be solved by international standardization. However, this will leave the great bulk of the technical literature published in the intervening years either untranslatable or translatable only with great extra difficulty. It is therefore suggested that this standardization be performed now, so that all technical literature published after, say 1960, would be translatable by machine. As a first step it is suggested that a universal font be established. For this purpose it will be necessary to make the following studies: (1) The readability of various fonts, from the human engineering point of view (accuracy and speed) and from the publisher's point of view (appearance and reader satisfaction). (2) The machine requirements. This will involve some crystal-ball estimates as to what the finally successful reading device will be like. Of course, such machines will eventually be able to cope with certain differences, but their task will be made enormously easier if they do not have to cope with the difference between K and K or between T and T.

It may be possible to standardize also on certain other things. For example, most equations are numbered, in parentheses, at either the beginning or end of the line. It might be possible to standardize on the beginning of the line, and to use the open-parenthesis sign, (, at the left to indicate any displayed matter. This could be a cue to the machine to photograph rather than translate.

Continued on page 41

Organisation and Method in Mechanical Translation Work

L. E. **Dostert, Institute of Languages and Linguistics, Georgetown University**

Certain postulates are posited as a basis for the orientation and organization of research in mechanical translation. They are the following:

1. The essential problem of mechanical translation is the establishment of acceptable correlation between the signs of one system (the source language) and those of another (the target language).
2. The signs of natural language, unlike the symbols of such systems as mathematics or chemistry, may be incomplete and multivalent.
3. The basic problem is the establishment of codes of systematic affixes to confer com pleteness and fixity in order to achieve acceptable correlation.
4. These systems of affixes must be such as to reflect the operations of translation and be programmable.

Based on these postulates, a group research program calls for certain organization and methods. Those in the Georgetown research project are as follows:

1. The recognition that at this phase of the research the primary problem is one of linguistic and translation analysis.
2. The essential direction of the research is placed in the hands of a group of scientific linguists of diverse competences.
3. The linguists meet regularly in a seminar in which specific problems are presented by a member of the group for discussion, review, and comments by the other members. As a result of the examination of specific problems, certain conclusions are formulated, some of them preliminary in character.
4. Under the guidance of the committee of linguists a group of research assistants with at least Master's standing in linguistics, who participate in the seminar, carry out the detailed research based on such conclusions, in conjunction with a group of bilingual trans - lation analysts.
5. The translation analysis is focused on material already translated in the field of chemistry. From this corpus, the material is analyzed systematically for eventual coding on three broad levels conjointly: 1) lexical, 2) morphological, and 3) syntactic.
6. The lexical material is culled from the actual translated text. The decision items within a given context are identified. The contextual cue or cues to the choice decision are indicated.
7. In addition to establishing solutions for lexical multivalence, procedures are being developed to handle problems in morphology and syntax as they arise in the material. At this stage in the research only preliminary formulations exist.
8. In carding the lexical and grammatical data, an attempt is made to symbolize the categorization by a code as follows:

1. ○ 1=1 One-for-one equivalence and identical linear sequence
2. ⊖ Deletion of source item(s) in output material
3. ⊕ Insertion of items in output
4. Choice of multivalent situation, based on item(s) following decision item
5. Choice in multivalent situation, based on item(s) preceding decision item
6. Combination of 4. and 5.
7. Inversion or change in sequential order of items at output
8. Choice in multivalent situation based on partial item in decision item

The seminar will work as required by emerging experience with the assistance of consultants in the fields of coding techniques, computational techniques, symbolic logic, and mathematics.

Report on Research

Cambridge Language Research Unit, Cambridge, England

THE CAMBRIDGE Language Research Unit is primarily concerned with analytic investigation of language, and in particular with a correlative study of the descriptive-linguistic, logical, algebraic and other notational characteristics of natural languages and of translation between natural languages. Much of this work is relevant to machine translation and the following four sections by members of the unit illustrate some of the applications that are being made. The first three are concerned with the possibilities of using a mechanical thesaurus; the fourth deals with mechanical translation via an interlingua.

Potentialities of a Mechanical Thesaurus
(M. Masterman)

The unit of a mechanical dictionary is the semantically significant "chunk", not the free word. From a logical point of view, uncoded MT dictionary entries form "trees", the paths of which can be determined, to a significant extent, by objective criteria. However, these, as they stand, are too complicated to be used directly for MT.

Attempts to construct multilingual MT dictionary entries show that the entries form, not trees, but algebraic lattices, with translation points at the meets of the sublattices. It also emerges that the complexity of the entries need not increase greatly with the number of languages, since translation points can, and do, fall on one another.

Such a multilingual MT dictionary is analogous, in various respects, to a thesaurus. A method of using such a thesaurus to refine the mechanical pidgin output of a bilingual mechanical dictionary has been devised.

Mechanical Translation Program Utilizing an Interlingual Thesaurus
(A.F. Parker-Rhodes)

The problem of setting the information contained in a fully general interlingual thesaurus into coded form for the use of an electronic computer would be formidable if not impossible of practicable solution, if it were necessary to include every entry as such. But if we could devise a system of coding such that each entry is represented by a numeral which would be calculated from information to hand respecting the current context-situation and the code-number given for the input word under consideration, then we could dispense with any actual thesaurus entries in the computer's storage, all the relevant information being contained in the input and output dictionaries which respectively provide the code-number of the input words and decode the numbers, calculated from these and the context, into the target language.

Mathematically, the problem is completely soluble, provided no limit is placed on the length of numerical symbols. If we limit ourselves to a practicable length of symbol, the question of adapting the general mathematical solution to actual use becomes one of ingenuity which can probably be solved, but which can only be assessed by practical effort. The mathematical procedure consists in finding a set of Boolean operations having certain prescribed properties which can be deduced from the conditions of the problem. These operations are few in number and could be built into a computer; being Boolean, they can be performed with very great speed.

A model solution, substantially simpler than that recommended for actual trial, will be described, and an example worked in it will be demonstrated. This will show up the sort of crossword-puzzle ingenuity required to devise a suitable context classification. The attraction of the method, despite this inevitable lack of elegance, is that it makes the computer actually calculate instead of merely looking things up in lists, and thus makes the whole procedure capable of sufficient speed to be feasible for a mechanical-translation program.

Linguistic Basis of the Thesaurus-Type Mechanical Dictionary and its Application to English-Preposition Classification.
(M.A.K. Halliday)

The thesaurus method of mechanical lexicography is an attempt to systematize the lexis in such a way that the "one-to-one word equivalence" principle can be maintained as the first stage in the dictionary, since the mechanical application of the concept of "primary meaning" implicit in this principle requires the arrangement of secondary translation equivalents into contextually determined systems. Each entry, consisting of a "key word" and its associates, constitutes one such system.

Multiple translation equivalence requires the specification of the conditions under which one of the terms in the closed system of a thesaurus entry is to be selected, these conditions being contextual features of the target language. This is illustrated by a "context-continuum" showing some word equivalence in non-technical railway terminology in four languages.

The thesaurus exploits the redundancy of the target language by handling its word classes without comparative identification. The autonomous treatment of the target language reduces the loss of determination involved in the translation process.

Among the word classes established for English as a target language, prepositions are particularly suited by their relatively low entropy to non-comparative treatment. Prepositions are classified as "determined" and "commutative". The former are listed as sub-entries of the determining word, having a single or multiple sub-entry according as they are wholly or partially determined. The latter constitute separate headings and are placed in closed commutation systems which differ from those set up for e.g. nouns in that they are in the first instance grammatically, not contextually, restricted.

General Program for Mechanical Translation between Any Two Languages via an Algebraic Interlingua. (R.H. Richens)

It has become clear that the amount of lexical and syntactical analysis required to produce a smooth and idiomatic mechanical translation from any base language into any target language is very great. It is interesting, therefore, to examine the possibilities of mechanical translation via a notational interlingua. With this approach, only one program is envisaged for translation between any two languages, with the addition of specific mechanical dictionaries for each input and output language.

The notational interlingua being studied is ideographic and constructed so as to represent the ideas of any base passage divested of all lexical and syntactical peculiarities; for which reason it is called Nude. The words in Nude are constructed of some fifty elements (Roman letters, capitals and lower case letters being regarded as different symbols), each of which denotes some basic idea such as plurality. animal, or negation. A word in Nude may consist of one letter only; the more complex a notion, the more letters are required. Each word in Nude is regarded as a relation, either 0-ad, 1-ad, or 2-ad; 1-ads are preceded by a point, 2-ads by a colon. Punctuation in Nude is used to indicate the concatenation of the words. The words linked by a 2-ad relation precede it and are separated by a comma, e.g., A, B:C; coordinate conjunction is expressed by a hyphen, e.g., A-B.C.

The translation program involves the following operations:

(1) Matching semantically significant "chunks" of the base passage against the Base-Nude dictionary.
(2) Reorganization of the syntax into Nude syntax by the method of cyclical reduction described at the 1955 Symposium of the Cambridge Language Research Unit, utilizing the word-class sequence entries of the Base-Nude dictionary (cf. MT III, 1).
(3) Treatment of chunk-chunk, chunk conjugation and chunk-semantic interactions by comparison with the appropriate interaction entries in the Base-Nude dictionary.
(4) Repetition of the above stages, using the Nude-Target mechanical dictionary. The potentialities of this method are to be illustrated by translation from a Japanese passage into English, German, Latin and Welsh.

List of Publications of the C.L.R.U.

1. Progress Report I (January, 1953), obtainable cyclostyled from C. L. R. U., 20 Milling-
ton Road, Cambridge, England.
2. Progree Report II, MT Vol. 3, No.1.
3. Annexe V to same is obtainable cyclostyled from Editors (MT).
4. The Potentialities of a Mechanical Thesaurus by Margaret Masterman.
5. A General Program for Mechanical Translation between any Two Languages via an Algebraic Interlingua, by R.H. Richens.
6. The Linguistic Basis of the Thesaurus-Type Mechanical Dictionary and its Application to English Preposition Classification, by M. A. K. Halliday.
7. An Algebraic Thesaurus, by A.F. Parker-Rhodes.

Stochastic Methods of Mechanical Translation

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

IT IS WELL KNOWN that Western languages are 50% redundant. Experiment shows that if an average person guesses the successive words in a completely unknown sentence he has to be told only half of them. Experiment shows that this also applies to guessing the successive word-ideas in a foreign language. How can this fact be used in machine translation?

It is clear that the success of the human in achieving a probability of .50 in anticipating the words in a sentence is largely due to his experience and the real meanings of the words already discovered. One cannot yet profitably discuss a machine with these capabilities. However, a machine translator has a much easier problem - it does not have to make a choice from the wide field of all possible words, but is given in fact the word in the foreign language, and only has to select One from a few possible meanings.

In machine translation the procedure has to be generalized from guessing merely the next word. The machine may start anywhere in the sentence and skip around looking for clues. The procedure for estimating the probabilities and selecting the highest may be classified into several types, depending on the type of hardware in the particular machine-translating system to be used.

It is appropriate to describe briefly the system currently planned and under construction. The central feature is a high-density store. This ultimately will have a capacity of one billion bits and a random access time of 20 milliseconds. Information from the store is delivered to a high-speed data processor. A text reader supplies the input and a high-speed printer delivers the output. The store serves as a dictionary, which is quite different from an ordinary manual type. Basically, of course, the store contains the foreign words and their equivalents. The capacity is so large, however, that all inflections (paradigmatic forms) of each stem are entered separately, with appropriate equivalents. In addition, in each entry, identification symbols are to be found, telling which part of speech the word is, and in which field of knowledge it occurs. Needless to say many words have several meanings, may be several parts of speech, and may occur with specialized meanings in different disciplines, and it is trite to remark that these are the factors which make mechanical translation hard.

