Machine translation methods and their application to an Anglo-Russian scheme

By I. K. Belskaya, Academy of Sciences of the USSR

An account is given of research which has lead to an algorithmic procedure for machine translation. The method has three stages: dictionary analysis, grammatical analysis and grammatical synthesis. It has been used for translations from English, German, Chinese, Japanese and Russian but the present paper concerns only translation from English to Russian.

The dictionary consists of 3000 English words and about the same number of Russian words, divided in two parts: mono-semantic and poly-semantic. The latter includes about one fifth the dictionary. The dictionary analysis uses five routines, of which the most important, both in theory and in practice, is the routine for the analysis of poly-semantic words.

There are six routines for the grammatical analysis, which is the most important section theoretically since it involves a detailed description of the structure of the source language. The routines treat verbs, punctuation marks, syntax of sentences, nouns and numerals, adjectives, and changes of word order. There are four routines for grammatical synthesis, which treat the dictionary, verbs, adjectives, nouns and numerals.

The dictionary and the routines were tested both using the BESM machine (1956) and also "by hand" by 10 laboratory assistants with no linguistic knowledge. Translations were made of about 100 English texts, some in applied mathematics, some literary. Structural transformations in the source text have been restricted to a minimum, such as the insertion or omission of a few "helping" words or punctuation marks and a few (local) changes of order. The translations thus obtained were quite adequate for understanding and did not require post editing.

1. Introduction

Research in the MT methods which are outlined below, was started late in 1954 on the initiative of Academician A. S. Nesmejanov, President of the USSR Academy of Sciences. The first experiments in MT from English into Russian were carried out in December, 1955 [1,2], and terminated the first stage of the research. Some of the principles on which our research is based were put forward in earlier publications, among which a paper published in RESEARCH, October 1957 [3], should be mentioned. Since then, considerable progress has been made towards adequate formulation of the method. We are now in a position to say that the second stage of the research has recently been completed, in which the method was extended to cover MT from languages as different in structure from English as Japanese, Russian, Chinese and German [4]. This research showed that the method was generally applicable.

The research on the Anglo-Russian scheme of MT has reached a stage where a complete grammatical analysis at a bilingual level, as well as the rearrangement of the most important types of English idiomatic construction, can be accomplished. Grammatical modification of the Russian translation (which indeed is the simpler part of the problem) is performed by an independent set of routines, termed Russian synthesis. In addition, the progress in Anglo-Russian MT has taken the form of a considerable growth of the volume of words now entered into the MT dictionary. More than 2000 words are stored in the English section of our multilingual MT dictionary, a still greater number of Russian equivalents being stored in the Russian section. The dictionary is thus made to cover different fields of applied mathematics[1]).

To complete this stage of research a large-scale test of the Anglo-Russian scheme has been carried out. 100 samples (which amounted to 3000 sentences) of "unknown" text were selected at random from different English authors, and translated into Russian in strict accordance with instructions provided by the MT dictionary and translation routines[2]). The ten persons chosen to carry out the experiment had no knowledge of English, nor had they any previous experience of the tasks required[3]).It emerged from the test that the scheme is very effective at dealing with all sorts of texts restricted lexically to applied mathematics. Grammatically no limitation as to type of the written text has been found necessary. One or two words per printed page is the average for "unknown" words with the present size of dictionary, which makes the translation quite understandable (see Tables 1, 2, 3, 4). For this reason, as well as to be consistent with the proposed series of strictly specialized MT dictionaries, we are not inclined to increase the volume of words in the present dictionary, but rather to proceed with compiling medium size (say 2500—3000 words each) dictionaries for various fields. This indeed will be our occupation at the next stage of research.

As the translation routines for Anglo-Russian MT are the principal achievement of the recent research it seems reasonable, in the present communication, to lay particular stress on the description of the routines for vocabulary and grammatical analysis of the English sentence. For the principles on which the MT vocabulary is based, the reader is referred to our earlier publication [3].

[1]) Participants in this work were G. A. Tarasova, whose contribution to the compilation of the Anglo-Russian dictionary is most valuable, and L. M. Bykova.

[2]) A. I.Martynova was engaged as supervisor in the testing procedure.

[3]) Several samples translated in this manner are given in Tables 1, 2, 3 and 4.

2. General considerations. Applicability of MT methods

Of the two most general MT problems—the possibility of machine translation and its applicability—the former has already been resolved, both theoretically and practically, whereas the latter problem still remains open for discussion. The object of the present research is to prove the applicability of MT methods to any sphere of language.

To date, it is only within the limited sphere of scientific writing that the applicability of MT methods has won general recognition. As to other uses of MT, most machine translators are inclined to feel very doubtful [4]. However, the majority of the restrictions imposed on MT application turn out, when analysed, to be due to a very strong inclination on the part of investigators to describe the translated language (the source language) in terms of its correspondences with some other system; for example, with another language or group of languages, or with a science other than linguistics, especially logic or a particular field in mathematics. The possibility of MT is discussed, then, in terms of the common elements of the systems compared. These elements may be more or less numerous, but the absence of complete correspondence between the systems, which is usually the case, inevitably suggest limitations in the scope of MT. Thus, the application of machine to translating literary works of art has more than once been declared to be absolutely ruled out ([4], p. 42).

In our opinion, it seems very reasonable to expect that these limitations can easily be eliminated should the problem be formulated in a different way; namely, "whether it is possible, within any existing language, to give a formal description of any of its multiple spheres, individual as they may seem." This comes to the same thing as saying that the applicability of MT depends on whether it is possible to identify the implicit set of rules governing this or that particular sphere of language applications, be it as narrow a sphere as say, Wordsworth's poetry, and further, on whether these rules can be formulated into a formal set. It is apparent that every piece of writing (insofar as written language is discussed) can be analysed on these lines within the sphere to which it belongs, and a set of rules for such analysis can be laid down. It is essential that these rules should be entirely formal. This is no obstacle, since language is only a formal system of specific character developed by man to give communicative expression to his mental activities. In consequence, it is immediately obvious that problems posed by stylistic peculiarities of literary works of art can satisfactorily be resolved, if they are treated on the lines suggested above, i.e. within the sphere to which they belong.

In this light, the supposed "principal informalizability" of poetry [4] should be rejected. On the contrary, poetry, as indeed any piece of literary art where formal elements are of no small importance, is particularly susceptible to machine translation, in this sense. This conclusion has been partially justified on empirical grounds, that is, by experimental translation of passages from Ch. Dickens[4]), J. Galsworthy, J. Aldridge and Edgar A. Poe. (Tables 2, 3 and 4). It is our firm belief, that further investigations will completely eliminate the restrictions now imposed on MT application.

An adequate description of a language, or of any particular sphere of it, should aim finally at establishing within the analysed system a set of correlations of the type

$$\text{means} \gtrless \text{effect}$$

which implies the correlation of linguistic means and their meaning (effect). In its most general sense, the problem

[4]) Illustration to be found in [3].

§1.4] Algebraic and Transcendental Equations 13

Repeated complex roots occur seldom in practical problems and are evaluated by trial and error, since the iteration process converges very slowly (if at all) in this case.

"Squaring the roots," or Graeffe's method,* is frequently more cumbersome than the methods outlined above, particularly in connection with complex roots, but is widely used.

Once all the roots x_i of an algebraic equation have been obtained, the results of the solution may be checked by means of Newton's relations:

$$\sum_{i=1}^{n} x_i = -\frac{a_{n-1}}{a_n}$$ (1.3.3)

$$x_1 \cdot x_2 \cdot x_3 \cdots x_n = (-1)^n \frac{a_0}{a_n}$$ (1.3.4)

For example, in Eq. (a) of this section:

$$\sum_{i=1}^{4} x_i = (1.3 + 1.49i) + (1.3 - 1.49i) + (-15 + 12.5i)$$
$$+ (-15 - 12.5i) = -27.4$$

$$x_1 \cdot x_2 \cdot x_3 \cdot x_4 = 1491,$$

indicating that the imaginary parts of the roots are probably slightly inaccurate.

