ACADEMY OF SCIENCES OF THE USSR

I. S. Mukhin

AN EXPERIMENT OF THE MACHINE TRANSLATION OF LANGUAGES CARRIED OUT ON THE B E S M

MOSCOW, 1956

The work in the machine translating from English into Russian was started in January, 1955, by the Institute of Scientific Information, of the Academy of Sciences of the USSR, in conjunction with the Academy of Sciences Institute of Precise Mechanics and Computing Technique. The research workers — participants in the work are Prof. Panov, D. Ju., Belskaja I., Korolev, L. N., Rasumovsky, S. N., Selenkevich, G., and the author of this paper.

The experiments were carried out on the high-speed electronic computer BESM, designed and constructed at the Institute of Precise Mechanics and Computing Technique by Prof. Lebedev, S. A., and his associates.

1. Introduction. The idea of machine translation of languages began to draw special attention of scientists of late in connection with recent advances in the development of the high-speed automatic electronic computers with programmed control. In 1948—1949 British and American scientists discussed the possibility of translation from one language into another by an electronic computer (1, pg. 2—3). Soon afterwards (in 1950—1951) various institutions both in Great Britain and the USA started working out the problem, and on January 7, 1954, the first public demonstration of translation from Russian into English was held in New York by the International Business Machines Co. For this experiment a special vocabulary was compiled, consisting of 250 Russian words in Latin script. The words were selected with a view that each of them had one, or at least two English equivalents.

The International Business Machines Co. experiment aroused great interest throughout the world. Numerous comments published in general and special periodicals gave the impression, however, that practical results of the experiment could hardly be expected in the nearest future. Most scientists were of the opinion that the problems to start with were those connected with the translation of scientific texts and the compilation of specialized vocabularies for different branches of science and technology was mentioned as one of them. A specialized vocabulary of the type is estimated, for English, at 1,000 general-purpose words and 1,000 special terms (3). But at present, in the opinion of most scientists, we can hardly afford such a large vocabulary, and hence a scientific book cannot be translated by machine as yet not to mention a work of art.

Having started work on automatic translation, we very soon came to the conclusion that it should be organized on lines different from those described in the report of the American experiment. To our minds, the excessively rigid connection between the translation programme and the vocabulary (ascription of the control codes directly to the words in the vocabulary) caused certain limitations in solving the machine translation problem. For this reason, we decided to try and develop such a system of sentence analysis that would enable us to find out the meaning of every word in the sentence (except for the case when it is impossible within a sentence*) as well as to determine its grammar characteristics. Linguistically, we proceeded from the assumption that all thoughts and ideas are expressed in language by means of words having very definite relations with each other. Experience showed that these relations could be defined, at least for the scientific texts. The system worked out for the analysis of an English sentence and synthesis of its Russian translation proved to be practically independent of the vocabulary.

The texts selected for translation were a number of excerpts from Miln's «Numerical Solution of Differential Equations» along with some more texts of different specification (e. g., an article from «The Times», etc.).

The following is a short description of the accepted vocabulary as well as of our system of analysis and synthesis; it should also give an idea of the way the BESM was made use of in the experiment.

2. Vocabulary. Translation requires a vocabulary, no matter whether it is done with or without the help of a machine. If, for man, each word is made up of letters, for a computing machine, which carries out operations with figures, letters must be replaced by figures.

If we substitute each letter of the Latin alphabet by a definite combination of figures, we shall be able to express any English word by a corresponding number. Thus. using the Baudot code (Fig. 2.1), we can change

* This happens when the sentence contains pronouns standing for words belonging to a previous sentence.

the words THE, EQUATIONS, METHOD, THEREFORE into the following numbers:

212608, 082320162112281505, 110821262830, 212608070814280708.

a — 16	v — 29	m — 11	t — 21
b — 06	z — 25	n — 15	u — 20
w — 13	i — 12	o — 28	f — 14
g — 10	j — 18	p — 24	h — 26
d —30	k— 19	r — 07	c — 22
e — 08	l — 27	s — 05	q — 23
		x — 09	y — 04

Fig. 2.1.

Our vocabulary includes 952 English words. Besides its numerical expression every word of the vocabulary has a definite ordinal number, that is, a special place-in-the vocabulary indication. Thus, for instance, the words BELOW (0608272813), DEVICE (300829122208), REGION (070810122815), WHOLE (1326282708), have respectively the following place-in-the-vocabulary indications: 110, 211, 570, 748.

A vocabulary compiled for the mechanical translation differs from the usual vocabulary in that it consists of two sections and contains, besides the Russian word corresponding to the English word sought, certain additional information (indications) concerning the Russian word.

One of the sections contains the English words, recorded as numbers and all the vocabulary indications of the corresponding Russian words. For example, in the case of nouns, the following information of the Russian word is given: gender, declension, soft or hard stem, presence or absence of sibilants in the stem, denotation of animate or inanimate objects, etc.; in the case of verbs, their conjugation, aspect, etc.; in the case of adjective, hard or soft stem, etc. We call this section of the vocabulary the English section.

The second sections consists of Russian words recorded as digit combinations in the order defined by their place-in-the-vocabulary indications given in the English section of the vocabulary (Fig. 2.2). The second section is called the Russian section of the vocabulary.