Further, in each entry there is, if necessary, a computing program which is to instruct the data processor to carry out certain searches and logical operations on the sentence.

In operation, each sentence is considered as a semantic unit. All the words in the sentence are looked up in the dictionary, and all the material in each entry is delivered to the high speed, relatively low capacity store of the data processor. This information includes target equivalent, grammar and programs. The data processor now works out the instructions given to it by the programs, on all the other material - equivalents, grammar and syntax belonging to the sentence - all in its own temporary store.

With these facilities in mind, we may now examine some of the procedures that can be mechanized to allow the machine to guess at a sequence of words which constitute its best estimate of the meaning of the sentence in the foreign language.

The simplest type of problem is "the unconscious pun" which a human may face in seeing a headline in a newspaper in his own language. He has to scan the text to find the topic discussed, and then go back to select the appropriate meaning. This can be mechanized by having the machine scan the text (in this case more than one sentence is involved), pick out the words with only one meaning and make a statistical count of the symbols indicating field of knowledge, and thus guess at the field under discussion. (The calculations may be elaborated to weight the words belonging to more than one field.)

A second type of multiple-meaning problem where the probability of correct selection can be increased substantially and can also be mechanized is the situation where a word has different meanings when it is in different grammatical forms, e.g. the two common and annoying French words: pas (adverb) "not", (noun) "step, pass, passage, way, strait, thread, pitch, precedence", and est (present 3rd singular verb) "is", (noun) "east". The probability of selecting the correct meaning can be increased by programming such as the following for pas: "If preceded by a verb or adverb, then choose 'not'; if preceded by an article or adjective, choose 'step', etc." Experiment shows this rule (and a similar one for est) has a confidence coefficient of .99 of giving the correct translation.

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

The formulas above use grammar (and we hope someday syntactical context) to increase the probability. The human mind uses in addition other types of clue. A fairly simple type, and hence one easily mechanized, is the association of groups or pairs of words (without regard to meaning). These are the well-known idioms and word pairs. In the system proposed the probability of correct translation of words in an idiom is increased almost to unity by actually storing the whole idiom (in all its inflected forms) in the store. The search logic of the machine is peculiar in that words, or word groups, are arranged in decreasing order on each "page", so that the longest semantic units are examined first. -Hence no time is lost in the search procedure. Available capacity is the only criterion for acceptance of a word group for entry in the dictionary. The probability that certain word groups are idiomatic is so high that one can afford to enter them in the dictionary.

In principle, the same solution applies to word pairs. For example état has several meanings, but usually état gazeux means "gaseous state". Can one afford to put this word pair in the dictionary? Only experiment, with a machine, can determine the probabilities of occurrence of technical word pairs. Naturally, there will be room for some, and not for others. The exceptions lie in the same ground that we cannot approach with grammatical clues, but which may be solvable with the syntactical approach, although at the moment the amount of information which would have to be stored seems to be much too large.

The choice of multiple meaning like "dream/consider" (Fr. songe) is not of first importance the ultimate reader can make his own choice easily. The multiple meaning merely clutters the output text.

The choice of multiple meaning of the so-called unspecified words like de (12 meanings), que (33 meanings) is much more important for understanding a sentence. The amount of cluttering of the output text by printing all the multiple meanings is very great, not only because of the large number of meanings for these words but also because of their frequent occurrence. Booth and Richens proposed printing only the symbol "z" to indicate an unspecified word; others have proposed leaving the word untranslated, and others have proposed always giving the most common translation. These seriously detract from the understandability. At the other extreme, one could give all the meanings. In the case of unspecified words, the reader can rarely choose the correct one. so he is given very little additional information at the expense of reducing the ease of reading.

The stochastic approach of printing only the most probable permits the best effort in making sense and prints only one word, so it is easy to read. What is the probability of successful translation?

Let us look at a few unspecified French words. Large samples of de have been examined. In 68% of the cases "of" would be correct; in 10% of the cases "de" would have been part of a common idiom in the store, and hence correct; in 6% of the cases it would have been associated as "de 1'", "de la" which are treated as common word pairs, and hence in the store. In another 6% of the cases it would have been correctly translated by the rule sent to the data processor from the store: "If followed by an infinitive verb, translate as 'to'." Another 2% would have been obtained by a more elaborate rule: "If followed by adverbs and a verb, then 'to'." The single example of de le + verb probably would not have been programmed or stored.

There remain then 8-10% of the cases where "in, on, from" should not be translated at all. In some of the cases "of" could have been understandable, just as in the title of this paper "Stochastic Methods of Mechanical Translation" and "Stochastic Methods in Mechanical Translation" are equivalent. Further study, of course, may reveal some other rules to reduce this incorrect percentage.

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

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

Contextual Analysis in Word-for-word MT

Kenneth E. Harper, Slavic Department, University of California, Los Angeles

EXPERIMENTS with word-for-word MT of Russian scientific literature have given results which, except for such limited purposes as indexing, are far from satisfactory. The difficulty is not so much one of word order as of syntactic and semantic ambiguity of individual words. Regardless of the treatment of the problem of inflected forms, for example, it is impossible in the majority of instances to identify the grammatical case of Russian nouns. In addition to syntactic ambiguity, multiple equivalents must be assigned to a large percentage of words (to an estimated 45% of the running words in a physics text). The chief disadvantage of word-for-word MT, then, is its prolixity: the reader is confronted with a burdensome multiplicity of potential equivalents (syntactic and semantic) for several words in each sentence.

The chief cause of this ambiguity is the fact that each word is examined in isolation, as a discrete item. The human translator operates with the tremendous advantage of something called "context". Broadly speaking, context signifies environment: surrounding words, sentences, and even the subject area itself. Investigation shows that restricted contextual analysis, performed routinely, can resolve most of the problems of ambiguity. Remarkable clarification is attained even when the comparison of a given ambiguous word x is limited to the immediately contiguous word in the sentence (the pre-x or post-x word). Without attempting to rearrange the word order of the Russian sentence, one can obtain the following by comparison of each ambiguous word with the coded grammatical features or semantic class of contiguous words:

a) Syntactic clarification. The ambiguity of case forms in nouns can be reduced to an insignificant percentage, and proper English equivalents can be supplied in the form of English prepositions as demanded by the genitive, dative, and instrumental cases. Such prepositions can be withheld in translation when the requirements of Russian grammar demand it. Participles and adverbs which are indistinguishable in form from adjectives, can, be given the correct equivalent; the comparative degree of adjectives and adverbs can be adequately handled. In general, there are no serious problems of syntax which cannot be resolved by reference to the grammatical features of pre- or post-words.

b) Semantic clarification. The correct English equivalents of most of the "glue words" (especially prepositions and conjunctions) can be found only through contextual analysis. The programming of such analysis should be based on the observed behavior of these words in actual conditions. Thus, the meaning of the conjunction "i", which has at least four equivalents (and, but, also, even) can be pinpointed in more than 90% of all occurrences by simple reference to the grammatical category of contiguous words; the pronoun-adjective "ikh", meaning "(of) their" or "(of) them", can be similarly resolved. It should be stressed that completely unpredictable and unexpected relationships can be found between structural context and meaning, and that the barest kind of routine comparison results in a high (although not absolute) degree of accuracy in the determination of meaning.

Non-structural clarification of meaning takes several forms. In the first place, techniques of MT lexicography need to be developed, i.e., the science of choosing the best "cover-all" target language equivalent from a group of relatively synonymous equivalents, and the selection of equivalents based on observed behavior, rather than upon the evidence of a dictionary. (Thus, in the area of physics the Russian izmenenie may always be found to equate with "change", although Bray's technical dictionary lists nine fairly distinct meanings.) In effect, what is needed are true ideoglossaries, based on actual, rather than potential, behavior.

The application of contextual analysis offers great potentialities for semantic clarification. Operating again on the basis of observation, we can construct and code word classes which cause contiguous words to behave in a predictable manner. Thus, the preposition po has ten potential possible equivalents when followed by a noun in the dative case; by reference to predetermined noun classes we can reduce the number of choices to one in a given instance. The necessity of treating each new combination as an "idiom" is eliminated. It is also possible to pinpoint the meaning of many nouns which are ambiguous even within an ideoglossary by reference to the class of the accompanying adjective, or to specified key words in the title or opening sentences of the text.

There is no question that the kind of study in syntax and semantics which can be realized with the aid of machine techniques will result in the discovery of usable principles of association, so vital in the operation of what is called "contextual analysis".

REIFLER *from page 33*

between words can be considered as letters of an extended alphabet and as part of a "semantic unit". This extension of the concepts of alphabet and word provides additional graphic and semantic distinctiveness which greatly improves the translation product.

Based on these points of view a program for machine translation has been devised which 1) provides for the translation of words and word sequences, 2) permits the dissection of compounds, and 3) permits the handling of prefixes and certain types of suffixes. Each unit of input is compared serially with the entries of the store to find the longest possible memory equivalent that matches an initial portion. This is accomplished by a logical ordering of the store to place any memory equivalent that is an initial portion of a longer one behind the longer one. Each entry consists of the memory equivalent of a "semantic unit" of the source language, its target language equivalent or equivalents, the control symbols for operating the machine, and the editing symbols intended to help the reader of the output text. In a more advanced machine the editing symbols become logical tags used in a computer to edit the information extracted from the memory and thus to supply a better translation product.

Since May 15 of this year our group has been working on a project for machine translation from Russian scientific texts into English by means of the photoscopic memory device being developed for the Air Force by the International Telemeter Corporation of Los Angeles. The project is based on a contract of the University of Washington with the International Telemeter Corporation. The term of the contract is one year.

KOUTSOUDAS *from page 34*

Application to non-Roman-alphabet languages (especially Russian) would be a possibility for the more distant future.

After a suitable standard font has been chosen, it will be necessary to convince the publishers of technical journals to use it. This should not present nearly so much difficulty as many proposals for international standardization, since these people are most likely to cooperate on such matters. Furthermore, the change will probably not involve any expense, since the printers of these journals have hundreds of fonts already and can continue to use the discarded fonts for non-technical publications.

On the Problem of Mechanical Translation †

D. Panov, The Academy of Sciences, Moscow, U.S.S.R.

HAVING STARTED WORK on mechanical translation, we arrived at the conclusion that both the lexical meaning and the morphological shape of the word can and should be utilized in analyzing the text, and that for purposes of translation it is impractical to omit the information which can be thus obtained. The utilization of the lexical meanings of words as well as of their contexts may also affect problems of coding. These questions are extremely important to automatic translation.

We based our work on the following principles:

1. Maximum separation of the dictionary from the translation program. This enables us to enlarge the dictionary easily without changing the program.
2. Division of the translation program into two independent parts: analysis of the foreign language sentence and synthesis of the corresponding Russian sentence. This enables us to utilize the same Russian synthesis program in translation from any language.
3. Storing all the words in the dictionary in their basic form. This enables us to design the program for synthesis of the Russian text according to the standard rules of Russian grammar.
4. Storing in the dictionary all the constant grammatical properties of words.
5. Determination of multiple meanings of the words from the context, whereas their variant grammatical characteristics are determined by analyzing the grammatical structure of the sentence.

These principles have proved quite reliable in the practice test to which they were subjected. Hence it seems to us that they constitute a reliable basis for the solution of the problem of MT.

The contents of the dictionary, for our experiments, were determined by an analysis of mathematical textual material, starting with Milne's "Numerical Solution of Differential Equations". For the practical experiments, which were carried out on the BESM (the USSR Academy of Sciences' high-speed electronic computer), a dictionary of 952 English and 1,073 Russian words was compiled.

For a number of English words (121 words, in our case), the place-in-the-vocabulary indication is replaced by special digit indication to show that these words have multiple meaning. The proper Russian word is chosen in this case by utilizing a special program of automatic translation, which we call "the Polysemantic Dictionary".

