[From: *Scientific American*, vol.194, no.1, January 1956]

Translation by Machine

Its wide study has been stimulated by the need of scientists to keep abreast of publications in several languages. Although a mechanical translator still does not exist, encouraging progress has been made

by William N. Locke

Suppose you became interested in working in a new field opening up in your line of work. Your first step would be to get all the background you could on the subject. To take a concrete example, let us say that the new field was the design of electrical switching networks. Looking through the literature, you would certainly find the pioneer 1938 paper by Claude Shannon on the theory of such networks, and a number of other, less important, papers. But how likely would you be to discover a Russian paper entitled приложеие матричой булевской алгебры к анализу и синтезу релейно-контакных схем? And even if you saw listed somewhere an English translation of its title ("The Application of Boolean Matrix Algebra to the Analysis and Synthesis of Relay Contact Networks"), how could you know that this article in the Russian language was the most important contribution to the field next to Shannon's original paper?

The question is not an idle one. Groups of people in several companies in the U. S. did in fact work for five frustrating years on the very points cleared up by this paper before discovering it. The article, by A. G. Lunts, was published in the journal of the U. S. S. R. Academy of Sciences in 1950. Even though this journal is available in the U. S., the article that would have saved so much time and work was overlooked until 1955, simply because most U. S. scientists and engineers cannot read Russian. Considering the time put in on the problems in question by a number of first-rate people, we can estimate that ignorance of the article cost the companies involved easily $200,000, not to speak of the five-year delay in certain switching-circuit developments.

What I have just cited is but one small example of the great cost to mankind of the language barrier—just in the fields of science and technology. The Russian example is not an exceptional case. Even in German and French, which theoretically a great many Americans can read, how many important papers await discovery, how many basic ideas have never been translated or recognized in this country?

Only 50 per cent of the world's scientific papers are published in English. More and more technical material is being published in more and more languages other than English. How are we going to get access even to just the high lights of this material? Translation is expensive—about $6 per page on the average. And good translators are not plentiful. Add to that the fact that a translator of scientific material must first of all know the subject he is translating; in order to translate papers in physics, for example, you practically have to be a physicist. Finally, even if we had plenty of expert translators, they would have an extremely difficult time choosing the material worth translating. The head of a government laboratory's translation section put it succinctly: "Our problem is to know what to translate."

There is the picture. What is the solution? We are practically driven to the answer that always suggests itself when we are faced with a need for mass production: Machines. To translate languages by machine is a little less easy than falling off a log, but the need is so great that in less than a decade since it was first seriously suggested many groups of people have gone to work on the problem.

In 1946 Warren Weaver of the Rockefeller Foundation read a sentence in an English physicist's report suggesting that computing machines might be adapted to translate languages. Weaver was so intrigued that he went to see the paper's author, A. D. Booth, and followed up with a letter to Norbert Wiener at the Massachusetts Institute of Technology. Weaver, having had some experience in deciphering codes, reasoned that languages are codes and should be capable of being decoded by a machine. Wiener's reply was disappointing: "I frankly am afraid that boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any quasi-mechanical translation scheme very hopeful."

Weaver was not discouraged. In 1949 he circulated to some 200 of his friends a memorandum, entitled "Translation," which directly inspired most of the work that has since been done in this country on machine translation. Projects to investigate the possibility were started, with Rockefeller Foundation support, in three universities. At the University of Washington Erwin Reifler looked into the basic semantic equivalents of languages. At the University of California at Los Angeles Victor A. Oswald and Stuart L. Fletcher, Jr., analyzed German syntax and in 1951 published the first paper devoted to machine translation: "Proposals for the Mechanical Resolution of German Syntax Patterns." At M.I.T. Yehoshua Bar-Hillel began an attempt to identify the universal grammar elements in various languages and also gave some thought to translating idioms.

Meanwhile Booth, collaborating with D. H. V. Britten at the Institute for Advanced Study in Princeton and later with R. H. Richens (a plant geneticist and linguist) in London, was working on a scheme of dictionary, or word-for-word, translation by a computer. Richens suggested that case and tense endings of words should be considered separately. Suppose, for example, that the word to be translated was *heiss.* This is the German word for "hot," but it is also the stem, and imperative singular, of the verb *heissen,* meaning variously "to call, to command, to be called, to mean." The computer would deliver the various meanings stored in its memory for *heiss,* and the reader of the output would choose the meaning that made most sense in the context. If the word to be translated was *heissen,* the machine would also give all the possible meanings of the ending *-en,* and the reader would have to select the one that made the most sense.

