Soviet Developments in Machine Translation: Russian Sentence Analysis †

T. M. Nikolaeva, Institute of Precise Mechanics and Computing Technique, Moscow, U.S.S.R.

The principles of Russian sentence analysis in MT are discussed. The description of various methods of receiving the grammar and vocabulary information for every word analyzed is given. The syntactic analysis of the Russian sentence is described.

Primary Analysis of the Russian Sentence

THE ANALYTIC PART of MT in translation from Russian is a system of routines that work out the grammatical and syntactic features of the words needed for translation into another language.

The entire analysis of Russian breaks down into two large, independent parts: a dictionary and a grammatical analysis. Each word in the sentence to be analyzed is examined in the dictionary, after which the word with the appropriate lexical information passes on to the grammatical analysis.

This information indicates the part of speech to which the word belongs, the characteristics of that part of speech, and finally the specific morphological properties of the word. For example, inherent features of the noun are: gender, membership in one of three declensions, relationship to the category of animateness, as well as stem type. Characteristics of the verb are conjugation, quality of stem, and the possibility of being transitive.

We have divided words having specific derivational characteristics into separate groups. An indication of the number of a group will constitute the dictionary information about the derivational properties of the word to be analyzed.

Our dictionary is peculiar in that each word includes only information about its grammatical nature and place in the grammatical system of the language as a whole. This is a special type of dictionary, one differing substantially from the familiar kinds.[1]

Our dictionary lacks information about the semantic side of the word, that is, about its particular meaning. This is explained by the unusual role played in our practical work by Russian, which serves as an intermediary language.

Since the analytic part of translation from Russian in our work is the same for translation into any language whatsoever, it would be useless to give the "translation" of a word in the Russian dictionary, inasmuch as it forms the content of the synthetic part of the dictionaries for the various languages.

In order to obtain the desired dictionary information, it is necessary to reduce the word being analyzed to the form in which it may be found in the dictionary, i.e., to the so-called dictionary form.

The dictionary form of the noun is the nominative case, singular number; of the adjective the nominative case, singular number, masculine gender; of the verb — the infinitive; of the numeral and the pronoun — the nominative case. The remaining, invariable parts of speech have only one form, which is also the dictionary form.

---

1.   Cf. L. V. Shcherba, "An attempt at a general theory of lexicography, " Izvestija   AN SSSR, Division of Literature and Language, no. 3, 1944.

If a sentence contains a word in the dictionary form, the word receives the sign FS (Forma Slovarnaja 'dictionary form') and for subsequent analysis passes on to the following routines in order to avoid premature conclusions about the "contextual" features of the given word.[2]

For example, the FS of the word dom 'house' may be either nominative or accusative case, singular number, while the FS of the word soldat 'soldier' may be either nominative case singular number or genitive case, plural number. Similar homonymic forms are distinguished in the appropriate routines.

In the routine designed to reduce a word to the dictionary form, the words are handled in accordance with their endings, which in Russian are adequately distinctive for the various parts of speech. The inflectional structure of Russian and the highly developed system of derivational suffixes along with virtually non-existent infixation and little homonymy of inflection contribute to the rather prompt recognition of the part of speech to which the word being analyzed belongs and help to supply it with the form needed to locate it in the dictionary.

For example, if a word has the ending -ymi or -emu in a given passage, presumably it can only be an adjective or participle.[3] Therefore, if after changing the inflection to -yĭ in the first case or to -iĭ in the second case the word still does not appear in the dictionary, the added ending is rejected and the ending of the remaining part checked for one of the participial suffixes. The word is then given the form of a verb, since the infinitive form of the corresponding verb is the dictionary form of a participle. It is less complicated to recreate the dictionary form of a verb found in context in the dictionary form since the endings of verbs in the personal form are almost non-homonymic.

However, inflectional homonymy is a rather complex phenomenon. Even in Russian where it is comparatively slight, it causes substantial

difficulties. For example, the ending -i may indicate: 1) the plural number of soft-stem short-form adjectives, e.g., sini 'blue'; 2) the imperative form of verbs, e.g., zhivi 'live'; 3) various cases of the noun koni 'horses', knigi 'book,' etc. In such cases the word undergoing analysis is treated in several blocks successively where the various endings of the dictionary forms are generated that are possible for a given ending of the parts of speech — until the word is found.