If the spelling of the word in the text coincides exactly with that of a word in the dictionary, i. e., their numerical codes coincide, this fact can easily be established by the operation of matching. This is the principle used for finding words in the dictionary.

In order to find words in the dictionary which possess an affix (say, 's' or 'ing' or 'ed'), the machine must discard these endings after which it must repeat the search for the word with the discarded affix.

To determine the meaning of a polysemantic word, the words surrounding it in the given sentence are analyzed. Both the semantic and the grammatical characteristics are established. The routines for determining the particular meaning of a polysemantic word are based on an elaborate analysis of a great body of concrete material and are placed together in a special part of the translation program called the "polysemantic dictionary". Idiomatic expressions are also included in this part of the program.

It should be noted that the establishment of the most simple and general criteria for determining a particular meaning of a word (or group of words) is the result of substantial preliminary work by our linguists on actual texts.

If a word in the sentence to be translated is not found in the dictionary, it is stored unaltered in the memory of the machine. When the translated sentence is printed out, such a word will be printed in Latin script.

Investigations in the area of the dictionary are fairly extensive. In our group they have been carried out by L.N. Korol'ev.

Of great importance is the space that a dictionary occupies in the memory. A method of "code compression" devised by L.N. Korol'ev

† Translated by M. Friedman and M. Halle, MIT.

considerably reduces this space.

The automatic translation program is divided into two main parts — analysis and synthesis.

In the first part, the form of the English words, their place in the sentence, and the grammatical information given in the dictionary are analyzed with a view to the determination of both the grammatical form of the corresponding Russian words and their place in the Russian sentence. The resulting information is recorded by means of indices, thereby permitting passage to the second part of the program -- "Synthesis of the Russian Sentence". Here, Russian words, taken from the dictionary in their basic form, acquire grammatical form in accordance with the indices obtained from the analysis.

Both English and Russian grammar is presented as a series of special schemes for the basic parts of speech: verbs, nouns, adjectives, numerals, etc. The working basis of each scheme is dichotomic analysis, i.e., a system of "checking" for the presence or absence of a certain grammatical (morphological or syntactical) characteristic of the analyzed word. In checking, only two answers are possible, either positive or negative. Each of these answers admits either a final conclusion and the development of the corresponding grammatical indices for the given word, or the continuation of the check for the presence of the next characteristic until a definitive answer is obtained together with an indication of which grammatical indices must be developed for the given word.

Different parts of the program are ordered in a sequence which ensures the development of the indices necessary to carry out further operations.

Starting with the input of the English sentence into the machine, the entire translation process has been carried out automatically with no human intervention whatsoever. To make the machine translate in the manner just described, an enormous amount of preliminary research work by philologists was required especially by I.K. Belskaya, our philologist-in-chief, and by the mathematicians I. S. Mukhin, L.N. Korol'ev, S.N. Razumovskii, G.P. Zelenkevich, and, in the early stages, N.P. Trifonov.

S.N. Razumovskii has been studying translation schemes and programs and their logical structure. He has developed a system of symbols that makes possible the recording of the details of the above mentioned schemes in an appropriate manner.

Our opinion is that the principles according to which machine translation of languages should be organized have been sufficiently clarified by now and that the time is ripe to undertake work on a large scale. We have started research work in automatic translation from German, Chinese, and Japanese into Russian.

In our discussions of machine translation from Chinese and Japanese, we thought that great difficulties would be presented by the input in these languages. However, this problem, apparently, will be solved easily by using the Chinese telegraph code.

The work on German is being carried out under the direction of Belskaya by G. P. Zelenkevich and E. A. Khodzinskaya; Chinese by A. A, Zvonov and V. A. Voronin; and Japanese by M. B. Efimov.

We also plan soon to take up the problem of translation from one foreign language into another. For this we intend to use Russian as the "inter-language".

Mechanical Translation Research at MIT †

Victor H. Yngve. Massachusetts Institute of Technology, Cambridge, Massachusetts

WORK IN THE FIELD of mechanical translation started at MIT in 1951 when Y. Bar-Hillel became perhaps the first full-time worker in the field. In 1952, he organized the first international conference on mechanical translation. It, too, took place at MIT. Compared to this second conference, the first one was small. We have contributions from over thirty people at this conference compared to thirteen four years ago — striking evidence of how fast the field is growing.

Bar-Hillel was interested in syntactic questions and was one of the foremost early exponents of the point of view that a machine would have to handle syntactic problems in order to provide adequate translations. Much of the work at MIT has continued from this point of view. The rather considerable difference in word order between German and English sentences indicates that a translation on a word-for-word basis without word order change would not be desirable.

Word order change, from German to English at least, often turns out to be phrase order change. For example, the order of adverbial expressions (phrases) of time and place frequently has to be reversed, or they have to be placed differently with respect to various objects or prepositional phrases. Routines for making such word order changes would seem to require some mechanical procedure for separating the sentence into phrases and clauses and identifying each of them as to kind or type. In other words, a machine routine for changing word order from German to English could be based on a routine for recognizing German sentence structure.

Besides word order changes, the correct meanings of the words must be selected. Many suggestions have been proposed for choosing among the various possible meanings of words. The method that we are exploring in detail promises to solve some of the multiple-meaning problems; namely, those that are connected with sentence structure. These very difficult multiple-meaning problems can be resolved by machine routines based on the same sentence structure analysis that the machine would make in order to initiate the word order changes. Examples of the sort of choices of meaning that could be made on the basis of sentence structure are not at all hard to find. Take, for example,the word "der". Various meanings are'the', 'of the', 'who', 'that', 'which', 'he', 'it'. If the sentence structure is known, then it is known whether "der" is an article, relative pronoun, etc., and many, if not all, of the incorrect meanings can be eliminated. Each of the other parts of speech offers many more examples.

An alternative to the use of rules based on sentence structure is the use of what may be called ad hoc rules. For example, if "der" follows a word that is capitalized without an intervening comma, the translation 'of the' will be right about 95 per cent of the time. But it will be wrong in those very cases in which the ad hoc rule does not correspond to the facts of sentence structure. The difficulty with ad hoc rules does not end with this 5-per-cent error; many times it is virtually impossible to find any satisfactory ad hoc rule for a situation that is quite clear on a structural basis.

In the course of our work, one thing has become very clear. If the machine is to recognize the sentence structure, we must have a description of sentence structure to serve as a basis for the recognition routine. Some sort of a description is required for any recognition routine. Even ad hoc rules are based on a description — not a proper description of sentence structure, but a description of certain statistical features of a sample text and its translations, such as: "In 95 per cent of the cases where 'der' should be translated by 'of the', 'der' is preceded, without an intervening comma, by a capitalized word." A few simple ad hoc rules can provide a rough translation that is better than a word-for-word translation, but the addition of more and more ad hoc rules is not the way to better and better translations, because these rules will become very involved and entangled one with another. A straightforward description of sentence structure may provide much simpler routines , and routines that will handle problems that ad hoc rules

† This work was supported in part by the U.S. Army (Signal Corps), the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U.S. Navy (Office of Naval Research); and in part by the National Science Foundation.

can't handle.

In spite of the wealth of material on German grammar and syntax, we have been unable to find an adequate description of the language on which recognition routines could be based. And this for perhaps two reasons: In the first place, there has probably been no pressing need prior to MT for such a grammar and syntax. In the second place, perhaps linguists have not known how to make such a grammar, or how to tell a good grammar from a bad one.

It is these problems that have been occupying N. Chomsky. He has been working on a theory of grammar that gives many new and powerful insights into the structure of language.

J. Applegate has been working on the detailed structure of the German noun phrase. His descriptive statements will take their place in a complete grammar of German. R. Lees and G.H. Matthews, in the short time that they have been with the project, have looked into the structure of the German verb phrase, and the range of applicability of certain recognition rules of the type proposed by Oswald and Fletcher.

It is hoped that our work will lead to an adequate description of the German and English languages and thus to accurate syntactic translations with the proper choice of word order and constructions in English from a German input. Many multiple-meaning problems will be solved at the same time. There will still be problems left, however. These are connected with the so-called meaning words. Perhaps these problems can be solved by utilizing a more sophisticated classification of the meaning of words than one has with a series of field glossaries.

Mechanical Translation and the Problem of Multiple Meaning †

A. Koutsoudas and R. Korfhage, Willow Run Laboratories, University of Michigan

THE UNIVERSITY OF MICHIGAN undertook research, late in 1955, in the analysis of language structure for mechanical translation. Emphasis was placed on the use of the contextual structure of the sentence as a means of reducing ambiguity and on the formulation of a set of operative rules which an electronic computer could use for automatically translating Russian texts into English. This is a preliminary report on the latter phase of the problem, stating the results and suggesting a practical method for handling idioms and the problem of multiple meanings.

It was decided that the first work would be done on Russian texts in physics, both because of the interest in this field and because of the general availability of texts. Some work has already been done in this field.[1] If this work proves successful, it will form a basis for work in other scientific, technical, and military fields.

A text was selected from a Russian journal on experimental and theoretical physics.[2] It was chosen to present most of the expected difficulties; i.e., stylistic, orthographical, grammatical, etc. On the basis of this text a vocabulary was set up and fifteen rules were established. (Subsequent work has altered the rules slightly to remove such obvious faults as the occurrence of "the" before proper names.) It should be realized, of course, that neither the vocabulary nor the rules were in generally applicable form. The vocabulary was simplified by applying a "one form, one meaning" rule whenever possible. Thus, inflectional endings were stripped from most word stems although in some cases a word was listed with two or three specific endings. Most words were given their scientific meaning only. Some words, however, occurred in more than one sense, or were combined with others to form idioms; in which case more than one meaning had to be listed. Finally, the words were listed in conventional grammatical categories; i.e., verb, noun, adjective, etc.

In the long run, we expect that the concept of conventional categories will be completely abandoned. What we hope to have, instead, are word groups the interaction of which will provide the grammatical and syntactical information needed. The need for such grouping has been made apparent.[3]

The rules were developed empirically by analysis of the essential processes undertaken by a human mind in translating a foreign text. It was found that most of the rules involved either word order or the grammatical functions which in Russian are indicated only by case endings and which in English might be classified by inserting a preposition. In most cases the rules concerning word order were sufficient to eliminate the necessity of referring to endings. To test the adequacy of the rules, several volunteers who had no knowledge of Russian were asked to translate the original text, using only our rules and vocabulary.* Except for random, minor stylistic faults, it turned out that the resulting translations were clear and accurate. Being convinced that the rules are as complete as is practicable for the text, we are currently enlarging the vocabulary in preparation for future tests on different texts.

Perhaps the most significant result thus far is the success in handling multiple meanings,

---

1. See K.E. Harper, "A Preliminary Study of Russian". Machine Translation of Languages, The Technology Press of the Mass. Institute of Technology and John Wiley & Sons, Inc., New York, 1955.

2. Zhurnal Eksperimental'noi I Teoretichesk'oi Fiziki. Vol.26, No.2, pp. 189-207, Feb., 1955.

† The work upon which this paper is based was performed under the Department of the Army, contract No. DA-36-039-sc-52654.

