CHAPTER IV

Machine Translation and Linguistics

1. General Remarks

Machine translation is a new and fast-developing area of linguistics in which exact methods of research are widely applied; indeed, they are necessary for progress.

An ineluctable part of the work in machine translation (MT) is the description of linguistic facts but in a unique form—namely, as rules for transforming a text in one language into an equivalent text in another.

These descriptions, consisting of the iteration of necessary operations, are so exactly drawn up that they can be "accepted" and used by an electronic computer. Thus, the immediate basic factor occupying our attention is the description of languages by exact methods, which can be verified experimentally by MT.

To avoid occupying ourselves with the complex theoretical question of the existence of a scientific description, we can stipulate that the construction of working models is a highly effective technique for creating and verifying a description of any system whatsoever. We shall explain just what this means.

Let us assume that we are considering a group of arbitrary objects generated by a mechanism unknown to us. This mechanism is not available for immediate observation, and we can draw conclusions about it only from the results of its activity, i.e., from the properties of the set of objects that it has generated. Here we are interested in the particular mechanism only in a strictly defined sense: It is important for us to know just those aspects of its functioning that cause it to generate the particular set. None of the concrete properties of the mechanism or of its functioning are relevant to us.

By analyzing the totality of objects available to us from the mechanism, we can create a hypothetical description of it. To verify this description, we can construct a model of the mechanism based on it. This would be only a model and not a copy of the mechanism, since very many concrete properties of the mechanism will not have been studied, and in some respects the model will not resemble the mechanism itself at all. But if this model in function can generate exactly the same objects as the mechanism studied, then we can conclude that our model is adequate in the relevant respects and consequently that our description is accurate. (Of course, this description does not have to be unique; other equally correct descriptions are possible.)

A model for a generative mechanism such as that in our example is directly relevant to linguistics. Actually, the aim of linguistics is the description of language, i.e., of the system generating speech. The system itself—language—is not manifest to the researcher; he must deal only with the results of its functioning—with speech. To verify one's descriptions, one can create working models corresponding to them—logical structures that can be realized in the form of electronic circuits and that must functionally generate speech objects. We think that a description can be considered accurate (although not natural) if we can create from it a working model capable of fulfilling any part of the functions of verbal communication.

If the problem of linguistics is defined to be the description of language as a structure producing speech, then the aim of MT is to embody this description in algorithms that are realizable on existing electronic computers. By the same token, MT provides linguistics with the experimental basis so necessary to it; in the course of MT, the description of linguistic facts is verified and made more precise, while the methodology of linguistic description itself is perfected. This is the value of MT to linguistics. MT specialists, in turn, should use the language descriptions created by linguistics. Thus, linguistics and MT cannot develop successfully without one another, or without a constant exchange of results and accumulated experience.

While this statement of the interdependence of MT and linguistics is line in theory, the actual situation is different. A paradox has arisen. On the one hand, MT work has received significant comment in linguistic circles. Special articles on MT are published in such linguistic journals as *Voprosy yazykoznaniya, Word, Modern Language Forum,* and *Babel.* Problems in MT are discussed at international linguistic congresses (the Eighth Congress of Linguists in Oslo, the Fourth Congress of Slavicists in Moscow); moreover, linguists take a considerable part in conferences on MT and related problems (the First Conference on MT, 1952, U.S.A.; the First All-Union Conference on MT, 1958, Moscow; the First Conference on Mathematical Linguistics, 1959, Leningrad, etc.). MT centers have been established at various linguistic institutes and also in universities in several countries, such as the U.S.S.R., the Chinese People's Republic, Czechoslovakia, the U.S.A., and England.

On the other hand, MT remains a highly specialized area that would seem, from its special problems and methods, to be quite separate from theoretical linguistics. At present in MT there is made almost no use of the achievements of contemporary linguistics; whereas pure linguists, while recognizing MT *de jure,* in developing their theories completely ignore MT *de facto.*

Yet MT is not just another special area of linguistics as are studies of Indo-European, Caucasian, and Semitic languages. A specialist in Paleoasiatic languages could easily know nothing about specialized research on Spanish, nor does a linguist studying lexicography need to deal with the problem of case in Caucasian languages. But MT touches equally on *all* specialized areas. The study of various languages and problems, using the approach and methods of MT that have been proven by experiment, will permit a future unification of the science of language. MT is simultaneously both a workshop, where the *methods* of precise linguistic research are perfected independently of the concrete sphere of application of these methods, and an experimental field, where the results are verified by experience. Therefore, it is very important for all linguists to learn as much as they can about MT.

In the following exposition, our purpose is not to describe precisely the bonds between specific questions of MT and the corresponding linguistic problematics; for that we need special research not yet undertaken. Our problem is purely *indicative*: to give a short description of some of the problems of MT that seemed to us to be of interest to linguists.

We have deliberately avoided aspects that, though particularly important for MT, are still too specific and technical at our present level of development. For example, the problem of MT dictionaries and of dictionary search, the problem of a morphological analysis of words during MT, etc., lie here.

2. Two Approaches to Machine Translation

The problems and methods of MT are variously understood by different researchers. Corresponding to differences in opinion in the MT field, there are two methods of approach, which in foreign literature are sometimes tentatively called the "95 per cent approach" and the "100 per cent approach."

In the first approach, the basic and final purpose of research is the realization of machine translation of scientific-technical texts with the least expenditure of time and effort. The quality of the translation may not be high; it suffices if the greater part of the translated text (hence, the name "95 per cent approach") is understandable to a specialist. For this reason the necessity of complete syntactic analysis is denied; a text is comprehensible to a specialist even with word-for-word translation (at least for certain pairs of languages). The structure of the language does not interest the researchers; the rules for translation are based on instances encountered in the texts analyzed and are gradually broadened by introduction of new texts and discovery of new cases. Such rules may not reflect actual linguistic regularities and may even contradict them. V. H. Yngve has called such rules "*ad hoc* rules."

In the second approach, the study of the general structural regularities of language that form the basis of concrete cases of translation are put foremost. In other words, the researcher tries to explain the possibilities and means used by language to express a particular thought. The rules for translation are formulated with regards to the possibilities explained. Realization of a translation on the machine is considered a means of facilitating knowledge of the structure of language (in the sense indicated above—as a group of laws according to which spoken sequences are constructed). This is necessary, since MT is not considered to be an end in itself but rather the first step in solving a more general problem: how to "teach" electronic computers a whole series of operations using speech, including editing and referencing and the introduction of bibliographic and other corrections to texts.

Much attention has been turned to syntactic and, more recently, to semantic analysis. It is proposed that the possibility of explaining the syntactic (and meaning) structure of a text will allow us not only to improve the quality of machine translations basically but also to automate the operations mentioned earlier that are connected with language.

An important place is assigned to purely linguistic studies of language. Thus, for example, the MT group at the Massachusetts Institute of Technology (U.S.A.) is working out a special structural grammar of German and a parallel, analogous grammar of English in order to determine the correspondence between these languages. "We are looking at language in a new light—through the prism of the 'memory' of a computer," write two members of this group, W. N. Locke and V. H. Yngve, "and we hope that our work on language structure will yield us new and interesting results" [40].

The "100 per cent approach" demands that, although he base his work on some limited text, the linguist use his knowledge of a language fully in formulating rules for translation, turning if necessary to special studies (i.e., introducing additional texts), and that he try to answer all questions cardinally, so that his solution may correspond to the *structural possibilities* of the language. Rules obtained in this way can be called "general rules" (as opposed to the "*ad hoc* rules" mentioned earlier).

The difference between "*ad hoc* rules" and "general rules" can be illustrated by an example from an article by A. Koutsoudas and A. Humecky, "Ambiguity of Syntactic Function Resolved by Linear Context" [38] In this article, rules are given for determining the syntactic function of Russian short-form adjectives in -o *(legko, bystro)* and of comparative adjectives in *–e, -ee (legche, bystree).* The rules are based on a large number of examples (approx. 700). They originate from analysis of "linear context" of three words (the short-form, one preceding word, and one following) and ensure correct analysis of nearly all initial examples.