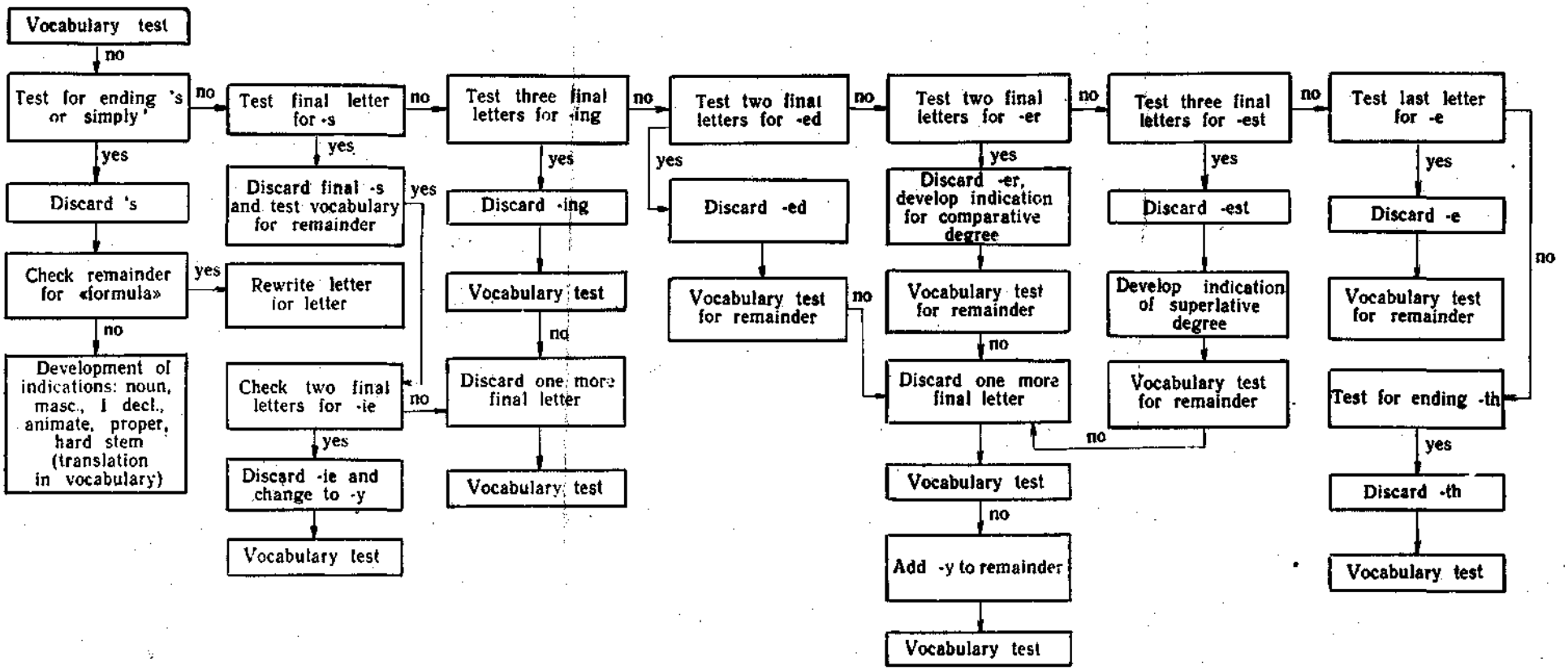
Vocabulary test
no
Test for ending 's or simply '
yes
Discard 's
Check remainder for «formula»
yes
Rewrite letter ior letter
no
Development of indications: noun, masc., I decl., animate, proper, hard stem (translation in vocabulary)
no
Test final letter for -s
yes
Discard final -s and test vocabulary for remainder
yes
Check two final letters for -ie
no
yes
Discard -ie and change to -y
Vocabulary test
Test three final letters for -ing
yes
Discard -ing
Vocabulary test
no
Discard one more final letter
Vocabulary test
Test two final letters for -ed
yes
Discard -ed
Vocabulary test for remainder
no
Test two final letters for -er
yes
Discard -er, develop indication for comparative degree
Vocabulary test for remainder
no
Discard one more final letter
no
Vocabulary test
no
Add -y to remainder
Vocabulary test
Test three final letters for -est
yes
Discard -est
Develop indication of superlative degree
Vocabulary test for remainder
Test last letter for -e
yes
Discard -e
Vocabulary test for remainder
Test for ending -th
yes
Discard -th
Vocabulary test
no

Fig. 5.1.

а — 16	ж — 29	и — 15	у — 20	щ — 23
б — 05	з — 25	о — 28	ф — 14	э — 17
в — 13	и — 12	п — 24	х — 26	ь — 09
г — 10	к — 19	р — 07	ц — 22	ы — 04
д — 30	л — 27	с — 05	ч — 23	ю — 01
е — 08	м — 11	т — 21	ш — 09	я — 03
				й — 18

Fig. 2.2.

3. **Vocabulary of polysemantic words**. For some of the English words of the vocabulary (121 words in our case) special digit indications, substituting the place-in-the-vocabulary indication of the Russian words, are used to show that these words have multiple meaning. The correct meaning of the Russian word in this case, comes as a result of the context analysis of the polysemantic word done by a special translation program, called the *vocabulary of polysemantic words.*

Example (Fig. 3.1). The word «true>> has two meanings in our vocabulary: «верный» and «выверить». Using the program of the polysemantic word vocabulary we find for the word «true» in our sentence (Fig. 3.1) the meaning «верный». The polysemantic word vocabulary gives the same kind of information on the word as the English part of the vocabulary.