1.4 Transcendental Equations

Any nonalgebraic equation is called a *transcendental equation*. A transcendental equation may have a finite or an infinite number of real roots, and may have no real roots at all. For example, the equation

$$\sin x = 2$$

has no real roots (Fig. 1.3) but an infinity of complex roots; the equation

$$\sin x = \frac{1}{2}$$

* This method is explained in detail in J. B. Scarborough, *Numerical Mathematical Analysis*, Johns Hopkins Press, Baltimore, 1930, pp. 198 ff., and in R. E. Dougherty and E. G. Keller, *Mathematics of Modern Engineering*, John Wiley & Sons Inc.. New York. 1936, pp. 98 ff.

КРАТНЫЕ КОМПЛЕКСНЫЕ КОРНИ РЕДКО ВСТРЕЧАЮТСЯ В ПРАКТИЧЕСКИХ ЗАДАЧАХ И ВЫЧИСЛЯЮТСЯ ПРИ ПОМОЩИ МЕТОДА ЛОЖНОГО ПОЛОЖЕНИЯ, ПОСКОЛЬКУ ИТЕРАЦИОННЫЙ ПРОЦЕСС СХОДИТСЯ ОЧЕНЬ МЕДЛЕННО (ЕСЛИ ВООБЩЕ СХОДИТСЯ) В ЭТОМ СЛУЧАЕ. "ВОЗВЕДЕНИЕ КОРНЕЙ В КВАДРАТ" ИЛИ МЕТОД GRAEFFE'A, ЧАСТО БОЛЕЕ ГРОМОЗДКО, ЧЕМ МЕТОДЫ, ОПИСАННЫЕ ВЫШЕ, ОСОБЕННО В СВЯЗИ С КОМПЛЕКСНЫМИ КОРНЯМИ, НО ШИРОКО ИСПОЛЬЗУЮТСЯ. ЕСЛИ ВСЕ КОРНИ x_i АЛГЕБРАИЧЕСКОГО УРАВНЕНИЯ ПОЛУЧЕНЫ, ТО РЕЗУЛЬТАТЫ РЕШЕНИЯ МОЖНО ПРОВЕРИТЬ ПРИ ПОМОЩИ СООТНОШЕНИЙ NEWTON'A:

$$\sum_{i=1}^{n} x_i = -\frac{a_{n-1}}{a_n}$$

$$x_1 \cdot x_2 \cdot x_3 \cdots x_n = (-1)^n \frac{a_0}{a_n}$$

НАПРИМЕР, В УРАВНЕНИИ (A) ЭТОГО РАЗДЕЛА:

$$\sum_{i=1}^{4} x_i = (1,3+1,49i)+(1,3-1,49i)+(-15+12,5i)$$
$$+(-15-12,5i) = -27,4$$

$$x_1 \cdot x_2 \cdot x_3 \cdot x_4 = 1491,$$

ПОКАЗЫВАЯ, ЧТО МНИМЫЕ ЧАСТИ КОРНЕЙ, ВЕРОЯТНО, НЕМНОГО НЕТОЧНЫ.

1.4. ТРАНСЦЕНДЕНТНЫЕ УРАВНЕНИЯ

ЛЮБОЕ НЕАЛГЕБРАИЧЕСКОЕ УРАВНЕНИЕ НАЗЫВАЕТСЯ ТРАНСЦЕНДЕНТНЫМ УРАВНЕНИЕМ. ТРАНСЦЕНДЕНТНОЕ УРАВНЕНИЕ МОЖЕТ ИМЕТЬ КОНЕЧНОЕ ИЛИ БЕСКОНЕЧНОЕ КОЛИЧЕСТВО ВЕЩЕСТВЕННЫХ КОРНЕЙ И МОЖЕТ НЕ ИМЕТЬ НИКАКИХ ВЕЩЕСТВЕННЫХ КОРНЕЙ ВООБЩЕ. НАПРИМЕР, УРАВНЕНИЕ

$$\sin x = 2$$

НЕ ИМЕЕТ НИКАКИХ ВЕЩЕСТВЕННЫХ КОРНЕЙ (рис.1.3), КРОМЕ БЕСКОНЕЧНОГО КОЛИЧЕСТВА КОМПЛЕКСНЫХ КОРНЕЙ; УРАВНЕНИЕ

$$\sin x = \frac{1}{2}$$

(SALVADORI, NUMERICAL METHODS IN ENGINEERING, p.13)

Table 2

He could not *say much more* about it, although he felt like telling them *exactly* what this *meant. If he had been a lesser* man he would have been *insulted* by this *telegram. Being Essex,* he could only assume it to be the *illicit interference* of someone *like B. Cooke,* or *A. Cutler.* They had not wanted him *on mission in the first place,* and no doubt they were doing their best to make it impossible for him. He wanted to explain this to *Katherine,* but not a word of it could reach her. J. Aldridge, "The Diplomat"	Он не мог рассказать значительно больше об этом, хотя ему хотелось сказать им в точности, что это значило. Если бы он был более мелним человеном, он был бы оснорълен этой телеграммой. Будучи Эссенсом, он мог только предположить, что это незаконное вмешательство ного-то, вроде Б. Кука или А. Катлера. Прежде всего, они не хотели, чтобы он ехал с миссией, и без сомнения они сделали все возможное, чтобы сделать это невозможным для него. Он хошел обяснить это Кэтрин, но ни одно слово из этого не могло дойти до нее.

Table 3

|	
It was the only way, probably, that such a proposition could have been made to *Soames.* He was *nonplussed. Conscience told* him to *throw* the whole *thing* up. But the *design* was good, and he knew it——there was *completeness* about it, and *dignity; the servants' apartments* were excellent too. He would *gain credit* by living in a *house* like that—with such individual features, yet perfectly *well-arranged.* Galsworthy, "The Man of Property," Part I, Chapter VIII.	Вероятно, это был единственный способ, чтобы такое предложение можно было сделать сомсу. Он был ошарашен. Рассудок велел ему отказаться от всей затей. Но чертеж был хорош, и он знал это — в нем была законченность, и благородство; комнаты для слуг были также отличные. Он приобрел бы уважение, живя в доме, подобном этому — с такими индивидуальными особенностями, и все же совершенно благоустроенном.

Table 4

From *childhood's* hour I have not been As other were—I have not seen As others saw—I could not *bring* My *passions* from a common *spring.* Edgar A. Poe, "Alone"	С часа детства я не был Как другие были — я не видел Как другие видели — не мог черпать Мои страсти из общего источника

Notes on Tables 1, 2, 3, 4

Samples of texts translated in strict accordance with the translational routines devised for MT are given in Tables 1, 2, 3, 4. Words italicised were either not found in the MT dictionary at all, or their meanings were different from those required in the present texts.

translation is, in fact, the problem of equating the correlations of one language with those of another. The procedure can be expressed symbolically by the following chart

$$\text{means}_1 \underset{\longleftarrow}{\overset{\longrightarrow}{\rightleftarrows}} \text{effect}_1$$
$$\parallel$$
$$\text{mean effect}_2 s_2$$
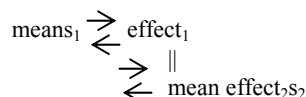
where 'effect$_1$' and 'effect$_2$' are identical, whereas 'means$_1$' and 'means$_2$' differ. In the course of this substitution of one language for another, transposition of semantic content from one language to another is realized.

In conclusion, a word must be added on the problem of the prerequisites of MT. These do not rest upon the existence of common basic elements in languages, as is often stated, but rather on the following two factors:

1) language in itself is only a system of formal means by which meaning is communicated;

2) all existing language systems are developed in order to express in their particular ways any shade of meaning, as well as various emotional effects.

In terms of our symbolization, this comes to saying that the number of "effects" in any two languages is equal, which makes the corresponding systems of "means" fully com-

parable, through their "effects." Since language systems are formal, any application of them can be provided with a description which is programmable on a machine.

3. A short outline of translational routines

The general procedure covered by the translational routines can be broken down into three independent steps, namely:

1) Vocabulary Analysis of the source language for which purpose an MT dictionary and a set of dictionary routines are used;

2) Grammatical Analysis of the source language for which purpose analysis routines are devised;

3) Grammatical Synthesis of the target language for which the same set of synthesis routines is applied to the texts translated from different source languages.

To make the outline concrete, the translation routines will further be described in their Anglo-Russian realization[5]).

[5]) A complete list of translational routines given in the order of their application to be found in Table 8.