However, D. S. Worth, in his critique of this article [62], cites examples contradicting nearly all of these formulated rules. This is explained, as Worth says, by the fact that Koutsoudas and Humecky had not studied the structural laws of Russian syntax but had simply lumped together the results of a series of translations from Russian to English. In this they used superficial facts—the character of two adjacent words. Study of a larger context would lead to magnification of the number of rules until they enumerated the many individual instances.

Worth's criticisms are valid. The fault does not lie in the fact that Koutsoudas and Humecky did not study some examples, or that they did not have enough examples. If they had tried to find a primary general solution, using only their own material, they would probably have obtained simpler and, moreover, more effective rules. Obviously, "general rules" must be based not on three-word or even larger "linear" context but on knowledge obtained about the whole sentence at the first analysis. (This knowledge is needed to answer many questions in translation and not just to find the syntactic function of the short-form adjective.) Thus, if there is an "obvious" predicate in a sentence (i.e., whether it be a verb in the personal form, or a short-form participle, or a short-form adjective, not neuter and not compared, or a so-called "predicative word" like *nyet* [there is no], *mozhno* [can], etc.), then the form in *-o* (or *-e, -ee)* under consideration can only be a modifier and must be translated as an adverb.[1] We note that it makes no difference in such an approach—as opposed to that taken by Koutsoudas and Humecky—where this "obvious" predicate is found, so no aux-

[1] For simplicity of illustration we have omitted the special case of the forms *budet* [will, will be], *bylo* [was] *(budet legko* [(it) will be easy], *bylo vozmozhno* [(it) was possible], etc.).

iliary rules are needed for handling all possible instances of inversion, substitution, etc.. Furthermore, if there is no "obvious" predicate in the sentence and the *-o* form under consideration is an adjective; of a definite semantic type (e.g., *legko* [(it) is easy], e*stestvenno* [(it) is natural], *nepravil'no* [(it) is incorrect]), and there is an infinitive present in the sentence but no possible infinitive-governing word, then the *-o* form is the predicate (translated into English by "it is" + adj.), and the infinitive is to be connected to the *-o* form (e.g., *legko videt' chto* ... [it is easy to see that ...]). Here again, the mutual word-order has no significance. Other rules for finding the syntactic function of short-form adjectives in *-o* are formed analogously.

Such "general rules" are based on a consideration of the principal possibilities (the semantic types of the short-form adjectives and the presence or absence of certain types of words in the sentence). These rules may be larger in volume than those of Koutsoudas and Humecky, but with a little increase in volume, they increase considerably in their effectiveness. In short, "general rules" can in every case ensure a selection that will be comprehensible (to a human being).

"General rules" are, of course, more interesting to a linguist. In the nature of things, their composition will lead to an exact description of the structure of language, i.e., to the discovery of laws such as those by which this or that content is expressed in language.

In general, the "100 per cent approach," with its broad view of MT, is more closely related to theoretical linguistics and is apparently able to function better in solving the latter's basic problems.

3. Syntactic Analysis in Machine Translation

In the first stages of MT's development, the researcher's attention was naturally drawn to the problems of word-for-word translation.

In word-for-word translation, the machine ascribes to each word or form of a word all possible translational equivalents, using a dictionary (or a dictionary and morphological tables).

Linguistic difficulties arising during such a translation are not great and are almost entirely reducible to technical problems. Therefore, it is entirely understandable that the history of practical work in MT began precisely with word-for-word translation.

During the past five or six years, in the U.S.S.R., in the U. S.A., and in England, several experiments in word-for-word translation have been conducted using machines; e.g.: Russian-English translation in the Computation Laboratory of Harvard University (Oettinger's group); French-English in Birkbeck College [23]; French-Russian at the Mathematics Institute of the Academy of Sciences, U.S.S.R. ([6], [7], [8]). (The French-Russian translations were not purely word-for-word; the algorithm employed contextual analysis to distinguish homonyms, etc., though not systematically.) The results have shown that word-for-word translation is suitable as a first approximation for definite pairs of languages and for specialized texts. In some cases, it is useful for direct application.[2] But even in these cases, word-for-word translation is in need of considerable improvement.

On the other hand, for certain pairs of languages (e.g., German-English and English-Russian), word-for-word translation is generally impossible; in such cases, it is necessary to base the translation on a syntactic analysis consisting of a determination of the bonds between words and between parts of the sentence.

Syntactic analysis gives machine translation an enormous potential for improvement.

The truth of this fact has long been recognized; one of the first publications on MT (in 1951!) was Oswald and Fletcher's remarkable article on the syntactic analysis of German text for translation into English [47]. The authors had formulated simple and, at the same time, quite effective rules for automatic analysis of the syntactic structure of German sentences. Their approach essentially foreshadowed the direction of research in this area.

In developing the ideas of Oswald and Fletcher, Victor Yngve proposed (in 1955) an interesting methodology that yields a very general solution to the problem of syntactic analysis (see

[2] See examples of French-Russian machine translation in [7].

[63]). Immediately after Yngve, there followed work on various aspects of syntactic analysis by many scientists abroad (the Cambridge MT group in England, the MT group of The RAND Corporation, the collaborators of the Georgetown group in the U.S.A., and others), and in the U.S.S.R. (the MT groups at the Mathematics Institute (MI), the Institute of Precise Mechanics and Computer Techniques (IPMCT), the Linguistics Institute (LI), Leningrad University (LU), and the Academies of Science of Georgia and Armenia).

We shall not give a detailed description of the activities of each group mentioned but shall limit ourselves to a survey of the general state of recent work on the automation of syntactic analysis, citing only the most interesting aspects.

We note especially that in MT the term "syntactic analysis" is rather widely understood and accepted, though insufficiently defined. Syntactic analysis includes the determination of bonds among words, the determination of the character of these bonds, the hierarchy of individual groups of words, the relations among the parts of a complex sentence, etc. Unfortunately, special research that would define the term exactly and establish the boundaries of syntactic analysis has not been undertaken by anyone. We shall, therefore, use the words "syntactic analysis" in the usual broad and rather fuzzy meaning (as primarily intending to determine the bonds among words of a complex sentence).

Many researchers base syntactic analysis on a *list of typical phrase-types* (or constructions). These typical phrases are described in terms of previously defined classes of words. To begin with, word-class attributes are ascribed to all the words in a text with the aid of a special dictionary. Then the machine, comparing the text with the list of phrase-types (i.e., with the list of minimal word-class sequences), finds specific phrases in the text, and thus determines the bonds among the words.

This method was proposed by Victor Yngve (U.S.A.) and, independently, by R. Richens (England). In the U.S.S.R., T. N. Moloshnaya was the first to apply it [17] for constructing an algorithm for English-Russian translation, using a dictionary of "configurations" (as typical phrase-types are called).

A dictionary (list) of elementary syntactic constructions (about 7,000 entries) is applied in Harper and Hays' Russian-English algorithm [35]. Dictionaries of configurations are applied by the majority of Soviet researchers (the MT groups at LU, IPMCT, LI, and the Georgian Academy of Sciences).

Several basic questions about syntactic analysis, as realized by cutting text into the simplest typical phrases, are considered in T.N. Moloshnaya's work [15] (for English) and in M. V. Sofronov's [20] (for Chinese).

The application of dictionaries of configurations permits the creation of a universal algorithm for syntactic analysis suitable for most, if not all, languages. Between languages, only the content of the configuration dictionary changes, while its general form and the rules for a search of configurations in text, using this dictionary, remain the same. (Analogously, rules for dictionary lookup do not change for various languages.) The general form of a configuration dictionary and a corresponding universal algorithm for syntactic analysis are being developed at LI.

In order to denote typical phrases, it is first necessary to classify words in a way that does not correspond with the traditional division into parts of speech. The number of such classes, in some algorithms, amounts to several dozen or even, in a few cases, to hundreds. Then, the number of typical phrases becomes several thousand.

But an approach is possible in which a single, constant distribution of words into classes in general does not obtain. In place of one class indicator, a series of indicators is written for each word, characterizing all words for all their interesting aspects. Word groupings can be constructed using any combination of indicators. When we need to define a class of words in order to apply some rule, we indicate that class by the necessary attributes and their meanings. Similar indications are included in the formulation of the rules (in the list of configurations); thus, word classes are formed specifically for each rule. This approach is used by LI in its algorithm for syntactic analysis.

