MECHANICAL TRANSLATION

DEVOTED TO THE TRANSLATION OF LANGUAGES WITH THE AID OF MACHINES

VOLUME FIVE, NUMBER THREE DECEMBER, NINETEEN FIFTY EIGHT

News

NATIONAL SYMPOSIUM

A National Symposium on Machine Translation will be held at the University of California at Los Angeles, February 2-5, 1960. The symposium will be sponsored jointly by the University of California, Berkeley; University of California, Los Angeles; University of Southern California, and the University of Washington. The purpose of the Symposium is to provide an accurate appraisal of the current state of progress and a description of methods now being used. There will be sessions in which reports on current research in the field are given, as well as sessions on methodology, grammatical studies, syntax, dictionary construction, semantic problems, information processing and its relation to linguistic analysis, programming for machine translation, and the design of equipment for information processing.

HARVARD UNIVERSITY

The Harvard Automatic Dictionary file, now containing about 30,000 stem entries representing approximately 15,000 Russian words or over 150,000 distinct inflected forms, can be made available on magnetic tape to responsible research workers. The entries are primarily from the fields of mathematics and electronics, but include an adequate supply of general words. The format of the entries, the details of the grammatical codes they contain, etc. are fully described in a series of reports. The file is now being used at Harvard as a tool for syntactic research.

Address inquiries to:

Anthony G. Oettinger
Harvard University
Computation Laboratory
33 Oxford Street
Cambridge 38, Mass.

SUMMER WORKSHOP

The M.I.T. group is planning a third Summer Workshop on German to English mechanical translation. Its purpose is to provide an informal and stimulating atmosphere in which experienced people as well as those who are new in the field can exchange ideas and work on various problems in German and English morphology and syntax from the mechanical translation point of view.

Those who are interested in obtaining further details about this program should write the undersigned.

Victor H. Yngve
Room 20 D 102
M.I.T.
Cambridge 39, Mass.

ADDITIONS AND CORRECTIONS

The news item in the last issue of MT (vol. 5, no. 2) reporting the development of an English-to-Japanese electronic reader and printer at Denki-Shikenjo, Tokyo, should have included the following information:

Anyone who wishes to obtain further information about the pilot models of the reader and printer should contact:
Denki Shikenjo, Denshi-bu
Nagata-cho 2-1
Chiyoda-ku
Tokyo, Japan

We wish to thank Mr. Ichiro Honda of the Department of Psychology at Kyoto University (Kyoto, Japan) who translated the reports of the development for calling our attention to this omission.

*The Work on Machine Translation in the Soviet Union **

Fourth International Congress of Slavicists Reports, Sept. 1958

V. Yu. Rozentsveig, First Moscow State Pedagogical Institute of Foreign Languages, Moscow, USSR

Problems of machine translation have been investigated in the Soviet Union since 1955.[1] A number of groups are carrying out theoretical and experimental work in the area of machine translation.

In the Institute of Precision Mechanics and Computer Technology of the Academy of Sciences of the USSR (ITM and VT) dictionaries and codes of rules (algorithms) have been compiled for machine translation from English, Chinese, and Japanese into Russian; and a German-Russian algorithm is being worked out. Experimental translations of individual passages have been made.[2] In the work of the ITM and VT group there is a marked striving for the rapid achievement of immediate, practical results. The efforts of this group are directed not so much toward a theoretical comprehension of the general problem of machine translation as toward a careful, detailed investigation of linguistic material, especially lexical. Dictionary routines, routines for analysis of the sentence in the source language, and routines for the synthesis of the sentence in the target language are being compiled in the ITM and VT on the basis of traditional methods of describing a language.

An essentially different course is being followed by the group working in the Steklov Mathematical Institute of the Academy of Sciences (MIAN). The problem of machine translation is being examined here as part of the larger problem of the automation of thought processes. The directors of this group regard the effective practical realization of machine translation only as the result of profound theoretical research in the area of mathematics and linguistics.

In MIAN three algorithms have been elaborated: French-Russian, English-Russian, and Hungarian-Russian.[3] During the compilation of the first of these algorithms in 1955-56, the workers in this group proceeded empirically, i. e. they extracted the rules for the translation of each word from a comparative analysis of French texts and their Russian translations. In the elaboration of the English-Russian algorithm, the MIAN group posed for themselves a more complex problem -- determination of the correspondences between the grammatical structures of two languages. The posing of such a problem was partially conditioned by the nature of the relationships of the English and Russian languages: although it was possible to build the analysis of a sentence on a morphological basis in translating a French mathematical text into Russian, such a method did not seem rational to the MIAN group in the case of English-Russian translations of similar texts. The problem was also partially conditioned by the theoretical goal of the director of the group. Professor A. A. Lyapunov: to work out strictly formal methods of describing languages in order to attain gradual automation of the whole process of machine translation.

---

* Translated by Lew R. Micklesen, Department of Far Eastern and Slavic Languages and Literature, University of Washington, December 1958.

---

1. The idea of machine translation was advanced even in the 30's by the inventor-technician, P. P. Smirnov-Troyansky.

---

2. I. K. Bel'skaya, "Concerning Certain General Problems of Machine Translation," Abstracts of the Conference on Machine Translation, Moscow, 1958, pp. 10-14, (hereafter referred to as Abstracts CMT).