Stem alternation in many Russian words constitutes another difficulty in reducing a word to its dictionary form. This is characteristic of verbs e.g., beru-brat' 'to take,' greb-gresti 'to rake,' etc. adjectives (dolog-dolgi 'long,' uzka - uzok 'narrow,') and numerals (vosem' - vos'mi ' eight').

However, despite their seeming variety the number of such variations is rather small, and the alternation affects a limited number of vowels (o-zero, e-zero, '-e, etc.). This makes it possible to use uniform methods of reducing such words to their stem form. In the case of alternation of consonants (archaic forms of the past tense of verbs) the number of such variations is even less since the infinitive of all verbs with irregular past tense endings can end only in -eret' (umer -umeret' 'to die'), -nut' (sokh - sokhnut' 'to dry'), or most commonly, in -sti. Therefore, in the following verb types. greb 'he raked,' nes 'he carried,' or mel 'he swept' ending in sti, the same simple command serves to reduce all these verbs to the dictionary form: "Reject the last letter and add the ending -sti."

Suppletion of individual forms of several verbs, pronouns, and non-pronominal substantives is taken into account by entering the suppletive form directly into the dictionary.

Accordingly, the entire routine breaks down into a number of separate parts more or less corresponding to the division in the parts of speech. We must mention the fact that practical necessity compelled us to make subgroups of nouns with endings in -ie, -iĭ, and -ija in the dictionary form, nouns in -mja, and degrees of comparison in -zhe and -she.

Since homonymic endings are analyzed in several parts, a word passes from one part to another until the final stage. The first attempt at a routine based on the endings themselves proved too clumsy to be practicable.

A word that cannot be found in the dictionary after going through the entire routine remains in Russian letters in the translation.

---

2. "Contextual" features are speech, not language, phenomena — for example, the particular case and number of a noun or tense, voice, mood, number, and person of a verb in every sentence.

3. Whether or not the ending refers to a pronoun (to nemu) is readily detected by checking for initial n before -emu.

The routine examined above makes it possible not only to obtain the dictionary form of a word by which it can be located in the dictionary, but also to find out its various contextual and grammatical features. For example, treatment of the word <u>delaesh'</u> ' you do' permits the word to be located in the dictionary and produces the following tags: second person, singular number, present tense, and predicate.

Thus, the contextual tags of words with non-homonymic inflections are ascertained in the very first routine.

The contextual features of words with homonymic inflections are ascertained in the subsequent routines on the basis of tactic and syntactic principles of context analysis.

The group of operations used in our work to reduce words to their dictionary form, obtain dictionary information, and analyze non-homonymic inflections is called the "primary analysis."

Determination of Morphological Tags

Certain features of the materials we studied — chiefly mathematical literature — make it necessary to devise a special routine to analyze the function of signs that are not words written in Russian letters.   We have agreed to call these signs formulas whether they are formulas in the usual sense of the word or symbols of something in non-Cyrillic letters.

The need to devise this routine arose from the fact that the so-called "formula" is not a "foreign" body within a unilingual flow of speech capable of being mechanically translated from one language into another like chapter numbers, figures, etc., but a full and equal member of the sentence performing the function of some part of speech.

Hence, the purpose of the routine to be described is to determine the part of speech to which any formula encountered in a text may functionally belong. A tactic principle underlies this routine, namely, confirmation of the meaning-differentiating role of word order in ascertaining the part of speech of the invariable word.

If the formula under analysis contains the sign $<$, $>$, $=$, $\neq$, or $\rightarrow$ functioning as a predicate, this formula receives the sign "sentence" and undergoes no further treatment. If the formula is preceded by an adjective or verb, it receives the sign "noun"; by a noun, the sign "invariable adjective." If directly followed by a noun, the formula receives the sign "numeral."

After going through this routine, all formulas lacking the sign "sentence" are examined for production of the required contextual signs just like ordinary words.

After being subjected to the two routines mentioned above, all words in a sentence will have an indication of the part of speech to which they belong and dictionary information: words that in the given context have non-homonymic endings will also have some contextual features. Subsequent analysis is to obtain the contextual tags of words with homonymic endings and invariable words and to ascertain the syntactic function of each word.