This plan for grouping words can be called a "sliding classification." A "sliding classification" is suitable wherever one could, in choosing various combinations of indicators, obtain a large number of word classes of any volume. One can select the indicators so that a class will consist of just one concrete word form; one can also, by taking another combination of indicators, construct a class that includes a very large number of forms. The same words can belong to one class with respect to one set of indicators and to another class with respect to another set.

The "sliding classification" permits a considerable decrease in the number of configurations, to several hundred instead of several thousand. "Sliding classification" is also of considerable interest from a theoretical standpoint. It is possible that the notorious problem of the parts of speech can be studied anew in the light of a consistent development of "sliding classification."

In syntactic analysis, many machine operations, and consequently much time, are spent searching the configuration dictionary. Configurations are compared with the text sequentially, one after the other, until one of them "fits" the phrase being analyzed. Such iteration of configurations seems uneconomical, and we would like to do away with it. An alternate method has been suggested by the collaborators of the Cambridge MT group [44].

Source-text elements that possess the characteristic of predicting groups of a certain type ("structures," as the Cambridge unit has decided to call such groups) are studied. These elements are called "operators." An "operator" has ascribed to it, in the dictionary itself, the number of the structure that it predicts and an indication of its position in this structure. Once the machine encounters this "operator" in text, it immediately turns to the proper structure (in a list of structures) and then searches the text for the remaining elements. Here, the machine does not have to search the whole list of structures.

Similar ideas are being developed by Garvin's group (U.S.A.) [28]. Here, special attention is devoted to so-called "decision points," or "fulcra."

A fulcrum is the element of a syntactic unit that determines the presence and character of this unit. The fulcrum of a sentence is a predicate group, while the fulcrum of the predicate is a verb in the personal form, or a short-form adjective (in Russian), etc. To each fulcrum correspond specific syntactic rules, which are only applied when that fulcrum is discovered. Fulcra are comparable to the operators of the Cambridge group.

In its algorithm for syntactic analysis (of Russian), LI applies a similar method. To each word are ascribed the "addresses" of the first (in list order) configurations into which this word can enter. There are two such "addresses." (Addresses are numbers, the ordinal numbers of the configurations.)

The first "address" is ascribed to the word's root in the dictionary; it is based on the nature of the root itself (its lexical meaning, its capacity to predict some word or form, etc.). The second "address" is produced during morphological analysis; it is based on the form of the word. Reference to the configuration list is always made through the "addresses" of words. Proceeding from left to right through the phrase being analyzed, each word is looked at in turn, and according to its first address, a particular configuration is selected for comparison with the phrase under analysis. For each configuration, "addresses" are indicated for the series of configurations to which the "operating" word must refer, depending on the results of comparison (whether the given configuration had "fit" the phrase being analyzed). After the comparison, the operating word is "readdressed," then the next word is taken, and the whole process is repeated from the beginning. In this way, search through the whole list of configurations is avoided.[3]

The consequences of syntactic analysis of complex sentences are of special interest. Analysis can proceed by splitting up the component parts of a complex sentence—simple sentences, independent elements, etc. For this purpose, punctuation and certain words, mainly subordinating, are noted specially. Syntactic analysis proper (determination of bonds among words) is conducted within each separate part. Oswald and Fletcher [47] proposed this method; the IPMCT algorithm [18] uses it in analyzing Russian.

Yet another approach is possible: The splitting of a sentence into parts is not effected initially but during determination of the connections among words; this splitting is not the beginning but the end of analysis. This approach is used in the LI algorithm. The phrase being analyzed is split into "segments" according to its punctuation and certain conjunctions (without

---

[3] Many details have been omitted for the sake of simplicity of presentation.

any special analysis of the punctuation or conjunctions themselves), so that the segments do not correspond initially to the meaningful parts of the phrase but are purely formally separated sections.[4]

Syntactic analysis is performed within each section so obtained with the aid of the configuration list. The initial splitting into segments is necessary to avoid forming false relations between words in one part of a phrase and words in another part. However, this splitting, while saving us from false correlations, hinders the determination of many true bonds, since connected words can belong to different segments at first.

Therefore, when a word is isolated, i.e., when there is no obligatory bond to be found for it within a segment, then the segment as a whole takes on a special designation: an indication of what bond has not been made for which word. Thus, for example, if a transitive verb (e.g., *peremeshchaet* [shifts]) is "separated" from its modifiers (e.g., *elementy* [elements]) as follows:

Segment I
"*Vse elementy*
[All elements],

Segment II
kotorye prinadlezhat A
[which belong to A],

Segment III
eto dvizhenie peremeshchaet v novoe polozhenie ..."
[this movement shifts to a new position],

then segment III will be marked with an indication that for its third word there is "missing" a substantive in the accusative, and segment I will be marked for the "absence" of a governing word; i.e., there is an "excess" in segment I of a substantive in the nominative-accusative case.

[4] Several weaknesses (the periods in abbreviations, etc.) were omitted to simplify the explanation.

The idea of using such designations was advanced by G. B. Chikoidze (in Tbilisi). It has proved fruitful. In its algorithm for analyzing Russian, LI uses only some twenty such designations, indicating the "absence" or the "excess" of words of a particular type in a segment.

When analysis within a segment is finished, segments are compared with each other for resultant designations, so that the "excess" words in certain segments can be connected with the corresponding "unsatisfied" words in other segments. As a result, some of the boundaries between segments are removed and a primary unification of segments obtains. Analysis is repeated, if necessary, with respect to the configurations within the enlarged segments and then a comparison is made of the segments for their designations, etc., until bonds have been established among all the words. Then, the segments will correspond to actual parts of the complex sentence. At this point, on the basis of a consideration of conjunctions and of knowledge of the structure of each segment obtained during analysis, the bonds among the segments and their hierarchy can be established. Here, analysis is completed.

The general organization of the analysis is a separate question.[5] In several projects, separate steps have been used following glossary lookup, consisting of morphological analysis, the finding of idioms, resolution of homographs, treatment of words with various peculiarities, etc. For example, the French-Russian ([6], [8]) and Hungarian-Russian [12] algorithms of MI and LI, and the Georgetown University algorithm, "SERNA" ("*S Russkogo Na Anglijskij*"—from Russian to English) [59], are so constructed. During later research it developed, however, that the indicated stages are not basically different from syntactic analysis. Actually, idiom determination in text is the same as the determination of phrase types, and resolution of homonyms is made on the basis of determination of the bonds among words. For this reason, the LI algorithm for syntactic analysis of Russian text includes not only idiom determination but also homonym resolution and treatment of special words. Idioms and the rules for resolving homonyms are

[5] About MT analysis, see p. 61 below.

simply special configurations in the general configuration list. This unique approach has allowed a reduction of all procedures to a small group of rote operations, which seemed convenient from the standpoint of constructing an algorithm and of programming.

4. The Problem of Meaning in Machine Translation

Since the purpose of machine translation, or translation of any kind, is transformation of text in such a way that its meaning is preserved (more or less), work on MT cannot omit a study of meaning and the level of content of languages. It is sometimes said that MT banishes meaning as an object of research, that the machine cannot make use of meaning characteristics. These assertions are simply untrue. The machine can use any characteristics, including those involving meaning, if only they are clearly described and enumerated beforehand. Isolation and description of the necessary meaning characteristics is, in fact, one of the most important problems in MT. However, the machine cannot at present actually make use of the various *extra-linguistic* factors connected with meaning (the correlation of language elements with the objects of real activity, psychological associations, etc.), since such questions have not been treated. The machine operates only with what is immediately contained in the text. Therefore, a purely linguistic description of meanings must be made for MT: The meaning of an element is describable by its substitutability (how it fits into synonymous series or into groups of translational equivalents in various languages)[6] and by its distribution (the appearance of the element in specified kinds of context). This approach is not the special property of MT; in fact, meaning must be studied by the same methods in linguistics as well. Here, of course, the productivity of other approaches is not denied, particularly the psychological approach. It is important only to distinguish clearly the linguistic and nonlinguistic approaches. MT forces this distinction to be made very logically.

