[*Information processing: proceedings of the International Conference on Information Processing*, Unesco, Paris 1959] 163

The use of machines in the construction of a grammar and computer program for structural analysis

By K. E. Harper and D. G. Hays, The Rand Corporation, Santa Monica, Calif. (USA)

The present paper describes progress made on the building of a descriptive grammar of Russian with the complementary efforts of man and machine. Linguistic research at the Rand Corporation begins with the collection on punched cards of a large quantity of raw text from Russian physics journals. As described elsewhere in detail, a total of 250,000 running words of text is being processed, in corpora of about 30,000 words each. Post editors supply codes to indicate (a) the structure of the Russian sentence and (b) its translation into English.

In this way the relative position of each word in the structure of the whole sentence is recognized and codified. Dependency codes are then punched back into the text cards. The entire corpus is then machine-sorted and listed according to the structural and morphological type of each item in the text, and according to lexical entries. Syntactic analyses of these listings lead to the identification of word classes according to function (the extension and modification of traditional grammatical classifications) and to identification of the relations between syntactic units of the sentence.

The word classes and functional relationships thus determined are imbedded in a computer program for sentence-structure determination that is now being tested. The program establishes a relationship between two words in a specific sentence when: (a) the words belong to classes that, in general, can be related, and (b) all intervening words in the sentence have previously been related to one or the other of the words in question. The sum of the word classes and functional relationships that can exist among them is a grammar for Russian physics texts, while the computer program for translation is a working statement of the grammar. The empirical questions now under test are: (a) What word classes and functional relationships are to be recognized for Russian ? (b) Do the computer-determined sentence structures match those given for the same sentences by linguists ?

1. Introduction

Grammars of natural languages are the product of human endeavor to classify words and their usage in sentences. It may be said that the accuracy and completeness of a grammar is largely a function of the grammarian's ability to perform generalizations upon a large number of language details. Of the two major subdivisions of grammar (morphology and syntax), it may be said that the former has been more adequately described than the latter. Since the frequency of occurrence of the forms of words is strikingly greater than the frequency of separate syntactic combinations, it may be conjectured that the inadequacies of syntactic theory are due, at least in part, to the inaccessibility of data to the grammarian. This suggests that data-handling machines may be usefully employed to improve the descriptive grammar of a language. The present paper describes, in brief, the adaptation of machines for this purpose.

Many researchers in the area of machine translation have considered the grammatical problem to be one of adaptation, so that if the existing grammar of the source language can be set down in computer terms, the grammatical analysis of any portion of that language can be satisfactorily performed. Researchers at The RAND Corporation have proceeded from a different point of view—namely that existing grammars of languages are incomplete, ill-adapted for purposes of machine translation, and prolix with regard to a specific area of discourse. In addition, we have felt that the determination of sentence structure would prove useful in dealing with both grammatical and semantic problems in machine translation. Our task has been, then, to build a grammar of our source language (Russian) in terms of structural analysis, and to use this grammar in a structure-determining program for machine translation of Russian to English. We have confined our research to Russian articles in the field of physics, without the pre-judgement of the identity of a subject-field grammar with the grammar of the whole language.

Although our interest here is in the building of a grammar, a brief description of the whole research procedure is relevant. Linguistic study at The RAND Corporation begins with the collection on punched cards of a large quantity of raw text from Russian physics journals. Approximately 250,000 running words of text have been collected. The research process consists of processing these texts in "corpora" of about 30,000 running words each, in four stages.

1) *Text preparation.* This consists of serialization and key-punching of the occurrences in a corpus (an occurrence being defined as a sequence of printed characters preceded and followed by either spaces or punctuation).

2) *Glossary development.* A second deck of cards is punched, including a card for every different spelling" of text occurrences. A machine listing of this summary deck is prepared, and to each item in the list pertinent linguistic information is manually attached. This information consists of morphological codes to indicate part-of-speech and inflectional characteristics, lexical tags (which we call word numbers) to group together inflectional variants of a given word, and tentative English equivalents of each item, listed in their uninflected form together with codes to specify their manner of inflection in English. This information is punched back into the summary cards and then transferred to the original text cards.