This is true certainly of the vast category of problems associated with force and motion.

Fig. 3.1.

4. The Input of the English text. The English text is put into the machine one sentence at a time. The text is preliminarily punched on a paper tape with a special puncher which has Latin letters and punctuation marks engraved on its keys.

Thus, the English text is represented on a paper tape in the form of groups of holes, according to the above mentioned code for Latin letters. Besides letters, the following supplementary denotations are employed: space between words — 00, period —31, comma —03. etc.

5. Finding words in the English section of the vocabulary. Those of the words in the English text whose spelling coincides exactly with that of the words in the vocabulary are easily determined by the operation of comparing carried out by the electronic computer. This process may be represented in a simplified way as follows. Let the number sought, i. e. the word in the text, be subtracted successively from each of the numbers representing word in the vocabulary. When the difference is zero our matching job is over: the word sought is found.

In some cases the words of the text do not coincide exactly with the corresponding words of the vocabulary, since they possess grammar affixes (s, 's, ing, ed, er, est, e, th).

If no exact coincidence between the word in the text and those in the vocabulary can be found, the word sought is verified for presence of one of the above mentioned affixes.

The affix found is then discarded, and the search for the text word in the vocabulary is repeated.

The entire process of finding the words of the text is carried out according to the scheme (Fig. 5.1).

One of the quickest ways of finding words in the vocabulary, in our opinion, is the following.

All the English words in the vocabulary, recorded as figure combinations, are positive numbers. Each word, depending on the number of letters it consists of, takes up one, three or four positions of the storage unit. Then the words are divided into groups: the first group consists of words occupying one storage position, the second — of words occupying two positions, the third one of those occupying three positions, etc.

The words within each group are arranged in the increasing order of the numbers representing the words.

Finding words in the vocabulary begins with a determination of the number of positions occupied by the word sought. This is followed by looking up the word in the corresponding word group of the vocabulary. For this purpose, the numerical value of the word in the middle of the group is compared to the value of the word sought. If the word sought is greater in value than that selected from the group, the half-group of words having greater values than that selected from the middle of the group

is considered. If it is smaller in value, the half-group having smaller values than that selected from the middle of the group is taken. This process of halving the groups of words in the vocabulary ensures rapid finding of the word needed.

The time required to find a word by this method is proportional to the logarithm of N, with a base of 2, where N is the number of words in the vocabulary.

A special programme is used to carry out the supplementation of the vocabulary and the ensuing rearrangement of the words.

6. Replacing the words of the sentence by equivalents. After the word has been found in the vocabulary, all the information on the word is taken from the vocabulary: the number of the word in the English section of the vocabulary, the number of the corresponding Russian word and the grammatical information recorded in the vocabulary on the Russian word.

This information is the numerical equivalent of the word, and all subsequent operations are carried out with this equivalent.

The numerical equivalent of each English word was stored in two positions of the memory. The use of two, and not one, three or any other number of positions is due to the particular characteristics of the BESM employed for our purpose.

7. Arrangement of the parts-of-speech indications in the storage positions. In order that the machine should be able to distinguish parts of speech automatically, the parts-of-speech indications, when transferred from the vocabulary to the positions, always occupy the same part of the position. Thus «1» in that position always means «noun», a «2» — «verb», «3» — «adjective», «4» — «numeral», «5» — «adverb», «6» — «preposition», «7» — «conjunction», etc.

It was mentioned above that each English word has two positions to substitute it. Table 1 shows the meanings of the indication digits of the words as well as their arrangement. They are placed in the first of the two positions substituting the English word. The number in the second position shows the number of the word in the Russian section of the vocabulary. If this number is 0000 after the entire programme is carried out, the cor-

responding English word will be omitted in the translation.

8. Division of the Automatic Translation Programme into Two Main Parts — Analysis and Synthesis. The first part of the programme deals with the analysis of the English sentence and includes the analysis of the indications taken from the vocabulary, the characteristic affixes of the English words, as well as their position in the sentence. The aim of the analysis is to make out the grammatical form and position in the sentence of the corresponding Russian words. The information thus obtained is expressed by means of a set of indications and makes it possible to pass over to the second part of the programme, which is the synthesis of the Russian sentence.

The second part of the programme makes necessary changes in the grammatical form and position of the Russian words, taking into account the set of indications the word has received.

9. Sequence of the Parts of the Programme. The sequence of the various parts of the programme of mechanical translation is shown diagramatically in fig. 9.1.

Separate parts of the program are applied in a sequence which in the vast majority of cases ensures development of the indications needed for the fulfilment of the subsequent operations. The role of the separate programmes is obvious from their names (Fig. 9.1) Only two parts of the programme need some explanations, they are the «syntax» and the «change of word order».

The «syntax» part of the programme breaks up complex sentences into clauses, by placing the following marks: beginning, end, end-beginning of clauses, punctuation marks.

The «change of word order» part rearranges the words in accordance with the rules of the Russian grammar.