Left section:

No.	Ref.		Operation
1	(29a, 2)		Search vocabulary for the whole word
2	(3a, 3b)		Check operand word for "formula"
3a	(15b, 4a)		Check two final letters for 'S
b	(15a, 4b)		
4a	(14p, 9b)		Check final letter for '
b	(14a, 5a)		
5a	(14b, 6a)		Check final letter (of remainder) for S
b	(14d, 30b)		
6a	(16b, 7)		Check three final letters (of remainder) for ING
b	(16b, 14c)		
7	(14c, 8)		Check two final letters for ED or ER
8	(15c, 9a)		Check three final letters for EST
9a	(14b, 10)		Check two final letters for TH
b	(15f, 35)		
10	(16c, 11)		Check three final letters for IAN
11	(14l, 12)		Check final letter for A
12	(14m, 13)		Check final letter for I
13	(17, 34)		Check three final letters for MEN
14a	(28c)	h (28s)	Discard final letter (of remainder)
b	(28b)	k (28t)	
c	(28d)	l (25)	
d	(28a)	m (27)	
e	(28n)	n (20)	
f	(28p)	p (30a)	
g	(28r)		
15a	(28a)	d (28m)	Discard two final letters (of remainder)
b	(30a)	e (28n)	
		f (31)	
c	(24)	h (26)	
16a	(21a)	d (19)	Discard three final letters (of remainder)
b	(28k)	e (28z)	
c	(21c)		
17	(28h)		Change next to the last letter to A

Right section:

No.	Ref.		Operation
18	(28e)		Add IS to remainder
19	(28h)		Add EX to remainder
20	(28g)		Add FE to remainder
21a	(28f)		Add Y to remainder
b	(28h)		
c	(28w)		
22a	(28l)		Add E to remainder
b	(28v)		
23	(28h)		Add IE to remainder
24	(28u)		Add VE to remainder
25	(28y)		Add UM to remainder
26	(28h)		Add ON to remainder
27	(28h)		Add US to remainder
28a	(29b, 30b)	n (29a, 14f)	Search vocabulary for remainder
b	(29a, 6b)	p (29a, 14g)	
c	(29b, 5b)	r (29a, 21b)	
d	(29a, 18)	s (29c, 14k)	
e	(29a, 16a)	t (29c, 22b)	
f	(29a, 14n)	u (29c, 16c)	
g	(29a, 16d)	v (29c, 15c)	
h	(29a, 34)	w (29c, 34)	
k	(29a, 22a)	y (29a, 15h)	
l	(29a, 15d)	z (29b, 34)	
m	(29a, 23)		
29a	(35)		Take information from vocabulary
b	(33)		
c	(31)		
30a	(32)		Develop indication "noun"
b	(33)		
31	(35)		Develop indication "adjective"
32	(35)		Develop indication "plural"
33	(35)		Develop indication "genitive"
34	(35)		Develop indication "unknown word" and store English word to be printed unaltered in Russian sentence
35	(0,1)		Pick up next word without indication "punctuation mark" (PM) and check it for "full-stop"

Table 5. Routine for affix discarding and vocabulary search.

Notes on Table 5

1. The routine of "affix discarding and vocabulary search" is given in full in Table 5. The routine does not include:

a) reconstruction of irregular verb forms, or of irregular forms of nouns, numerals, adjectives and adverbs stored in the MT dictionary;

b) analysis of contracted forms, such as ".. .'ll", ".. ,'d", ".. .'s" in "I'll", "we'd", "he's", etc., since these are not characteristic of scientific texts.

2. As soon as the operand word is found in the MT dictionary it is replaced by its vocabulary form, the indication "inflected" being developed if the discarded affix belongs to the class of starred affixes.

3. The indication PM given only to those punctuation marks which are meant to be analyzed by the punctuation routine.

4. Dictionary analysis

Dictionary analysis of the English sentence starts with a search for every word of the text in the MT dictionary. The first dictionary routine to be used is that for transforming words of the text into the standard forms listed in the MT dictionary (Table 5). Thus "wanted" will be transformed into "want," "stopped" into "stop," "coming" into "come," "lying" into "lie," "copies" into "copy," "bigger" into "big," etc.

When the dictionary search is complete, another routine is applied which treats the words which for various reasons have not been found in the dictionary. These are termed "unknown words" because their lexical equivalents remain unknown throughout the translation procedure. Yet, for the forthcoming grammatical analysis, it is essential that grammatical qualification of the "unknown words" should be obtained. It is impossible to foresee every word in every text of a language, even in one particular sphere, since some of them may be occurring for the first time in the language, not to mention quite a number of more trivial reasons. However, the 'unknown words' do not affect the translation, provided that they have been classified grammatically. To meet this problem a very important routine, that of classifying "unknown words" into "part of speech," has been devised in which extensive use is made of the morphology and syntactic significance of the words.

Another category of sentence constituents which undergoes preliminary grammatical analysis in accordance with a dictionary routine, is the so-called "formula" by which various symbols used different sciences are understood. The syntactic function of every "formula" in the sentence is defined in accordance with a special routine. So much for the words and symbols not found in the MT dictionary.

In addition to lexical equivalents, the words found in the dictionary are provided with information (termed "invariant characteristics") which is partly grammatical, partly semantic in character. For a more detailed description of this information the reader is referred to our earlier publication [3]. The only thing that needs to be mentioned here is that within the 'invariant characteristics' obtained from the dictionary a distinction is made between final and preliminary information. Information is considered to be final for the dictionary cycle when the lexical equivalent of the word is included. Otherwise preliminary information of the word is restricted to the indication "homonymous" or "polysemantic." Special routines have been devised to deal with homonymous and polysemantic words, the analysis of the former words preceding that of the latter.

The four types of homonym analysed by the routine, are adjective-noun" (homonym1), "noun-verb" (homonym2), verb-adjective" (homonym 3) and "preposition-adverb"

(homonym 4). Among homonyms 1, a more complicated sub-type is distinguished, that of "adjective-noun-verb" (homonym 1 complicated). Examples of homonyms are shown below:

CHECK: 1) adjective — контрольный
        2) noun — контроль
        3) verb — "polysemantic"

SQUARE: 1) adjective — квадраный
        2) noun — "polysemantic"
        3) verb — возвести в квадра

In specifying homonyms 1, 2 and 3 a combination of morphological and syntactical analysis of the word is used. Thus, any inflection (except for ER or EST) identified in homonyms 1 or 3 makes "adjective" an impossible alternative, just as ED or ING inflexions in homonym 2 rule out the "noun" solution. These morphological criteria do not, however, find as wide an application as syntactic analysis does in view of the scarcity of inflexions in English. The information which homonyms acquire in the course of this analysis may or may not be final for the dictionary cycle, since some of them are provided here with the indication "polysemantic" instead of with a lexical equivalent.

The total number of polysemantic words stored in our dictionary is about 500. Determination of multiple meaning is performed by specifying typical contexts for polysemantic words, in accordance with a special routine which concludes the dictionary analysis of the English sentence. Since the basic principles of this routine have been discussed elsewhere [3], we do not propose to dwell on them here. However to make this paper comprehensible on its own, two illustrations of context analysis of polysemantic words have been included (Tables 6 and *7*). For details, and for the background of the method, the reader is referred to our earlier publication [3].

5. Target language synthesis

The grammatical processing of a sentence is broken up into two independent steps: analysis and synthesis. The latter is the simpler of the two, and for this reason it is not here that the interest of the problem lies. Therefore the discussion of synthesis will be restricted to a few general comments.

The synthesis routines provide rules for grammatical modification of the translated text in accordance with grammatical information obtained in the course of the analysis of the English original. The most important feature of these routines is their non-comparative nature, which means that both the rules of wordchanging and also certain rules of word-building, are formulated strictly within a particular target language. Because of this, the same synthesis routines can be applied to sentences translated from different source languages. However, synthesis requirements are inclined to increase whenever the routines serve multi-lingual MT purposes. With multi-lingual MT in view, synthesis routines should be:

1) *exhaustive* in describing the target-language word-changing system, since grammatical rules with no application in MT from one language may become vitally important when the source language is changed;

2) *reliable* in carrying through any instruction obtained from the analysis of the source language. This makes it necessary to provide every "nonproductive" category of the target language with a "productive" grammatical equivalent. So far, the problem of grammatical equivalents within a language stands out as the most important theoretical requirement for the synthesis in MT.