3. See V.H. Yngve, "Sentence for Sentence Translation", MT. Vol.2, No.2, Nov., 1955.

* The Russian text with the vocabulary and rules based on this text will be found on pp.48 to 49. A standard translation and a translation made with the help of the rules by a volunteer who had no knowledge of Russian are on pp.50 to 51.

which has given us an insight into the problem of idioms. Although the problem of ambiguity as exemplified by this situation was greatly reduced by the use of a highly specialized vocabulary, the situation still occurred and a means for solving it had to be found. Published results on this problem have, generally, involved either a post-editor or a separate idiom dictionary.[4] These methods seem undesirable particularly in view of the additional computer time required for translation. Consequently, a method was developed which, it is felt, is widely applicable. The assumption was made that the specific meaning of a word could be determined from its context. It developed that not only is this assumption valid, but in fact we need not consider sequences of more than four words. The method used is the following:

All possible meanings of a word are listed, consecutively, in the order (1), (2),..........(n). In general, in order to have corresponding meanings mesh, it will be necessary to list some meanings for each word more than once, and to include some blank translations. When a word with multiple meanings is encountered, the number (n) of meanings is noted and translation is postponed. Subsequent words are examined for the number of possible meanings of each, until a word (X) with a single meaning is encountered. If there is only one word in the sequence preceding X, then the first listed meaning is assigned to this word. If there is more than one word in the sequence preceding X, we determine (M), the minimum of all (n) noted in the sequence. Let us denote by (i) [A] the i-th meaning of a word A, and by 0 a blank (null) translation.

Given a two-word sequence, A B, we consider (M) [A] and (M) [B] . If neither of these are blank, we translate, assigning meaning (M) to each word. If either of these is blank, we consider (M-l) [A] and (M-l) [B] and apply the same test to these. In this way, we find the highest numbered meaning which is not blank for either A or B and assign this meaning to each.

Given a three-word sequence, ABC, we consider (M) [B]. If (M) [B] is 0, we consider successively meanings M-l, M-2,….., as above, and assign finally to all three words the highest numbered meaning which is non-blank for all. If (M) [B] is not 0, then if (M) [A] and (M) [C] are both 0, we assign meaning (M) to the three words; otherwise we search meanings M-l, M-2, ...of all three words, applying the above rule.

In a four-word sequence, ABCD, (M) [B] is again considered. The procedure followed is that used for a three-word sequence, except that (M) [D] must be considered along with (M) [A] and (M) [C] .

In all cases, if no translation is found by the above procedure, we assign to each word meaning (1).

By properly ordering the meanings for each word (listing some meanings several times if necessary), it has been found possible to obtain valid translations for over 96% of the two-word sequences [The two exceptions which occurred, по делу and цель в, were easily handled by separately listing дел in the form делу , and цел in the form цель .] and for over 90% of the three-word sequences which might occur. These figures are based on the possible sequences without reference to their relative frequency of occurrence in actual use. It is not known how the difficulties in "properly" ordering the meanings will multiply as the vocabulary is increased. With each new word (or meaning) added, the order of the meanings previously listed may have to be changed *so* as to maintain consistency as much as possible.

In this system an idiom is handled as merely an additional meaning which is possible. A study of the structure of three-word idioms showed that generally the second word had the least number of meanings. On this basis it was decided to assign to the second word the entire idiomatic meaning, and to supply corresponding 0 translations for the other two words. Thus, for example, the Russian idiom по сути дела ("actually") would appear as по = 0, сут = actually, дел = 0. (Note the dropped inflectional endings.)

To illustrate this method, let us consider the eight Russian words том, дел, сут, цел, по, в, о, and теори. From these eight words it is possible to form 56 two-word sequences and 336 three-word sequences. However, of these only 29 two-word and 106 three-word sequences are linguistically possible. It is assumed, of course, that the appropriate inflectional endings are supplied in each case. (The list of sequences, with translations, is available on request.) By working with these 135 sequences it was found that the arrangement of meanings given in Table I is the best possible. There seem to be no algorithms for ordering the meanings, other than that the idiomatic meaning, if any, be

4. See, for example: "The Treatment of Idioms" by Y. Bar-Hillel, typewritten, 8 pages; "A Study for the Design of an Automatic Dictionary" by A.G. Oettinger, doctoral thesis, Harvard University, 1954.

the last meaning listed for at least one of the words.

TABLE I

том	сут	теори	цел	в	по	о	дел
1. that 2. 0	1. essence 2. actually	1. theory 2. 0 3. theory	1. purpose 2. 0 3. 0 4. order to 5. target	1. in 2. 0 3. in 4. in 5. 0	1. by 2. 0 3. according to 4. 0 5. at	1. about 2. 0	1. fact 2. 0

It may be noted that on the basis of only the three words по, сут, and дел the shorter arrangement of meanings given in Table II suffices,

Table II

по	сут	дел
(1) by (2) 0	(1) essence (2) actually	(1) fact (2) 0

It will be observed that there is a certain amount of redundancy inherent in this system. However, it is felt that this is a minor fault; first, because the percentage of redundant meanings in the entire vocabulary appears to be small (around five per cent) and second, because this plan does not require a separate idiom dictionary or other special devices which tend to increase computer translation time. Although further research is necessary for the complete development of this method, we believe that the theory used is valid and that it eventually will lead us to the solution of most multiple-meaning problems.

VOCABULARY AND RULES

NOUNS

Буссин - Boussinet
врем - time
времен -(1) time (2) the period
вычитани - subtraction
движени - movement
действительност - reality
дело - (1) fact, (2) 0
значени - value
значениями - values
интервал - interval
корреляци - correlation
Крутков - Kroutkov
малост - shortness
момент - instant
некоррелированност - uncorrelativity
обобщени - generalization
Орнштейн - Ornshtein
основани - reason
Планк - Plank
последействи - after-effect
предполозкени - assumption
промехутк - interval
приращени - increment
приращений - Increments
процесс - process
работ - work
рассмотрени - examination
результат - result
результатам - results
релаксаци - relaxation
сил - force
скорост - velocity
создали - (1) formulation
(2) formulate
сравнени - (1) comparison
(2) as compared
Стокс - Stokes
сут - (1) essence
(2) actually
теори - (1) theory

(2) on the theory
(3) in the theory
течени - (1) course
(2) during the
удар - collision
уравнена - equation
ускорени - acceleration
Фоккер - Fokker
формул - formula
формулой - by the formula
функци - function
цел - (1) purpose
(2) in order to
частиц - particle
частот - frequency
частност - (1) particularity
(2) in particular
Эйнштейн - Einstein

VERBS

был — а — was
был — и — were
выражать - to express
оказыва - ется - proves to be
описыва - ет - describes
отсутству - ет - is absent
предполага - лась - was assumed to be
предполага - лись -were assumed to be
привед - ет - will lead
создать - to formulate
явля - ется - is

ADJECTIVES

больш - large
броуновск - Brownian
выражающ - expressed
гидродинамическ - hydromatic
законн - legitimate
корреляционн - correlated
мал - small
марковск - Markov's
меньш - smaller
небольш - small
независим - independent
некоррелированн - uncorrelated
несправедлив - incorrect
неупорядоченн - random
остающ - remaining
перв - first
подобн - such
полн - complete
пригодн - applicable
применим - applicable
протекакщ - taking place
различн - various
рассматриваемым - observed
сделан - made
случайн - random
справедлив - correct
сравним - comparable
том - (1) that (2) 0
указанн - indicated
упорядоченн - correlated
физическ - physical

ADVERBS

более - a more
больше — more
всё-таки - nevertheless
достаточно - sufficiently
правильно - correctly
после - after
поэтому - therefore
соотвественно - accordingly
статистически - statistically
также - also
точнее - more precisely
учитывая - by taking into account

MINOR PARTS OF SPEECH

а - and
в - (l) in, (2) 0, *(3)* 0
даже - even
для - for
если - if
и - and
к - to
когда - when
лишь - only
между - between
не - not
но - but
о - (1) about, (2) 0
однако - however
по - (1) by, (2) 0
порядка - within
при - at
с(о) - with
также - also
то - then
что - (1) that, .(2) that
этому - 0

ABBREVIATIONS

ДР - others
см - see
т.е. - i.e.

PRONOUNS

её - its
она - it

RUSSIAN TEXT

В первых работах по теории броуновского движения /[1]/ (см. также /[22]/) значения скорости частицы в различные моменты времени предполагались по сути дела статистически независимыми. Соответственно этому была применима формула Эйнштейна

$$M\ (x - x_0)^2 = 2 \qquad (1)$$

а также уравнение Эйнштейна-Фоккера--Планка, справедливое для марковских процессов. В действительности, однако, корреляция между значениями скорости отсутствует лишь при достаточно больших интервалах времени между рассматриваемыми моментами. Поэтому формула (1) оказывается несправедливой для малых интервалов времени (порядка времени корреляции для скорости).

В целях создания более полной теории, пригодной для меньших интервалов времени, были сделаны предположения (Орнштейн, Крутков и др., см. также /[3]/) о том, что некоррелированной случайной функцией является не скорость, а ускорение, т.е. сила. Точнее, некоррелированной предполагалась неупорядоченная сила, остающаяся после вычитания гидродинамической силы, выражающейся по формуле Стокса. Если, учитывая гидродинамическое последействие, упорядоченную силу выражать формулой Еруссине, то предположение о некоррелированности неупорядоченной силы приведет, в частности, к результатам работы. Физическим основанием предположения о некоррелированности неупорядоченной силы является малость её времени корреляции по сравнению со временем релаксации скорости для больших броуновских частиц (большая частота ударов). Для небольших частиц, когда время корреляции сравнимо с временем релаксации, подобные теории не применимы. Но даже если указанное предположение законно и теория правильно описывает процессы, протекающие в промежутки времени порядка времени релаксации (и больше), то она всё-таки является не пригодной для рассмотрения приращений скорости в течение времен порядка времени корреляции неупорядоченной силы.

STANDARD TRANSLATION

In the first works on the theories of the Brownian movement (see also #2) the values of the velocity of a particle at various instants of time were actually assumed to be statistically independent. Accordingly, Einstein's formula $M(x-x_0)^2 = 2\ldots.(1)$ was applicable as well as the Einstein-Fokker-Plank equation, which holds true for Markov's processes. In reality, however, the correlation between the values of the velocity is absent only at sufficiently large intervals of time between the observed instants. Therefore, formula (1) proves to be incorrect for small intervals of time (of the order of magnitude of correlation time for the velocity).

In order to formulate a more complete theory which would be applicable for smaller intervals of time, assumptions were made (Ornstein, Kroutkou and others; see also #3) that the uncorrelated, random function is not the velocity, but the acceleration, i.e., the force. More precisely, it was assumed that the random force which remains after the subtraction of the hydrodynamic force, expressed by Stoke's formula, is uncorrelated. If by taking into account the hydrodynamic after-effect, the correlated force, is to be expressed by Bousett's formula, then the assumption of the uncorrelativity of the random force will lead, in particular, to the results of the work (perhaps he means to the satisfying results?). The physical reason of the assumption about the uncorrelativity of the random force, is the shortness of time of its correlation as compared to the relaxation time of the velocity of the large Brownian particles (high frequency of collisions). For the small particles, when the time of correlation approximates the relaxation time, such theories are not applicable. But even if the indicated assumption is legitimate and the theory correctly describes the process which takes place in the interval within the relaxation time (and longer), the theory still is not applicable for the observed increments of velocity during the periods within the time of correlation of the random force.

SIMULATED MECHANICAL TRANSLATION

In the first works on the theory of the Brownian movement (see also) the values of the velocity of the particle in the various moments of the time were assumed to be actually statistically independent. Accordingly, was applicable the formula of the Einstein and also the equation of the Einstein-Fokker-Plank, correct for the Markov's processes. In reality, however, the correlation between the values of the velocity is absent only at sufficiently large intervals of the time between the observed instants. Therefore, formula (1) proves to be incorrect for the small intervals of the time (within the time of the correlation for the velocity).