Repetition of the «verbs» part is necessary because the verb indications cannot be worked out completely until the «syntax», «numericals», «nouns» and «adjectives» parts of the programme are fulfilled. Application of the «verbs» part before these parts of the programme is necessary because some of the information on verbs, obtained as a result of the first application of this part of the programme, is needed for the «syntax», «numerals», «nouns» and «adjectives» parts.

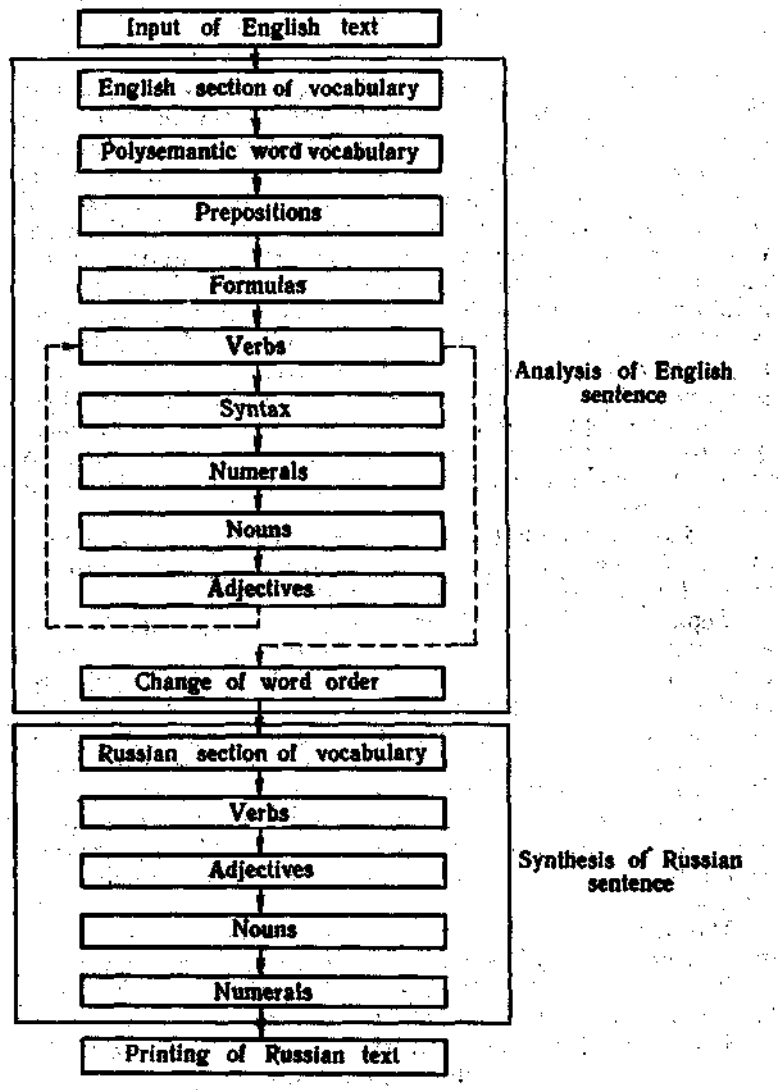
Input of English text
English section of vocabulary
Polysemantic word vocabulary
Prepositions
Formulas
Verbs
Syntax
Numerals
Nouns
Adjectives
Change of word order
Analysis of English sentence
Russian section of vocabulary
Verbs
Adjectives
Nouns
Numerals
Printing of Russian text
Synthesis of Russian sentence

Fig. 9.1.

10. **An Illustration.** Let us see in a simplified form how an English sentence is put into the machine, taking as our example the sentence in Fig. 3.1. The sentence is punched on a paper tape in the form of a number consisting of one hundred and seventy six digits (Fig. 10.1).

21261205001205002107200800220807211612152704002814002126080
02916052100221621081028070400281400240728062708110500160505
28221216210830001312212600142807220800161530001128211228l5

Fig. 10.1.

After the sentence is entered, the machine breaks up this one-hundred-and-seventy-six-digit number into separate number-words.

Then begins the work with the English section of the vocabulary and the polysemantic word vocabulary.

11. Substitution of Words by Equivalents in the Sentence in Fig. 3.1. As a result of the application of the English section of the vocabulary and the polysemantic word vocabulary, each word in our sentence is substituted by two numbers located in two positions. The first number contains all the indications transferred from the English section of the vocabulary and the polysemantic word vocabulary, according to the designation in Table 1. The second number is the ordinal number of the Russian word. If the Russian number equals 0000, the Russian meaning of the word it stands for has not been found in the vocabulary. The Russian meaning of such a word can be found through the subsequent parts of the programme. If the Russian number of the word remains equal to 0000 by the end of the analysis of the English sentence, the English word is omitted in the translation.

Let us see how our sentence is transformed here. The words of the sentence are substituted by the follow ing equivalents (Fig. 11.):

This	1000100300000100011156327
is	2000011000200000001038 0000
true	3000000000000000001204 6344

certainly	510132 2257
of	600472 0000
the	3000000000000000001161 0000
vast	300000000000000000729 4410
category	120000000000001000130 2253
of	600472 0000
problems	121000020001001000529 3620
associated	2010000400000000030085 2140
with	600749 0000
force	1200000200000000000312 3012
and	71001028 6470
motion	110000030000001000441 3367

Fig. 11.1.