3) *independent* of analysis, since the latter may be very different for different languages.

Table 6

ADVANCE	4 (a, c)	Check the operand word for indication "verb"
Homonym 2— 1. Noun Polysem 2 Verb Polysem	a (b, 1)	Check the operand word for indication "inflected", the affix for ED and the following word for indication "noun"
	b (II, III)	Check the selected noun for group "textbook"
	c (IV, V)	Check the preceding word for group "in"
	I (0)	

Table 7

DESIGN Homonym 2 —Verb Polysem	69 (I, a)	Check the nearest following noun (*2a), or the nearest preceding noun (*2b) for group "construction" or group "method"
	a (II, III)	Check the nearest following preposition for indication PPV, or PD, or PA+)

$^+$) PPV: Preposition of passive voice
PD: Preposition of direction
PA: Preposition of aim

Notes on Tables 6, 7

1. Two examples of context analysis carried out by the polysemantic word analysis routine are given in tables 6 and 7.

2. The following symbols are used in describing the routines: A (B, C) means that the next operation is B if the answer is positive and otherwise is C. A (B) means passing to B in both cases, and A (C) means that the final result is reached and that no further search is necessary. The quantities A, B and C may be letters or figures, or a combination of both.

6. Source language analysis

Unlike synthesis, "independent" analysis cannot be recommended, for this would not help to make it economical. Analysis problems are both numerous and important scientifically, and they indeed deserve the special discussion which is given below. English analysis is covered by six routines, which are applied in the order indicated in Table 8. In view of the limitation on the length of the present communication, the discussion is restricted to a general outline of the most important analysis routines, among which the "verb" and "syntax" routines stand out as playing the key part in the whole procedure of the analysis.

6.1 Verb analysis

The verb analysis routine is divided into five sections, the first section being compulsory for every verb of the sentence, whereas only one of the remaining sections is employed for each type of operand verb.

In Section 1 a selection for further analysis is performed. Among the words picked out for analysis in this routine are those which possess the indication "verb," provided that they do not have any of the following indications: "to be disregarded" (D), "not to be changed" (NCh), or the (Russian) indications "participle," "verbal adverb" or "verbal noun." The check for these indications is meant to exclude from further analysis those of the verbs that have been elsewhere provided with characteristics that satisfy the synthesis routines.

In addition to selection, correction of certain verb indications is included in Section 1. Among verb indications liable to correction are tense within the verb-predicates of IF-clauses and case government associated with link-verbs, as well as some more particular indications. Analysis of homonymous forms, such as the past indefinite and subjunctive, of irregular verbs also belongs here. Checks for grammatical context, which imply the possibility of a correction, are performed both when one of the above indications is ascribed to the operand verb in the dictionary and also when the verb is about to be developed in the course of further analysis.
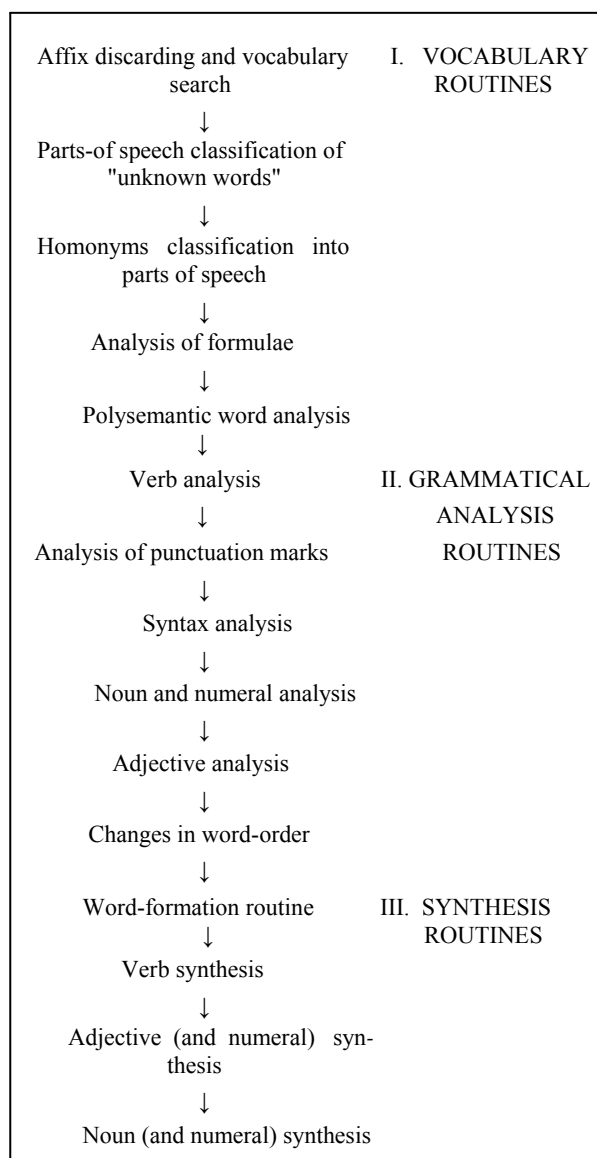
The preliminary checks of Section 1 are followed by the verb analysis proper, for which purpose the operand verb is sent to one of the four different sections, according to its morphological structure. Thus, verbs with S-endings[6]), are sent to Section 2, verbs with ED-endings, as well as certain forms of irregular verbs, to Section 3, verbs with ING-endings, to Section 4, while uninflected verbs are analyzed in Section 5.

Grammatical qualification of "S-verbs" in Section 2 depends on whether S stands out as the only ending of the verb, or whether another ending (usually ING) is associated with it. In the latter case, the following indications are developed for the Russian equivalent verb: "verbal noun; neuter; plural," which implies further analysis by the "noun" routine at the proper time. When S is the only ending, the English characteristics of the verb (predicate in the present indefinite form) are transformed into Russian indications, but not without checking for correction conditions (see above). The resultant characteristic is "predicate," associated with either "present" or "future" tense; the number and person (or gender for the past tense in other cases) of the Russian predicate remain undefined until the subject of the Russian sentence has been determined. The analysis of "ED-verbs", i. e. verbs with ED-endings and certain groups of irregular verbs, is performed in Section 3;

_____
[6]) The term "ending" is applied to affixes following the stem of the word, whereas affixes preceding the stem are called "prefixes."
Note: the term "affix" is restricted to those formatives that are used in wordchanging, whereas formatives used for derivation are termed "suffixes."

Table 8. MT Routines in the order of their application

Affix discarding and vocabulary search	I. VOCABULARY ROUTINES
↓	
Parts-of speech classification of "unknown words"	
↓	
Homonyms classification into parts of speech	
↓	
Analysis of formulae	
↓	
Polysemantic word analysis	
↓	
Verb analysis	II. GRAMMATICAL ANALYSIS ROUTINES
↓	
Analysis of punctuation marks	
↓	
Syntax analysis	
↓	
Noun and numeral analysis	
↓	
Adjective analysis	
↓	
Changes in word-order	
↓	
Word-formation routine	III. SYNTHESIS ROUTINES
↓	
Verb synthesis	
↓	
Adjective (and numeral) synthesis	
↓	
Noun (and numeral) synthesis	

where syntax definitely takes precedence. The four main patterns which are used here for the grammatical analysis of the context of the verb are indicated in Table 9 as Patterns 1 (la, 1b, 1c, 1d), 2 (2a), 3 (3a, 3b), and 4 (4a, 4b, 4c). It should be noted that in all cases context analysis of a word implies the observation of "Rules of Word Selection." These rules are based on classifying all the words in a sentence in three categories, which are:

1) words of third-degree structural significance, which include particles, adverbials, parentheses, and coordinated[7]) parts of the sentence;

2) words of second-degree structural significance, which are words and word groups placed in the attributive position with respect to some word of the sentence;

3) words of first-degree structural significance, which include words not identified as belonging to either of the two previous categories.

By applying the 'Rules of Word Selection' in the course of a search, all words of lower category than the operand word are omitted. Thus the chief constituents of the required grammatical pattern are isolated.

---

[7] The term "coordinated" is applied to those parts of the sentence, which are introduced by a coordinating conjunction or punctuation mark.