In order to create a more complete theory, applicable for the smaller intervals of the time, assumptions were made (Ornshtein, the Kroutkov, and others, see also) that the uncorrelated random function is not the velocity, and the acceleration, i.e., the force. More precisely, it was assumed that the random force, remaining after the subtraction of the hydrodynamic force, expressed by the formula of the Stokes is uncorrelated. If, by taking into account hydrodynamic after-effect, correlated force is to be expressed by the formula of the Boussinet, then the assumption about the random force will lead, in particular, to the results of the work. The physical reason of the assumption about the uncorrelativity of the random force is the shortness of its time of the correlation as compared with the time of the relaxation of the velocity for the large Brownian particles (large frequency of the collisions). For the small particles, when the time of the correlation is comparable with the time of the relaxation, such theories are not applicable. But even if the indicated assumption is legitimate and the theory correctly describes the process, taking place in the interval of the time within the time of the relaxation (and more), then it is, nevertheless, not applicable for the examination of the instants of the velocity during the period within the time of the correlation of the random force.

INSTRUCTIONS: 0 blank translation

("ending" means entire ending - not just final letter.)

1. Compare word with dictionary: If there is exact equivalence, translate. If there is multiple meaning, then this will be true for several consecutive words. In this case, choose the highest meaning common to all of the words. E.g., if there is a sequence of two words, the first having two meanings and the second three, then choose the second meaning for both.
2. If there is no exact equivalent, then remove as many letters from the end as is necessary to obtain a correspondence, and translate using the following rules. If there is no rule applicable to the ending, translate the word and ignore the ending.

RULES: The placement of "the". Place "the":

1. Before all nouns after a punctuation mark and before all adjectives when they begin a sentence.
2. Before nouns preceded by minor parts of speech and before adjectives also preceded by minor parts of speech except не .
3. After the verb, if the noun follows the verb or it is separated by one word.

Nouns preceded by adjectives:

1. If the adjective ending is ые , ых , их, и, then the noun is plural: otherwise sing.
2. If the word preceding the adjective is a noun, and if there is no punctuation mark between the first noun and the adjective, then place "of the" before the adjective.

Nouns preceded by pronouns:

1. Precede the pronoun by "of".

Nouns preceded by nouns:

1. If there is no punctuation mark between the nouns, then preface the second noun by "of the".

Continued on page 61

Mechanical Translation of French †

L. Brandwood, Birkbeck College Research Laboratory, University of London

IN THE NEAR FUTURE an attempt to translate from a foreign language by machine will be made at the computational laboratory of Birkbeck College. It will differ from previous experiments in that the sentences or passages of French to be translated will not have been specially chosen or "doctored" in any way beforehand: on the contrary, they will be constructed by French scholars, who will be invited to do their best to fault the machine. What follows is an account of the method, or program, which makes such an experiment possible.

First of all, however, I think a brief explanation of the general procedure in handling words by electronic computer will not be out of place. To a layman like myself the information that such a computer works in terms of electronic impulses conveys very little, but as I suspect that the majority of readers are in the same boat, I shall no doubt be forgiven for using terms which though strictly not correct have at least the advantage of being more readily intelligible.

We may say then that the computer works, like any other calculating machine, with digits. It cannot recognize letters; at any rate not as they stand. In order that it may recognize them the alphabet has first to be coded, that is to say, each letter represented by a particular number. Since a word is merely a collection of letters and we have a number for each letter, we can now put any words we choose into the computer to be stored on a magnetic drum in terms of impulses. These constitute the machine's "dictionary", which may be looked upon as exactly like an ordinary French-English dictionary, but with the numbers representing the French words arranged in order of either ascending or descending magnitude. This is important for the rapid identification of the words in the text being translated, which takes place as follows. The machine identifies an incoming word by comparing the number which represents it with the numbers contained in its dictionary. The comparison takes the form of a simple subtraction. Starting with a number at the middle of the dictionary it subtracts from this the number of the word being translated: if the result is zero, the word has been identified and the English equivalent, which is marked alongside the French can be output; if the result is not zero, then the number sought lies in one or other of the two halves of the thus divided dictionary according to whether the result is positive or negative. So, taking the number at the middle of the relevant half, the machine carries out the same comparison as before, and if -- as is more than likely -- this too proves unsuccessful, it continues halving the numbers that remain until the one required is located.

This ingenious process of elimination, which was devised by Dr. A.D. Booth, ensures the speedy identification of words even in a very large dictionary. In fact the size of the latter can be greatly increased without appreciably lengthening the look-up process. To be precise, the number of comparisons required is doubled as the number of words in the dictionary is squared. This means that since on an average seven comparisons are required to identify any word in a dictionary of a hundred, only fourteen will be necessary to cope with a dictionary of ten thousand or twenty-eight to cope with one of a hundred million.

The capabilities of the look-up method might seem rather wasted when it is revealed that the maximum storage capacity of the Birkbeck computer is at present only 250 dictionary items (i.e. 250 foreign words each with one English equivalent), though this will shortly be increased to 2000. It must be borne in mind, however, that this is not the limit of all computers, even of those existing, and storage capacities will continue to grow larger. But that lies in the future, and for the time being all our attempts at translation have to be scaled down and done, so to speak, in miniature.

The machine's capacity allowed us 250 dictionary items. We could not use all of these for the actual dictionary, however, because we wished to include a few elementary syntactical instructions to enable the machine to produce

† An earlier version of this paper appeared in Babel, Vol.11, No. 3, 1956, a special issue devoted to mechanical translation.

something a little better than a word-for-word translation. We did not know beforehand in what proportion the dictionary space should be divided between vocabulary and grammar, but reckoned that with a vocabulary of about a hundred words there would be enough space left for the instructions in mind and a little over for subsequent modifications. What was wanted therefore, was a vocabulary which, though very small, would yet be capable of producing good French sentences in a reasonable variety; in short, the proportion of the various parts of speech which it contained had to be correctly balanced. This was achieved in a very simple fashion by selecting a passage from a book and extracting the first hundred different words: to these were added any missing forms of the personal, possessive and relative pronouns encountered. The result was a vocabulary consisting of 25 verbs, 2 adverbs, 23 nouns, 10 adjectives, 3 numerals, 8 prepositions, 6 conjunctions, all forms of the unstressed personal pronoun (except en and y), all forms of the simple possessive pronoun, the relatives qui, que, dont, the definite and indefinite articles and the negative ne. . . .pas. Volunteers will be presented with this vocabulary and asked to construct sentences or short passages for the machine to translate. They will be permitted to use the singular and plural forms of all nouns and adjectives, the masculine and feminine of the adjectives, and all parts of the verb except the 1st person imperative and the subjunctive mood. The former is excluded because it requires a rather periphrastic translation, which combined with its comparative rarity makes it hardly worth bothering about, the latter because many of its forms are identical with those of the present and imperfect indicative, and could only be distinguished from these by taking into account the syntactical context. This, however, is out of the question with the present storage capacity. Nevertheless, disregarding the compound tenses, which are accounted for by the separate consideration of avoir or être and the perfect participle of each verb, there still remain about 30 different forms to each regular verb, the actual figures for the four conjugations being: donner 33, finir 29, vendre 32, recevoir 31. For an irregular verb, unless it is defective, the number is usually a little higher. In short, the 25 verbs (13 regular, 12 irregular) produce well over 800 different forms, all of which the machine must be able to translate. Obviously, such a large number will not fit into our machine's dictionary as they stand. How then are they to be incorporated? The solution thought of as long ago as 1947 by Booth and Richens, is that a word should be divided into a stem and an ending.

Here we must pause in order to define our terms. A stem is not necessarily a stem in the sense that grammarians know it. Generally speaking, it is the longest part of a word common to all forms (inflections) of that word. The stem of donner, for example, is donn-; what remains of each inflection of donner after this has been subtracted constitutes an ending. This, however, is not a universal rule, since there are instances where a single stem would account for all forms of the verb, but more than one is chosen for the sake of utility (e. g. the stem acqu- would cover all forms of acquérir, but in addition acquier- has to be used in order to comply with the general system covering all verbs). Utility might in fact be said to be the only principle of grammar in mechanical translation.

Verbs are divided, then, into stems and endings, and the following results cover each of the Four Regular Conjugations:

I. Donner (stem: donn-)

	Person					
	Singular			Plural		
Tense	1st	2nd	3rd	1st	2nd	3rd
Present	e	es	e	ons	ez	ent
Imperative		e	e		ez	ent
Future	rai	ras	ra	rons	rez	ront
Conditional	rais	rais	rait	rions	riez	raient
Imperfect	ais	ais	ait	ions	iez	aient
Past Hist.	ai	as	a	âmes	âtes	erent

Infinitive -er
Present Participle ant, ante, ants, antes
Perfect Participle é, ée, és, ées

II. Finir (stem: fin-)

	Person					
	Singular			Plural		
Tense	1st	2nd	3rd	1st	2nd	3rd
Present	is	is	it	issons	issez	issent
Imperative		is	isse		issez	issent
Future	irai	iras	ira	irons	irez	iront
Conditional	irais	irais	irait	irions	iriez	iraient
Imperfect	issais	issais	issait	issions	issiez	issaient
Past Hist.	is	is	it	îmes	îtes	irent

Infinitive -ir
Present Participle issant, issante, issants, issantes
Perfect Participle i, ie, is, ies

III. Recevoir (stem: rec-) †

	Person					
	Singular			Plural		
Tense	1st	2nd	3rd	1st	2nd	3rd
Present	ois	ois	oit	evons	evez	oivent
Imperative		ois	oive		evez	oivent
Future	evrai	evras	evra	evrons	evrez	evront
Conditional	evrais	evrais	evrait	evrions	evriez	evraient
Imperfect	evais	evais	evait	evions	eviez	evaient
Past. Hist.	us	us	ut	ûmes	ûtes	urent

Infinitive -evoir
Present Participle evant, evante, evants, evantes
Perfect Participle u, ue, us, ues

IV. Vendre (stem: vend-)

	Person					
	Singular			Plural		
Tense	1st	2nd	3rd	1st	2nd	3rd
Present	s	s	-	ons	ez	ent
Imperative		s	e		ez	ent
Future	rai	ras	ra	rons	rez	ront
Conditional	rais	rais	rait	rions	riez	raient
Imperfect	ais	ais	ait	ions	iez	aient
Past. Hist.	is	is	it	îmes	îtes	irent

Infinitive -re
Present Participle ant, ante, ants, antes
Perfect Participle u, ue, us, ues

† ç is treated as c.

If now the French verb stems and their English translations are entered into the machine's dictionary together with an instruction that each word fed in for translation is to be identified with the stem in the dictionary which it most closely resembles, we will be supplied with the basic meaning of every verb in the passage to be translated. The tense, however, is still lacking and for this we have to refer to the endings. These too need to be included in the dictionary, though at the same time they must be kept apart and distinct from the stems; otherwise confusions will arise (e.g. between the singular present indicative of avoir and the endings denoting the singular Past Historic of the 1st conjugation). Nor should they be included as they appear above for two reasons: Firstly the use of different inflections to indicate difference in person and number is on the whole a linguistic redundancy, so that where two identical endings occur we need in fact make only one entry in the dictionary: secondly, inasmuch as the endings are used merely to distinguish tense, there is no necessity to enter the whole ending but only sufficient to enable this distinction to be made (e.g. the -rai part of the ending -evrai in recevoir is enough to indicate that the tense is future). These distinctive endings are 46 in number, i.e.

Present	e	s	z	t	ois	oit		=stem	(is)	(it)
Future	rai	ras	ra	rons	rez	ront				
Conditional	rais	rait	rions	riez	raient					
Imperfect	ais	ait	ions	iez	aient					
Past Historic	ai	as	a	mes	tes	rent	is	it	us	ut
Perfect Part.	é	ée	és	ées	i	ie	(is)	ies	ue	ues
Infinitive	r	re						u	(us)	
Present Part.	ant									

One or two explanatory remarks are necessary here. The "=stem" means that there is no ending (e.g. il vend). The feminine and plural forms of the present participle are omitted because they belong strictly speaking to its adjectival function. The forms in parentheses, though characteristic of the tense under which they are marked, are also found in another tense of which it is preferable to regard them as primarily characteristic. This is because the limited capacity of the machine prevents the inclusion of instructions (assuming that they could be formulated) which would enable it to distinguish these formally identical tenses; consequently, for the time being one translation must be made to serve for both, and it is naturally desirable that this translation should be determined by the tense in which this disputed form has the wider occurrence. The ending -is, for example, represents the present indicative in finir type verbs, the past historic in finir, vendre and sentir type verbs. Therefore we take it as characteristic primarily of the latter tense; it is just a question again of utility.