As can be seen in (11.1) the word PROBLEMS is substituted by the numbers 121000020001001000529 and 3620. The digits of these numbers mean (Fig. 11.2): The second number 3620 is the number of the Russian word «задача»

1 — noun.
2 —2-nd declension.
1 — stem ends in sibilant or г, к, х.
0 — word has no flexion.
0 — plural.
0 — is not a predicate.
0 — case not determined.
2 — feminine.
0 — word denotes animate object.
0 — word is not proper noun.

0 — number indication of word not developed.
1 — English word has -s ending.
0 — is not verbal noun.
0 — is not subject.
1— word has soft stem.
0 — person of word not determined.
0 — absence of « omitt» indication.
0529 — word number in English section of vocabulary.

Fig. 11.2.

This word was not been found in the vocabulary at first (since it has the ending «s») and was detected only after its ending was discarded. At the same time the presence of the ending «s» was recorded as one of the indications.

The contents of the equivalents of the rest of the words can easily be determined with the aid of Table 1.

12. **The Results of the Application of the Parts of the Programme Doing the Analysis of the English Sentence to the Phrase (3.1).** After the words of the sentence have been substituted by their equivalents, they are subjected to the parts of the programme concerned with the analysis of the English sentence.

We mean to explain the principle of the operation of different parts of the programme by considering two examples.

Example 1. In the sentence (3.1) the word OF occurs twice. In the first case the equivalent of the word OF undergoes changes according to the scheme in Fig. 12.1, as follows:

1—2 — 3 — 4 — 5,

where the figures 1 to 5 denote the individual logical elements of the «prepositions» program. As a result, we get the following value of the equivalents

620472
5046

which means:

6 — preposition,

2 — takes Genitive case,

0472 — word number in the English section of the vocabulary, the number 5046 is the Russian number for the word «для».

In the second case the equivalent of OF changes according to the same scheme in the following manner:
1---2 —3 — 4 —5 —6 —7 —8 —9—10—11 —
— 12—13
This gives the following value of the equivalent:
620472
0000
where: 6 means «preposition»,
2 means «takes Genitive case»,
0472 is the word number in the English section of the vocabulary.
The number 0000 shows that in this case OF is not translated.

Example 2. The equivalent of the word CATEGORY undergoes changes according to the scheme (Fig. 12.2).

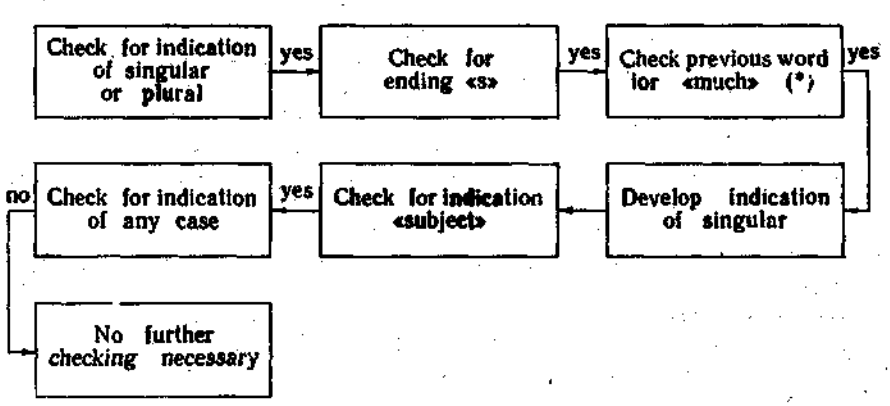

Fig. 12.2.

which is part of the «nouns» programme, and acquires the following value:
12001022010001000130
2253

The equivalent of the other words are changed in a similar manner.

After the programmes of the English sections are fulfilled, we shall have the following equivalents of the words in the sentence (3.1) (Fig. 12.3):