3) *Translation.* This process begins with the translation text deck, now containing glossary information and sorted into textual order. A computer program produces a listing of the text as a rough translation; linguists (whom we call post-editors) convert this listing into a smooth English version of the original. This conversion is done by the assignment of codes for each occurrence in text to indicate:
(a) variations in syntax in English (inflection, insertion of connectives, auxiliary words, changes in word order, etc.), (b) the correct choice of equivalent, based on the requirements of context (including recognition of idioms and homographs), and (c) the structure of the Russian sentence. The latter deserves a few words of explanation. We think of the structure of a sentence as a tree-like set of relations among the words in the sentence. One word in every sentence is independent. Except for the unique independent word, every word in a sentence depends on one and only one other word in the same sentence. (Double dependency is allowable in special instances.) This concept of sentence structure recognizes Ajdukiewicz as its lineal ancestor, and is closely akin to the immediate constituent analysis of Harris, and to the approach suggested by Oswald for German in the early days of machine translation research [1, 2, 3]. The concept of dependency is partly syntactic, partly semantic. Syntactically, one occurrence depends on another if the inflection of the first depends on the nature of the second; semantically, one occurrence depends on another if the meaning of the first complements or modifies the meaning of the second. These definitions are related in any real language.
Given these rules and definitions, post-editors supply codes for each occurrence in the text sentence, identifying its dependency relationship. (The code assigned to a dependent word is merely the sequence number of the occurrence which serves as its governor.) In ambiguous situations, as with ellipses and idiomatic combinations, arbitrary dependency procedures are established for sake of consistency. Coding is checked at least three times for accuracy and conformity to the rules. The result is a coded description of the structure of each sentence. All codes (referring to English syntax, equivalents, and Russian structure) are punched back into the text cards.

4) *Analysis.* This is fourth step in the research process. Here, the post-edited text is studied with the goal of refining both the glossary and the computer program for translation. At the conclusion of each corpus, or cycle, a machine sorting produces information on the relative frequency of equivalents of multivalent Russian words, and information about the utilization of English syntactic codes by the post-editors. The embodiment of this information in the glossary and computer program results in an improvement of the machine-translated text in the succeeding corpus; the result of cyclic improvement is gradual elimination of the post-editor.
In summary, the problem of machine translation is treated empirically and cyclically. Structural analysis provides a convenient method for collecting and relating large masses of data about the interplay of words in sentences.