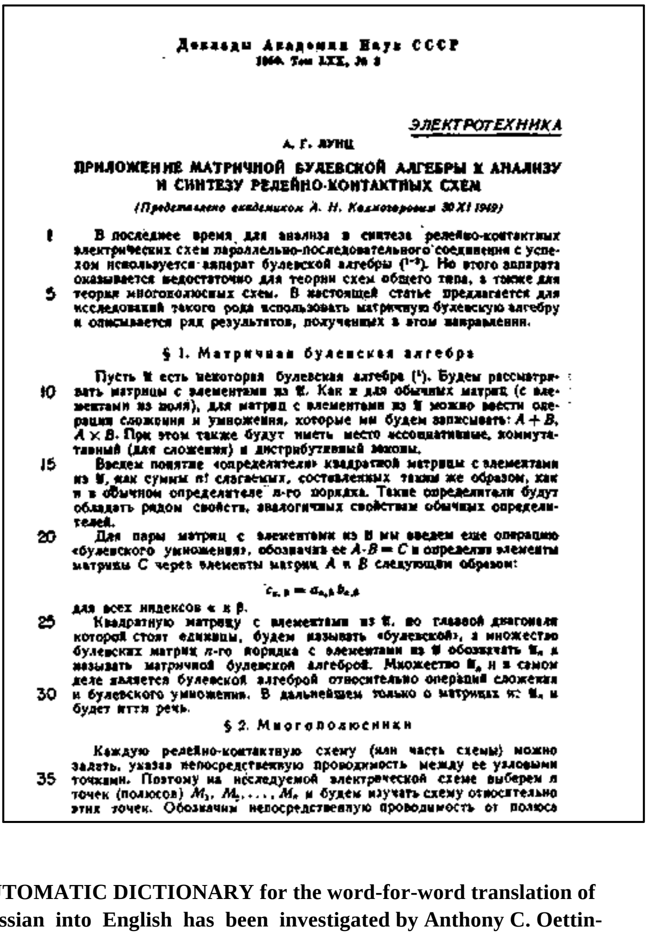
The multiplicity of possible meanings is an obvious weakness of any word-by-word translation system. Still it is only fair to mention that the *-en* ending is one of the most versatile in the German language. If we had taken *heisst,* the *-t* would have been much easier to handle. It has been estimated that German stems have an average of about one and a half meanings each.

In his famous memorandum Weaver put forth the suggestion that a machine might select the correct meaning of a word by taking into account one or more words on each side of it. Examining this proposal, Abraham Kaplan of the Rand Corporation later found that maximum information about the meaning of a word comes from the first two words on either side of it.

By 1952 so many people were interested in machine translation that Bar-Hillel organized a conference on the subject at M.I.T., financed by the Rockefeller Foundation. The Conference itself proved to be a study in the difficulties of communication. Gradually, however, the specialists in different fields—computer engineers, linguists, logicians and mathematicians—learned one another's language and came to a realistic view of the problems to be solved. Few had realized the costs involved. I remember the stunned silence that followed the statement that a computer such as was needed for translation might rent for around $30,000 a month.

After three days everyone felt that further research was certainly worthwhile and that limited objectives could be accomplished. It looked as though the best approach would be to start with the automatic dictionary idea, translating text a word at a time. Such translation would be crude in the extreme, but many scientists believe it would be intelligible to specialists in the field of the article. This is, of course, just the way most human beginners go about translating.

Since the 1952 conference a journal, *Mechanical Translation,* has been founded at M.I.T. by Victor H. Yngve and the author of this article, and Booth and I have edited a book, *Machine Translation*



Доклады Академии Наук СССР
1950. Том LXX, № 3

ЭЛЕКТРОТЕХНИКА

А. Г. ЛУНЦ

ПРИЛОЖЕНИЕ МАТРИЧНОЙ БУЛЕВСКОЙ АЛГЕБРЫ К АНАЛИЗУ И СИНТЕЗУ РЕЛЕЙНО-КОНТАКТНЫХ СХЕМ

(Представлено академиком А. Н. Колмогоровым 30 XI 1949)

В последнее время для анализа и синтеза релейно-контактных электрических схем параллельно-последовательного соединения с успехом используется аппарат булевской алгебры (1-3). Но этого аппарата оказывается недостаточно для теории схем общего типа, а также для теории многополюсных схем. В настоящей статье предлагается для исследований такого рода использовать матричную булевскую алгебру и описывается ряд результатов, полученных в этом направлении.