More than the dictionary information about the words and their ending in context is needed in order to obtain this information.   The place of each word in the text as a whole, morphological features of the surrounding words, and   syntagmatic connections must also be determined. For purposes of investigation such an analysis   requires larger semantically self-contained units than the individual words thus far discussed. We call these units "clauses" (<u>Predlozhenie</u>).

We call "sentences" (<u>Fraza</u>) segments of written text divided by several marks of punctuation.   Smaller portions of sentences known in   conventional grammar as "subordinate clauses" (<u>Pridatochnoe Predlozhenie)</u> or parts of "compound sentences" <u>(Slozhnosochinenoe Predlozhenie)</u> are, in our terminology, "clauses" proper.

In order to analyze the relations of words within a clause we must first isolate it, i.e., locate the beginning and end of each self-contained semantic unit.

We cannot use the existing punctuation marks as dividers since they do not always   indicate the beginning or end of a complete thought. They may accompany so-called "parenthetical" words, which carry no syntactic load, or introduce homogeneous members or even simply stress a given word,   as in the case of the so-called "sense dash" (<u>Smyslovoe</u> <u>Tire</u>).

We have therefore divided all the marks of punctuation into two large groups — homonymic and non-homonymic. The non-homonymic marks are the period, question mark, and exclamation point, which always   separate individual sentences or entire   segments of meaning. The other marks may separate either whole clauses or individual words.

Accordingly, we have designed for the analytic part of our work a special routine for punctuation mark analysis to isolate clauses as self-

contained units. Each mark is checked for its relation to homogeneous and heterogeneous con-junctions, the presence of parenthesis at the words next to the mark, and the presence of a verb with the sign "personal form" to the right or left of the word in question. Depending on the presence of such signs, each mark of punc-tuation is provided with one of the following signs: "heterogeneous" (Neodnorodnyĭ) (i.e., introducing a subordinate clause), "parenthetical" (Vvodnyĭ) (introducing parenthetical words, par-ticiples, and gerunds), "homogeneous complex" (Odnorodno-Slozhnyĭ) connecting parts of a com-pound sentence) or "homogeneous simple" (Odnorodno-Prostoĭ) separating homogenous members of a sentence). The break-up of a sentence into its individual units of meaning follows the generation of these signs. Analysis by the routines then continues within the clauses thus obtained.

As mentioned above, several contextual signs are ascertained in the first routine. Verbs, the forms of which are for the most part non-homonymic, get the largest number of tags. On-ly the voice and mood must be ascertained since the features of tense, number, and person of verbs in the personal form have already been determined by the primary analysis.

Analysis of the mood of verbs presents no special difficulty. The imperative is deter-mined by analyzing verb conjugations, while the endings of the subjunctive mood are identi-fied by the presence of the subjunctive particle -by in the clause.

The most complicated problem is that of for-mal demarcation of the active and passive voices. We have distinguished two types of passive voice: processual - imperfective aspect (e.g., dom stroitsja 'the house is being built') and resulta— tive - perfective aspect (e.g., dom postroen 'the house has been built'). The passive voice of the perfective aspect, which is formed by the short passive participle of the past tense can be readily distinguished, whereas the passive voice of the imperfective aspect, which is formed by addition of the particle -sja to the personal form of the active voice, is sometimes formally almost indistinguishable from the active voice of verbs that have the particle -sja in the infini-tive . Various studies of the problem of voice distinctions in Russian usually refer only to the "polysemy" of the particle -sja, but fail to pro-vide criteria for determining the cases where -sja gives the verb purely passive meaning.

The "classical" passive construction in Indo-European languages is a combination of passive subject with verb in the passive voice and agent subject in the instrumental case (e.g., kniga chitaetsja studentom 'the book is read by the student'). This type of construction is rather uncommon in Russian where the active voice predominates and the absence of a subject is expressed, for example by an indefinite -personal sentence. Moreover, even if such a "classical" passive construction occurs, there are cases where structural and sentence homo-nymy arise. Let us consider, for example, two structurally identical sentences: mal'chik prichesyvaetsja shetkoĭ 'the boy brushes his hair with a brush" and fraza obrabatyvaetsja skhemoĭ 'the sentence is treated in accordance with the routine.' In the first sentence the pre-dicate has the form of the active voice, while in the second the verb, which is externally si-milar in form to the other verb, obtains the sign of the passive voice.