2. Construction of the grammar

Our present concern is with the "grammar" constructed for the texts at hand. As we have indicated above, the fourth stage of the research process consists of analysis of the decisions made by post-editors during the translation stage. It is here that the machine proves its great usefulness: Since the post-editors' decisions are represented by codes punched in each of the text cards, machine sorting of these codes makes it possible to retrieve and analyze the thousands of bits of linguistic information scattered at random throughout a large body of text. It is particularly convenient to sort this data according to the various structural categories. Thus, for any given corpus, all occurrences are sorted once according to morphological code and once by lexical tag (i.e., by word number). As an example of the former, all nouns in the dative case are grouped together with their governors (which may, morphologically, include nouns, verbs, prepositions, and adjectives). These governing words will then receive an additional grammar code (in the glossary), indicating their function as governors of the dative case.
For the sake of convenience in analysis, a separate listing is made of all words which have served as governors, together with their dependents. If desired, the codes indicating the preferred English equivalent and English syntactic usage are also listed for each occurrence. In addition, special listings are prepared for such structural combinations as governors of infinitive verbs, subordinate and coordinate conjunctions, relative pronouns, and relative adverbs. Because of the special difficulty in translating prepositions, separate listings are prepared of both the governors and dependents of all prepositions, sorted according to their English equivalents for each occurrence in text. In one corpus, for example, a total of 5,677 prepositions were listed in this manner.
The prime objective of analysis of these listings is the identification of word classes by function. In effect, these word classes are modifications or extensions of the classifications of traditional morphology. Traditionally, adjectives are a part-of-speech class possessing a certain inflectional pattern and functioning as the modifiers of nouns. An inspection of the listings reveals additional functions for certain adjectives: they may also serve as governors of prepositions, the dative case of nouns, or the infinitive; they may depend in the short form, neuter, only on other adjectives, etc. A transitive verb has the traditional, or "primary" function of governing a direct object; we find such "secondary" functions as governing a given preposition, an instrumental-of-agency noun or an indirect object in the dative case; in its reflexive or participial form, the verb may vary its pattern of governorship. For our purposes, each of these functions is a characteristic of the word, as deserving of codification in the glossary as gender, number, or tense. The morphological code is then supplemented by other codes indicating a given pattern of structural behavior. The sum of these codes (carried in a five-column field of a punched card) is called a word class.
It is obvious that many of these functional word classes are semantically determined. The syntactic functions of certain semantic classes are, of course, well known. (Words of modality, communication, perception, etc. are cited such studies of syntax as the *Grammatika rnsskogo yazyka* of the Academy of Sciences of the USSR [4]). It should be said, however, that these classes are rarely satisfactorily defined, and that their functions are sometimes inaccurately or incompletely described. On the other hand, in many instances the members of a functional word class have little or nothing in common semantically.
It should be said that great emphasis is placed here upon empirical evidence. Thus, a certain morphological type may according to traditional grammar, serve two or more

syntactic functions; if we find that, in fact, it has served only one of these functions in the text analyzed, it is coded for only this function. A variation in this procedure is the prediction of the functional behavior of morphological and semantic types.

Governors of the subordinate conjunction "что" can be pre-coded with a high degree of accuracy, for example. With all deference to human imagination and intuition, we have generally resisted the temptation to borrow from the experience and guesses of native Russians or grammarians. The syntactic part of our grammar codification is, thus, quite accurate for the few hundred thousand running words of text analyzed. Its only limitation is, in theory, the size of the sample tested. The fact that some of the word classes may apply only to texts in physics does not invalidate its accuracy for the area of discourse under investigation.

By way of summary, we may compare in two ways our grammar of Russian physics with traditional grammars of Russian. In the first place, the procedure for identifying word classes results in the unification and standardization of the syntactic functions referred to above as "primary" and "secondary." These terms are used here to distinguish between functions which are conventionally attributed to words in grammars or dictionaries, and those which are not. The latter constitute what is generally referred to as "usage," a comprehensive term which signifies that the codification of grammar has broken down. The grammarian and the lexicographer consider that the capability of a transitive verb to govern a certain case of the direct object is minimum, indispensable information. Why is the capability of such a verb to govern other parts of speech ignored, or suggested only in randomly selected examples? The answer probably lies in the limited ability of the linguist to collect language data.

In the second place, the identification of word classes is exhaustive, at least for the text analyzed. A grammarian would not consider it his proper task to make an inventory of the members of a given word class, perhaps because he does not possess the tools to perform this task. The resulting impoverishment to lexicography is not serious to the native Russian, whose familiarity with "usage" more than compensates for the shortcomings of grammar. The resultant loss to grammar is more serious, since it is on the basis of detail that more powerful generalizations about syntax, usage, and "context" must be formulated. Incomplete and restricted as our grammar may be, it suggests a methodology applicable to the far more complex task of writing a descriptive grammar of the whole Russian language.

The identification of word classes constitutes only a part of the work of analysis. For purposes of machine translation of new text, we must provide for the machine recognition of the joint occurrence of these classes, and a specification of the relationship between them. One convenient way of dealing with this problem is to make a table of word class pairs, in which, for example, is specified the relationship between all word classes which can govern the dative case of nouns and all dative nouns. At present, a table of some 7,000 entries has been written, expressing the relationship between one word class as governor and another class (chiefly morphological) as dependent. Each pair-entry in this table is called a dependency type. The relative position of the two elements of the pairs is indicated, both in Russian and in English, as well as a resultant code for each element.