This is not the place to give a detailed description of all the processes involved in the analysis of verb patterns in Section 3. For this reason, the discussion will be restricted to just a few comments on those patterns which lead to the most interesting solutions. They are Patterns 1a, 1b, and 2a. Among the different solutions of Pattern 1a we may note that of transforming the English construction of modal passive,

i. e.   modal verb + selected verb, + operand verb
                                    denoted as
                                    "auxiliary 1"
                                    (BE)

into the Russian active compound predicate,

i.e.   modal verb;   + operand verb;
                impersonal   infinitive

the transformation being associated with the conversion of the English subject into the Russian direct object (Tables 10 and 12).

Pattern 2: 2a is provided, among other solutions, with that of transforming an English complex object construction into a Russian subordinate clause.

The resultant characteristics developed for the verbs analysed in Section 3 include both morphological and syntactical information. Of the syntactical indications only "predicate" and "attribute" are fixed here. The indication "predicate" is associated with the morphological indications of mood (indicative, subjunctive and infinitive are developed here), of tense (present or past) and voice (both active and passive are developed). "Attribute" is accompanied by the morphological indications of "participle", of tense (present or past) and voice (active or passive).

ING-forms of verbs are analysed in Section 4. Here the same verb patterns are used, though important changes in their value affect the order in which the analysis is performed. Pattern 1b disappears, whereas Patterns 1d, 3a, and 3b are used much more extensively. Pattern 1d is complicated in order to differentiate quite a number of semantic groups of verbs which are significant for the analysis. There are some modifications in Patterns 4a, and 4b, and another two patterns are introduced, namely Patterns 3c and 5a.

The resultant characteristics of the Russian equivalent verb include one of the following indications: a) "verbal noun, neuter;" b) "participle, present tense, active voice; attribute;" c) "verbal adverb, present (or past) tense"; d) "not to be translated (NT), to be disregarded (D)." In addition to these, "infinitive," "subjunctive" or "indicative mood," with the corresponding set of indications, may be added if the characteristics of the operand verb derive from some other words of the sentence, called the "selected" (helping) words.

Uninflected verbs are analyzed in Section 5. Verb Patterns 1a, 1b, 1c, 1d are replaced here by 1e, 1f, 1g; Patterns 2b, 2c are added to Pattern 2 a, which, itself, is greatly enlarged. Patterns 4a and 5a disappear; instead, Pattern 4c increases considerably, and Patterns 5b-l, 5b-2, 5b-3 are introduced. Among different solutions of Patterns 5b-l and 5b-2 at least two are worth special mention, which deal with the transformation of the English constructions of complex subject and of attributive infinitive into the Russian complex sentence or subordinate clause respectively. The resultant information includes the indication of infinitive, imperative, subjunctive or indicative mood, with the indications of tense (present, past or future) and of voice (active or passive) attached in case of the indicative mood. The only syntactical indication fixed here is "predicate." Verb analysis in the different sections of the routine is based on a classification of the verbs which was devised to characterize English verbs both within the English system and also with regard to the Russian translation traditions. Within the English language, verbs are classified

into "modal" and "half modal" (help, dare); "auxiliary" and 7 sub-classes of "half-auxiliaries;" "causative" (cause, enable, make, order, command, etc.); "declarative" (declare, call, label, report, etc.); verbs taking two objects (give, offer, permit, etc.); etc. To meet the requirement of the Russian translation traditions, verbs are divided into classes and semantic groups. To date, 53 groups of verbs have been established. These are summarized into three classes, the first two classes comprising verbs having translational peculiarities in finite (class I) or infinite (class II) forms; class III covers more complicated cases.

The verb analysis routine is applied until every verb of the sentence is provided with all the grammatical information required in the synthesis routines, except for the indications of number, person (or gender) which are not defined until the subject of the Russian sentence is established. It may be noted that the information obtained in this routine is not restricted to the operand verb, but extends also to some information which is available at this stage of analysis concerning "selected" (helping) words (verb, nouns, adjectives, etc.) and punctuation marks. Moreover, quite a number of transformation in sentence structure are introduced here, such as changes of word-order, insertion of necessary conjunctions and other words, punctuation marks, etc. These transformations are associated with the translation of complex subjects, complex objects, attributive infinitive and gerundival subjects as well as some other verb constructions.

6.2 *Punctuation analysis*

The verb analysis routine is followed by the routine devised to analyze the punctuation marks of the English text, with the exception of those terminating an utterance[8]). This analysis serves to establish both the "English function" (i.e. the function within the English text) of each punctua-

tion mark, and also its "Russian function", by which we imply the corresponding Russian indication. These functions may or may not coincide. If they do not, both the English and the Russian indications are developed, For example, commas marking a prepositional phrase obtain the English indication "CP" (comma, parenthetical), associated with the Russian indication "CD^R" (comma to be disregarded in Russian), as a result of which this comma will not appear in the Russian text. There are also cases in which English punctuation marks should be neglected in the course of the English analysis, although they are rendered by the same marks in the Russian text. This is achieved by developing the indication CD (comma, to be disregarded in English).

6.3 *Syntax analysis*

The syntax analysis routine follows the punctuation routine, since it is essential that the information which can be obtained in both previous routines should be available here. The analysis is carried out in three cycles. In Cycle I, parentheses, comparative AS-phrases, and attributive word-groups with a participle, verbal adverb or adjective as the chief constituent, are marked out by means of appropriate qualification (and insertion when necessary) of punctuation marks. This qualification includes the development of the indication CP$^{b/e}$ (comma, parenthetical, beginning/end) and CA$^{b/e}$ (comma, attributive, beginning/end).

---

[8]) Note that our application of the term "utterance" is at variance with its usual applications. A piece of text, terminated by a full-stop, exclamation or question mark, we call an "utterance" in order to distinguish it from a "sentence" by which we mean only a simple sentence, i. e. a sentence containing not more than one non-coordinated predicate.

Table 9. Patterns of grammatical context used in verb analysis

Preceding word (selected) is 1.VERB	+ operand verb	1a 1b 1c 1d selected verb is 1e 1f 1g	auxiliary I auxiliary II half-auxiliary notional modal half-modal auxiliary III	
2. NOUN	+operand verb	2a verb + 2b verbal particle + 2c CI +	selected noun	
3. ADJECTIVAL	+ operand verb	3a , 3b selected , 3c adjectival	is	pronominal adjective preposition numeral
4.CONJUNCTION	+ operand verb	4a 4b 4c selected 4d conjunction is 4e 4f	CI CH PM PM PM indicated PM	"bracket" "comma" "semicolon" "full-stop"
5. ADVERBIAL	+ operand verb	5a selected 5b adverbial	is	adverb indefinite { 5b—1 verb / 5b—2 noun / 5b—3 adjective } + selected adverbial particle

Test 951   **Table 10**

E	They	insisted	on	a	parallelism	between
N*E	1E	2E	3E	4E	5E	6E
e	Noun Pronominal Personal Plural Nominative	Verb	Homonym 4 \ Polysem.	Adjective Pronominal Indefinite article	Noun Abstract	Preposition
R1	ОН	НАСТАИВАТЬ	НА	N T	ПАРАЛЛЕЛИЗМ	МЕЖДУ
r	Noun Pronominal / Subject Plural Nominative	Verb + Accusative / Predicate Past Plural 3^d Person	Preposition + Prepositional Case	Adjective D_R	Noun Masculine / Singular Prepositional Case	Preposition + Ablative
N*R	1R	2R	3R		4R	5R
шо						
R2	ОН	НАСТАИВАТЬ	НА		ПАРАЛЛЕЛИЗМ	МЕЖДУ
R	ОНИ	НАСТАИВАЛИ	НА		ПАРАЛЛЕЛИЗМЕ	МЕЖДУ

Test 951 (continued)