We started with the dictionary properties of the verbs themselves in our attempt to solve this problem, dividing all verbs capable of re-ceiving the formant -sja into four main cate-gories .

1) Verbs in which the particle -sja consti-tutes an integral dictionary feature. These verbs, which have only active meaning, make up a group III-ag (e.g., gordit'sja 'to take pride in,' ochutit'sja 'to find oneself' etc.)[4]

2) Verbs in which the addition of -sja is a method of producing a passive meaning. These verbs make up group III-vg (e.g., vyrabatyvat' - vyrabatyvat'sja 'to manufacture - to be man-ufactured,' stroit' - stroit'sja 'to build - to be built').

3) Verbs in which the addition of -sja is a method of producing a reflexive meaning. These verbs make up group III-bg (e.g., myt' - myt'-sja 'to wash - to wash ourself').

4) Verbs in which the addition of -sja indi-cates a total change in lexical meaning. These verbs make up group III-gg (e.g., risovat' - risovat'sja 'to draw - to show off,' tronut' - tronut'sja 'to touch - to spoil').

---

4. The first letter of the symbol for this group represents the serial number, the second, the initial letter of the name of the part of speech. (Transl. note: g = glagol 'verb')

Thus when the verbs of group No. III-ag have the particle -<u>sja</u> they automatically receive the sign "active voice," whereas the verbs of group No. III-vg receive the sign "passive voice."

The main difficulty, therefore, comes from the verbs of group No. III-bg and No. III-gg where the distinction in voice is produced on the basis of an analysis of adverbial complements and adjectives of circumstance and the relationship to the category of animateness in the subject and object of the action.

The various signs for the nouns, as discussed above, are produced in the primary analysis routines. This pertains to non-homonymic inflections. For example, a noun in a clause with the case ending -<u>jakh</u> immediately gets the signs plural number, prepositonal case. A noun with the ending -<u>u</u> gets the appropriate signs as follows: if directly after rejection of this ending the word is immediately found in the dictionary, it is naturally a first declension noun, masculine gender. But in a given context it can represent the genitive, dative, or prepositional (more precisely, locative) case, singular number. If the noun is found in the dictionary after replacing the contextual ending with -<u>o</u>, it is a neuter noun and can immediately obtain the signs dative case, singular number.

If it is necessary to replace the contextual ending with -<u>a</u> in order to find this word in the dictionary, the given word, which is a feminine noun, obtains the signs accusative case, singular number.

However, we are far from being able in all cases to produce the signs by using the primary analysis method. We need a special analysis because of widespread homonymy in the genitive and accusative cases of animate, masculine nouns, homonymy in the nominative and accusative cases of inanimate, masculine nouns, homonymy in the locative and dative cases of certain nouns (e. g., <u>les</u> 'forest' - <u>v lesu</u> 'in the forest,' etc.) It is also difficult to recognize the case of third declension nouns which do not distinguish between the endings of the genitive, dative, and prepositional cases or the case of nouns ending in -<u>ie</u>, -<u>ija</u>, or -<u>ii</u>.

The differentiation of homonymic forms in these nouns is effected by the following analysis. The verbal predicate is checked to see whether it is transitive, for the presence of a certain person and number, whether it belongs to a group governing a specific case, and for its position with respect to the noun being analyzed.

Besides analysis of the predicate a check is made for a preposition from a certain group before the given noun (skipping adjectives and adverbs standing before the noun as well as extended attribute sequences.)

The presence or absence before the word being analyzed of a noun or numeral requiring a certain case is also very important in distinguishing between the accusative and genitive cases.

This series of checks makes it possible in the majority of words to determine quite accurately the case of a noun that has homonymic inflections in the sentence under analysis. To illustrate, we shall describe the handling of the word <u>dom</u> 'house', which in the text is in the accusative case, plural number and preceded by <u>na</u> 'in, to,' (preposition group No. I-vpr).