Clearly, however, the fewer there are of these identical tense forms, the lesser the confusion will be. Hence the stem fin- instead of fini- for the finir type verb in the present scheme. If the stem had been fini- the following difficulties would have arisen:

1. Endings -s and -t
   1st, 2nd, 3rd pers. sing. pres. indic.
   =
   1st, 2nd, 3rd pers. sing. past historic
2. Ending -*s*
   1st, 2nd, pers. sing. pres. indic. and past historic
   =
   perf. particip. masc. plural
3. No ending: bare stem
   Sing. perf. particip. masc.
   =
   3rd pers. sing. pres. indic. of vendre type verbs
4. Ending -e and -es
   Perf. particip. fem. sing. and plural
   =
   1st and 2nd pers. sing. respectively pres. indic. donner type verbs

With the stem fin- on the other hand, only the first two of these are encountered. Fin-, therefore, is the best stem, but only, it must be repeated, in the present scheme. As the dictionary is enlarged and takes in more irregular verbs we may find that there are ambiguities which can be solved by making the stem fini- instead of fin- (e.g. the verb saisir will have a stem identical in form to the 1st and 2nd pers. sing. pres. indic. of savoir, unless the 'i' is added). The important thing to remember is that there can be no one optimum system for all translation programs: their general principles may very well be the same, but the details must always vary according to the size of the dictionary and the nature of its contents. This applies particularly to stems of verbs, which will continually have to be modified in order to avoid ambiguities (e. g. 1st, 2nd, 3rd pers. sing. pres. indic. of vivre = 1st, 2nd, 3rd pers. sing. past historic of voir, which in its 1st and 2nd persons may again be confused with the stem of the verb viser). Since correspondences of this type are far from being uncommon, it may prove advisable in a larger dictionary to revert to the conjugational classification each with its separate procedure, in order to reduce the confusion.

Returning to our explanation of the table of endings, the reason for using these particular forms to represent the present indicative is to be found in the rule previously mentioned, according to which the machine identifies any incoming word or part of a word with the nearest equivalent (by which is meant the nearest shorter equivalent) in its dictionary. Therefore, if we enter, for example, -s in the dictionary as a present indicative ending, the machine will identify with it not merely the bare -s of vendre and sentir type verbs, but also the -es of the 1st and the -ons of all conjugations, simply because there is no other entry in its dictionary with which it can identify them. It is for this reason that the 1st/2nd and 3rd pers. sing. pres. indic. of recevoir type verbs (-ois, -oit) has to be entered in full, since otherwise they would be identified not with the present forms -s and -t, but the past historic -is and -it.

The system described above covers not only the regular but also the irregular verbs in the majority of their forms. The only difference is that many of the latter require several stems instead of one; aller, for example, the following: vais, va-, vont, all-, i-. Of course, difficulties do arise, just as they do in the regular verbs, but their number is relatively small. There are five main types.

1. One or two verbs such as valoir and vouloir have a 1st and 2nd pers. sing. pres. indic. ending in -x, and one or two such as prendre and asseoir a 3rd pers. sing. in -d. This is solved by adding -x and -d to the list of endings denoting the present tense.

2. There are some verbs with a past historic form which resembles none of the other tenses (except in one or two instances the perfect participle), and consequently requires a stem peculiar to itself. One would think, since this stem is peculiar to a past tense, that its translation in the dictionary could be in the past tense without reference to the ending. For eus, the past historic of avoir, for example, the stem would be eu- and the translation "had" This, however, will not work. The singular with its "present" endings -s and -t comes out all right, but the plural endings -mes, -tes, and -rent are marked in our system as characteristic of the past historic, so that the translation of the plural will be the past tense of "had", which is not at all right. No matter whether we give the stem the equivalent "have" or "had", there will be an anomaly between the translation of singular and plural forms. The solution is to make the stem not eu-, but e-. The endings will now be found all in the past historic list instead of half in this and half in the present indicative list, and a completely correct translation will be ensured. Besides avoir this solution is applicable to only a few other irregular verbs (e. g. croire, croître, plaire) and not to all those to which it would be desirable to apply it (e.g. être, voir, savoir, mouvoir, pouvoir, boire, faire, lire, prendre, mettre). This is because any stem devised for the past historic of these verbs will be ignored by reason of the rule that incoming words must be identified with the longest similar stem in the dictionary. Let me make this clear with an illustration: for the past historic of voir, i.e. vis, vis, vit, vîmes, vîtes, virent, we can only set up a stem to cover vit (i. e. vi-) vîmes and vîtes (i.e. vî-); whatever we do about the other three they are bound to be confused with the stem of either the verb vis-er or the verb vir-er. In cases like this a possible solution is to enter the ambiguous forms in the dictionary as they stand together with an instruction that if they are found without an ending they belong to a different verb and have a different translation than if they are found with an ending. So when vis occurred alone, it would be recognized as belonging to voir, when it occurred with an ending, that is in a longer form, it would be ascribed to viser. The 3rd pers. virent would no longer be ambiguous, if

entered in entirety.

Other verbs with past historic forms which cannot be treated in the manner of avoir are tenir and venir (tins and vins) and those which, like finir in the regular verbs, have forms identical in the singular to those of the present tense (e.g. dire, rire, suffire, confire, circoncire, conclure). The answer for tenir and venir is again to enter the offending forms into the dictionary in their entirety; for dire, rire, etc., as for finir, only an examination by the machine of the context in which they occur will succeed. This, however, lies beyond the scope of our present modest program, which must perforce allow the ambiguity to remain unresolved.

3. Several verbs (e.g. dire, faire, écrire, conduire, craindre) have a perfect participle identical in form to the 3rd pers. sing. pres. indic. This causes no difficulty in the compound tenses, where the accompanying auxiliary serves to distinguish the participle as such, but when it is used adjectivally, then not only the masc. sing. but also the masc. plur. and the feminine forms will all be interpreted by the machine as present indicatives. A solution by consideration of form is possible only for the unique ending of the masc. plur. -ts (e.g. condui -ts), those of the others being common to other tenses, i.e. the masc. sing.(-t), as we have said, to the 3rd pers.sing. pres. indic., the fem. sing. (-te) to the 1st and 3rd pers. sing. pres. indic. of 1st conjugation verbs like jeter which double the 't' (i.e. jet-te), and the fem. plur. (-tes) to the 2nd pers. sing. pres. indic. of these verbs (i.e. jet-tes) or the 2nd pers. plur. past historic of all verbs. An extra stem (i.e. jett-) for the jeter type verbs helps a little, but a complete solution is possible only if the machine can take account of the syntactical context of these adjectival perfect participles.

4. Two of these verbs, faire and dire, present a slightly more complicated problem, since not only is the fem. plur. of their perfect participle identical in form with the 2nd pers. plur. of the pres. indic. (faites and dites), but this form itself has a past historic ending (-tes). This means that faites and dites have to be entered into the dictionary as they stand, which gets round the second part of the difficulty, leaving the first to be dealt with in the way suggested in no. 3 above.

5. a) The masc. sing. perfect participle of asseoir (assis) and acquérir (acquis) are identical with the 1st and 2nd pers. sing. past historic, whilst the fem. forms of the same have endings which are marked in the above scheme as primarily characteristic of the present indicative.

b) All perfect participles ending in -u or -i in the masc. sing. will end in -us and -is in the masc. plur. -forms characteristic of the past historic. The solution is again that mentioned in no. 3. In our present program, however, no difficulty is encountered over this similarity of perfect participle and past historic, since we are confining ourselves to indicating in translation only whether a verb is past, present, or future. -- As regards meaning no distinction is made between the various forms of the past tense.

These, then, are the main difficulties which have been encountered in drawing up a system for translating verbs. Some may have been overlooked, important ones at that -- I cannot be certain. Of one thing, however, I am quite certain; that is that there is not a single difficulty anywhere which is incapable of being surmounted.

In the method described above, first the meaning of a verb is obtained from the stem, then the tense is discovered from the ending. This is only one way. There is another, as the reader may already have observed for himself in perusing the table of endings, namely by a particular letter in the ending instead of by the whole ending. The most obvious example is the letter 'r', which is sufficient to indicate that the tense is either conditional or future indicative. In place of an ending dictionary, therefore, we can have a series of instructions like this:

1. If the ending ends in -r or -re, then the verb is an infinitive, and "to" should be inserted before the translation of the stem. The parentheses denote that the "to" will not always be needed to make good English in the translation, e.g., in puis-je sortir? (may I get out), whereas il va partir (he is going to leave) does require it. When larger storage capacities are available, it will be possible to incorporate instructions enabling the machine itself to make this distinction.

2. If the ending ends in -ant, we are dealing with the present participle and "-ing" should be added to the translation of the stem. Again it would be quite easy with a larger storage at one's disposal to include an instruction requiring the machine to deduct the mute -e from words ending so, before adding the "-ing".

3. If the ending contains the letter r and this is not followed by -ent (in which case the tense is the past historic), then the verb in question is either future or conditional. If now the letter i occurs after (but not necessarily immediately after) this r and is not the last letter of the ending, the verb is conditional, and the word "would" must be inserted in front of the stem meaning. Otherwise it is the future tense, and the word to be inserted is "will". (The English idiom of "shall" with the 1st person and the interchange of this and "will" with a change of emphasis could likewise be included in a machine with a larger dictionary space).

4. If the ending begins with e (not followed by a), s (not followed by a or i), t (not followed by anything), x, o, n, l, d, or if there is no ending (i. e. the stem is bare) and avoir or être does not precede, then the verb is in the present indicative, and the stem meaning is an adequate translation. (The same can be said of the suffixed -s in the 3rd pers. sing. of the English present as of the other refinements mentioned above.) If there is no ending and avoir or être does precede, the verb is in the perfect tense (see no. 5).

The reason why e, s, x, o, d, t indicate the present tense will be clear. The reasons why n and l do, and why e should not be followed by a, s, by a or i, or t by any letter are as follows:

n- enables certain irregular verbs which double the final consonant of their stem in the 3rd pers. plur. (e. g. venir - viennent) to be comprehended under one stem (i. e. vien-) instead of two.

l- does the same for 1st conjugation verbs in -eler (e.g. atteler - attelle).

t is not followed by another letter since -tes is an ending characteristic of the past historic. (Note this means that 1st conjugation verbs in -eter [e. g. jeter - jette] need two stems).

e is not followed by a or â because 1st conjugation verbs in -ger (e. g. manger) retain the e before a and o, so that, but for this precaution, the imperfect and past historic of these verbs would be translated as the present tense.

s is not followed by a or i, as happens in the imperfect of finir type verbs, if the stem fini- is chosen instead of fin- (i. e. -ssais, -ssait, -ssions, -ssiez, -ssaient.)

(Note: with the present method the stem fini- is preferable to fin- for this reason: if the i were not included in the stem, it would come at the beginning of all forms of the present indicative, which would consequently be translated as a past historic. As it is, only the singular forms of the past historic are confused with those of the present indicative.)

5. Everything not accounted for by the previous instructions is in the past tense, and "-(e)d" should be added to the translation of the stem. (Note: this will result in some odd looking forms in the case of the English strong verbs, but as we have said so often before, it is only the restricted storage capacity of the present machine which forbids the incorporation of these strong forms (e.g. "sang, sung") alongside the regular form ("sing") together with instructions for their correct application.)