---

3. See O. S. Kulagina and I. A. Mel'chuk, "Machine Translation from French to Russian, " Voprosy Yazykoznaniya, 1956, No. 5; T. N. Moloshnaya, "Some Problems of Syntax in Connection with Machine Translation from English to Russian, "Voprosy Yazykoznaniya, 1957, No. 4.

The theoretical basis for the isolation of typical sentence structures was the concept of the syntagma (according to de Saussure) or of the construct (according to Fortunatov). Machine translation, however, requires a certain modification of this system. In the structural syntactic analysis proposed by the author, T. N. Moloshnaya, of the English-Russian algorithm worked out at MIAN, constructs consisting not only of two members but also of many members (constructions with an absolute participle, etc.) are isolated. Such elementary structures were called configurations. They are composed of words classified according to formal signs. The analysis consists in reducing each configuration to its basic word, that is, shortening it. In this way, syntactical links are established between the words of a sentence. Synthesis of the Russian Sentence is made by means of substituting for it a given English configuration which corresponds to the Russian configuration and completing it with Russian words on the basis of the data of the dictionary, more precisely, of the Russian part of the dictionary, and on the basis of the corresponding morphological rules. The dictionary for machine translation, as compiled at MIAN during work on the French-Russian algorithm consists of two parts: (1) the foreign, containing the words of the given language (more precisely their stems, i.e. the graphically invariable parts of a word) with their corresponding tags indicating part of speech, idiomatic relationships, government by preposition and grammatical characteristics and (2), the Russian, containing Russian stems and the corresponding information about them. The Russian part of the dictionary is independent of the foreign part; so it may be used in translating from various languages. The rules for the morphological form of a Russian word are also independent of the language from which the translation is made.

The significance of the MIAN English-Russian algorithm lay in the fact that in contrast to all preceding algorithms in which the analysis of the text under translation was realized in terms of a translation into Russian (a category of the Russian language was ascribed to a foreign word), in T. N. Moloshnaya's algorithm the structural-grammatical analysis of an English sentence proceeded, in principal, independently of the language into which the text was being translated. This is extremely important, for an independent analysis opens the way for the realization of machine translation not only from one concrete language to another, but also from many languages to many others.

Several scientific groups are now working along this path opened up by the efforts of the MIAN Group. In the division of applied linguistics of the Institute of Linguistics of the USSR directed by A. A. Reformatsky, rules for the analysis and synthesis of a text and an abstract system of lexical and syntactic correspondences between various languages are being worked out independent of a translation into a concrete language by I. A. Mel'chuk. All of this should allow us to do machine translation from several languages into several other languages (the model of such an intermediary language is being made on the basis of an analysis of Russian, English, Chinese, French, and Hungarian). Syntactic analysis lies at the basis of the translation system being developed by I. A. Mel' chuk — morphological data are employed only as auxiliary data in the establishment of configurations, i.e. in bringing out the relationships between words in the source language and the expression of these relationships by means of the target language.

In this connection one should mention the research on the isolation and cataloguing of the system of relationships in the Russian language carried out in close collaboration with I. A. Mel'chuk in the Laboratory of Electrical Modelling of the Ail-Union Institute of Scientific and Technical Information of the State Scientific-Technical Committee in the Soviet of Ministers of the USSR and of the Academy of Sciences of the USSR(LE). In Russian mathematical texts the workers of this laboratory, Z.M. Volotskaya, E. V. Paducheva, I. N. Shelimova, and A. L. Shumilina isolated and described about 200 syntagmas (two-membered constructs in a subordinate relationship) which are essential in both the analysis and the synthesis of a Russian sentence.

A substantial contribution to the theory of translation algorithms and their programming was made by O.S. Kulagina (MIAN). She developed a system of so-called elementary operators of the simplest steps of which any translation process may consist and of programs corresponding to these steps. As a result, significant generalization and standardization in the process of making algorithms can be attained, all of which allows us to pose the problem of automation of the programming of algorithms and then the problem of their automation and construction.

The Experimental Laboratory of Machine Translation of the Leningrad State University (ELMP) under the directorship of N.D.Andreyev is also endeavoring to realize the idea of developing completely independent methods of analysis and synthesis and of some abstract logical system making it possible to go from analysis to synthesis, i.e. a system that will serve as an intermediary language. In this laboratory extensive material from various linguistic systems is being investigated; Indonesian-Russian, Arabic-Russian, Japanese-Russian, Burmese-Russian, Norwegian-Russian, English-Russian, Spanish-Russian and Turkish-Russian algorithms are being developed. The intermediary language which N. D. Andreyev is attempting to create is an artificial language constructed by averaging the phenomena of various languages. It is regarded as a material language with its lexicon, its morphology, and its syntax, but with the one peculiarity that it consists of symbols *. In the selection of the categories at the basis of his symbolization, N. D. Andreyev considers the most frequent phenomena and also the international prestige of each language.[4]

The system of signs developed in ELMP for the recording of the intermediary language can be used also for the recording of information in information machines.

Along with work on the algorithms of machine translation from foreign languages into Russian and from Russian into foreign languages being conducted in the Gorki State University, the following algorithms are being elaborated: Armenian-Russian and Russian-Armenian (in the Computation Center of the Academy of Sciences of the Armenian SSR), Georgian-Russian and Russian-Georgian (in the Institute of Automation and Telemechanics of the Academy of Sciences of the Georgian SSR).