E	'arithmetical'	and	'general'	algebra	so	rigid	
N*E	7E	8E	9E	10E	11E	12E	
e	Adjective \ Polysem.	Conjunction	Adjective	Noun Mathematical term	Adverb D	Adjective	
R1	„АРИФМЕТИЧЕСКИЙ"	И	„ОБЩИЙ"	АЛГЕБРА	ТАК	ЖЕСТКИЙ	,
r	Adjective / Feminine Singular Ablative	CH	Adjective / Feminine Singular Ablative	Noun Mathematical term Feminine / Singular Ablative	Adverb D	Adjective / Adverbial D	CI SD
N*R	6R	7R	8R	9R	10R	11R	12R
шо							
R2	„АРИФМЕТИЧЕСКИЙ"	И	„ОБЩИЙ"	АЛГЕБРА	ТАК	ЖЕСТКИЙ	,
R	„АРИФМЕТИЧЕСКОЙ"	И	„ОБЩЕЙ"	АЛГЕБРОЙ	ТАК	ЖЕСТКО	,

E	that	,	if	it	could	be	maintained	,	it	would
N°E	13E	14E	15E	16E	17E	18E	19E	20E	21E	22E
e	Adjective* Pronominal	Conjunction	Noun Pronominal Impersonal Singular	Verb Modal Past Predicate	Verb Auxiliary I	Verb	Com	Noun Pronominal Impersonal Singular	Verb Auxiliary of Modal Predicate
	Polysem.			Polysem.		Polysem.			Polysem.	
R1	ЧТО	,	ЕСЛИ	ЭТОТ	МОЧЬ	БЫТЬ	СОХРАНИТЬ	,	ЭТОТ	NT
z	CI	CI SB	CI	Adjective Pronominal Neuter Singular Accusative [Subject*]	Verb Predicate Present Impersonal D	Verb Subjunctive Impersonal Predicate D	Verb + Accusative / Perfective Subject to be converted into direct object / Infinitive Predicate	CI SE	Adjective Pronominal Neuter Singular Nominative Subject	Verb Predicate D
N°R	13R	14R	15R	16R	17R	18R	19R	20R	21R	
wo										
R2	ЧТО	,	ЕСЛИ	ЭТОТ	МОЧЬ	БЫТЬ	СОХРАНИТЬ	,	ЭТОТ	
R	ЧТО	,	ЕСЛИ	ЭТО	МОЖНО	БЫЛО БЫ	СОХРАНИТЬ	,	ЭТО	

E	effectively	destroy	the	generality	;	and	they
N°E	23E	24E	25E	26E	27E	28E	29E
e	Adverb* D	Verb	Adjective Pronominal Definite article	Noun Abstract	Semicolon	Conjunction	Noun Pronominal Personal Plural Nominative
			Polysem.			Polysem.	
R1	СУЩЕСТВЕННЫЙ	РАЗРУШИТЬ	NT	ОБЩНОСТЬ	;	И	ОН
z	Adverbial Adjective D	Verb + Accusative / Predicate Subjunctive Neuter Singular	Adjective D R	Noun Feminine / Singular Accusative	Semicolon Stop	CI	Noun Pronominal / Subject Plural Nominative
N°R	22R	23R		24R	25R	26R	27R
wo							shift
R2	СУЩЕСТВЕННЫЙ	РАЗРУШИТЬ		ОБЩНОСТЬ	;	И	КАЗАТЬСЯ
R	СУЩЕСТВЕННО	РАЗРУШИЛО БЫ		ОБЩНОСТЬ	;	И	КАЖЕТСЯ

Test 951 (continued)   **Table 10 (continued)**

E	never	seem		to	have	realised	fully		that
N°E	30E	31E		32E	33E	34E	35E		36E
e	Adverb D	Link- -verb		Homonym 4 Polysem.	Verb Auxiliary II Polysem.	Verb	Adverb D		Adjective* Pronominal Polysem.
R1	НИКОГДА	НЕ	КАЗАТЬСЯ	, ЧТО	NT	ОСОЗНАВАТЬ	ПОЛНОСТЬЮ	,	ЧТО
z	Adverb D	Par- ticle D	Verb +Ablative Predicate Present Singular 3ᵈ person	CI SD	CS	Verb ——— Predicate Present D	Verb +Accusative ———— Predicate Past Plural 3ᵈ person	Adverb D	CI SD	CI
N°R	28R	29R	30R	31R	32R		33R	34R	35R	36R
wo	shift	shift	PUT COMBINATION (30R)+(31R)+(32R)before(27R)		shift					
R2	,	ЧТО	ОН		НИКОГДА	НЕ	ОСОЗНАВАТЬ	ПОЛНОСТЬЮ	,	ЧТО
R	,	ЧТО	ОНИ		НИКОГДА	НЕ	ОСОЗНАВАЛИ	ПОЛНОСТЬЮ	,	ЧТО

Test 951 (continued)

E	a			formula		true	with	one
N°E	37E			38E		39E	40E	41E
e	Adjective Pronominal Indefinite article Polysem.			Noun Mathematical term		Homonym 3 Polysem.	Homonym 4 Polysem.	Numeral* Polysem.
R1	NT			ФОРМУЛА		ВЕРНЫЙ	ПРИ	ОДИН
z	Adjective D_R		CI SD	Noun Mathematical term Feminine ———— Subject Singular Nominative	CAᵇ	Adjective ———— Feminine Singular Nominative	Preposi- tion +Prepo- sitional Case	Adjective ———— Feminine Singular- Prepositional case
N°R			37R	38R	39R	40R	41R	42R
wo	SHIFT	SHIFT OF COMBINATION (46E-49E) PRECEEDS INSERTION OF COMMA (37E)						
R2	ВПОЛНЕ	ВОЗМОЖНЫЙ	, ЧТО	ФОРМУЛА	,	ВЕРНЫЙ	ПРИ	ОДИН
R	ВПОЛНЕ	ВОЗМОЖНО	, ЧТО	ФОРМУЛА	,	ВЕРНАЯ	ПРИ	ОДНОЙ

E	interpretation	of	its	symbols	is	quite	likely	to
N°E	42E	43E	44E	45E	46E	47E	48E	49E
e	Noun Abstract	Prepo-sition Polysem.	Adjective Pronominal Possessive	Noun	Verb Auxiliary I Predicate Present Singular 3^d person Polysem.	Adverb D	Adverb* D	Homo-nym 4 Polysem
R1	ИНТЕРПРЕТАЦИЯ	NT	ОН	СИМВОЛ	NT	ВПОЛНЕ	ВОЗМОЖНЫЙ	ЧТО
z	Noun Feminine ----- Singular Prepositional case	Prepo-sition +Ge-nitive	Adjective Prono-minal Genitive ---- Feminine Singular	Noun Masculine ---- Plural Genitive	Verb + Nominative ---- Predicate Present Singular 3^d person	Adverb D	Adjective Adverbial D	CI
N°R	43R		44R	45R		46R	47R	48R
шо					PUT COMBINATION (46E) + (47E) + (48E) + (49E) BEFORE (38E)			
R2	ИНТЕРПРЕТАЦИЯ		ОН	СИМВОЛ		ВПОЛНЕ	ВОЗМОЖНЫЙ	ЧТО
R	ИНТЕРПРЕТАЦИИ		ЕЕ	СИМВОЛОВ				

E		be	false	with	another	
N°E		50E	51E	52E	53E	54E
e		Verb Auxiliary I Polysem.	Adjective	Homonym 4 Polysem.	Adjective Pronominal	Full-stop
R1		БЫТЬ	НЕВЕРНЫЙ	ПРИ	ДРУГОЙ	
z	CA^e	Verb +Ablative ---- Predicate Future Singular 3^d person	Adjective ---- Feminine Singular Ablative	Preposition + Preposi-tional case	Adjective ---- Feminine Singular Prepositional case	Full-stop
N°R	49R	50R	51R	52R	53R	54R
шо						
R2	,	БЫТЬ	НЕВЕРНЫЙ	ПРИ	ДРУГОЙ	.
R	,	БУДЕТ	НЕВЕРНОЙ	ПРИ	ДРУГОЙ	.