22/23, 3/ [5] – Check the given noun for rejected ending -<u>a</u> or -<u>ja.</u>

23/VII, 24/ – Check the given noun for the sign "feminine gender."

24/XI, 28/ – Check for a noun or numeral before the given word (skipping adjectives and adverbs.)

28/XI, 29/ – Check for a preposition from group No. I-apr before the given word (skipping adjectives and adverbs).

29/30, 27/ – Check the given word for the sign "neuter gender" or whether it belongs to group No. I-as. [6]

30/VI, 31/ – Check for a preposition from group No. I-vpr before the given noun.

VI/0, 0/ – Produce signs "accusative case, plural number" for the given noun.

---

5. For an explanation of the symbols, cf. D. Ju. Panov, <u>Avtomaticheskii Perevod</u> (Automatic Translation), Academy of Sciences USSR Publishing House, Moscow, 1956, and I. S. Mukhin, <u>Opyty Avtomaticheskogo Perevoda na Elektronnoi Vychislitel'noi Mashine BESM</u> (Experiments in automatic translation with the BESM Electronic Computer), Academy of Sciences USSR Publishing House, Moscow, 1956.

6. The word <u>dom</u> belongs to group No. I-as, which includes masculine nouns ending in -a in the nominative case plural.

The occurrence of homonymic endings is much more frequent in participles than in nouns. For example, feminine adjectival endings coincide in the genitive, dative, instrumental, and prepositional cases. The case of the adjective is determined on the basis of the syntagmatic connections of the given word, i.e., by the noun to which it is related, with cognizance taken of the possibility of a so-called extended attribute occurring before the noun.

The oblique cases of personal and parts of indefinite-personal pronouns are included directly in the dictionary with an indication of the number and case as well as the stem form of the given word since the grammatical significance of these pronominal forms is in most cases expressed lexically.

The case of forms of indeclinable adjectives and nouns are determined after concluding the analysis of the morphological signs of the entire sentence because a knowledge of the case and number of the adjacent words may help resolve this question.

Moreover, the case and number of indeclinable nouns are determined by analyzing the form of the predicate, prepositional government of this predicate, and adjectives and participles modifying the given word.

Syntactic Analysis of the Sentence

In order to conclude the analysis of a Russian sentence we need more than the dictionary information about the words and their contextual morphological features. The correct transmission of the total meaning of the sentence requires information about the function of each word in the expression of the complete thought and interrelations between the members of the sentence. It is precisely this information that is furnished by the syntactic analysis of each word.

As the science of linguistics developed during the past century, two viewpoints on the essence of the language material studied by syntax have become clearly discernible. Supporters of the one theory regard the sentence as a self-contained complex of words united not only by close internal interconnections but by the fact that they belong to the entire sentence within which each word has its strictly determined place.

Adherents of the second theory treat the sentence as an aggregate of groups. The so-called phrases or syntagmas, each of which is an in-dependent linguistic entity susceptible of syntactic investigation.

Such prominent scholars as F.F. Fortunatov, A.M. Peshkovskiĭ, M.N. Peterson, and others in Czarist Russia dealt with the problems involved in investigating phrases. Their work was continued by A. A. Reformatskiĭ, O. S. Akhmanova, and other Soviet linguists. In recent times the theory of word combinations has been steadily kept in the foreground of linguistic research.[7]

The clear distinction between phrases and sentences, already set forth in the works of the well-known German linguist J. Ries, was continued and extended by Acad. A. A. Shakhmatov.[8]

The compilers of the Academy's Grammatika Russkogo Yazyka (Grammar of the Russian Language) devoted one volume in the section on "Syntax" to an analysis of the types of phrases [9] and another to an analysis of the kinds of sentences.[10]

Development of the new branch of science known as automatic translation, which demands of linguists working in this field maximum precision in defining linguistic categories and concepts, has raised anew the question of the proper unit of context analysis — phrase or sentence.

It is our view that the sentence should constitute the unit of investigation, even though a study of word interconnections in a phrase is of great interest, but, it seems to us, for other purposes.

The focus of attention in our work is the sentence as a whole, which is analyzed not as an aggregate of individual syntagmatic patterns but as a complex entity with individually interrelated parts.