In the First Moscow State Institute of Foreign Languages (I MGPIIYa) where under the directorship of I.I.Revzin theoretical investigations of the problems of machine translation and of related problems of linguistic theory of translation and methodology of foreign language teaching have been carried out, the elaboration of Russian-English, Russian-French, and Russian-Spanish translation algorithms for foreign policy texts has begun . At the Institute, the Machine Translation Society has been created at whose meetings theoretical problems are discussed and an exchange of ideas about the practical problems of the compilation of the algorithms takes place. In the bulletin published by the Society are published both theoretical and experimental work connected with the problem of machine translation. In May, 1958, the Society convened the First All-Union Conference on Machine Translation. Seventy-nine institutions were represented at the conference, including twenty-one institutes of the Academy of Sciences of the USSR and eight institutes of the Academies of Science of the Union Republics, eleven universities, and nineteen other institutions of higher learning in the country. Linguists, mathematicians, and technicians took part in the work of the conference. At the plenary and sectional meetings of the conference there were discussions of more than seventy reports and communications devoted to general linguistic problems arising in connection with the use of language in present-day automatic devices as well as to special problems of construction of algorithms for machine translation. [5]

The central problem now confronting linguists working in the field of machine translation is that of the methods of formal description of linguistic structures. Structural methods, particularly the methods elaborated by descriptive linguistics, offer much of value for the formal description of language — it was not by accident that the work of Fries in the structure of the English language proved useful in working out English configurations. It has become clear, however, that these methods are inadequate for the formal description of language to the extent that this is demanded in automatic translation. In connection with this a search for means of applying mathematical methods to the analysis of language was begun. With this in mind the Department of Philology of the Moscow State University initiated a seminar on mathematical linguistics in 1956, joining mathematicians and linguists under the direction of P.S. Kuznetsov, V. V. Ivanov, and V. A.Uspensky. Here, as well as at the meetings of the Machine Translation Society the idea, suggested by Academicians A. N. Kolmogorov and A. A. Lyapunov, of applying the methods of mathematical logic and of set

---

* Translator's note: The author obviously means symbols different from the conventional symbols of language.

4. N. D. Andreyev, "Machine Translation and the Problem of an Intermediary Language," Voprosy Yazykoznaniya, 1957, No. 5.

5. See Abstracts CMT, M., 1958

theory to the study of language was discussed. Thus, for example, A. N. Kolmogorov's idea about the possibility of a strict formal definition of the category of case (the work of V.A. Uspensky and, in part, also of R. L. Dobrushin) was expounded and developed. It is interesting to note that eight cases can be counted in the declensional system of the Russian substantive according to this definition.

A method for defining grammatical categories, worked out by a student of Professor Lyapunov, O. S. Kulagina (MIAN), was discussed at the seminar. This method of definition allows one to obtain, independently of the concrete features of the language, a classification of words and a determination of their syntactic relationships. Language in this conception is regarded as a set of elements — words, or more exactly — word forms. A finite number of words arranged in a definite order is called a sentence. Certain sentences are assumed to be marked — these are sentences constructed according to the norms of the given language — others are unmarked. According to the criteria of mutual substitutability of words in the marked sentences the entire set of words is broken down into groups of mutually equivalent words.

In terms of this system a series of definitions corresponding, in general, to certain traditional morphological categories, for example, parts of speech, was successfully obtained. The advantage of this classification lies, however, in the fact that it has been deduced on the basis of an exact and strictly formal system of definitions. It is particularly effective for languages with a rather symmetrical system of word forms (for example, French). In languages like Russian that do not possess this symmetry, the method of defining a grammatical category proposed by R. L. Dobrushin can be utilized.

By making use of the criterion of equivalency, the relationships between the classes of words isolated are also determined. Moreover, the concept of configuration, mentioned earlier, gets a more exact definition: a configuration is defined by O. S. Kulagina as that combination of not less than two words belonging to various non-intersecting subsets, which can be reduced to one element without any marked sentence containing this configuration losing its marked quality. Thus the combination of the words "thick book" in the sentence "the thick book lies on the table" can be reduced to the element "book" or can be replaced by the element "thing" or the element "it" without the sentence ceasing to be marked. The isolation of the configurations allows one to determine the syntactic structure of the sentence.

The set-theory concept of language is strictly deductive and formal. This is just what determines its importance both for general linguistics and for machine translation. Naturally the formal description of language is possible only to a limited extent. Thus, the concept of the marked quality of sentences, without which it is impossible to determine the equivalence of elements and configurations of a language, will have little effect if it is extended to all functional areas of language. But in a limited sphere of language — and machine translation at the present time is being considered only within the limits of scientific and technical prose — this concept is sufficiently exact and effective. Thus, all sentences in a given language which are met in a given field of scientific literature can be considered marked.

The set-theory conception of language is important in yet another respect. Since it allows us to construct and investigate a grammatical model, i.e. a simplified analog of actual linguistic relationships, this theory opens one of the possible ways for logico-semantic investigations of language. In this connection we should point to the ideas of V. V. Ivanov about the possibility of applying mathematical methods to the definition of the lexical meaning of words. I note that, contrary to wide-spread opinion, the theory of machine translation is not limited to the investigation of language in its formal aspect alone. The search for methods of objective, precise description of the system of meanings in language has begun.