§ 1. Матричная булевская алгебра

Пусть 𝔄 есть некоторая булевская алгебра (4). Будем рассматривать матрицы с элементами из 𝔄. Как и для обычных матриц (с элементами из поля), для матриц с элементами из 𝔄 можно ввести операции сложения и умножения, которые мы будем записывать: $A+B$, $A \times B$. При этом также будут иметь место ассоциативные, коммутативный (для сложения) и дистрибутивный законы.

Введем понятие «определителя» квадратной матрицы с элементами из 𝔄, как суммы $n!$ слагаемых, составленных таким же образом, как и в обычном определителе n-го порядка. Такие определители будут обладать рядом свойств, аналогичных свойствам обычных определителей.

Для пары матриц с элементами из 𝔄 мы введем еще операцию «булевского умножения», обозначив ее $A \cdot B = C$ и определив элементы матрицы C через элементы матриц A и B следующим образом:

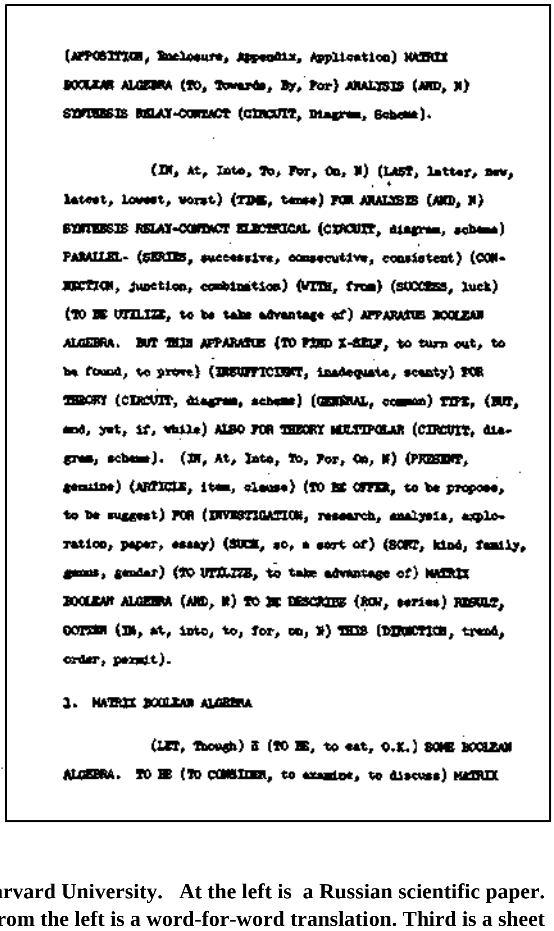
$$c_{\alpha,\beta} = a_{\alpha,\beta} b_{\alpha,\beta}$$

для всех индексов α и β.

Квадратную матрицу с элементами из 𝔄, по главной диагонали которой стоят единицы, будем называть «булевской», а множество булевских матриц n-го порядка с элементами из 𝔄 обозначать $\mathfrak{A}_n$ и называть матричной булевской алгеброй. Множество $\mathfrak{A}_n$ и в самом деле является булевской алгеброй относительно операций сложения и булевского умножения. В дальнейшем только о матрицах из $\mathfrak{A}_n$ и будет идти речь.

§ 2. Многополюсники

Каждую релейно-контактную схему (или часть схемы) можно задать, указав непосредственную проводимость между ее узловыми точками. Поэтому на исследуемой электрической схеме выберем n точек (полюсов) $M_1, M_2, \ldots, M_n$ и будем изучать схему относительно этих точек. Обозначим непосредственную проводимость от полюса



(APPOSITION, Enclosure, Appendix, Application) MATRIX BOOLEAN ALGEBRA (TO, Towards, By, For) ANALYSIS (AND, N) SYNTHESIS RELAY-CONTACT (CIRCUIT, Diagram, Scheme).