---

[6] See below, pp. 65-66, on the "thesaurus method."

In the light of MT studies, we can consider anew such classic linguistic questions as that of homonymy and synonymy. Thus, from the MT point of view, one can speak of homonymy when the same sequence of elements (e.g., letters) must, for the sake of satisfactory translation, be treated variously. The distinction between homonymy and polysemy is not made at this time, since it makes no difference to the machine at MT's present stage of development whether or not there is any connection in meaning between two possible translations of a particular word. Later, when we have more complete systems of "semantic factors" (see p. 66), this distinction will become essential, and its value will be exactly measured by a group of general "semantic factors" constituting the meanings of the two words compared.

Unfortunately, general theoretical questions connected with research on the meaning aspect of language for MT purposes have not been treated at all. For this reason, we shall limit our discussion to one of the practical aspects of the broad theme: "meaning in MT." We have in mind the problem of multivalence.

Elimination of the multivalence of language elements (words, grammatical indicators), in its broadest sense (including the various cases of homonymy, see above), is a basic problem of MT in its more general form. Multivalence on the MT level means the presence of several translations; the removal of multivalence is the choice of the necessary equivalent from among the several possible ones. If multivalence did not exist, and the machine did not have to make such a choice, then MT would be reduced to very simple transformations.

The problem of multivalence of language elements (mainly that of words) is constantly being discussed in MT studies. Many suggestions have been made concerning automatic elimination of lexical multivalence. They can be grouped as follows:

(1) Limitation of multivalence according to subject-matter. It is proposed to apply special idioglossaries in which words are given only the meanings applicable to them within a given field. One could also furnish each translation of a word with a code indicating the area in which it is applicable.

(2) Reducing the number of translations by choosing the most general translations (i.e., those that can be stretched to fit all instances and still not confuse the meaning of a text, though weakening the style) or the most probable (the most frequent) translations.

(3) Context analysis. Interesting research by A. Kaplan [37] has shown experimentally that context, even when understood to be simply *adjacent words,* possesses considerable "force" for removing multivalence. Obviously, if by the context of a multivalent word we mean "words immediately connected syntactically with the given word," then the "differentiating force" of such context will be still greater. For just this reason, V. H. Yngve proposed a solution of the problem of lexical multivalence based on a previously developed syntactic structure for the sentence being translated [64]. This solution seems to be the most productive. First, the attributes of various meaning-categories (object, person, action, condition, organization, etc.) are ascribed to words; the translation of the multivalent word is chosen using rules indicating which of these attributes in words syntactically connected with the given word correspond to the choice of this translation. Something similar is done in applying the "thesaurus method" (see pp. 65-66).

A special case of the use of context for removing multivalence is the discovery of idioms having a special translation.

(4) The most "powerful,2 but at the same time an extremely complex, means of removing ambiguity consists of giving the machine so many designations of meaning and the connections among them that it can "understand" the content of a text (in the broad sense of the word). Then, besides syntactic bonds, the machine can in translating make use of the meaning relations—rules showing the permissible combinations of semantic designations. Given such a capability, the machine can correct faulty text (with typographical errors, omissions, faults) by the meaning.

Special work is being done for transition to such semantic analysis with the purpose of obtaining a sufficiently full collection of the simplest semantic elements, such that through combinations of these, one can represent the meanings of any language units. Such elements have been called "semantic factors" [3]. Semantic factors are necessary not only for MT but also for many other operations on text, especially referencing and correction, as well as for encoding scientific-technical information to be stored and operated upon by so-called information machines.

Several groups are working on extracting semantic factors for texts in various fields of knowledge. We cite in particular J. Perry and A. Kent's group in the U.S.A., the Cambridge group in England, and the MT Laboratory at the First Moscow State Pedagogical Institute of Foreign Languages.

We shall not treat in detail the question of a method for expanding meanings into semantic factors. Basically, this method consists of defining semantic factors by determining the correspondences among the various elements both within one language and between languages. Later, when we discuss interlingua and, in particular, the specification of the semantic elements of an interlingua, we shall describe one of the methods applied—the so-called "thesaurus method" (see pp. 65-66).

The construction of sets of semantic factors is especially valuable for linguistics because it permits the study of meanings as systems, i.e., as units formed by definite rules from a small number of simpler elements.

5. Interlingua

The problem of interlingua for MT, formulated at an earlier stage of MT's development, is frequently discussed in the literature and in MT publications.[7] Nevertheless, it is far from a final solution; moreover, complete clarity has not as yet been attained in several general representations of interlingua. We shall confine ourselves to a short résumé of some of the ideas expressed on this subject.

In nonliteral MT (and frequently also in word-for-word MT see V. H. Yngve's remarks on p. 64), the translation process is separated into two stages: analysis and synthesis.

In *analysis,* specific data about the text being translated (information about the translations of words, their morphological forms, the connections among words, etc.) are extracted from

7 . Reifler's paper at the first MT conference, 1952 [51].

it. These data express the same meaning[8] as the input text but explicitly and unambiguously, unlike the language elements, which are connected with the meaning inexplicitly and ambiguously (meaning may, for example, be expressed by the relative distribution of the language elements). The set of data we can obtain from analysis is so arranged that, by referring to it, we can construct an output text. Constructing texts from analysis data is the converse of analysis and is called *synthesis.*

For every language, data are collected consisting of the characteristics needed for a unique and explicit expression of the meaning of texts in this language. These characteristics are, on the one hand, the goal and result of analysis and, on the other, the raw material for synthesis. The set of characteristics is developed for a concrete language with the introduction of its grammatical categories and others necessary and convenient for translation of the information. This set is, in fact, the unique "intermediary language."

In binary translation (from one language to another in a given direction), analysis of the input language is performed immediately in terms of the characteristics of the output; this is so-called "dependent analysis." For example, in French-Russian translation, the cases of nouns are immediately determined during analysis of the French text, since these characteristics are needed for synthesizing the Russian text.

But in multiple translation (from many languages to many others in any direction), such an approach is not very useful; as many analysis algorithms are needed for each input language as there are output languages proposed (each algorithm leads from the text in the input language to the characteristics of one of the output languages). Thus, in "dependent analysis" ten languages would need ninety analysis algorithms (nine "dependent analyses" for every language) and ten synthesis algorithms (since synthesis is always independent).

In order to avoid a large number of algorithms, we can apply "independent analysis": For each language there is just one analysis algorithm leading from the text in the input language to the characteristics of this language, and one synthesis algorithm performing the converse operation. In addition, there is

[8] Or rather, almost the same; some loss of information may occur.

a set of rules by which the characteristics of the input language derived from analysis are transformed into the characteristics of the output language needed for synthesis. This set of rules is also an interlingua. For example, the interlingua of the MT group at M.I.T. ([63], [64]) can be understood in this way.

There exists yet another approach, as follows: After the necessary correspondences have been made between the sets of characteristics of concrete languages, these sets are united in a particular manner into one maximal set (macroset) that suffices for the unique expression of the meaning of a text in any of the input languages. This universal set of characteristics is regarded as an interlingua. Then, analysis will always lead directly from the input text to universal characteristics, and synthesis begins immediately with these characteristics. In this approach, a special stage of transformation (between analysis and synthesis) is apparently practically nonexistent, because of the inclusion of aspects of transformation in analysis and synthesis.

The interlingua, in this sense, is nothing other than a notational system applicable for a unique, explicit, and sufficiently suitable expression of the meaning contained in texts in languages subjected to translation.

This position is entirely in agreement with the principles of the "100 per cent approach" to MT mentioned above, which requires that translation be realized "by the meaning," i.e., that the meaning be extracted from the text being translated, written in a special, standard form, and then that the output-language text be constructed only according to this meaning, independent of the input text.

Before proceeding to the question of the form of an interlingua, we shall touch, in passing, on the necessity for an interlingua that has arisen in the literature.