Test 530   **Table 11**

E	Use	the	Lagrange	variation	–	of	–	parameters
N°E	1E	2E	3E	4E	5E	6E	7E	8E
e	Verb	Adjective Pronominal Definite article	Noun Animated Proper Masculine Singular	Noun Mathematical term	Hyph	Preposition	Hyph	Noun Mathematical term
	Polysem.	Polysem.		Polysem.		Polysem		
R1	ПРИМЕНИТЬ	N T	ЛАГРАНЖ	ВАРИАЦИЯ		N T		ПАРАМЕТР
r	Verb Accusative ___ Predicate Imperative	Adjective D_R	Noun Animated Proper Masculine ___ Singular Genitive	Noun Mathematical term Feminine ___ Singular Genitive	Hyph CD_R	Preposition + Genitive	Hyph CD_R	Noun Mathematical term Masculine ___ Plural Genitive
N°R	1R		2R	3R				4R
шо			PUT (2R) AFTER (4R)	PUT COMBINATION (3R) + (4R) + (2R) AFTER (5R)				
R2	ПРИМЕНИТЬ		МЕТОД	ВАРИАЦИЯ				ПАРАМЕТР
R	ПРИМЕНИТЕ		МЕТОД	ВАРИАЦИИ				ПАРАМЕТРОВ

Test 530 (continued)

E	method	to	solve	the	inhomogeneous	n th	–	order
N°E	9E	10E	11E	12E	13E	14E	15E	16E
e	Noun	Homonym 4	Verb	Adjective Pronominal Definite article	Adjective	Adjective Ordinal Formula	Hyph	Noun
		Polysem.		Polysem.				Polysem.
R1	МЕТОД	К	РЕШИТЬ	N T	НЕОДНОРОДНЫЙ	$n^{ый}$		ПОРЯДОК
r	Noun Masculine ___ Singular Accusative	Preposition +Dative	Verb +Accusative ___ Verbal Noun Singular Dative	Adjective D_R	Adjective ___ Masculine Singular Genitive	Adjective ___ Masculine Singular Genitive	Hyph CD_R	Noun Masculine ___ Singular Genitive
N°R	5R	6R	7R	8R		9R		10R
шо	SHIFT					PUT COMBINATION (9R) + (10R) AFTER (12R)		
R2	ЛАГРАНЖ	К	РЕШИТЬ	НЕОДНОРОДНЫЙ		ДИФФЕРЕНЦИАЛЬНЫЙ		
R	ЛАГРАНЖА	К	РЕШЕНИЮ	НЕОДНОРОДНОГО		ДИФФЕРЕНЦИАЛЬНОГО		

E	differential	equation		$y^{(n)}+a_1y^{(n-1)}+...+a_ny=f(t)$ (a)	.
N°E	17E	18E		19E	20E
e	Homonym 1	NOUN		Formula	Full-stop
R1	ДИФФЕРЕНЦИАЛЬНЫЙ	УРАВНЕНИЕ		$y^{(n)}+a_1y^{(n-1)}+...+a_ny=f(t)$ (a)	.
z	Adjective Mathematical term / Neuter Singular Genitive	Noun Neuter / Singular Genitive		Formula D_E	Full-stop
N°R	11R	12R		13R	14R
wo	SHIFT	SHIFT			
R2	УРАВНЕНИЕ	n ый	ПОРЯДОК	$y^{(n)}+a_1y^{(n-1)}+...+a_ny=f(t)$ (a)	.
R	УРАВНЕНИЯ	n ого	ПОРЯДКА	$y^{(n)}+a_1y^{(n-1)}+...+a_ny=f(t)$ (a)	.

E	The	examples	of	Chapters	V	and	VI	,	in	which	Lagrange's
N°E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E
e	Adjective Pronominal Definite article / Polysem.	Noun	Preposition / Polysem.	Noun	Formula	Conjunction	Formula	Conj	Homonym 4 / Polysem.	Adjective Conjunctive	Noun Animated Proper Masculine Singular
R1	NT	ПРИМЕР	ИЗ	ГЛАВА	V	И	VI	,	В	КОТОРЫЙ	ЛАГРАНЖ
z	Adjective D_R	Noun Masculine / [Subject *] Plural Accusative	Preposition + Genitive	Noun Feminine / Plural Genitive	Formula D_E	CH D_E	Formula D_E	CI SB	Preposition + Prepositional case	Adjective Conjunctive / Substantivized Feminine Plural Prepositional case / i	Noun Animated Proper Masculine Singular Genitive
N°R		1R	2R	3R	4R	5R	6R	7R	8R	9R	10R
wo											PUT (10R) AFTER (11R)
R2		ПРИМЕР	ИЗ	ГЛАВА	V	И	VI	,	В	КОТОРЫЙ	УРАВНЕНИЕ
R		ПРИМЕРЫ	ИЗ	ГЛАВ	V	И	VI	,	В	КОТОРЫХ	УРАВНЕНИЯ

Test 885 (continued)   Table 12 (continued)

E	equations	are	used	,	can	be	treated	without
N°E	· 12E	13E	14E	15E	16E	17E	18E	19E
e	Noun	Verb Auxiliary I Predicate Present Plural Polysem.	Homonym 2 Polysem.	Com	Verb Modal Predicate Present Polysem.	Verb Auxilia-ry I. Polysem.	Verb -	Preposition Polysem.
R1	УРАВНЕНИЕ	NT	ИСПОЛЬЗОВАТЬ		МОЧЬ	NT	РАССМАТРИВАТЬ	БЕЗ
z	Noun Neuter / Plural Nominative Subject	Verb +Nominative Predicate Present Plural D	Verb +Accusative Predicate Present Passive Plural 3d person	CI SE	Verb Predicate Present Impersonal D	Verb Predicate Infinitive D	Verb +Accusative Predicate Infinitive Subject to be converted into direct object	Preposition +Genitive
N°R	11R		12R	13R	14R		15R	16R
wo	SHIFT							
R2	ЛАГРАНЖ		ИСПОЛЬЗОВАТЬ	,	МОЧЬ		РАССМАТРИВАТЬ	БЕЗ
R	ЛАГРАНЖА		ИСПОЛЬЗУЮТСЯ	,	МОЖНО		РАССМАТРИВАТЬ	БЕЗ

Test 885 (continued)

E	reference	to	the	Lagrangian	method	.	
N°E	20E	21E	22E	23E	24E	25E	
e	Noun	Homonym 4 Polysem.	Adjective Pronominal Definite article Polysem.	Noun Animated Proper Masculine Singular	Noun	Full--stop	
R1	ССЫЛКА	НА	NT	ЛАГРАНЖ	МЕТОД	·	
z	Noun Feminine / Singular Genitive	Preposi-tion +Accu-sative	Adjective DR	Noun Animated Proper Masculine Singular Genitive	Noun Masculine / Singular Accusative	Full--stop	
N°R	17R	18R	19R	20R	21R		
wo			PUT(19R) AFTER(20R)		SHIFT		
R2	ССЫЛКА	НА		МЕТОД	ЛАГРАНЖ		
R	ССЫЛКИ	НА		МЕТОД	ЛАГРАНЖА		

Attributive groups are not isolated until preliminary checks for certain patterns of grammatical context have been carried out. Among these patterns are the following three:

1) preceding word is CA [b];
2) preceding word is a noun; (provided that it is associated pattern 3);
3) following word is: 3a) preposition
                          3b) conjunction (or conjunctive word)
                          3c) punctuation mark (PM)
                          3d) verb indicated "participle, short form."

In practice, isolation of these word-groups reduces to establishing their right word limit ("end"), since the left hand limit ("beginning") in these cases can easily be associated with the chief constituent of the construction. The 'end' of the isolated word-group is found by searching to the right of the chief constituent until the nearest of the following is encountered:

a) CA [e]
b) noun with indication (or condition of) "subject"
c) verb with indication "predicate"
d) conjunction without the indication "coordinating"
e) punctuation mark without indications CP or "b" ("beginning")

It is essential that the search should be performed in the order indicated above[9]).

In Cycle 2, sentence limits are established by checking the utterance[10]) for the presence of:

1) conjunctions with the indication "non-homogeneous" (CI)
2) conjunctive words (nouns or adjectives)
3) words with indication "initial"
4) two (or more) predicates within a passage terminated by sentence limits already established
    a) immediately following each other
    b) not immediately following each other
5) two nouns following each other, but not joining in a "lawful" combination.

A very detailed analysis of every pattern is carried out in the order indicated above. Normally, sentence limits do not acquire the indications SB (sentence beginning) or SE (sentence end) in this stage of the analysis, except in two cases:

a) when the pattern analyzed is
    non-homogeneous   adjective   (absence
    conjunction   +   marked as   +   of ;
    "of condition"   "predicative"   noun)
b) when two conjunctions follow each other, the second being provided with a correlative conjunction or conjunctive word.