If it is true that complete formal description of an actual language is hardly accessible, that it is necessary to attain only formal approximations to actual language, then a statistical evaluation of the probability of this approximation acquires special importance [6]. On the other hand, certain phenomena of language do not yield, for the time being, to structural description and can be formally described only statistically.

---

6. See V. A. Uspensky, "Conference on the Statistics of Speech," Voprosy Yazykoznaniya, 1958, No. 1, p. 173.

The quantitative aspect of linguistic phenomena, both lexical and grammatical, has been considered, as a rule, in all the algorithms formulated. One should point particularly to the statistical investigations carried out on Russian language material in the Laboratory of Electrical Modeling. I have already mentioned the cataloguing of Russian syntagmas. This work was accompanied by a statistical investigation of the language of Russian mathematical texts. The results of this work conducted by I. A. Mel'chuk, T. N. Moloshnaya, A. L. Shumilina, Z. M. Volotskaya, and I. I. Shelimova, were, along with other works, announced at the conference on the statistics of speech convoked in October 1957 by the Section of Speech of the Commission on Acoustics of the Academy of Sciences of the USSR and by Leningrad University. This work is of interest not only in a practical respect. Its value consists in a true solution to the problem of combining statistical and structural methods: a count of linguistic elements was carried out by the authors on the basis of a clear-cut definition of such concepts as "syntagma", "type of syntagma", etc. As I. I. Revzin showed in his report presented at the conference mentioned, the correlation of structural and statistical methods has a two-sided nature: statistics aids in specifying the structure of language and an exact structural definition of units, the number of which are counted, insures the proper conduct of the statistical investigation.

A frequency count of dictionary units is important not only in connection with machine translation. No longer speaking about statistical investigations of problems of general and particular linguistics [7], which have already become traditional, we shall point to recent works connected with the use of language in various devices for the storage, processing, and transmission of information. In reference to the Russian material we can call attention to the use of methods of machine translation for the coding of telegraphic and telephonic messages.

It has been established (V. I. Grigor'ev and G. G. Belonogov) that the size of a telegraph message in Russian can be diminished by 3-4 times if the telegraphic communication is translated from a letter code into a dictionary (lexical) code. Statistical investigations have shown that in the case of such coding 4, 000 common words would be sufficient in order to insure the transmission of 97.5 percent of a general-language text.

The problem examined here is connected, for the most part, with an analysis of the text under translation. For the Soviet specialists the elaboration of effective methods for analysis presented special difficulties: they dealt primarily with morphologically poor languages. It would be erroneous, however, to assume that the synthesis of the Russian sentence did not present any serious difficulties to them. By way of illustration we may cite the difficulties arising in the synthesis of Russian aspectual forms, inasmuch as the category of aspect permeates the entire Russian verbal system.

Here two problems of principle arise. In the first place, it is necessary to find a principle of classification of Russian verbs which will allow us to obtain for each verb in an absolutely regular way (by adding or taking away the same letters) all forms of the perfective as well as of the imperfective aspect. Such work was done by Z. M. Volotskaya (LE), who obtained three breakdowns of the whole Russian verbal complex according to method of formation: a) of present tense forms; b) of past tense forms; and c) of the perfective stem from the imperfect stem. [8]

In the second place — and this task is much more difficult — it is necessary to work out the rules for the choice of one or the other aspectual form. Inasmuch as the tendency towards carrying out the operations of synthesis independently from those of analysis has already been noted, these rules must be constructed on the basis of contextual data, considering, for example, the presence in the sentence of adverbs, the character of the combination, etc. In a series of cases one must limit oneself only to a probable solution, based on statistics.

The problem of machine translation from Russian, of course, occupies Soviet investigators less than the problem of translation into Russian. But investigative work connected with the analysis of the Russian sentence has already begun (chiefly in the Laboratory of Electrical Modeling, the Division of Applied Linguistics of the Institute of Linguistics of the Academy of Science of the USSR and in ITM and VT). From the point of view of general linguistics the work reveal-

7. In this connection one should recall the works in the statistical investigation of Russian literary works, carried out in the 20's and 30's by A. I. Peshkovsky, M. Peterson, et al.

8. See Abstracts CMT, p. 87

ing the redundancy of certain categories of the Russian language is most interesting. Thus, for example, the category of gender in the Russian verb, expressed only in the forms in -1 of the singular of the past tense and of the conditional mood, is redundant, unnecessary from the standpoint of analysis. It is clear (V. N. Vinogradova, the Institute of Linguistics of the Academy of Science of the USSR) that in scientific texts the number of verbs with the expressed form of gender comprises from four to thirty percent and that in the majority of sentences the verb can be related only to the subject — the only substantive in the nominative case. Nor is it necessary, in most cases, to consider the inflection of the Russian adjective and determine the relationships of the adjective to the substantive with which it agrees on the basis of the position of the adjective in the sentence. (N. N. Leont'eva and G. H. Vavilova, the Institute of Linguistics).

Interesting also is the work on the determination of syntactic links for the preposition-case groups of the Russian language (I. N. Shelimova) and also the work on the elaboration of the syntactic links for formulas in Russian mathematical texts (M. M. Langleben) — by formulas the author means all elements not found in the machine dictionary during the processing of the text (mathematical formulas, foreign-language citations, surnames, etc.)