The opponents of interlingua have indicated that its advantages (reducing the number of analysis algorithms) can only become effective for a rather large number of languages, while for three or four—and especially for only two—languages, the interlingua is not at all necessary, since it yields little advantage in the number of algorithms and complicates each of them. However, as we have said earlier, in binary translation, too, a certain "intermediary language" is applied—e.g., the characteristics of the output text obtained from analysis. V. H. Yngve has shown that nearly all algorithms apply an "intermediary language" even if inexplicitly and unconsciously. For example, in the French-English algorithm of Birkbeck College (in England), the dictionary is divided into French and English parts. Each French word has stored with it not its English equivalent but only the address of the location of its equivalent. The set of addresses in fact represents the "intermediary" or transitional language, as Yngve has called it. Such a "language" permits the writing of language information in the machine in the most economical form and is convenient in machine operations. Since these "intermediary languages" exist, they must be applied deliberately. It now becomes apparent that interlingua is necessary both in binary translation and in multiple, and Yngve's group (at M.I.T.) is occupied with developing an interlingua for German-English translation.

Of course, there remains the purely terminological question: Should one call just any "intermediary" (transition) language an interlingua?

Still another argument is used against interlingua: Interlingua, while decreasing the general number of analysis algorithms from $n + n (n — 1)$ to $2n$, i.e., in the ratio $n^2 : 2n = n/2$, (for twenty languages, a tenfold reduction), seems to lead to greater complexity of the algorithms. But this assertion is rather indefinite, for there does not exist at present a way of evaluating the "complexity" or the "simplicity" of algorithms. Moreover, no one has yet compared algorithms constructed in conjunction with an interlingua with algorithms in which interlingua is not used at all (if the latter, in fact, exists; see above).

At present, the need for interlingua as such is recognized by all groups in the U.S.S.R., by the researchers in the Cambridge group in England, by V. H. Yngve's group in the U.S.A., and by others. However, the form of the interlingua is not as yet decided upon.

In the literature four types of interlinguas are discussed:

(1) One of the natural languages may be used as an interlingua (e.g., the language of the country in which particular MT algorithms are being created). But since the interlingua must ensure a monovalent, explicit, and maximally economical

notation for meaning extracted from the input text, and no natural language satisfies these requirements, this method apparently is not being followed consistently by anyone in practice.

(2) The interlingua may consist of a standardized and simplified natural language. An example of this is the "Model English" proposed by Stuart C. Dodd [41].

(3) The interlingua may be one of the artificial international languages, such as Esperanto or Interlingua. The use of Interlingua as an interlingua has been studied by A. Gode [31].

(4) However, a method more likely to be useful is the creation of specially adapted artificial languages for MT. Pioneer groups dealing directly with the problem of interlingua (at Cambridge, at Leningrad University, and at the MT Federation in Moscow) have all come to the same conclusion: construction of an interlingua as a system of correspondences among natural languages (for simplicity in presentation, we shall not touch upon the differences existing among the approaches used by the groups mentioned). This viewpoint is most fully presented in the publications of the Cambridge group in presenting the so-called "thesaurus method" ([42], [43]).

A thesaurus is a particular kind of dictionary in which words are grouped into thematic classes that are divided into sections and, further, into categories.[9] In the most famous dictionary of this kind—*Roget's International Thesaurus of English Words and Phrases*—there are six classes, twenty-four sections, and more than 1,000 categories. For example: The class "Space" includes the sections "General," "Measurement," "Form," "Motion"; the section "Motion" is divided into the categories "Change of Position," "Rest," "Land Travel," "Flying (air travel)," "Traveller," "Sailor," "Aeronaut," etc. In addition to being joined into thematic groups, the words are listed alphabetically, and each is assigned numbers (or headings, called "heads") for the thematic groups to which it belongs.

A word can belong simultaneously to several groups, as in the case of homonyms ("rock," as *skala* [cliff, crag], or as *kachat'*

[9] The term "thesaurus" is also used to refer to dictionaries in which the lexical system of a language is presented quite thoroughly.

[to rock]), or in the case of polysemy ("rod," as *sterzhen'* [stirring rod], or as *rozga* [birch rod]).
The entry for the word "flat" from *Roget's Thesaurus* is:

flat 172—inertia
191—story, level
207—low
213—horizontal
223—color
etc.

In other words, every word has assigned to it series of synonyms with which it is associated (in various meanings); a series of synonyms (or rather, the group nearest to the word in meaning) forms a thematic category.

Thesauri can also be interlingual. In that case, groups of words from several languages, similar in meaning, are joined into the same thematic category.

Translation of lexical content is done in two stages when an interlingual thesaurus is used:

(1) There may be several thematic-category numbers with the word to be translated, and the necessary number (that most suitable in the given context) is chosen first; for this purpose, sets of such numbers are compared for syntactically connected words, and common numbers are selected.

(2) All words in the output language that are near in meaning, and might *in a particular context* be the equivalents of a given word, are pulled according to their thematic-category number. The choice of the proper equivalent from among several possible ones is made according to special rules belonging entirely to the output language.

In the specially constructed thesaurus, where groups of words in various languages are, taken as a whole, mutually and uniquely related to one another, thematic-category numbers may be thought of as *the words of an interlingua.*

The relations among semantic elements (words) in the interlingua can be expressed by the same indexes, with symbols for the related elements [55], or with parentheses grouping pairs of elements—the defining and the defined—so that a pair included in parentheses may be thought of as a single element

[49]. The interlingua of the Cambridge group does not have grammar in the general sense (number, case, tense).

We shall not describe in detail the approaches of the Leningrad and Moscow groups to the problem of interlingua but shall refer the reader to the relevant publications: [1], [2], [3], [13]. We shall only note that workers in the Leningrad group have already obtained practical results. They have developed an experimental version of interlingua for a series of natural languages (Russian, Czech, English, Indonesian, and others), and soon an experimental machine translation should be realized from any one of these languages to another, using interlingua. Along with the interlingua created by determination of the correspondences among natural languages, another type of interlingua is possible: purely logical, developed from analysis of the content of some science but without introduction of data from natural languages. Apparently, the members of Perry and Kent's group in the U.S.A. and of the Electromodeling Laboratory of VINITI[10] in the U.S.S.R. are following this method.

6. Formalization of Algorithm Notation

In conjunction with the problem of interlingua, much attention has been drawn to the question of a specialized "language" for MT algorithm notation. Because such a "language" permits a generally known standardization of algorithms, it simplifies their construction and control and, most of all, essentially simplifies their programming by permitting a transition to automatic programming. Formal notation for algorithms presupposes the use of a small number of precisely defined expressions (commands, questions, etc.). A standard program is made for the realization of each such expression. Then, since all expressions have a standard form, the machine can decipher these expressions and replace them with the corresponding programs. In other words, automatic programming is nothing other than a machine translation of the MT algorithm itself from the

[10] [VINITI = All-Union Institute of Scientific and Technical Information.—Tr.]

language in which it was written by the analyst to the internal language (the so-called "order code") of a particular machine. Naturally, the more standardized and regular the initial notation of the algorithm, the more simply the corresponding translation is realized.

Several MT groups apply a logico-mathematical symbolism as algorithm notation for finding predicates, augmented by a series of conditional designations (the Harvard and Georgetown groups in the U.S.A.). A special symbolic language, which includes designations of language elements and of the operations being performed, has been developed by the Leningrad group. This language has been proved in practice—for writing several algorithms (11).

The language presented by Yngve for writing algorithms (his *programming language)—COMIT* [65]—has still another structure. The essence of Yngve's idea is that a single standard form is used for writing the rules composing the program. Each rule has five parts. The number of the rule is written in part I, and part V contains the number of the rule to which to proceed after carrying out the operations required by the present rule. In part II are indicated the elements (words, parts of words, etc.) or attributes on which to perform the operation; what is to be done with these elements or attributes (substitution, erasure, or addition of elements; ascription or erasure of attributes; etc.) is shown in part III. Part IV defines the boundary of the algorithm to which the particular rule applies, and sometimes contains an indication about a transition to this or that subrule of the rule (this indication to be used by a special part of the algorithm, called the "dispatcher").

COMIT is used by the MT group at M.I.T. for writing algorithms, in particular, a German-English algorithm. COMIT is also beginning to be applied by several other groups in the U.S.A.