At this stage of the analysis certain changes in the structure of the Russian text are also provided. In this connection, mention should be made of the insertion of the conjunctive word (with appropriate indications) when it is omitted in the English attributive clause (Patterns 4 and 5). In Cycle 3 the information so far obtained is used to qualify sentence limits by indicating "beginning" or "end" for any sentence inserted within the limits of another sentence. Among other more particular cases, sentences in which the subject and predicate are separated by an attributive or other subordinate clause are analysed here.

[9]) Note that wherever following or preceding words are examined the "Rules of Word Selection" are strictly observed in the analysis routines.
[10]) See above for our definition of the term.

The position (initial, middle, or final) of the nonpredicative piece of the broken sentence determines the order in which the parts are treated. The limits of the sentence which have not been recognized as "insertions" within other sentences are qualified as SD (sentence division), since neither "beginning" nor "end" indications are considered necessary here.

The information obtained by application of the "syntax analysis" routine is extremely valuable, for as long as syntactical units are marked out for further analysis Nouns, numerals and adjectives are analyzed within these units, the order in which syntactical units are treated being as indicated below:

1) a sentence (minus all parenthetical and comparative or attributive word-groups)
2) comparative and attributive word-groups (minus parenthetical word-groups) within the sentence
3) parenthetical word-groups within the sentence
4) next sentence (minus all parenthetical and comparative word-groups).

Step 4 is again followed by Steps 2, 3, 4 etc., until the last sentence is studied.

6.4 *Noun analysis*

The noun analysis routine is devised so as to cover the analysis of two word classes, nouns and numerals. The ordinal numbers are regarded as adjectives [5], so only cardinal numbers are termed numerals here. They are not entered into the class of nouns, owing to their morphological peculiarities. The routine is divided into two parts, the development of case indication being the object of Part I. whereas in Part 2 the indication of number is developed. The two parts differ in scope as well as in method. In Part 1, in which both nouns and numerals are treated, syntactical methods are used, since qualification of nouns inflected with "S", " ' " or "IAN" has been achieved at an earlier stage (Table 5). The grammatical context of the operand noun or numeral is checked for the presence of some "governing" or "coordinating" element preceding the operand word. Prepositions, verbs, verbal nouns and numerals belong to the "governing" group, whereas conjunctions and punctuation marks with the indication "homogeneous", and conjunctions of comparison are considered to be "coordinating". In all cases the indication required is taken from one of the preceding words, either governing, or coordinated with, the operand. If there is no governing or coordinating word, other patterns are applied, special attention being given to conjunctive nouns. Nouns are the only class of word analysed in Part 2 of the routine, in which morphological methods provide the most important information for developing the indication of number. If the word is inflected the number is defined as "plural," otherwise syntactical methods are applied. It should be noted that patterns of grammatical context are introduced at this point so as to reflect the peculiarities of number and case forms in the English and Russian languages. Thus with number indications, differences between English and Russian in classifying nouns into "countables" and "uncountables" are taken into account. Among other idiomatic constructions in Russian, that of the combination "numeral and noun" in which the numeral has retained the old "dual" government, should be pointed out. Certain peculiarities in Russian verb government are also considered. Syntactically, nouns and numerals, are classified only when used in the function of an attribute or subject of the sentence.

6.5 *Adjective analysis*

The adjective analysis routine is the last of the series English analysis routines which develop grammatical indications to be used in the Russian synthesis routines.

Information acquired from the adjectives (and participles) of the text includes the indications of gender, number, case, degree of comparison and "short/full" form. In addition to these, the indications of "substantivized" or "adverbial adjective" are developed.

Preliminary checks for the absence of the indications required are followed by an examination of the morphological structure and syntactical environment of the operand word. The main interest of the testing procedure lies in determining whether the operand is placed in the attributive position towards some noun of the sentence[11]). If the result of the search is positive, it becomes very important to pick up the right noun, which in some cases is not a very easy task. Another important search is aimed at establishing a predicative position for the operand. Finally, if both searches are negative, the word is qualified as 'substantivized.'

6.6 *Change-of-word-order*

The change of word order routine is designed to give a final touch to the translated text before the synthesis routines are applied. English patterns of word-order which do not correspond to Russian patterns are recomposed. It is interesting, however, that these rearrangements are usually of local character. The most important word-order changes performed in this routine[12]) are due to the difference in the position of attributes expressed by nouns or noun combinations (Tables 11 and 12) and to differences in the expression of negation in the English and Russian languages. Other changes are of no particular importance.

7. Conclusions

The heart of the whole method suggested above lies in the most careful description of every language included in the MT system followed by a very detailed comparison of these descriptions. The latter is the basis of MT research. The comparison of the English and Russian languages in the course of MT studies has proved to be more fruitful than has been expected, and the structure of these languages was found to be strikingly alike, even in a great many of details. For this reason, an attempt was made to work out an Anglo-Russian MT scheme which would make maximum use of similarities of structures in the two languages. For this reason, structural transformations of the translated text have been restricted in the present scheme to MT to such insignificant changes as omission and insertion of just a few helping words and punctuation marks, and a few (local) word-order changes. Nevertheless, the translations thus obtained are quite adequate for understanding, and do not require post-editing, as can be seen in the samples cited (Tables 1, 2, 3, 4).

8. References

[1] PANOV, D.: *Concerning the Problem of Machine Translation of Languages.* Moscow: 1956.

[2] MUKHIN, I. S.: *An Experiment of the Machine Translation of Languages Carried Out on the BESM.*

[3] BELSKAYA, I. K.: *Machine Translation of Languages. Research* vol. 10, October 1957. pp. 383-389.

---

[11]) Sometimes it can be a noun of another sentence, as in the case of conjunctive adjectives.

[12]) Certain more specialized word-order changes are performed at an earlier stage of analysis (6.1 and 6.2).

[4] Тезисы конференции по машинному переводу. Москва, 1958 г.

[5] Сборник статей по машинному переводу. ИТМ И ВТ АНСССР, (в печати).

9. Discussion

O. F. Koulagina (USSR): Miss Koulagina talked about the "Conference on Mathematical Linguistics" which took place in Leningrad on April 15-21, 1959. In her speech she said that there were 78 participants from Leningrad, Moscow, Kiev, Gorki, Erevan, Tbilisi, Prague, and Peking, and that a total of 59 papers was presented. She also outlined briefly the general topics which were discussed at the conference.

S. Gorn (USA): I should like to comment on the negative answer I have been receiving from most of the members of the panel to the following fundamental question: Suppose that, during the process of translation, your program picks up a sentence of the following type: "We will now define the word X to mean ..." Does your program recognize that the words "define" and "mean" are not merely to be translated but must also signal the program to change itself, and to change the dictionary by automatically adding the word X to it ?

The answer has uniformly been "no". Where the significance has been recognized, the statement was still made that such a requirement was a luxury and had a low priority. This seems to me to be wrong, because the one power that general purpose computers and their command languages have is this ability of self-reference.

Let me explain by a diagram:
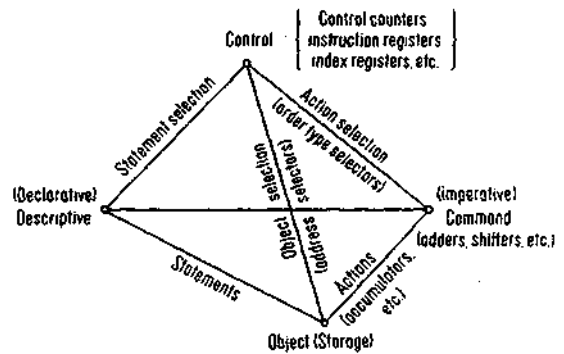
The areas of a mechanical language

In the diagram, the "object" area of a language (or machine), which is the area in which all words are nouns, there are two types of syntax sub-languages: the descriptive syntax, and the command syntax.

The reason why such a language system does not need to be stratified, is that there is a "Control syntax" which changes the context of words by moving them from one area to another, and even to and from the "Control" itself. In a machine this is done by transfer and control commands and the common storage. In a language the Control syntax includes all words referring to the relationship between symbols and their meaning. Examples are the words "define", and "mean", and quotation marks. Why should MT not be taking immediate advantage of a fundamental capability its tool already possesses?