For the analysis of a Russian sentence it is necessary to characterize the marks of punctuation. Only in such a way can one find the limits of a simple clause within a sentence, isolate its similar members, aid the further clarification of the co-relationships of the individual parts of a sentence with complex punctuation, determine a group of similar members. T. N. Nikolayeva (ITM and VT) conducted an analysis of polysemantic marks of punctuation (comma, dash, colon) in Russian[9].

Thus the realization of machine translation presupposes serious theoretical investigations, which, in turn enrich the problems of general and applied linguistics.

---

9. See Abstracts CMT, pp. 104-107

A Survey of Soviet Work on Automatic Translation

Anthony G. Oettinger, Computation Laboratory, Harvard University, Cambridge, Massachusetts

The Soviet literature on Automatic Translation is surveyed with the objective of acquainting English-speaking workers in the field with sources of information about relevant Soviet work. The survey is complemented by a bibliography, which we believe to be comprehensive. Limited machine facilities appear to have restricted the range of concrete accomplishment, but much of the theoretical work is excellent.

RESEARCH ON THE linguistic and technological problems of automatic translation has been carried on for several years, chiefly in the United States, Great Britain and the Soviet Union. The journal Mechanical Translation, and the series Current Research and Development in Scientific Documentation issued by the Science Information Service of the National Science Foundation already provide a ready guide to anyone interested in following work in the United States and Great Britain. The purpose of this article is to survey Soviet work to date, and to call to the attention of English-speaking workers the sources of published information on Soviet work in automatic translation.

The general impression left by a survey of the Soviet publications is that serious work of high quality is under way on a significant scale. Limited machine facilities appear to have restricted the range of concrete accomplishment but much of the theoretical work is excellent.† The available examples of accomplished translations are of such high quality, and reflect such small samples as to suggest that they are the results of contrived "experiments" of the same kind that received such extensive publi - city in this country. There is not now, and never was, any question that automatic translation could be achieved if only proper rules could be developed. Anyone who wishes to take the trouble can program a machine to produce perfect translations of any small set of sentences. The crucial problem is the formulation of rules and procedures adequate for the accurate and economical translation of large volumes of new material. It is plain that this is recognized by the Russians themselves. For example, Bel'skaja (A 3), after describing "some of the recent achievements of the machine translation research group of the USSR Academy of Sciences" concludes "the principles have been proved reliable, experimentally; it is therefore no exaggeration to state that the time has come to consider the opportunities for practical large-scale work in this field." The principles in question are sound, generally recognized as such, but not earth-shaking, and anyone now working on automatic translation must agree with the final statement.

In analyzing the progress of automatic translation, two dates seem particularly significant: May, 1957, and October, 1956. [1] In May, 1957, a "Conference on the Problems of Development and Construction of Information Machines" (A 12) took place in Moscow. In October, 1956, at a Session of the Academy of Sciences of the USSR on the Problems of Automation of Production (A 23), considerable attention was given to automatic language translation.

From the brief report of the 1957 conference given by Mel'chuk in Voprosy Jazykoznanija (A 12), it appears that great attention is now being given to "an automatic dictionary of unusually great capacity," (presumably to the development of large capacity storage devices) to problems of automatic character recognition, and to other questions of importance in information organization and retrieval, as well as in automatic translation. It appears that Soviet workers in automatic translation considered

† For a survey of recent Russian work, see V. Yu. Rozentsveig, "Fourth International Congress of Slavicists Reports, " of this issue of MT (Ed.).

1. More recently (May, 1958) a Conference on Machine Translation took place in Moscow. The translated titles of the papers presented there are available from the author or from this journal. The abstracts of these papers, in the original and in translation, are on deposit in the libraries of Harvard and M.I.T.

themselves hampered by the relatively small storage capacity and primitive input-output devices available with the BESM and Strela Machines,[2] although the BESM has an auxiliary magnetic drum and four magnetic tape mechanisms.[3] While these limitations need not hamper, and from all appearances have not hampered, conventional numerical computations where the input-output problem is less serious, the demands of commercial data-processing have probably given the U. S. a lead in experimental work, although much of the theoretical basis of Soviet effort is of first-rate quality.

Mel'chuk also reports a stress on the creation of more educational and communication facilities in the areas of mathematical linguistics and automatic translation: "In the resolutions of the conference it is indicated that the development of work in mathematical linguistics and in machine translation, as well as the timely and regular publication of results, have great significance for Soviet science and for the economy. Therefore it was considered essential to establish a special publication (similar to the journal Mechanical Translation in the U.S.A.), to organize a corresponding sector in the Institute of Linguistics of the Academy of Sciences of the USSR, and to prepare specialists in this area at linguistic faculties of the universities."[4] In this respect, the Russians are following the lead of the United States, where graduate courses in mathematical linguistics have been offered at Harvard University since 1954. More recently, Harvard established an undergraduate field of concentration in Linguistics and Applied Mathematics, and the University of Michigan introduced graduate degree programs in the area of "Language Models and Logical Design," to cover problems which "on the one hand. . . involve language, and on the other. . . require knowledge of the results and techniques of mathematics and engineering."