The so-called "operator notation" for MT algorithms developed by O. S. Kulagina ([4], [5]), in addition to introducing a standard form of rules, contains a whole list of allowable operations—operators. An operator is a small algorithm handling one precisely specified part of a problem: e.g., verifying the presence of an attribute, noting an attribute, searching for words with particular attributes. The operator has a *fixed* internal structure but *variable* parameters; thus, one and the same operator can, for example, verify the presence of various attributes for various words. Kulagina's operators are like standard details [i.e., components] from which the MT algorithm is formulated.

On the basis of the analytic part of the French-Russian algorithm ([6], [8]), Kulagina selected seventeen operators: three different verification operators, two different search operators, an erasure operator, an operator for inserting words, etc. These operators are not all bound to the specifics of the French language and can be applied in algorithms for a number of other languages.

Thanks to the application of operators, the logical structure of the algorithms becomes quite explicit, and their construction is thus simplified. Operator notation permits a transition to the automatic programming of algorithms. Kulagina has perlormed an experiment in automatic programming of part of the Hungarian-Russian algorithm [12]; in five minutes, the machine constructed five programs that would have taken twenty to thirty man-days.

The idea of operator notation seems highly productive; at present, and as a continuation of Kulagina's work, a compilation of so-called algorithm operators is being made [14]. Operators connected with programming technique, with peculiarities of realization, are excluded from this compilation, and new operators resulting from the creation of a more complex type of algorithm are introduced.

7. The Interaction of Man and Machine during MT

This question has many interesting facets of which we shall mention several here.

Man can participate in the process of MT either by initially preparing the text to aid the machine in handling multivalence, etc. (pre-editing), or by the necessary polishing of the rough translation made by the machine (postediting). The question of the usefulness of pre- or postediting (or of both) still remains unsolved. Most researchers are inclined to prefer postediting, though there are no exact figures on this. Evidently, Y. Bar-Hillel vas right [22] in emphasizing the importance of pre- or postediting and in indicating that, since high-quality, *fully* automatic translation is not at first achievable, it would be desirable to organize an intelligent interaction between man and machine and to arrive as quickly as possible at partially automatic mass translation. This would permit the accumulation of valuable experience for the further development of machine translation.

Electronic computers can be successfully applied to assist humans in varied research on language. During the 1950's, several experiments were conducted in which the machine helped to produce, with minimal expenditure of time and effort, listings ("concordances") of large quantities of text: of the Bible, of the preachings of Thomas Aquinas, of the Dead Sea Scrolls, etc. (see papers by Cook [25], Tasman [57], and Ellison).

All of these experiments demonstrated the usefulness of computers in various kinds of lexicographic work (extracting dictionary materials from text, sorting these materials, etc.), and for all sorts of statistical counts: machine-aided calculation of the frequencies of letters and morphemes, words, and even syntactic constructions; thus, the National Bureau of Standards produced a frequency count for various kinds of syntactic constructions for English using the SEAC [56]. Such application of machines has great value not only for MT but also for linguistics as a whole.

Experiments involving "learning machines" are especially interesting; "learning" is used here in its broadest conditional sense. The simplest such experiment involves a machine's completing its own dictionary independently during the translation. A word in the text to be translated, but not in the dictionary, is pulled from the text along with the defining context and an indication of its text location; then it is placed in the dictionary in alphabetic order. A man then writes the necessary dictionary information (in the MT groups of Harvard University, U.S.A., and at Birkbeck College, England).

In the MT studies being conducted by the group at The RAND Corporation (U.S.A.), the machine is expanding the list of elementary syntactic constructions available to it.[11] Sequences of words not corresponding to any in the list are printed out by the machine along with their text location and are classified by specific characteristics for later study by linguists.

We should note experiments in applying the machine for automation and even for automatically writing MT algorithms. For example, a plan developed at the Computation Laboratory of Harvard University is as follows. A word-for-word Russian-English translation is made with the aid of the machine. This translation is corrected by a posteditor using special instructions prescribing definite actions and the writing of changes introduced in a standard form. The postedited translation is again input to the machine, which compares it with the initial (word-for-word) translation, discovers the differences, collects and classifies them, and then, on the basis of an analysis of these differences, constructs an algorithm capable of introducing into the word-for-word translation the same changes that had been written in by the posteditor. This algorithm is included in the initial stage of the translation, and initial translation improves. Now the posteditor receives something better than a word-for-word translation. Once again he corrects the text, which is again input to the machine, and the cycle is repeated until the quality of translations output by the machine satisfies the posteditor. Thus the machine is able, as it were, to "learn" by analyzing and imitating the actions of the posteditor ([30], [36], [45]).

8. Some Facts about Work in MT

In the preceding sections no exhaustive characterization of all the basic problems of MT is to be found. These sections are meant only to give the reader a general idea of the state of machine translation.

Machine translation is a little over ten years old. The idea of mechanizing translation from one language to another was expressed by the Soviet inventor P. P. Troyansky as far back as

---

11. [In fact, the machine has not done more than aid in the expansion.—Tr.]

1933; in that year Troyansky obtained a patent for his translating machine.[12] But at that time Troyansky's ideas did not receive the necessary development. After the invention of high-speed electronic computers, the idea of mechanizing translation with their aid arose once again (1946, Weaver and Booth); in 1949, the first research was begun (in the U.S.A.). In 1952, the Massachusetts Institute of Technology called the First Conference on MT, and from then on the number of publications dedicated to MT questions has risen steadily. In the beginning of 1954, IBM conducted an experiment in Russian-English translation on the IBM 701. Thus, the possibility of MT was proven in practice. In the U.S.S.R., work on MT began in 1954, and in 1956, English-Russian and French-Russian translations were realized. Since 1955, more new groups have joined in MT research. The scope of the work has been increasing steadily.

At present, machine translation is being pursued in the following countries: the U.S.S.R., the U.S.A., England, Japan, China, Czechoslovakia, Italy, France, Sweden, Israel, Mexico, and India. Only the U.S.A. has more than ten groups participating. These groups are concentrated in the larger research centers, such as the universities—Harvard, Georgetown, Washington, Chicago, M.I.T., and others; and in corporations—RAND and Ramo-Wooldridge; etc. The largest of the groups includes dozens of workers. There are two groups at work in England (Birkbeck College and Cambridge University[13]). In the U.S.S.R., five groups in Moscow are working on MT and related problems (at the Pedagogical Institute of Foreign Languages and at four institutes of the Academy of Sciences: the Mathematics Institute, the Institute of Precise Mechanics and Computer Techniques, the Electromodeling Laboratory of VINITI, and the Institute of Linguistics); and there is one group in each of five other cities: Leningrad (Leningrad University), Kiev (Kiev University and the Computational Cen-

[12] See the brochure "Perevodnaya mashina P. P. Troyanskogo" ["The Translation Machine of P. P. Troyansky"], published in 1959 by the Izd-vo Akademii nauk SSSR, Moscow, pp. 1-40 (translated in JPRS 3532, U.S. Joint Publications Research Service, July, 1960, pp. 1-39).

[13] [The Cambridge Language Research Unit is actually independent of the University.—Tr.]

ter of the Academy of Sciences), Erevan (Computational Center of the Armenian S.S.R.), Tbilisi (Institute of Automatics and Telemechanics of the Georgian Academy of Sciences), and Gorky (Radiophysical Institute).

In the U.S.A. and in the U.S.S.R., special journals are published on MT: *Mechanical Translation* (M.I.T.) and *Mashinnyj perevod i prikladnaya lingvistika* [*Machine Translation and Applied Linguistics*] (Moscow Institute of Foreign Languages).

The group of languages being machine translated has greatly increased. Whereas attention at first was primarily concentrated on Russian and English, work is now being conducted on MT in the following languages as well: French, German, Italian, Chinese, Hindi, Japanese, Indonesian, Arabic, Hungarian, Czech, Georgian, Armenian, and others.

From 1957 to 1960, quite a few experimental machine translations were made both in the U.S.S.R. and abroad. At the Mathematics Institute, French-English translation experiments have been conducted that include translations of selected running texts; examples of phrases translated by the machine appear in [7] and [23]. Recently, English-Russian translation experiments have been begun there, too.

Experimental Russian-English translations have been made by various groups in the U.S.A. The Harvard and Georgetown groups and that at The RAND Corporation conduct these experiments more or less regularly.