---

7. Cf. F.Mikush,"A discussion of structuralism and the syntagmatic theory, " Voprosy Jazykoznanija (Problems in Linguistics), 1957, No. 1.

8. J. Ries, "Was ist ein Satz?" Prague, 1894; J. Ries, "Was ist Syntax?" Marburg, 1931.

9. Mention should be made of the fact that the concept of phrases held by the compilers of the Academy's Grammar differs from that supported by representatives of the school of Acad. F.F. Fortunatov.

10. Grammatika Russkogo Jazyka, Vol. II, Syntax, Part I and II, Academy of Sciences USSR Publishing House, Moscow, 1954.

Sentence analysis, therefore, consists of isolating the principal members of the sentence (subject and predicate) and the secondary members . We have included among the secondary members, objects, attributes, and adverbs, although, as A.B. Shapiro correctly points out in his interesting article, this "classical" three-fold division does not correspond to the complex interrelations arising between the members of a sentence.[11]

Most textbooks of Russian cite the "free" word order in the Russian sentence as an indisputable fact. However, examples of such "freedom" do not at all correspond to reality. The order in ja kupil knigu 'I bought a book' cannot be changed into knigu ja kupil without damaging the sentence and ignoring the normative function of the word order, which rests on a centuries-old language tradition. In the second example the sentence tends to enlarge owing to the defining adverb: e.g., kupil vchera 'I bought yesterday,' kupil na ulitse Gor'kogo 'I bought on Gorky Street,' kupil s udovol'stviem 'I bought with pleasure.' Moreover, these textbooks usually ignore the stylistic peculiarities of the various branches of literature, which also make possible a more precise differentiation of the members of the sentence. Thus, the sentence my uravnenie reshili 'We solved the equation,' lit. 'We the equation solved' is wholly acceptable word order in other branches of literature, but is virtually impossible in mathematical texts.

Recognition of the fact that word order in Russian is by no means free, that, on the contrary, it functions largely to distinguish meaning, enabled us to devise a routine to effect the formal isolation of the members of a sentence.

In referring to the rigidity of Russian word . order we have in mind more than the ordinal succession of the members (i.e., which member of the sentence comes first, which second, etc.). It is also important for us to know in what order certain members of the sentence and parts of speech may precede each member of the sentence.[12]

There were two possible ways of setting up this routine:
1) To search for each of the potential members of the sentence throughout the sentence.
2) To analyze each word in turn.
We chose the second way as being less clumsy.

Before analyzing a sentence the "parenthetical" (Vvodnye) parts are automatically skipped. These "parenthetical" constructions are separated on the basis of information obtained from the "punctuation-marks" routine. Gerunds and participles are analyzed independently.

If the sentence is complex, the remaining portion is broken down into individual, simpler elements; if the sentence is simple, it is treated as a whole. The division also proceeds on the basis of data produced by the "punctuation-marks" routine.

Verbs with the sign "personal form" and modal adverbs of Group No. II-bn, which more or less correspond in traditional terminology to the "category of state" (Kategorija Sostojanija) receive the sign "predicate" (skazuemoe) during the primary analysis. Gerunds receive the sign "adverb of manner" (obstoyatelstvo obraza deistvija) also during the primary analysis. Verbs in the infinitive receive one of these signs: "part of compound verbal predicate" (chast' sostavnogo glagol'nogo skazuemogo), "object" (dopolnenie), or "non-agreeing attribute" (nesoglasovannoe opredelenie), the presence of another verb in a personal form and the contextual-morphological environment of the given infinitive serve as criteria for the choice of sign.

During the analysis of adjectives the presence of nouns syntagmatically related to the given adjectives is revealed. The adjective then receives the sign "agreeing attribute" (soglasovannoe opredelenie). If, however, it does not modify any noun, the adjective passes on to another part of the routine for analysis.

An adjective of this type that does not relate to a noun is first checked for the sign "instrumental case." If the answer is in the affirmative, the verb is checked to see if it belongs in the byt' 'to be', schitat'sja 'to be considered,' etc. group of verbs, which require the predicate in the instrumental case. Verbs of this type are found in group No. II-gg. If the number of this group is indicated in the dictionary information for the given verb, the adjective receives the sign "part of compound predicate" (chast' sostavnogo skazuemogo).