The latter is particularly helpful in determining the structure of the sentence, since it is a means of reducing the ambiguity so characteristic of Russian forms in isolation. Typical resolutions of ambiguity in the dependency table are: clarification of the case and number of nouns through their relationship to governors and dependents, and clarification of the pronominal-adjectival ambiguity of demonstratives. Resultant entries in the table may also indicate relationships which may or may not require the addition of a connective in English translation. Other routines involving conjunctions, relative adverbs, and pronouns specify inter-clausal relationships.

In summary, the construction of a grammar for Russian physics texts consists of (a) recognition and codification of functional word classes, and (b) codification in a table of the relationships which exist between these word classes. The dependency table, of course, merely indicates the relationship of word classes in a vacuum; for purposes of machine translation we also need routines that will specify which relationships are true for any given sentence of test.

3. Sentence structure determination

Our machine-translation system consists of three distinct parts: glossary look-up, determination of input sentence-structure, and creation of an English sentence for output. These three stages in the translation process are well-known, and we need not dwell on them. The first step is entirely straightforward, and the third nearly so. We devote the remainder of this paper to a discussion of our techniques for sentence-structure determination, with just a word or two on the relation between this step and the following. The most characteristic feature of our method for sentence-structure determination is its segregation of the two principal modes of syntactic symbolism, the modes that we call "agreement" and "word order." (Punctuation is little used in our present method.) Agreement means any relation between the grammatic types of two occurrences that must hold if the occurrences are to be connected in a sentence structure. Agreement is tested by a simple table-look-up routine, using the extensive table of dependency types mentioned above. Word order means any relations of two occurrences that must hold if the occurrences are to be connected in a sentence structure. The importance of segregating these two phases of sentence-structure determination will be discussed below, after the word-order routine has been explained.

We use the term "precedence" to describe a specific relation between occurrences in a sentence; roughly, one occurrence precedes another if the word-order rules of the language permit them to be connected by a dependency connection. The difficulty inherent in linguistic analysis, at this level, is that the precedence relation and the dependency relation are inextricably confounded; they *must* be treated simultaneously within a given sentence. Hence precedence is defined in terms of the dependencies that have already been established, and dependency is defined in terms of the conditions, including precedence, under which a dependency may be established.

If X, Y, ... are occurrences in a sentence, XpY means that X precedes Y and XdY means that X depends on Y. If WdXd ... dYdZ, we say that W "derives from" X, ..., Y, and Z; no occurrence depends on itself, but every occurrence derives from itself.

Precedence is defined implicitly by the following postulate:

a) XpY if the occurrence number of X is less than the occurrence number of Y, if every occurrence whose number is greater than X and less than Y derives from X or Y, if neither X nor Y derives from the other, and if X, Y, or both are independent.

The condition on intervening occurrences means, linguistically, that phrases are compact. In the English sequence,

"In Flanders field a poppy grows,"

it is not conceivable that "field" is the subject of the verb and "poppy" the object of the preposition; it is advantageous to avoid consideration of the pair "field-grows," so no precedence is established. Note that a precedence is cancelled when both its members find governors, or when one is found to derive from the other.

A dependency is established according to the following rule, which is given as an implicit definition of the relation:

b) XdY is established if g(X), g(Y) is listed in the table of dependency types, if XpY or YpX, and if X has not previously been found to depend on any other occurrence.

(Here g(X) and g(Y) are the grammatic types of X and Y respectively.) One complication is that the table of dependency types may specify that XdY is implied only if XpY, only if YpX, or in either case. Once established.a dependency is cancelled only by rules that we cannot attempt to expound here.