Mel'chuk's summary of a paper by Ljapunov presented at this conference supports the idea that Soviet results to date have been limited to small samples. Ljapunov is reported to have suggested the use of automatic programming techniques for deriving translation algorithms which "would permit 'teaching' a machine independently to develop rules for translation, using parallel texts in two languages and a previously prepared dictionary." This, in place of "empirical search (for rules) in every individual case." The importance of developing generalized methods for formal linguistic analysis is stressed. Stripped of the "teaching" verbiage, this idea has considerable merit, and some independent work in a similar direction is in progress also in this country.

Mel'chuk also mentions "one such generalized method" presented by O. S. Kulagina. This method is based on "set-theoretic concepts, and permits the assignment of words in a language to some equivalence classes (analogous to 'parts of speech')." Kulagina apparently presented a paper along the lines of one she gave three months earlier at a meeting of the Seminar on Problems of Machine Translation (A 9). The latter paper has considerable expository merit, and it is clearer and more sensible than similar papers on set-theoretic concepts in language which have sprouted like ungainly weeds in the lawn of our information-retrieval literature. The work is along somewhat different lines, and of lesser extent but of caliber comparable to that of the excellent theoretical work of Chomsky in this country.

However, as in Chomsky's work, establishing a fruitful relation between the mathematical model and any concrete, natural language is left, so to speak, as an exercise for the reader. It is in this exercise that the major problems (other than that of educating linguists) lie, for, as Kulagina herself points out "as a matter of fact, nothing new is done here, it is simply that a different form of exposition is given, with a carefully drawn set of postulates." The anguish of Soviet linguists when facing incomprehensible and not obviously useful mathematical apparatus has a familiar ring. Witness the final comment of the chairman (V. Ju. Rozentsvejg) presiding over the presentation of Kulagina's paper: "Not everything presented earlier is clear to me, indeed, one would like to have a translation of this terminology into our plain language."

More significant, perhaps, are the remarks of G. V. Kolshanskij following Kulagina's paper, to the effect that all this is well and good for a finite vocabulary and a limited number of phrases, but that the problems of a living, open-

2. Konferentsija "Puti Razvitija Sovetskogo Matematicheskogo Mashinostroenija i Priborostroenija" Plenarnye Zasedanija, VINITI Moskva, 12-17 Marta, 1956 g.

3. Ibid, p. 34.

4. The contents of the first seven issues of an informal Bulletin of the Seminar on Problems of Machine Translation are listed in Part D of the Bibliography.

ended language go beyond this. This point is amplified in a paper by Barxudarov and Kol - shanskij (A 2), apparently based on a presentation by Kolshanskij at the following meeting of the Seminar on Problems of Machine Translation. These writers feel that "the problem of machine translation at present is basically a linguistic problem, and that it is essential to seek its solution in this realm only." While not wishing to deny the potential practical value of automatic translation, they say "however, it is necessary to keep in mind that this mode of translation will perform only an elementary function, that it will play only a subsidiary role in translation, moreover only of special texts. When it is necessary to obtain a really good translation, a machine translation can be only raw material for a human translator-editor." This echoes comments of Zhirkov (A 24) to the effect that it will be difficult for a machine to choose, for example, between the synonyms lilovyj and fioletovyj for 'violet'.

Zhirkov also is concerned about the magnitude of the tasks facing linguists: "As soon as the question of the actual practical application of a translating machine arises, we must realize the necessity of developing detailed programs for those linguistic conditions and problems which the machine is expected to resolve. This is far from being a simple matter, especially since the task will devolve on linguists. Linguistic programs must be prepared in accordance with the properties of machines, and linguists until now have had no experience in this realm. There is no point in concealing the fact that they will find this a difficult task. "

All of these remarks support the notion that fully automatic translation, as an operating proposition, is probably as far from realization in the USSR as it is in this country. However, one must take into account the fact that criticism and pessimism on the part of linguists may not accurately reflect the progress made by groups led by mathematicians and engineers. Barxudarov, Kolshanskij, and Zhirkov persist in raising questions about 'violet', idioms, poetry and semantics, much in the manner of linguists such as Joos[5] in this country. The concepts of successive approximations, the "fail-safe" principle, and the tolerance toward stylistic inelegance of which scientists and technicians are capable, are utterly foreign to some linguists.

Besides the papers already mentioned, several others of interest appeared in 1957. All merit careful reading by American workers in the field. While none give indications of spectacular, achievements, indeed modest disclaimers are the rule, they reflect considerable thoughtful work. The occasional novel ideas, and new formulations of known ideas, are generally lucidly expressed, but with gaps suggesting the absence of really significant large-scale machine experimentation.

Andreev, in "Machine Translation and the Problem of an Intermediary Language" (A 1), trots out the old chestnut about the number of algorithms needed to translate pairwise among n languages growing as n^2, while use of an intermediate language would reduce the number of algorithms to 2n. He discusses, quite elegantly, some criteria for a good intermediate language. Some interesting distinctions are introduced among "semantic," "formal," and "tectonic" symbols, which put into rather neat form a problem handled earlier but more obscurely by the Wundheilers. [6] He then defines measures for the lack of congruence between the structure of a string in one language and that of its translation in some other language. However, like most theoretical skeletons to date, this one has little flesh on it, and it is difficult to take translation among n languages seriously until some satisfactory results have been obtained for at least one pair.