---

11. "In investigation of secondary members in the Russian sentence," Voprosy Jazykoznanija, 1957, No. 2.

12. Similar positions. Cf. Ch. C. Fries, The Structure of English, Harcourt, Brace and Company, New York (1952).

Adjectives without the sign "instrumental case" and not attributes are examined later like nouns.

Nouns to be analyzed are divided into two large groups: with preceding preposition and without preceding preposition. (The possibility of an extended attribute occurring before a noun undergoing analysis is also taken into consideration.)

The analysis within each group takes place by cases. If there is a nominative case, the possibility of two heterogeneous nouns in the nominative is considered. (Should there be more than two such nouns, homogeneous members must necessarily be found among them.)

Depending on the pronominality of one of these substantives, the presence of certain marks of punctuation, or the sequence of words with respect to one another, one of these nouns receives the sign "subject" and the other "nominal part of compound predicate" (imennaya chast' sostavnogo skazuemogo).

A noun in the genitive case may be a non-agreeing attribute or an object. To verify the latter a check is made to see whether there is a negation in the sentence or the predicate belongs to group No. II-ag (verbs requiring the genitive without a preposition.

In determining the function of a postsubstantival noun, it is important to check for the verbal nature of the noun in front of it. If the preceding noun is verbal the noun in the genitive case will always be an indirect object.

Most nouns in the accusative case without a preposition receive the sign "direct object," while nouns in the dative case without a preposition receive the sign "indirect object. "

Substantives in the instrumental case without a preposition receive the tag "adverb" or "indirect object" depending on the pronominal nature of the given substantive, whether animate or not, and its position in the sentence. Our routine for the indirect object expressed by the instrumental case distinguishes between the so-called "instrumental agent" (tvoritel'nyi dejatelja) and the "instrumental instrument" (tvoritel'nyi orudija).

This precise definition of functions of the instrumental case is important in translating, for example, from Russian into English where the two cases will be translated differently.

Nouns with preceding preposition may be defined as "adverbs," "indirect object," or "non-agreeing attributes."

Homonymic endings are analyzed in several parts. Therefore, a word passes from one part to another until the final treatment. Varying this by creating a routine based on the endings themselves rather than on the possible parts of speech proved to be considerably more clumsy.

The routine takes cognizance of the peculiarities of mathematical texts, where adverbs are used as non-agreeing attributes. In such cases the adverb is checked for "quotation marks." If the answer is in the affirmative, the adverb receives the sign "non-agreeing attribute"; if in the negative, it receives the sign "adverb."

Numerals are equated with nouns in the analysis for function.

Thus, after passing through this last grammatical analysis routine all the significant words obtain syntactic tags indicating their role in the sentence as a whole. The accurate determination of the syntactic function of the words is highly important in translation from Russian.

Creation of an analytic part for translation from Russian testifies again to the complete possibility and feasibility of automatic translation. Furthermore, it seems to us that it is premature to speak of the so-called "limits" of automatic translation since as developmental work proceeds newer possibilities emerge for eventual formalization in language description, even for the use of similar methods to solve stylistic problems.

Research on the structural description of language as a whole is opening up broad avenues not only for perfecting machine translation but also for achieving a deeper understanding of language itself. It is also helpful in looking at a given language from another "viewpoint," requiring a detailed and precise solution of the linguistic problems that arise. The usual statement that various linguistic anomalies exist, without explanation of the cause, is clearly inadequate in our efforts to achieve machine translation.

The method of grammatical and lexical analysis used in MT requires clear criteria for the circumstance producing each linguistic phenomenon.

In conclusion, we should like to list briefly the principal problems that we encountered in our work on a structural description of the Russian language:

1) Problems of word formation.

2) The significance of lexical environment in showing grammatical category.

3) The function of word order in delimiting homonymic inflections.

4) The interrelation and mutual influence of word order, sentence members, and the various parts of speech to which they belong.

5) The problem of equivalence (e.g., non-standard adverbial correspondences and the verbal noun).

6) Problems related to the specific character of language norms.

7) The problem of lexical units and interlingual phraseological correspondences.

8) Information about scientific literature as an independent linguistic style.