(IN, At, Into, To, For, On, N) (LAST, latter, new, latest, lowest, worst) (TIME, tense) FOR ANALYSIS (AND, N) SYNTHESIS RELAY-CONTACT ELECTRICAL (CIRCUIT, diagram, scheme) PARALLEL- (SERIES, successive, consecutive, consistent) (CONNECTION, junction, combination) (WITH, from) (SUCCESS, luck) (TO BE UTILIZE, to be take advantage of) APPARATUS BOOLEAN ALGEBRA. BUT THIS APPARATUS (TO FIND X-SELF, to turn out, to be found, to prove) (INSUFFICIENT, inadequate, scanty) FOR THEORY (CIRCUIT, diagram, scheme) (GENERAL, common) TYPE, (BUT, and, yet, if, while) ALSO FOR THEORY MULTIPOLAR (CIRCUIT, diagram, scheme). (IN, At, Into, To, For, On, N) (PRESENT, genuine) (ARTICLE, item, clause) (TO BE OFFER, to be propose, to be suggest) FOR (INVESTIGATION, research, analysis, exploration, paper, essay) (SUCH, so, a sort of) (SORT, kind, family, genus, gender) (TO UTILIZE, to take advantage of) MATRIX BOOLEAN ALGEBRA (AND, N) TO BE DESCRIBE (ROW, series) RESULT, OOTTEN (IN, at, into, to, for, on, N) THIS (DIRECTION, trend, order, permit).

1. MATRIX BOOLEAN ALGEBRA

(LET, Though) E (TO BE, to eat, O.K.) SOME BOOLEAN ALGEBRA. TO BE (TO CONSIDER, to examine, to discuss) MATRIX

AUTOMATIC DICTIONARY for the word-for-word translation of Russian into English has been investigated by Anthony C. Oettinger of Harvard University. At the left is a Russian scientific paper. Second from the left is a word-for-word translation. Third is a sheet

of Languages, with essays contributed by practically every active worker in the field. Besides the book, 56 articles on the subject have been published to date.

Now that we have quickly reviewed the history of this idea from its birth in 1946 to the present, let us take a close look at the concepts involved in machine translation. The process must involve five basic steps: (1) feeding the original text—written or spoken—into the machine; (2) transforming this text into symbols the machine can handle; (3) translating the meaning from one language to another; (4) turning the translation back into conventional words or other units in the new language; and (5) presenting the translated text in readable or audible form. Various groups of workers have concentrated on one or another step, and we can conveniently consider the steps one at a time.

Some of those working on step 1 have felt that the limitations of machines demand that the text be reduced to a digestible form before it is fed into the machine. Translation would be easy if authors of scientific papers in all languages would write in a universal syntax, so that only the words needed to be translated. But it is more or less generally agreed that you simply couldn't force authors all over the world to change their style of writing because their work might be translated by machine. Various workers, notably Reifler, have suggested instead that pre-editors be employed to rewrite or code the texts of articles before they are fed into the machine. The main trouble with this idea is that the salary of a pre-editor plus the expense per word of the machine would almost surely be as costly as a human translator's services.

The input problem may, however, be made easier by developments in step 2—the reading of the text by the machine itself.

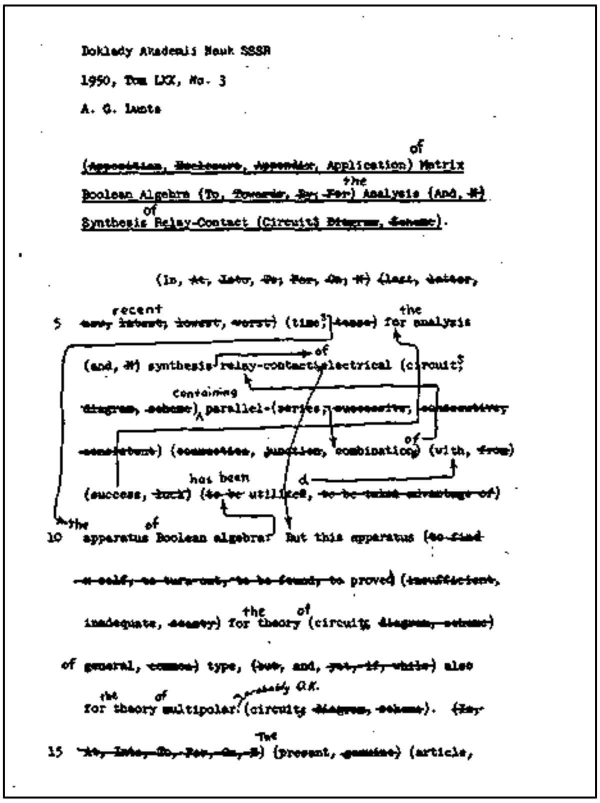
A number of investigators in the U. S. and abroad are working on the problem of direct recognition of written or spoken text by a machine, not for translation but for commercial purposes. At least three companies are developing automatic check-reading machines for banks, and one or more such machines will soon be on the market. There is a great financial incentive here: a New York bank has estimated that it could save more than $2 million a year if it could mechanize the tabulation and sorting of checks.