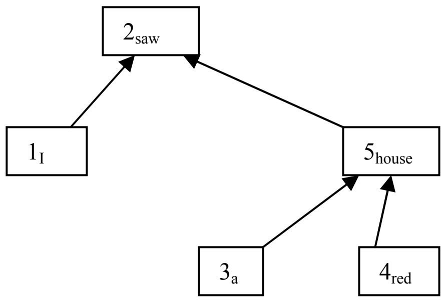
It is obvious that precedences and dependencies must be determined alternately in a sentence; a program with this effect is now in operation on the IBM Type-704 computer at The RAND Corporation.

Consider, as a simple example, an English sentence: "I saw a red house." Number the occurrences of the sentence consecutively:

1_I 2_{saw} 3_a 4_{red} 5_{house}

According to postulate (a), we have 1p2p3p4p5; the first step in the computer program is tabulation of these precedences. Now, using postulate (b), the program searches for dependencies. The pairs for which precedence holds are considered in turn; first, 1p2. The grammatic types of these words are available from the glossary. The pair of types is sought in the table of dependency types and, in this case, an entry is found; hence 1d2—"I" depends on "saw."

The next dependency to be discovered is 4d5; the adjective modifies the noun. But now the precedence list must be altered. If 4d5 then 3p5, for the only intervening occurrence is 4, and it depends on 5. The new precedence is added to the list, and eventually it is tested for dependence. Since 3d5, it follows that 2p5; again a precedence is added. When the grammatic types of 2 and 5 are tested against the table of dependency types, and 5d2 is determined, a tree structure is established for the whole sentence:



The computer program contains a rule directing it to proceed to the next sentence when this stage has been reached. (When the structure cannot be determined completely, other stop rules apply.)

A second illustration indicates the complexity that may develop at intermediate stages in the analysis. This time the sentence is Russian:

1 2 3 4 5 6
"наблюденный рост поверхностого сопротивления с час-
7 8 9 10 11
тотой может быть обьяснен возрастанием эффективной
12 13
глубины пронинновения."

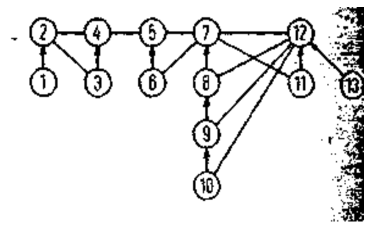
We wish to translate this sentence, "The observed increase in surface resistance with frequency can be explained by an increase in the effective depth of penetration." References to the Russian sentence will be made via the occurrence numbers.

The initial precedences are listed: 1p2p ... p12p13.

Next the table of dependency types is brought to bear on the sentence. Each pair of words for which precedence has been established is tested against the entries in the table. Thus, since 1p2, the grammatic types of "наблюденный" and "рост" are obtained from the glossary and sought in the table of dependency types. Since there is grammatic agreement, an entry in the table asserts that "наблюденный" depends on "рост." Next, since 2p3, "рост" and "похерхностного" are tentatively paired, but no table entry is found for them. When "поверхностного" and "сопроивления" are treated in the same way, since 3p4, the adjective is asserted to depend on the noun, and the precedence relation is modified.

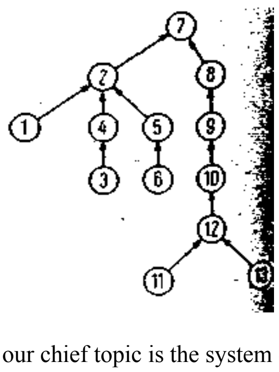
If word 3 depends on word 4, no independent word intervenes between 2 and 4, so 2p4 must be added to the list of precedences. This precedence is added to the list following 12pl3, and the search for dependencies continues with 4p5.

When the entire sentence has been scanned once, it has the following form:



(In this diagram, arrows represent dependencies, while simple line segments represent precedences.) The precedences remaining from the original list are 2p3, 4p5, and 6p7. New precedences have been added to the list, as they were established, in the following order: 2p4, 5p7, 7p11 10pl2, 9pl2 8pl2, and 7pl2. The order is important, since otherwise a distant governor might supplant one close at hand.