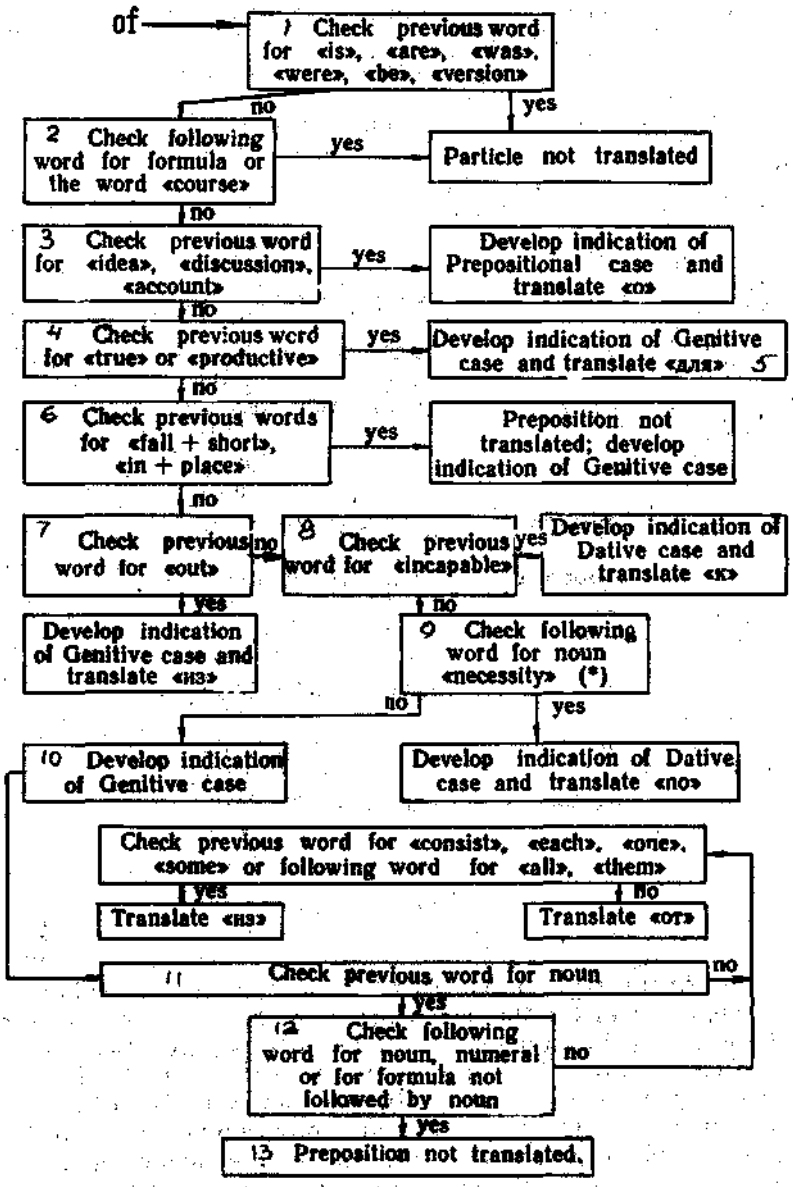
of
1 Check previous word for «is», «are», «was», «were», «be», «version»
no
yes
2 Check following word for formula or the word «course»
yes
Particle not translated
no
3 Check previous word for «idea», «discussion», «account»
yes
Develop indication of Prepositional case and translate «о»
no
4 Check previous word for «true» or «productive»
yes
Develop indication of Genitive case and translate «для»
no
6 Check previous words for «fall + short», «in + place»
yes
Preposition not translated; develop indication of Genitive case
no
7 Check previous word for «out»
no
8 Check previous word for «incapable»
yes
Develop indication of Dative case and translate «к»
yes
no
Develop indication of Genitive case and translate «из»
9 Check following word for noun «necessity» (*)
no
yes
10 Develop indication of Genitive case
Develop indication of Dative case and translate «по»
Check previous word for «consist», «each», «one», «some» or following word for «all», «them»
yes
no
Translate «из»
Translate «от»
11 Check previous word for noun
no
yes
12 Check following word for noun, numeral or for formula not followed by noun
no
yes
13 Preposition not translated.

Fig. 12.4.

This	100110130010010001115 6327
is	2100000000000000001038 0000
true	3000010030110000001204 6344
certainly	510132 2257
of	620472 5046
the	3000010220010000001161 0000
vast	300001022001000000729 4410
category	120010220010001000130 2253
of	620472 0000
problems	121000220011001000529 3620
associated	3100000020103004100852140
with	650749 5030
force	120010520010000000312 3012
and	71001028 6470
motion	110010530010001000441 3367

Fig. 12.3.

13. **Application of the Part of the Programme Doing the Synthesis of the Russian Sentence to the Sentence (3.1).** The next step in mechanical translation is the treatment of the Russian words recorded in the denotations (2.2). The correct grammatical form of the Russian word is derived by using certain parts of the programme of synthesis of the Russian sentence, which are made up in accordance with the requirements of the Russian grammar, the indications of the word equivalents being obtained as

a result of the analysis of the English sentence. The sequence of the parts of the programme can be seen in Fig. 9.1 As an example we may consider the changes in the Russian word «категория», which take place according to the scheme (Fig. 13.1).

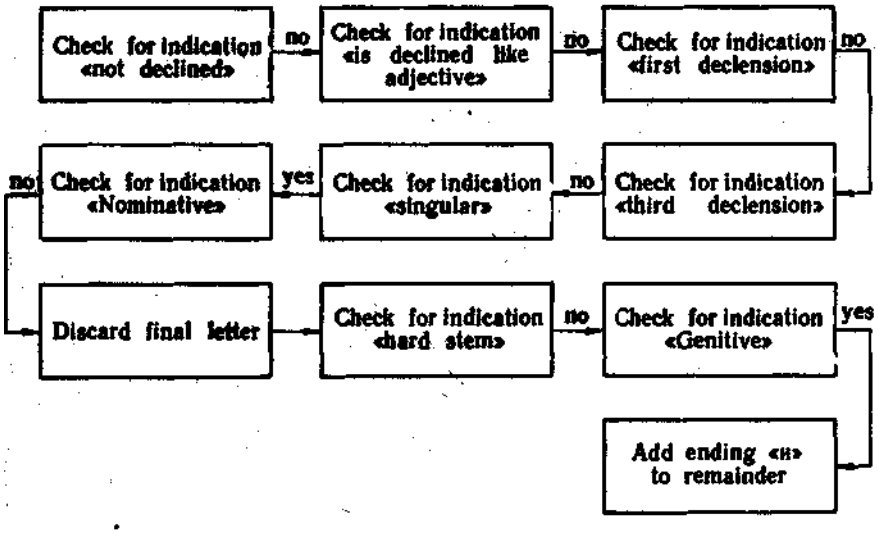

Fig. 13.1.