This task involves the reading of numbers. Recognition of letters or words by a machine is a more complex problem, but work is going forward on this too. When the feat is accomplished, as it should be within a few years, it will be possible for a translating machine to read text directly from the printed page, without any operator or pre-editor.

Of course letter recognition will have to cope with the many different fonts of type, headings, italics and special symbols that are used in printed publications. But untranslatable symbols, and even diagrams and illustrations, can simply be reproduced unchanged in the output, just as a human translator copies them as they appear in the original.

Translation of the sounds of spoken language into a machine code (which is closely comparable to a secretary translating dictation into shorthand) is another active field at present Millions of dollars are being spent by the Bell Telephone Laboratories and others trying to build a machine to do this. It may be that by analysis of the meaningful elements in spoken language we shall be able to write equations to program machine translation of speech either into the written form of the same language or into the written or spoken form of a different language.

This brings us to step 3—the actual translation by the computer. It is clear that computer components already



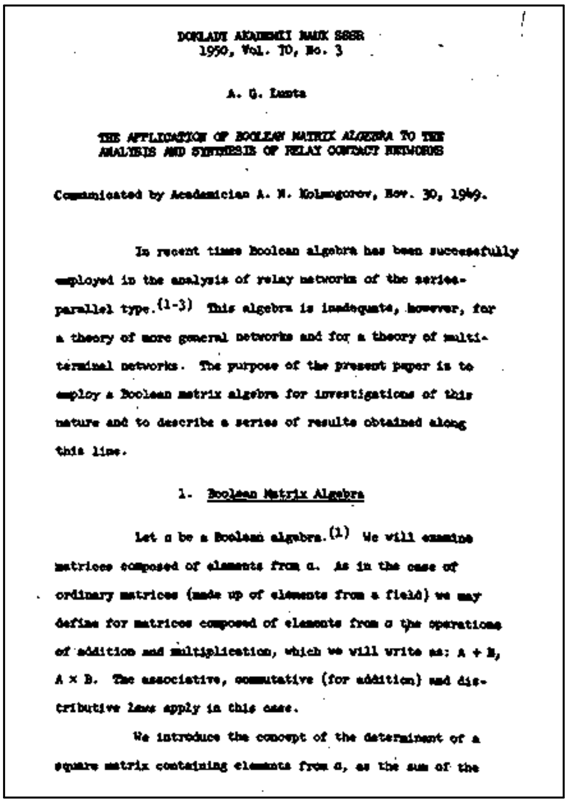
Doklady Akademii Nauk SSSR

1950, Tom LXX, No. 3

A. G. Lunts

(Application) of Matrix Boolean Algebra (To, Towards, For) the Analysis (And, ...) of Synthesis Relay-Contact (Circuits, Diagram, Scheme).

(In, ...) recent (time; ...) the for analysis (and, ...) synthesis relay-contact electrical (circuits; diagram, scheme) containing parallel-(series; successive, ...) (connection, junction, combination) (with, from) (success, luck) has been (utilized, ...) the apparatus of Boolean algebra. But this apparatus (... proved) (insufficient, inadequate, ...) for the theory (circuits; diagram, scheme) of (general, common) type, (but, and, ...) also for the theory of multipolar (probably O.K.) (circuits, diagram, scheme). The (present, genuine) (article,



DOKLADY AKADEMII NAUK SSSR
1950, Vol. 70, No. 3

A. G. Lunts

THE APPLICATION OF BOOLEAN MATRIX ALGEBRA TO THE ANALYSIS AND SYNTHESIS OF RELAY CONTACT NETWORKS

Communicated by Academician A. N. Kolmogorov, Nov. 30, 1949.