Moloshnaja's paper "Certain Questions of Syntax Connected with Machine Translation from English to Russian" (A 14), sketches in some detail a method for analyzing the structure of sentences, akin to the parsing once taught in the schools. A broad outline of this approach was given earlier by Panov, Ljapunov, and Muxin (A 17, pp. 182-192). Success in this direction would be of tremendous value, and one would expect to hear of further activity. Significant work, however, probably requires the prior compilation of a substantial automatic dictionary. The state of affairs at the time of publication is perhaps best summed up in Moloshnaja's own words: "The problem of choosing an index, i.e., of eliminating lexico-grammatical

5. M. Joos, "Meaning in Relation to M.T.," Report of the Eighth Annual Round Table Meeting on Linguistics and Language Study, Monograph Series on Language and Linguistics, Georgetown University Press, 1957, pp. 13-18, and "Review of Machine Translation of Languages", Language, 32, 1956, pp. 293-298.

6. Luitgard and Alex Wundheiler, "Some Logical Concepts for Syntax, " Machine Translation of Languages, Wiley, New York, 1955, pp. 194-207.

homonymy in every given case, is very complicated. It requires additional work and more precision. . . . All the foregoing does not claim to be exhaustively worked out and complete. For the time being, it is a working hypothesis, whose verification on a large amount of material and with the aid of a machine is contemplated." It is reasonable to guess that until more has been heard about an automatic dictionary, the likelihood of successful verification remains small. This area bears watching, as capable of bearing fruit much earlier than the similar work of Kulagina (A 9), which is less directly coupled with reality.

The paper "Machine Translation from French to Russian" by Kulagina and Mel'chuk (A 10) is based on the analysis of 20,500 words of running text, and it is probable that the test material that was translated came from the same source.[7] The analysis of syntactic problems is generally sound and lucid, and in most cases should lend itself to generalization and verification on a large scale. However, one guesses that some important techniques described in the paper will not be adequate for large-scale work. This guess is supported to some extent by Kulagina's later switch to working in the mathematical stratosphere, in marked contrast to the down-to-earth approach in this paper, and by Ljapunov's concern with methods for generating translation algorithms. One feels that the mathematicians in the Ljapunov group became appalled by the wealth of detail to be considered in achieving practical large-scale results. This could account for the stand taken more recently by Barxudarov and Kolshanskij. However, the conceptual quality of the work is such as to suggest that the Russians have nothing to fear but fear itself. This paper is well worth reading. Like that reported by Moloshnaja, it is a product of the group working at the Steklov Mathematical Institute under the general leadership of Ljapunov. An outline of the same work may also be found in reference (A 23, pp. 182-192).

The paper by Razumovskij, "On the Problem of Automatizing the Programming of Translation from One Language into Another" (A 20), is interesting as one indicator of growing concern among workers on automatic translation with sophisticated machine methods. The strong Russian interest in automatic programming is well known from manifestos by Ljapunov (A 23, pp. 118-121), from the attention paid to the subject in Kitov's book (A 6), from interesting papers by Janov (B 4, 5), and from descriptions of new books slated for publication (C 1, 2, 3). Razumovskij's paper is mathematically elegant, but contains nothing that is novel to anyone familiar with mathematical logic and its applications to computer programming. Its significance lies chiefly in the indication it gives that Russians are thinking along lines also considered potentially fruitful by workers in this country. It should be of value in research such as that contemplated by Moloshnaja, and in the practical execution of algorithms developed as a product of this research. Razumovskij is also credited with the development of the notation used by the Panov group to write down their translation algorithms. The notation is quite good, especially if used in conjunction with flow-charts.

Korolev's paper "Coding and Code Compression" (A 7) is much less significant. One may guess that it represents a very belated publication of results obtained as a consequence of great concern at early stages over limited storage capacity. Most of Panov's early papers mention this work. It is reasonable to guess that once the resolutions of the May 1957 meeting bear fruit, the tricks described by Korolev no longer will be used.

Many of the papers discussed in the preceding paragraphs were published after the October 1956 session of the Academy of Sciences, although some of the work had been done prior to that session, and reported there. Most of the work done under Panov at the Institute of Exact Mechanics and of Computing Techniques was published prior to the session, and will be discussed after some comments on the session itself.

In the proceedings of the Plenary Session of the Academy of Sciences on the Scientific Problems of Automation of Production (A 23) considerable space is devoted not only to a description of the work accomplished by the Panov and Ljapunov groups, but to discussions of the merits of efforts on automatic translation.

In the introductory remarks by A. N. Nesmejanov, the central theme of urgency about mechanization and automation is first sounded. Throughout the Proceedings, there are numerous allusions to the directives of the Twentieth Party Congress, in which automation of production is set up as a major task. Nesmejanov says: "Even such an area as translation from

7. On this delicate point, which may also be raised about some work in this country, see the exchange between Denis Williams and I. S. Muxin, in the Journal of the I. E. E., January, 1957, p. 34 and February, 1957, p. 110.

one language to another is already an object (objective) of automation. Thus even for linguists the subject of our session should not be altogether strange."