The required grammatical forms of the rest of the words in the sentence are derived in a similar way. As a result we get the final translation of the sentence, recorded with the aid of the designations in (5.2).

14. **Output of the Russian Text** The last step of mechanical translation is printing the text by the printer. The numbers are changed into Russian letters according to the designations in (2.2) and printed on a paper tape. The phrase (3.1) will be printed in the form of the Russian sentence, given in the Fig. 13.2.

172128 1308071528 06082520052728131528 302703 2806021207152818
191621081028071212 2516301631 061303251615150426 05 0512272818
12 301312290815120811

Fig. 13.2.

15. **A Few Examples of Texts Translated by Machine**.

In conclusion we present a few examples of the texts translated from English into Russian by the BESM in the latter part of 1955 and early in 1956. The translation is given exactly as it was obtained from the machine, without editing (Fig, 15.1).

Of course, the first experiments in mechanical translation carried out at present in the USSR and other countries are far from being practical realization of the machine translation of languages on a large scale. But our firm belief is that new achievements are to be expected in the nearest future, at least as concerns machine translation of scientific and technical texts.

When a practical problem in science or technology permits mathematical formulation, the chances are rather good that it leads to one or more differential equations. This is true certainly of the vast category of problems associated with force and motion, so that whether we want to know the future path of Jupiter in the heavens or the path of an electron in an electron microscope we resort to differential equations. The same is true for the study of phenomena in continuous media, propagation of waves, flow of heat, diffusion, static or dynamic electricity, etc., except that we here deal with partial differential equations.

Если практическая задача в науке или технике допускает математическую формулировку, шансы довольно велики что это приводит к одному или более дифференциальным уравнениям. Это верно безусловно для обширной категории задач связанных с силой и движением, так что хотим ли мы знать будущий путь Юпитера в небесах или путь электрона в электронном микроскопе мы прибегаем к дифференциальным уравнениям. То же верно для изучения явлений в непрерывной среде, распространения волн, потока тепла, диффузии, статического или динамического электричества, и т. д., за исключением того что мы здесь будем рассматривать дифференциальные уравнения в частных производных.

In problems of this type numerical methods become a necessity due to absence of other methods for getting the requisite information out of the differential equations.

В задачах этого типа численные методы становятся необходимостью обусловленной отсутствием других методов для получения необходимого сведения из дифференциальных уравнений.

It is often impossible, however, to perform the actual eliminations, and hence this transformation is of theoretical rattier than practical interest.

Часто невозможно, тем не менее, выполнить действительные исключения, и следовательно это преобразование имеет теоретический скорее чем практический интерес.

Suppose that both equations actually contain all the possible partial derivatives of second order.

Допустим, что оба уравнения действительно содержат все возможные частные производные второго порядка.

This was based on an expensive experiment done by myself and Dr. R. H. Richens, of Cambridge University, in which we worked out a method of translating small sections of selected text in foreign languages. We gave an account of this at a conference in Massachusetts in 1952, after which the International Business Machines Company, in conjunction with Georgetown University, applied our methods to give a popular demonstration which was limited to translating a few sentences from Russian into English. There is no possibility at present of translating a book as a work of art.

Это было основано на дорогом эксперименте проведенном мной и доктором R. H. Richens, от Кэмбриджского Университета, в котором мы разработали метод перевода малых отрывков выбранного текста на иностранные языки. Мы дали отчет о этом на конференции в Massachusetts в 1952, после которого I. B. M. компания в сотрудничестве с Джорджтаунским Университетом применили наши методы чтобы дать наглядную демонстрацию, которая была ограничена переводом нескольких предложений с русского на английский. Не имеется возможности в настоящее время перевода книги как произведения искусства.

REFERENCES

1. Machine Translation of Languages. Edited by W. Locke and A. Booth, 1955, p. 243.

2. N. Macdonald. Language Translation by Machine. «Computers & Automation», 3, No. 2, pp. 6—10.

3. A. Booth. Calculating Machines and Mechanical Translation. «Discovery». 1954, 15, No. 7. pp. 280-285.

TABLE 1.

Sequence of digits	Indication digits	Meaning of indication digits
		Noun
firts digit		1 — noun
second digit		0 — is declined like adjective. 1 — belongs to first declension 2 — belongs to second declension 3 — belongs to third declension
third digit		0 — stem does not end in sibilant or г, к, х. 1 — stem ends in sibilant or г, к, х
fourth digit		0 — word is declined 1 — word is not declined
fifth digit		0 — plural 1 — singular
sixth digit		0 — is not predicate 1 — is predicate
seventh digit		0 — case not defined 1 — Nominative case 2 — Genitive case 3 — Dative case 4 — Accusative case 5 — Instrumental case 6 — Prepositional case
eighth digit		1 — masculine 2 — feminine 3 — neuter
ninth digit		0 — word denotes inanimate object 1 — word denotes animate object
tenth digit		0 — word in not proper noun 1 — word is proper noun
eleventh digit		0 — number indication not developed 1 — number indication developed

twelfth digit	0—English word has no -s, -ing, -'s ending 1—English word has ending -s 2—English word has ending -ing 3—English word has ending 's
thirteenth digit	0 — is not verbal noun 1 — is verbal noun
fourteenth digit	0 — is not subject 1 — is subject
fifteenth digit	0 — word has hard stem 1 — word has soft stem
sixteenth digit	1 — first person 2 — second person 3 — third person
seventeenth digit	0 — «omit» indication absent 1 — «omit» indication present
last four digits	word number in English section of vocabulary