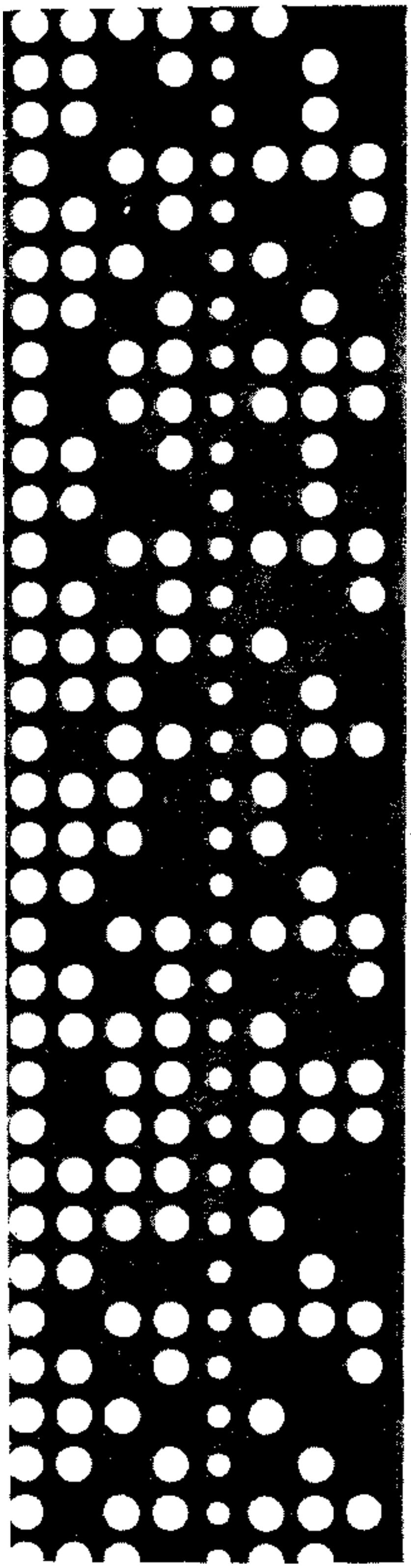
In recent times Boolean algebra has been successfully employed in the analysis of relay networks of the series-parallel type.(1-3) This algebra is inadequate, however, for a theory of more general networks and for a theory of multi-terminal networks. The purpose of the present paper is to employ a Boolean matrix algebra for investigations of this nature and to describe a series of results obtained along this line.

1. Boolean Matrix Algebra

Let α be a Boolean algebra.(1) We will examine matrices composed of elements from α. As in the case of ordinary matrices (made up of elements from a field) we may define for matrices composed of elements from α the operations of addition and multiplication, which we will write as: A + B, A × B. The associative, commutative (for addition) and distributive laws apply in this case.

We introduce the concept of the determinant of a square matrix containing elements from α, as the sum of the

in which the comprehension of a reader was tested. Fourth is an expert translation. These samples appear in *Machine Translation of Languages,* published by John Wiley & Sons, Inc., and the Technology Press of the Massachusetts Institute of Technology.



PUNCHED TAPE was used to program the Massachusetts Institute of Technology computer *Whirlwind* for the demonstration detailed on the opposite page. A foreign language and instructions for translating it might be fed into the machine by the same means. Because this would require the cumbersome translation of written language into the language, students of machine translation hope for the development of devices which could read from the printed page.

existing or under development will be able to handle the job of translation once the input can be fed in in suitable form. General-purpose computers now on the market can be programmed to do translation; in fact, more than one already has been. In January, 1954, the IBM 701 translated a number of Russian sentences as a test. It used a 250-word vocabulary and five Russian syntax constructions. The words were translated and their order was changed automatically to make the output acceptable English. At Harvard University Anthony Oettinger has programmed the Mark IV computer to split Russian words into stems and endings and derive the grammatical meanings from the endings. Booth, in London has done similar work. At Washington Reifler has had a small special computer built to test translation procedures.