When the sentence has been scanned once, the precedences added to the list are considered. The first pair in this list is 2p4. Since our rule now calls for "skipping over" 3, an adjective, the genitive noun "сопроиления" is now compared with the noun "рост," and the table of dependency types asserts that the genitive depends on the preceding noun. The precedence 2p5 is added to the list when the new dependency is found. Later, when the new pair is considered, the semantic type of "poor" makes it a candidate to govern the phrase "с частотой," and the pair 2p7 is added. Thus the subject and predicate of the sentence are finally brought together in a precedence pair, and a connection is established. Meanwhile, "глбины" has been found to depend on "возрастанием,"and the structure of the sentence is completely determined. It may be represented by the following diagram:



Although our chief topic is the system of sentence-structure determination that we are developing, we may turn aside for a moment to note the usefulness of this system for machine translation. One approach to MT is that of contextual search, which marks words as "problems" and "clues." A genitive noun is a problem, since it *may* need to be connected to its governor, in English, either directly or through a preposition (e. g., "in" or "of"). When a genitive noun occurs in text, context is searched for a clue to the function of the genitive; search may be limited to one or two preceding words, or it may extend over a whole sentence. Among the clues to be sought, of course, are such words as prepositions governing the genitive, nouns, cardinal numbers, etc. If the right clue word is located by the search procedure, the translation is made correctly. Sentence-structure determination obviates the need for contextual search at the stage of output construction. The "problem" words may be treated in one of several ways, all of which hinge on the fact that the relevant "clue" word is either the governor, or a dependent, of the "problem" word. One technique uses the resultant codes mentioned earlier. When a cardinal number (nominative or accusative) is coupled with a genitive noun, a resultant code of the noun is obtained from the table of dependency types; it shows that the noun phrase functions as nominative or accusative. The "problem," so to speak, is eliminated before it arises.

An alternative treatment of "problem" words lets them stand as such until the structure of the sentence is entirely determined. Then the sentence is scanned once, and each problem is handled as it arises. In the illustrative sentence above, for example, a genitive noun depends on "РОСТ", which we translate as "increase." When this genitive is treated, the insertion routine tests the grammatic code of its governor; words like "рост" are coded to bring about insertion of "in" before certain genitive noun dependents, whereas most noun governors would require "of." Note that no search of context is required.

Several other problems in the illustrative sentence were handled in similar fashion: "to" was not inserted before the infinitive, and "c" was given the translation "with." Both syntactic transformations between languages, and semantic resolutions of ambiguity, are made more accurate and less expensive when contextual analysis is supplanted by inspection of governors and dependents. Proximity is a clue to sentence structure, but not the only clue; finding the right clue words for a given problem is so much like determining the structure of the sentence that it seems best to segment the problem as we have done. Our treatment of homographs may be mentioned briefly. A homograph is a form that has two distinct grammatical types. Determination of sentence structure proceeds simultaneously with both types until a governor is found for one of them; then the other is expunged. As our study of Russian text proceeds, we hope to find criteria for a more sophisticated decision process.

Punctuation is not much used in our system; we hypothesize that its proper place is in the routine for writing the list of initial precedences. Thus, for example, the last occurrence before a parenthetic insertion should be listed as preceding the first mark after that insertion. Again, a hyphen is a mark of priority. Ordinarily the only priority-recognized is that initial precedences are analyzed first, and other precedences are handled in order of discovery. When two occurrences are connected by a hyphen, it seems that their precedence pair should be moved to the head of the list. (Other similar examples could be mentioned.)

Finally, interclausal connections may be noted; here we find the only exception to the rule of "at most one" governor. A relative pronoun (который) has two governors, one within the clause it introduces and one outside. Inside the clause, of course, it acts as a pronoun, being the subject of a verb, object of a preposition, etc.; this is its "ordinary" governor. Outside the clause, "который" depends on a noun; this is the interclausal connection that establishes the structure of a complex sentence. Other clausal connectives, such as "что" and "где," are similarly treated. Such simple procedures may contain the seed of a technique for structure determination over longer spans than single sentences.