These are two ways of ascertaining the tense of a verb, which we may call the stem-ending and the eliminating method respectively. The question now arises of which is the better, but this is not easily answered, depending as it does to a great extent on the machine, the size of the dictionary employed, the nature of the text to be translated and the quality of translation required. The eliminating method, for instance, must assume that the person receiving the translation will be content to have all past tenses translated simply as such and prepared to use his own common sense and judgement to distinguish where the sense requires a perfect (i.e. has done), imperfect (i.e. was doing), or past historic (i.e. did). If the more precise translation is desired, then there is no doubt that the stem-ending method is to be preferred.

As regards the nature of the text, if the tense in which it is written is the past, the stem-ending method will probably be quicker than the eliminating method, which arrives at this tense only at its last stage; if, however, the tense is the present and future, then it may well be that the eliminating method will prove quicker.

All this smacks too much of "perhaps" and "maybe", however, as would any remarks which we might hazard on the amount of space which each method would occupy in the machine's dictionary. A definite answer to all such questions will, after all, be forthcoming when the machine starts on its program, since this will be tried with both methods. In the meantime two facts are quite certain:

a) the eliminating method is superior to the stem-ending method when it comes to translating perfect participles, because it needs to consider only the first letter of the ending, which is the one distinctive of the perfect (i.e. é, i, u) whereas the stem-ending method has to have all the forms (fem. and plur.) written down for it

to choose from.
b) the stem-ending method is more economical for treating the irregular verbs, because the fact that it can ignore the middle part of some such verbs enables it to account for them with one stem less in each case than the eliminating method. The verb asseoir, for example, requires three stems in the latter method, i.e. assied- (for the singular of the pres. indic., past historic and conditional), assey- (for the imperf. and plural of the pres. indic.), and ass- (for future, past historic and conditional), but only two in the former method which can ignore the -ied- in the sing. pres. indic. and so use ass- for this too. Out of about seventy irregular verbs examined, this economizing applied to seventeen, (i. e. acquérir, boullir, tenir, sentir, sortir, partir, servir, asseoir, savoir, pouvoir, écrire, joindre, craindre, mettre, battre, suivre). To these must be added their compounds and any similarly conjugated verbs, though some of the latter are misleading because a verb with a similar stem prevents the reduction of the number of stems. For instance, the only stem that would be needed for dormir would be dor-, but for the existence of a verb dorer; so the stem has to be dorm-, as for the eliminating method, and the sing. forms of the pres. indic. (dors, dort) have to be entered in the dictionary as they stand.

That completes our treatment of verbs, and we can pass on now to explain the program for dealing with the other parts of speech. Nouns and adjectives are taken together, since both may have feminine and plural forms, and these can be satisfactorily accounted for by the same stem-ending procedure, thus:

Type 1. If the feminine form is created merely by adding extra letters to the masculine form without involving any change in the latter, then the masculine form (so marked) is entered in the stem dictionary, the extra letters denoting the feminine form (so marked) in the ending dictionary, (e.g. chien -ne, joli -e).

Type 2. If the feminine form is created by first altering the masculine form, then adding to it, (e.g. chameau, chamelle), then the letters which are common to both forms (i.e. chame-) are entered in the stem dictionary, those peculiar to the masculine and feminine in the ending dictionary (i.e. -au, -lle). The gender of a noun which has only one form for both masculine and feminine forms must be determined by other means (e.g. by reference to the definite or indefinite article or an adjective, if it is accompanied by any of these).

The plural form of nouns and adjectives, both masculine and feminine, is easily recognized--as long as there is in fact a distinct plural form to be recognized--inasmuch as the ending will have either -x or -s added to it. Apart from irregulars the only class of nouns and adjectives which does not comply with this "rule" is that ending in -al, which drops the 1 before adding -ux (e.g. cheval, chevaux). This is resolved in the same way as the Type 2 feminine form above, that is, by putting the letters common to both singular and plural (in this case cheva-) in the stem dictionary, the letter 1 which is dropped in the ending dictionary.
So we have the following three types of regular nouns and adjectives:

		Stem	Ending	Plural
Type 1	noun	chien (m)	-ne (f)	-s
	adj.	joli (m)	-e (f)	-s
Type 2	noun	chame- (	-au (m)	-x
			-lle (f)	-s
	adj.	bre-	-f (m)	-s
			-ve (f)	-s
Type 3	noun	cheva-	-1 (m)	-ux
	adj.	principa-	-1 (m)	-ux
			-le (f)	-s

In short, we can identify all plurals which are genuine forms and not identical with the singular by the three suffixes -s, -x, -ux. The only difficulty is in getting the machine to recognize where these suffixes occur after the ending and where directly after the stem. The latter is true of the masculine forms of Type 1 nouns and adjectives and masculine adjectives and all nouns of Type 3; the former of Type 2 nouns and adjectives and Type 3 feminine adjectives. It can be done quite easily if the machine proceeds according to the following instructions, which do not even require it to distinguish between the three suffixes:

1. If when the machine looks in its stem dictionary,
a) the incoming noun or adjective can be identified exactly with one of the stems, then it is a Type 1 masculine noun/adjective and can be translated immediately.
b) the incoming noun or adjective cannot be identified exactly with a stem, because an extra letter or letters is left over, it cannot be translated immediately;
2. then the machine looks in its ending dictionary.

a) If the extra letter(s) can now be identified exactly with any ending contained in the dictionary, then the incoming word is either a feminine singular noun/adjective of Type I, a masculine or feminine singular noun/adjective of Type II, or a masculine or feminine singular noun/adjective of Type III, and can be translated accordingly (see below);
b) the extra letter(s) can be identified with an ending contained in the dictionary, but if there is still an extra letter left over, then the incoming noun/adjective is the plural of the stem + ending word and its gender is that marked after the ending. The plural is denoted by adding -s to the English translation.
c) the extra letter(s) cannot be identified with an ending contained in the dictionary, then the incoming noun/adjective is the plural of the stem word and its gender is that marked after the stem. (Note: the identification of number in adjectives is, of course, irrelevant for their actual translation, but as was noted above, it can be useful for determining the gender and number of nouns with identical forms for the masculine and feminine or the singular and plural.)

This scheme will provide rapid identification of all regular adjectives and nouns, but it has the fault of being somewhat uneconomical as regards storage space, since the ending dictionary needs to be of the same size as the stem dictionary, or even larger, most stems having one and often two endings. A possible modification offers itself, however, when we realize that the number of different endings for regular adjectives and nouns is only 15 (5 masculine and 10 feminine) i.e. masculine: -eur, -au, -l, -f, -x; feminine: -euse, -rice, -lle, -ne, -te, -sse, -e, -ve, -se. It would seem feasible, therefore, to have an ending dictionary consisting of only these 14 endings which would be capable of producing a correct translation of all regular French adjectives and nouns. And, in fact, it would be, but for English often having a completely different feminine form of a noun where French has only a partially different form (e.g. fils, fille -- son, daughter), or an irregular plural where French has a regular one (e. g. homme, -s -- man, men). A reasonable solution is to divide all nouns and adjectives into two classes:

I. those which have a separate word for masculine and feminine, e.g. cheval (horse) - jument (mare), those which are without gender in English, e.g. le village (village), abstract nouns -- all in fact, which do not require a distinction of gender in translation, (Consequently all adjectives can be included in this class.),

II. those which have a feminine or plural form that does require a different translation from the masculine or singular form.

The first class, in which the ending is required only for identification not translation, can be dealt with in the manner suggested above, that is by having an ending dictionary for the whole class of only 14 distinct endings (or 28, if we wish to include the plurals and so get immediate recognition of these, rather than provide the machine with instructions). This will retain the speed of identification while at the same time improving the economy.

For the second class the unmodified system will be retained, and each noun will be accompanied by its feminine and plural endings, alongside each of which will be the appropriate translation where different from that of the masculine singular. When the feminine form is merely a lengthened form of the masculine, the two translations will appear with the stem and ending respectively, for example, stem: chien = dog, ending -ne = bitch (dog). When it involves a change in the masculine form, the two translations will accompany the two endings, thus: stem: act- (no translation), endings: -eur = actor, -rice = actress.

In closing the section on nouns and adjectives the observation made previously with verbs must be repeated, namely that if the stem-ending method anywhere involves a confusion of two words with different meanings, the stems chosen for these words should be altered, where possible, to create a distinction. Where it is not possible, the offending forms will have to be regarded as irregular and entered in the dictionary in full.

The only other part of speech in the program which requires separate instructions for its identification is the pronoun, to be more explicit the personal (unstressed) and possessive pronouns. Four difficulties arise, namely those of distinguishing the accusative of the 3rd personal pronoun (le, la, l', les) from the definite article, the dative plural of the 3rd personal pronoun (leur) from the identical possessive form, and the nominative nous and vous from the accusative/dative, and lastly that of removing the oblique forms of the personal pronoun from their French position before the verb to their normal English position after the verb. The first two difficulties are surmounted by assuming that whenever le, la, l', les are followed by a pronoun or verb, they are forms of the personal pronoun not the definite article, and similarly that whenever leur is followed by a verb it is the personal pronoun. The last two are also solved simultaneously in the following way:

Type 1. the sequence is
pronoun^1 + pronoun^2 + pronoun^3 + verb
if nous or vous is the first pronoun, then it is nominative: translate in order -
pronoun^1 + verb + pronoun^2 + pronoun^3
(e.g. nous le leur donnons = we give it them)

Type 2. the sequence is
pronoun^1 + pronoun^2 + verb
If
a) pronoun^1 is a nominative form (je, tu, il, elle, ils, elles), pronoun^2 is oblique. Therefore translate in order: pronoun^1 + verb + pronoun^2 (e. g. je le donne = I give it)
b) pronoun^1 is oblique, pronoun^2 is also oblique. Translate in order: verb + pronoun^1 + pronoun^2 (e.g. le berger le lui donne = the shepherd gives it him)

c) pronoun^1 is a form which is not recognizable per se as nominative or oblique (i.e. nous or vous), then the verb must be examined to see whether it agrees with nous or vous. Thus: 1) if the verb ends in -ons or -mes, nous- is nominative
2) if the verb ends in -z or -tes, vous is nominative
Then translate in order given in a). Otherwise nous and vous are oblique and the translation order is that of b).

Type 3. the sequence is - pronoun + verb. The procedure is exactly the same for as Type 2.

This, then, is the program which will be used for translating from French mechanically. The objection might well now be raised that it is devoted only to the grammatical side, the parsing of words, and gives no attention to syntax. Apart from the fact that an instruction will, as it happens, be included to reverse in translation the order of adjectives following their nouns in French, this objection is quite valid. In defense we must make the plea made so often already that the restrictions imposed by the present limited storage capacity of the computer render any comprehensible scheme for dealing with syntactical problems academic. When the storage capacity is enlarged, as it soon will be, there will be the possibility of making use of such a scheme.

Nevertheless, it is probably true to say that, since its word order is more or less identical with that of English, the main difficulty in translating French lies precisely in what we have endeavored to achieve in the program described, namely the successful identification of the various forms of words. This being so, the program will prove capable of producing an adequate translation of such French prose as has no literary pretensions — for example, scientific publications, for which indeed it is primarily intended. Whether this belief is justified or not will be decided by the actual results of the experiment. These will be published.

KOUTSOUDAS *from page 51*

Nouns preceded by punctuation:

1. If the noun ends in я , then hold translation until the verb is translated. If the verb is plural, then the noun is plural, otherwise the noun is singular.