One of the most difficult problems, engaging the concentrated efforts of a number of investigators, is the inflection of words. In an inflected language such as Russian or German, practically all the important information-bearing words are varied in meaning by prefixes, suffixes or even infixes. Dictionaries usually list a word in one of its inflected forms—the nominative singular of a noun, the infinitive of a verb. Several workers on the machine translation problem have suggested that for a machine we shall need instead of the ordinary dictionary a dictionary of word stems. These stems (e.g., the German *heiss)* would be listed in the machine's memory. A word fed into the machine would be identified immediately if it was the same as a stored stem. If it did not match any stored entry, the machine would strip off its letters one by one until an identifiable stem was found. The stripping process would start backward from the last letter of the word, and if that did not work, it would begin again with the first letter to remove a possible prefix.

Oettinger has carried this approach to an elegant conclusion. He employs the following procedure. A Russian word is fed into the machine. The machine has built-in circuits for identifying inflectional endings immediately, and if the word has one, it strips the ending off. Then the machine looks up the remaining stem in its memory. If it is listed there, the machine can give its English meaning. Meanwhile, to complete the meaning of the word, the machine also hunts up in a separate memory the sense of the stripped-off inflectional ending.

Irregular words will be no problem at all for the machine, although they give a human translator the most trouble. They will simply be entered in toto in the memory and translated directly. For example, the German *war* and the French *était* will be translated at once as the English "was."

Some years ago there was considerable worry about whether a computer could have a large enough memory to store all the stems, plus the various endings, plus the irregular words, plus the grammar rules, plus the programming instructions. But it looks now as though computers will soon have plenty of fast storage capacity in the form of magnetic drums, tape or photographic film.

When all is said and done, word-by-word rendition will be only a halfway house toward satisfactory translation. To give really usable performance, translating machines will have to consider a whole sentence, at least, as a unit. This means that it will have to be concerned with the structure of words in groups.

Yngve, who is in charge of the machine translation project at M.I.T., has developed an original and very promising approach to this problem. Like other workers in the field, he was struck by the fact that the words in a sentence fall into two general classes: high-frequency words which carry comparatively little meaning (such as "the," "of," "by" and so on) and the lower-frequency verbs and nouns which convey most of the information. The meaning of the latter can be translated word for word. But a sentence makes sense only when they are related to one another by the high-frequency structural words. In other words, the high-frequency words of comparatively little meaning provide the structural framework of a sentence: e.g.,
"By the (law) of (Archimedes), the (weight) of a (submerged object) must (equal) the (weight) of the (displaced water)." Following up this idea, Yngve and a group of four linguists at M.I.T. are endeavoring to write rules for making the information contained in a language's syntax (or sentence structure) completely explicit. Combining such sets of rules for two languages with word dictionaries, we should have the linguistic basis for translation by machine.

The fourth and fifth steps of the machine translation process have received less attention than the first three, but they seem to present no great difficulties. Present computers already print their output in forms which appear to be satisfactory for our purposes, and undoubtedly there will be improvements in speed. As for the last step—the polish-

CATHODE-RAY TUBE displays three Russian words and their word-for-word translation. The tube is one output of *Whirlwind.* which was especially programmed to construct these symbols. Similar programming might be used in the output of a translating machine. As each set of symbols (which would be smaller and more numerous than they are here) appeared they could be photographed.

ing of the machine's output into the final text in the new language—there is a difference of opinion. Reifler has suggested a post-editor to make the text more readable. Whether a human editor will enter the picture will depend on how badly we need him and what he will cost.

How good a translation could be produced by machine? The perfect translation would be one in which all the ideas (and esthetic values) of the original text were reproduced faithfully in the new language. How closely it will be possible to approach that ideal we are unable to say at present. The indications are, from the work of Yngve and others, that we shall succeed almost immediately in getting better than a word-for-word rendition. Even a crude translation may be good enough to enable specialists in the same field, who already have a considerable common background of understanding, to communicate with one another. For this reason, as well as the great need, everyone interested in machine translation is concentrating on scientific and technical material. As the quality of the machine output improves, the translations will become understandable to wider and wider circles of readers. Eventually it may become possible to advance from expository technical writing into narrative and other types of literature.

Although it is only nine years since the idea was born, many people are hard at work on specific features of the design for a translating machine. On the "hardware" side, engineers are developing devices for recognition of written characters and the sounds of speech. large and rapid computer memories, logical circuits, high-speed printers and automatic composing machines. On the linguistic side, experts are analyzing vocabulary and grammar as they have never been analyzed before.

In answer to the question "When shall we see a machine translate?" my best guess is within five years. By that time there should be in operation one or more models turning out a good deal better than a word-by-word translation.