Verb

firts digit	2 — verb
second digit	0 — word has no number indication 1 — word has number indication
third digit	0 — second conjugation 1 — first conjugation
fourth digit	0 — stem does not end in sibilant or г, к, х 1 — stem ends in sibilant or г, к, х
fifth digit	0 — word is conjugated 1 — word is not conjugated
sixth digit	0 — singular 1 — plural
seventh digit	0 — is not predicate 1 — is predicate

eighth digit	0 — verb does not take definite case 1 — verb takes Nominative case 2 — verb takes Genitive case 3 — verb takes Dative case 4 — verb takes Accusative case 5 — verb takes Instrumental case 6 — verb takes Prepositional case
ninth digit	0 — verb has no gender 1 — masculine 2 — feminine 3 — neuter
tenth digit	0 — imperfect aspect 1 — perfect aspect
eleventh digit	0 — verb has no tense 1 — past tense 2 — present tense 3 — future tense
twelfth digit	0 — absence of «omit» indication 1 — presence of «omit» indication
thirteenth digit	0 — is not verbal adverb 1 — is verbal adverb
fourteenth digit	0 — active voice 1 — passive voice
fifteenth digit	0 — is not subject 1 — is subject
sixteenth digit	0 — indicative mood 1 — imperative mood 2 — oblique mood 3 — indefinite mood
seventeenth digit	1 — first person 2 — second person 3 — third person

eighteenth digit	0 — English word has no -s, -ing, or -ed ending 1 — English word has ending -s 2 — English word has ending -ing 3 — English word has ending -ed
last four digits	Word number in English section of vocabulary

Adjective.

firts digit	3 — adjective
second digit	0 — word has hard stem 1 — word has soft stem
third digit	0 — second conjugation 1 — first conjugation
fourth digit	0 — word stem does not end in sibilant or г, к, х 1 — word stem ends in sibilant or г, к, х
fifth digit	0 — word is declined 1 — word is not declined
sixth digit	0 — plural 1 — singular
seventh digit	0 — is not predicate 1 — is predicate
eighth digit	0 — case of word not defined 1 — Nominative case 2 — Genitive case 3 — Dative case 4 — Accusative case 5 — Instrumental case 6 — Prepositional case
ninth digit	0 — gender of word not defined 1 — masculine 2 — feminine 3 — neuter
tenth digit	0 — word denotes inanimate object 1 — word denotes animate object

eleventh digit	0 — adjective is in complete form 1 — adjective is in short form
twelfth digit	0 — word has no indication of number 1 — word has indication of number
thirteenth digit	0 — indication of degree or participle absent 1 — adjective is in superlative degree 2 — adjective is in comparative degree 3 — word is participle
fourteenth digit	0 — past tense 1 — present tense
fifteenth digit	0 — word is not subject 1 — word is subject
sixteenth digit	0 — word does not take definite case 1 — word takes Nominative case 2 — word takes Genitive case 3 — word takes Dative case 4 — word takes Accusative case 5 — word takes Instrumental case 6 — word takes Prepositionll case
seventeenth digit	0 — «omit» indication absent 0 — «omit» indication present
eigtheenth digit	0 — English word does not end in -ed 1 — English word ends in -ed
last four digits	Word number in English section of vocabulary
	N u m e r a l
firts digit	4 — numeral
second digit	0 — plural 1 — singular
third digit	0 — is not predicate 1 — is predicate

fourth digit	0 — case of word not defined 1 — word is in Nominative case 2 — word is in Genitive case 3 — word is in Dative case 4 — word is in Accusative case 5 — word is in Instrumental case 6 — word is in Prepositional case
fifth digit	0 — word has no gender 1 — masculine 2 — feminine 3 — neuter
sixth digit	0 — word has no indication of number 1 — word has indication of number
seventh digit	0 — is not subject 1 — is subject
eighth digit	0 — «omit» indication absent 1 — «omit» indication present
last four digits	Word number in English section of vocabulary

Adverbs, parenthetic words, particles.

firts digit	5 — adverb
second digit	0 — adverb 1 — parenthetic word 2 — particle
last four digits	Word number in English section of vocabulary

Preposition

firts digit	6 — preposition
second digit	1 — takes Nominative case 2 — takes Genitive case 3 — takes Dative case 4 — takes Accusative case 5 — takes Instrumental case 6 — takes Prepositional case
last four digits	Word number in English section of vocabulary

Conjunction

first digit	7 — conjunction
second digit	1 — co-ordinative conjunction 2 — subordinative cojunction
third digit	0 — no indication of beginning or end of clause 1 — indication of beginning of clause 2 — indication of end of clause 3 — indication of end of one clause and beginning of another
fourth digit	0 — is not subject 1 — is subject
last four digits	Word number in English section of vocabulary