Nouns preceded by verbs:

1. If the word preceding the verb is not a noun, then invert the verb - noun word order.

Verbs preceded by nouns:

1. If the noun ends in у , then replace the "to" associated with the verb by "is to be".

Adjectives:

1. If the ending is ы , then precede the adjective by "are".

2. If the ending is о , then precede the adjective by "is".

Verbs preceded by adjectives:

1. Preface the adjective by "is" and place at the end of the sentence; enclose the verb in "it --- that".

Bibliography

Victor H. Yngve 71
The Outlook for Mechanical Translation
Le Linguiste. No.4, (1956), pp.5-7.

Discusses the direction of current MT research and its possible effect on the translation profession. In the future, the abundance of rough mechanical translations of technical articles may call attention to many that are of merit and thereby increase the demand for accurate translations by competent translators.

Author

L. Brandwood 72
The Translation of a Foreign Language by Machine
Babel. Vol.2, No. 3, October 1956, pp. 111-118.

See Brandwood, Mechanical Translation of French, in this issue of MT.

Edmond Cary 73
Mécanismes et Traduction.
Babel, Vol. 2, No. 3, October 1956, pp.102-107.

A philosophical essay about the effect of mechanical translation on people. There is some discussion of the development of special terminology in various fields and the effect that it has had on modes of thought. One of the chief benefits of mechanical translation, according to the author, will be its contribution to our understanding of the nature of language and thought.

J.R. Applegate

M.V. Heberden 74
Teaching the Machine Grammar
Babel, Vol.2, No. 3, October 1956, pp. 119-124.

A summary of the work of the Italian Operational School based on Silvio Ceccato's article "La grammatica insegnata alle macchine". (See abstract 70, MT, Vol.3, No. 1.)

J.R. Applegate

Andrew D. Booth 75
Present Objectives of MT Research in the United Kingdom
Babel. Vol.2, No. 3, October 1956, pp. 108-110.

A presentation of the principles which form a basis for research being done in England. A discussion of the problems encountered in developing machine techniques for the translation of French to English is given. The problems include: reduction of look-up time for dictionary items, inflected forms, idioms and rearrangement.

J.R. Applegate

Victor H. Yngve 76
The Outlook for Mechanical Translation
Babel, Vol. 2, No. 3, October 1956, pp.99-101.

See abstract 71 in this issue.

L. Brandwood 77
Previous Experiments in Mechanical Translation
Babel, Vol.2, No. 3, October 1956, pp. 125-127.

A comparison of three attempts at mechanical translation (Georgetown, Birkbeck College and the USSR Academy of Sciences) is given. The author does not attempt to evaluate the experiments because of the inadequacy of the reports on which the article is based.

J.R. Applegate

V.H. Yngve 78
Terminology in the Light of Research on Mechanical Translation
Babel, Vol.2, No. 3, October 1956, pp. 128-132.

Discusses how research aimed at resolving multiple-meaning problems in mechanical translation might help us to understand how words with several meanings can be used unambiguously. Points out six kinds of contextual clues that a reader uses in resolving ambiguity.

Author

D. Panov 79
Concerning the Problem of Machine Translation of Languages
Academy of Sciences, USSR, (1956), 35 pages.

This is a report of research being done at the USSR Academy of Sciences in the field of mechanical translation. The basic assumption that mechanical translation requires more than determining the meanings of words in the source language leads to a division of the translation program into parts: (1) analysis of the source-language sentence to determine the meaning and grammatical characteristics of each word in it and (2) synthesis of the target-language sentence according to previously formulated rules that are applied to the information obtained from the analysis of the source-language sentence.

J. R. Applegate

D.Y. Panov 80
Automatic Translation
Academy of Sciences of the USSR, Moscow, (1956), 45 pages. (In Russian).

A description of the methods used in programming the BESM computer to translate from English to Russian. It is considerably more detailed than the Mukhin article. (See abstract 69, MT, Vol.3, No.l.)

V.H. Yngve

Paul Garvin 81
Some Linguistic Problems in Machine Translation
For Roman Jakobson, Mouton and Co., (1956), The Hague, pp.180-186.

The steps in translation can be discovered by a reconstruction of the translation process from a comparison of the original text with its translation. The results of this translation analysis must be stated explicitly and unequivocally, with each logical step in terms of the yes-no decision required by the binary operation of electronic circuits.

V.H. Yngve

Lew R. Micklesen 82
Form Classes: Structural Linguistics And Mechanical Translation
For Roman Jakobson, Mouton and Co., (1956), The Hague, pp.344-352.

Discusses the differences between the form classes of structural linguistics and MT. In structural linguistics the initial division of the word material into the most inclusive word classes (parts of speech) and the possible subsequent subdivision of these larger classes into subclasses are made on the basis of the structural or grammatical meanings signalled in each case. In MT the initial division of the word material is made on the basis of the grammatical and non-grammatical meanings of each word. The criterion of potential mutual pinpointing alone supplies the basis for the determination of both intended grammatical and non-grammatical meaning; therefore, as far as possible, no MT form class should contain both pinpointers and pinpointees liable to occur as IC's of the same pinpointing problem. Further division of MT classes is governed by engineering considerations that essentially render more frequently occurring words more accessible.

In comparing structural linguistic and MT form classes the factor of non-grammatical meaning has, of course, also to be considered. If we ignore the modifications necessary because of engineering expediency, it would appear that MT classes are not only more inclusive because they are based on a consideration of the total meaning of linguistic forms but also more realistic in view of the fundamental semantic and formal nature of every linguistic sign.

Author

J.R. Firth 83
Linguistic Analysis and Translation
For Roman Jakobson, Mouton and Co., (1956), The Hague, pp.133-139.

A plea for building bridges between languages on the basis of linguistic analysis rather than on the basis of "naked ideas" as an interlingua or on the basis of some universal grammar.

V.H. Yngve

Robert E. Wall, Jr. 84
Engineering Process in Machine Translation
The Trend. Vol.8, No. 3, July 1956.

A review of recent developments in computer machines, coding, and memory devices as they pertain to or can be adapted to MT with examples taken from Russian and French. Some attention is also given to the cost of translating machines.

G.H. Matthews

Lew R. Micklesen 85
A Machine Translates from one Language to Another
The Trend, Vol.8, No. 3, July 1956.

A translation into English of a newspaper report on the English translation program for the BESM at the Academy of Sciences of the USSR containing a short description of the coding, dictionary, some logical operations, and the output of a given English text.

G.H. Matthews

Geoff Douthwaite 86
The UW Automatic Language Translator
Washington Engineer, Vol. 8, No. 4 (Feb. 1956)

Brief sketch, with general schematic, of a demonstration "translating" unit using "time delay sequential relay circuits", a drum containing a 35-relay German vocabulary, a corresponding English drum, and teletype input and output, which will transfer selected German texts with no syntactic operations into a word-for-word English equivalent, printing "internationals" in their unaltered German form, but searching its memory for vocabulary item in hand by successively stripping off the ending a letter at a time. Author emphasizes contrast between MT and computer design and necessity for syntax program.

R. B. Lees

Victor H. Yngve 87
The Translation of Languages by Machine
Information Theory, (Third London Symposium), Butterworth's Scientific Publications, London, pp.195-205.

See abstract 62, MT, Vol.2, No. 3.

Yehoshua Bar-Hillel 88
Can Translation be Mechanized?
Mada'. Vol.1, No. 2, April 1956, (In Hebrew).

Any operation that can be decomposed into a sequence of definite basic operations, even if these number in the millions, falls within the range of electronic computers. It is clear that the operation of translation can be broken down in this way, at least in part -- especially with respect to scientific texts. But the hitherto accepted methods require a rapid access mechanical "memory" with storage capacity greatly in excess of that of available electronic computers. The central problem of mechanizing translation, therefore, is the preparation of methods that permit a more restricted memory. Several such methods are described here.

Author's abstract translated by
A. N. Chomsky

Enrico Maretti 89
Adamo II
Civiltà delle macchine, No. 3, (1956), pp. 25-32.

This article is primarily a description of the mechanical model of mental processes constructed by the author and exhibited at the International Exposition of Automation at Milan in April 1956. Much of the article is devoted to a discussion of the Italian Operational School and the work of Silvio Ceccato. [See abstract No.70 in MT, Vol. 3, No. 1.] There is also some discussion of neurological theory which was used in planning the model. A note by Ceccato on the significance of the model is included.

J. R. Applegate

V.H. Yngve 90
The Technical Feasibility of Translating Languages by Machine
Electrical Engineering, Vol.75, No. 11, (1956), pp.994-999.

This is a general discussion of the problems involved in mechanical translation. The machine requirements — size, speed, economy— for word-for-word translations are considered. In addition, the linguistic problems arising in mechanical translation on a sentence-for-sentence basis are introduced, and possible avenues of solutions for these problems are presented.

Author

R.E. Wall, Jr. 91
Some of the Engineering Aspects of the Machine Translation of Language
Communication and Electronics (AIEE), No. 27, November 1956, pp. 580-585.

Suggests various, presently feasible design characteristics for a Russian-to-English translator, yielding better than 10 words/sec. to compete economically with human translator: internal machine code for Cyrillic employing both 5 and 10 digit codes for the 80 distinct characters; omission of "internationals" from dictionary (about 10% of running scientific text); use of only end-of-word letter stripping from Reifler's compound-separation scheme to avoid skipping to new addresses in memory; use of alphabetic listing in dictionary, requiring one unique address per spelling; use of King, Brown, and Ridenour's optical storage on photographic plates; indexing scheme to locate approximate row address before scanning individual 6500-bit rows; separate storage of about 1200 most frequent words if dictionary search is longest (limiting) time factor, and possibly a similar separate idiom dictionary; and Bliss and Ruedy's cathode-ray screen high-speed teleprinter output. Design yields about 42 1/2 megabits of storage with 0. 065 sec/word access time. Author asserts need for further immediate research on linguistic problems attending use of grammatical "tags" in dictionary.

R.B. Lees

A. Koutsoudas and R. Korfhage 92
The Computer as a Translator
Michigan Alumnus, Vol.63, No. 10, (Dec. 1956), pp.34-37.

Gives some of the history of Mechanical Translation and of the University of Michigan project. Indicates that most of the Russian endings are inconsequential in translation and that the problem of word order is a minor one. Words should be entered in the dictionary alphabetically by the first two or three letters and then in order of frequency. Endings should be discarded in most cases, and a few rules should be used to deal mainly with word order. Multiple meaning problems can be approached with a context of from two to six words. The whole meaning of an idiom is attached to a central word, with no translation of the other words when they occur in an idiom.

V.H. Yngve

R.A. Crossland 93
Graphic Linguistics and its Terminology
MT, Vol.3, No. 1, pp. 8-11.

This paper represents a proposal for developing a procedure and appropriate terminology to be used in the analysis of written texts analogous to the procedures and terminology used in the analysis of spoken languages. The author proposes scientific analysis of written texts without reference to the spoken form of the language. He also suggests that the written and spoken forms of a language represent different realizations of concepts and should, therefore, be treated separately.

J. R. Apple gate

A.F. Parker-Rhodes 94
An Electronic Computer Program for Translating Chinese into English
MT, Vol.3, No.l, pp. 14-19.

This is a discussion of the problems that arise in preparing a computer program for translation. Although the language pair considered is English and Chinese, general problems, symbols, dictionary ordering, etc., are considered The contrast between the "lexical" and "algorithmic" approaches is explored. The first requires listing all forms of the source language with target-language equivalents. The second means prescribing rules which permit discovery of word orderings represented by numerical symbols with rules for subsequent conversion of these symbols to significant forms in the target language.

J. R. Applegate

R.H. Richens 95
Preprogramming for Mechanical Translation
MT, Vol.3, No.l, pp. 20-25.

In this paper the author gives a discussion of the steps required before a program for mechanical translation can be prepared. Translation is defined as a type of communication involving a change of symbols between utterance and reception. The operations of transferring meaning, structure and ambiguity as well as the insertion of necessary information and prevention of excessive semantic analysis are discussed.

J. R. Applegate