MT experiments have been conducted successfully from French to English, from Russian and English to Chinese, from English to Japanese, and from English to Czech in England, China, Japan, and Czechoslovakia.

The experience accumulated as a result of these experiments has permitted the serious undertaking of *mass* MT. Further development of the theory of MT needed here will lead to the presentation of new and interesting problems and will have considerable influence on linguistics as a whole.[14]

---

[14] The author expresses his sincere gratitude to V. V. Ivanov, A. A. Reformatsky, O.S. Kulagina, and L. N. Iordanskaya for their valuable notes and advice.

BIBLIOGRAPHY FOR CHAPTERS III AND IV

1. Andreyev, N. D., "Mashinnyj perevod i problema yazyka-posrednika" ("Machine Translation and the Problem of an Intermediary Language"], *Voprosy yazykoznaniya,* Vol. 6, 1957, pp. 117-121.
2. ———, "Meta-yazyk mashinnogo perevoda i ego primenenie" ["A Metalanguage of Machine Translation and Its Use"], in *Materialy po mashinnomu perevodu,* Vol. I, Izd-vo Leningradskogo gosudarstvennogo universiteta, Leningrad, 1958, pp. 40-60. (Translated: JPRS 2150, U.S. Joint Publications Research Service, February, 1960, pp. 26-40.)
3. Ivanov, V. V., "Lingvisticheskie voprosy sozdaniya mashinnogo yazyka dlya informatsionnoj mashiny" ["Linguistic Problems in the Creation of a Machine Language for Information Machines"], in *Materialy po mashinnomu perevodu,* Vol. I, Izd-vo Leningradskogo gosudarstvennogo universiteta, Leningrad, 1958, pp. 10-39. (Translated: JPRS 2150, U.S. Joint Publications Research Service, February, 1960, pp. 6-25.)
4. Kulagina, O. S., "Ob operatornom opisanii algoritmov perevoda i avtomatizatsii protsessa ikh programmirovaniya" ["Operator Description of Algorithms for Translation and Automation of the Process of Programming Them"], *Problemy kibernetiki,* Vol.2, 1959, pp. 289-302.
5. ———, "Operatornoe opisanie algoritmov perevoda" ["Operator Description of Algorithms for Translation"], *Mashinnyj perevod i prikladnaya lingvistika,* No. 2(9), 1959, pp. 6-22, and No. 3 (10), 1959, pp. 3-34. (Translated: JPRS 3599, U.S. Joint Publications Research Service, August, 1960, pp. 5-15.)
6. ———, "O mashinnom perevode s frantsuzskogo yazyka na russkij, I" ["French-to-Russian Machine Translation, I"], *Problemy kibernetiki,* Vol. 3, 1960, pp. 181-208.
7. ———, and G. V. Vakulovskaya, "Opytnye perevody s frantsuzskogo yazyka na russkij na mashine 'Strela'" ["Experimental French-Russian Translations on the Machine 'Strela'"], *Problemy kibernetiki,* Vol. 2, 1959, pp. 283-288.
8. Kulagina, O. S., and I. A. Mel'chuk, "Mashinnyj perevod s frantsuzskogo yazyka na russkij" ["French-to-Russian Machine Translation"], *Voprosy yazykoznaniya,* Vol. 5, No. 5, 1956, pp. 111-121.
9. Liu Yüng, T''uan, "Issledovatel'skaya rabota v oblasti MP v

Kitajskoj Narodnoj Respublike" ("Research on Machine Translation in the Chinese People's Republic"], *Voprosy yazykoznaniya,* Vol. 8, No. 5, 1959, pp. 102-101.

10. ————, "Vopros o poryadke slov i ego reshenie pri MP s russkogo yazyka na kitajskij" ["Problems of Word-Order in Russian-Chinese Machine Translation and Their Solutions"], *Yu-yen Yen-chiu,* No. 4, 1959, pp. 107-116. (Translated: JPRS 3356, U.S. Joint Publications Research Service, June, 1960.)
11. "Voprosy statistiki rechi" ["Questions on the Statistics of Speech"], *Materialy po mashinnomu perevodu,* Vol. I, Leningrad State University Press, Leningrad, 1958.
12. Mel'chuk, I. A., "O mashinnom perevode s vengerskogo yazyka na russkij" ["Machine Translation from Hungarian to Russian"], *Problemy kibernetiki,* Vol. 1, 1958, pp. 222-264. (Translated: JPRS 646-D, U.S. Joint Publications Research Service, April 10, 1959.)
13. ———, "Raboty po mashinnomu perevodu v SSSR" ["Work on Machine Translation in the USSR"], *Vestnik AN SSSR,* No. 2, April 24, 1959, pp. 43-47. (Translated: JPRS 662, U.S. Joint Publications Research Service.)
14. ———, "O standartnykh operatorakh dlya algoritma avtomaticheskogo analiza russkogo nauchnogo teksta" ["Standard Operators for Automatic Analysis of Russian Scientific Text"], not yet published.
15. Moloshnaya, T. N., "Nekotorye voprosy sintaksisa v svyazi s mashinnym perevodom s anglijskogo yazyka na russkij" ["Certain Questions of Syntax in Connection with Machine Translation from English into Russian"], *Voprosy yazykoznaniya,* Vol. 6, No. 4, 1957, pp. 92-97.
16. ———, "Voprosy razlicheniya omonimov pri mashinnom perevode s anglijskogo yazyka na russkij" ["Problems in Distinguishing Homonyms in Machine Translation from English into Russian"], *Problemy kibernetiki,* Vol. 1, 1958, pp. 216-221. (Translated: JPRS 646-D, U.S. Joint Publications Research Service, April, 1959.)
17. ———, "Algoritm perevoda s anglijskogo yazyka na russkij" ["An Algorithm for Translating the English Language into Russian"], *Problemy kibernetiki,* Vol. 3, 1960, pp. 209-272. (Translated: JPRS 6492, U.S. Joint Publications Research Service, December 29, 1960, pp. 41-123.)
18. Nikolayeva, T. M., *Analiz russkogo predlozheniya* [*An Analysis of Russian Prepositions*], Institut tochnoj mekhaniki i

vychislitel'noj tekhniki, Izd-vo Akademii nauk SSSR, Moscow, 1958.

19. Rozentsvejg, V. Yu., "Raboty po mashinnomu perevodu s inostrannykh yazykov na russkij i s russkogo na inostrannye v Sovetskom Soyuze" ["Work in the Soviet Union on Machine Translation from Foreign Languages into Russian and from Russian into Foreign Languages"], *Reports of the Fourth International Conference of Slavicists,* Moscow, 1958. (Translated by Lew Micklesen, University of Washington.)
20. Sofronov, M. V., "Obshchie printsipy mashinnogo perevoda s kitajskogo yazyka" ["General Principles of Machine Translation from the Chinese Language"], *Voprosy yazykoznaniya,* Vol. 7, No. 2, 1958, pp. 116-121. (Translated: JPRS DC-319, U.S. Joint Publications Research Service, November 14, 1958.)
21. Bar-Hillel, Y., "The Present State of Research on Mechanical Translation," *American Documentation,* Vol. 2, 1951, pp. 229-237.
22. ———, *Report on the State of Machine Translation in the United States and Great Britain* (mimeographed), Hebrew University, Jerusalem, Israel, February 15, 1959.
23. Booth, A. D., "General Applications of Digital Computers," *Journal of the Institution of Electrical Engineers,* New Series: Vol. 3, No. 36, December, 1957, pp. 629-636.
24. ———, L. Brandwood, and J. Cleave, *Mechanical Resolution of Linguistic Problems,* Butterworth and Co., Ltd., London, 1958.
25. Cook, C. M., "Automation Comes to the Bible [Univac Helps Compile Concordance]," *Computers and Automation,* Vol. 7, March, 1958, pp. 16-18.
26. Delavenay, E., *La Machine à Traduire,* "Que sais-je?" series, No. 834, Presses Universitaires de France, Paris, 1959.
27. Garvin, P., "Linguistic Analysis and Translation Analysis," *Report of the Eighth Annual Round Table Meeting on Linguistics and Language Studies,* Monograph Series on Language and Linguistics, No. 10, Georgetown University Press, Washington, D.C., 1957, pp. 19-38.
28. ———, D. Lochak, M. Mathiot, and C. Montgomery, *Report of Group II—The Georgetown University Project in Machine Translation Research,* Seminar Work Paper MT-73, Georgetown University Press, Washington, D.C., 1957.
29. Giuliano, V. E., "An Experimental Study of Automatic Language Translation," *Mathematical Linguistics and Auto-*

matic Translation (Report No. NSF-1), The Computation Laboratory, Harvard University, Cambridge, Massachusetts, January, 1959.