Our technique for sentence-structure determination differs from others, as we have said, by its segregation of word-order analysis from agreement analysis. The table of dependency types, on which agreement analysis is based, is large, specific to Russian, and not fully known; furthermore, it is subject to change with time. Material of this type, we feel, is best relegated to a table whose entries can be modified easily and cheaply. The program that alternately determines precedences and dependencies is, on the other hand, extremely simple; rules of word order, as embodied in this program, are not specific to Russian and appear to be fully known. Thus, by making the program independent of grammatic types, we seem at one stroke to have obtained program simplicity and stability. We hope, in fact, that the program will be applicable to any Indo-European language for which glossary, grammatic classification, and table of dependency types are written.

Obviously the virtue of this program is not its solution of a complex problem. The problem, if it has been solved at all, was complicated only because it was intertwined with that of agreement. Every language presents its own formulae for agreement, on which we can now concentrate without distraction, and with the greatest possible help from computational tools.

4. Summary

The products of The RAND Corporation's machine-translation research program are: a method of research, a computer program for sentence-structure determination, a morphological and functional code for Russian grammar, a table of dependency types, and a glossary of Russian physics. The grammar that we have devised is limited by the subject matter and size of our corpus; within that limit, the method assures us that it is precise and complete. We consider such a grammar superior for the purposes of machine translation. The glossary and grammar that we have developed cannot be transferred directly to any other language; however the general method of research that we characterize as cyclical and empirical, and the computer program for word-order analysis and sentence-structure determination that we have described, could be applied at once to the study of any inflected language. We believe that such application would lead, rapidly and efficiently, to systems of machine-translation for new language pairs.

5. References

[1] AJDUKIEWICZ, K.: *Die Syntaktische Konnexität.* Studia Phil. 1, 1935 pp 1-27.
[2] HARRIS, Z. S.: *Methods in Structural Linguistics,* Chicago: University of Chicago Press 1951.
[3] OSWALD, V. A. and FLETCHER, JR. S. L.: *Proposals for the Mechanical Resolution of German Syntax Patterns.* Modern Lang. Forum 36, 1951 pp 1-24.
[4] Academy of Sciences of the U.S.S.R., Institute of Linguistics: *Grammatika Russkogo Yazyka.* Moscow: Izdatelstvo Akademii Nauk SSSR 1954.

6. Discussion

Margaret Masterman (UK): It has been my feeling during this discussion that we have not been sufficiently careful in every case, to make clear what we understand by the phrase "mechanical translation." I should like to propose the following distinctions which will, I think, help us to be clear what each speaker means when he uses the words, and which will also serve as a convenient yardstick against which to judge the aims and achievements of workers in the field.

a) The one-for-one substitutions of the very early days of research in this field which I shall refer to as "transliteration." This results not in a natural-language output but in a "logical" construct which R. H. Richens called "mechanical pidgin."
b) "Insufficiently transformed mechanical translation" which is the category in which most of the output at present being obtained must be placed. In general this output is intelligible but requires post-editing.
c) "Full mechanical translation" (called FAHQMT by Bar Hillel) producing a high-quality output and requiring no post-editing. This is the aim of the Cambridge Language Research Unit and of some workers in Leningrad. There has so far been no output in this category.

Before the step from "insufficiently transformed mechanical translation" to FAHQMT can be taken, a mathematical theory of meaning such has not been discussed in this conference must come into being. It must be a comprehensive theory not deriving from any corpus of texts, but generalised to the whole of language. It is towards the establishment of this theory that my "efforts and those of my have been and are being directed.

P. L. Garvin (USA): Linguistic description per se, even if based on modern structural methods, is inadequate for MT purposes because of the difference in objectives. In linguistics, the basic question is to determine what are the representatives of a particular category (e.g., the allomorphs of a given morpheme). In MT, the question is the reverse, i.e., to determine how many categories are represented by a given item. (e.g., of how many morphemes can this morph be an allomorph ?)