30. ———, "A Formula Finder for the Automatic Synthesis of Translation Algorithms," *Mathematical Linguistics and Automatic Translation* (Report No. NSF-2), The Computation Laboratory, Harvard University, Cambridge, Massachusetts, March, 1959, pp. IX: 1-41.
31. Gode, A., "The Signal System in Interlingua—A Factor in Mechanical Translation," *Mechanical Translation,* Vol. 2, No. 3, December, 1955, pp. 55-60.
32. Harper, K. E., "The Mechanical Translation of Russian, A Preliminary Study," *Modern Language Forum,* Vol. 38, No. 3-4, 1953, pp. 12-29.
33. ———, "Semantic Ambiguity," *Mechanical Translation,* Vol. 4, No. 3, December, 1957, pp. 68-69.
34. ———, "Contextual Analysis," *Mechanical Translation,* Vol. 4, No. 3, December, 1957, pp. 70-75.
35. ———, and D. G. Hays, *The Use of Machines in the Construction of a Grammar and Computer Program for Structural Analysis,* The RAND Corporation, P-1588, 1959. [Published in the *Proceedings of the International Conference on Information Processing,* UNESCO, 1960.]
36. Jones, P. E., "A Feedback System for the Harvard Automatic Translator," *Mathematical Linguistics and Automatic Translation* (Report No. NSF-3), The Computation Laboratory, Harvard University, Cambridge, Massachusetts, August, 1959, pp. XIV: 1-60.
37. Kaplan, A., "An Experimental Study of Ambiguity and Context," *Mechanical Translation,* Vol. 2, No. 2, November, 1955, pp. 39-46.
38. Koutsoudas, A., and A. Humecky, "Ambiguity of Syntactic Function Resolved by Linear Context," *Word,* Vol. 13, No. 3, December, 1957, pp. 403-414.
39. Lehiste, J., "Order of Subject and Predicate in Scientific Russian," *Mechanical Translation,* Vol. 4, No. 3, December, 1957, pp. 66-67.
40. Locke, W. N., and V. H. Yngve, "Research in Translation by Machine at M.I.T.," *Reports for the Eighth International Congress of Linguists,* Vol. 11, University of Oslo Press, Oslo, Norway, pp. 315-318.
41. Locke, W. N., and A. D. Booth, eds., *Machine Translation of Languages* (14 essays, with historical review), John Wiley

& Sons, Inc., New Yolk, 1955, and M.l.T. Technical Press, 1955; Chapman and Hall, Ltd., London, 1955; Izd-vo inostr. lit., Moscow, 1957.

42. Masterman, M., "The Thesaurus in Syntax and Semantics," *Mechanical Translation,* Vol. 4, No. 1-2, November, 1957, pp. 35-43.
43. ———, "What Is a Thesaurus?," *Essays on and in Machine Translation by Cambridge Language Research Unit,* Cambridge University Press, Cambridge, England, 1959.
44. ———, A. Parker-Rhodes, and M. Hoskyns, "Skeletal Program for the Translation of *'dan c liang ze'* (The Comparative Analysis of 'However, these two, in appearance and solubility, are slightly unlike' [original Chinese]): An Account of the Pilot Project of the Cambridge Language Research Unit," *Progress Report II, Annexe V,* Vol. 2, Cambridge University Press, Cambridge, England, 1956, pp. 26-38.
45. Mattingly, I. E., "Post-editing for Feedback," *Mathematical Linguistics and Automatic Translation* (Report No. NSF-3), The Computation Laboratory, Harvard University, Cambridge, Massachusetts, August, 1959, pp. I:1-26.
46. Micklesen, L. R., "Russian-English MT," *American Contributions to the Fourth International Congress of Slavicists, Moscow, September, 1958,* Mouton & Co., 's-Gravenhage, Netherlands, 1958.
47. Oswald, V. A., Jr., and S. L. Fletcher, "Proposals for the Mechanical Resolution of German Syntax Patterns," *Modern Language Forum,* Vol. 36, No. 3-4, 1951, pp. 81-104.
48. Oswald, V. A., Jr., and R. H. Lawson, "An Idioglossary for Mechanical Translation," *Modern Language Forum,* Vol. 38, No. 3-4, 1953, pp. 1-11.
49. Parker-Rhodes, A. F., "Some Recent Work on Thesauric and Interlingual Methods in Machine Translation." (Presented at the International Conference for Standards on a Common Language for Machine Searching and Translation, September 6-12, 1959, Cleveland, Ohio.)
50. Perry, J. W., "Translation of Russian Technical Literature by Machine; Notes on Preliminary Experiments," *Mechanical Translation,* Vol. 2, No. 1, July, 1955, pp. 15-24.
51. Reifler, Erwin, "The First Conference on Mechanical Translation," *Mechanical Translation,* Vol. 1, No. 2, August, 1954, pp. 23-32.
52. ——, "Mechanical Determination of the Constituents of German Substantive Compounds," *Mechanical Translation,* Vol. 2, No. 1, July, 1955, pp. 3-14.

53. ———, "Outline of the Project," *Linguistic and Engineering Studies in the Automatic Translation of Scientific Russian into English,* University of Washington Press, Seattle, Washington, June, 1958.

M. Reynolds, A. C., "The Conference on Mechanical Translation," *Mechanical Translation,* Vol. 1, No. 3, December, 1954, pp. 47-55.

55. Richens, R. H., "Interlingual Machine Translation," *Computer Journal,* Vol. 1, No. 3, October, 1958, pp. 144-147.

56. National Bureau of Standards, "Syntax Patterns in English Studied by Electronic Computer," *Computers and Automation,* Vol. 6, No. 7, 1957, pp. 15-17, 32.

57. Tasman, P., "Literary Data Processing," *IBM Journal of Research and Development,* Vol. 1, No. 3, July, 1957, pp. 249-256.

5S. Thomas, R., "The Use of SEAC in Syntactic Analysis," *Report of the Eighth Annual Round Table Meeting on Linguistics and Language Studies,* Monograph Series on Language and Linguistics, No. 10, Georgetown University Press, Washington, D.C., 1957, pp. 151-161.

59. Toma, P., *SERNA System,* Georgetown University Press, Washington, D.C., June, 1959.

60. Wall, R., "Use of the IBM 650 Computer for the Study of Syntax in the Solution of the Problem of Multiple Meaning," *Linguistic and Engineering Studies in the Automatic Translation of Scientific Russian into English,* University of Washington Press, Seattle, Washington, June, 1958.

61. ———, and Udo K. Niehaus, "Russian to English Machine Translation with Simple Logical Processing," *Linguistic and Engineering Studies in the Automatic Translation of Scientific Russian into English,* University of Washington Press, Seattle, Washington, June, 1958. (Presented at the AIEE Fall General Meeting, Chicago, Illinois, October 7-11, 1957.)

62. Worth, D. S., "Linear Content," *Word,* Vol. 15, No. 1, April, 1959, pp. 183-191.

63. Yngve, V. H., "Sentence-for-Sentence Translation," *Mechanical Translation,* Vol. 2, No. 2, November, 1955, pp. 29-37.

64. ———, "A Framework for Syntactic Translation," *Mechanical Translation,* Vol. 4, No. 3, December, 1957, pp. 59-65.

65. ———, "A Programming Language for Mechanical Translation," *Mechanical Translation,* Vol. 5, No. 1, July, 1958, pp. 25-41.

66. ———, "A Model and an Hypothesis for Language Structure," *Proceedings of the American Philosophical Society,* Vol. 104, No. 5, 1960, pp. 444-466.