V. V. Karibskij, in a commentary (pp. 59-65) on a paper by V. A. Trapeznikov entitled "The Role of the Technical Sciences in the Development of Automatic Control and of the Technical Means of Automation" (pp. 38-55) strongly criticizes the scientific organizations of the USSR for their lame effort in automation, in the face of reports from abroad of great progress in this area. He declares, for example, that "workers concerned with the design and construction of electronic calculating machines declare that they can apply any machine to the purposes of automation of production, but that they must have a mathematical description of the process to be automated. However, they consider that the mathematical description of these processes is a task for the technicians who are closer to production problems and who are more familiar with the technology and the equipment." He points out that the technicians do not have the necessary mathematical knowledge, but that meanwhile the scientists sit on their hands: "As a result, it happens that the scientists who have created the splendid contemporary mathematical machines in fact hold back the application of these machines towards the goals of automation of production." The parallel between these remarks and the more recently expressed attitude of Barxudarov and Kolshanskij is interesting.

Some remarks of I. S. Bruk in his paper "The Outlook for Applications of Control Machines in Automation" (pp. 131-148) indicate a slightly different attitude toward reports from abroad:

> "The theme of 'thinking machines' is hardly ever off the pages of popular publications abroad. Some speeches of well-known figures, some articles on the correspondence between the structure of the nervous system and calculating machines, were taken up by a press susceptible to sensationalism. The imagination of readers was taken by descriptions of guided missiles and other kinds of armament, about which so much is written abroad.
>
> "Doubtless all this helped to create false conceptions of the role of 'thinking machines', and of a fully automated industry without people. "

Bruk goes out of his way to indicate "how erroneous notions about the role of contemporary calculating machines can lead to an incorrect evaluation of the prospects for their applications." He says: "It is necessary to say that in our country not everything in this area is satisfactory." He then devotes two paragraphs to pointing out that he considers automatic translation as an outstanding example of "the incorrect evaluation of the role of calculating machines, as it strikes the fancy of the public." The gist of his argument is that automatic translation is a poor illustration of the rational use of computers, principally because the job can be done well enough by people, with more complicated texts than "primitive, simplified and specially selected ones for machines" and also because the machine is not likely to replace people altogether.

A rebuttal by V. S. Pugachev (pp. 151-152) dwells chiefly on the importance of the algorithmization of speech in the development of automation, particularly with respect to self-organizing systems, speech-controlled automata, and the like. Pugachev dreams of using translating machines to permit conversations between two people speaking different languages. "The establishment of such lines of communication will have tremendous significance for the development of mankind."

A. A. Dorodnitsyn agrees that translation is an interesting problem, but says (p. 157), "I cannot agree that this task is one to which it is now essential to devote large forces and means." He does not think that it is an important job from the economic point of view, especially because he believes that the Soviet Union has adequate numbers of conventional translators, and even greater numbers of specialists who can read literature on their own speciality in the significant languages. He further believes that the distribution by the Institute of Information of copies of foreign journals will be more helpful to scientific progress in the USSR than the creation of specialized translating machines. "Besides," he says, "the problem of automatic translation is not sufficiently close to solution to warrant serious talk about the construction of translating machines."

A. N. Kolmogorov (p. 161) agrees with Bruk that for practical communication live translators or knowledge of foreign languages will scarcely be replaced by machines in the foreseeable future. However, he stresses the importance of the effect of work on automatic translation on the builders of calculating machines, especially with respect to broadening their ideas about the logical powers of ma - chines, and on linguists, as a practical criterion so important in guiding the development

of theory. The precision and completeness with which rules for automatic translation must be formulated leave no room for the vague phrases which humanists are inclined to call "laws."

These arguments have a very familiar ring. The presence of opposition, and knowledge that the session of the Academy would stress the need for achievement in automation, may also account for the multiple publication, loud drum beating, and excessive emphasis on American "achievements" associated with the relatively meager results achieved prior to the session. Publications since the session have been much more subdued, possibly because of greater confidence in support. In this connection, the following quotation[8] may be illuminating.

> "And what about these — friends of ours? Are they alive and well?" Lopatkin asked.
> "Well enough. . . .They are constantly praising their machine in the newspapers. I think they are building a factory. Shutikov has twice already been abroad. "
> "You say they are praising their machine in the papers? How can that be? It means they are hiding something. It can't be that everything is going smoothly with those machines. So we can expect some more trouble, Nadezhda Sergeyevna."

This also has a familiar ring.

In the communication by M. V. Keldysh, A. A. Ljapunov, and M. R. Shura-Bura (A 23, pp. 100-130) entitled "Mathematical Problems in the Theory of Calculating Machines" there is a brief mention of automatic translation as one among a number of applications which "at present are in the initial stages of investigation." Brief mention is made of some systems of automatic programming, some of which are described in greater detail by Kitov (A 6) and about which more will very likely be said in forthcoming publications. Automatic translation is mentioned in somewhat greater detail on pp. 121-124. The IBM-Georgetown "public demonstration" of 1954 is mentioned briefly, and described as having been conducted with "a program of rather limited possibilities." The material is of no great import, since more technical detail is given in other papers already cited. The work of the Institute of Exact Mechanics and Computing Techniques is characterized as empirical, while that of the Steklov Mathematical Institute is described as based on structural linguistics.

The paper by Panov, Ljapunov, and Muxin (A 17), presented at these sessions, summarizes work done by both groups, and with the exception of a few points of detail, adds nothing to the contents of the more specialized papers by members of these groups represented in the bibliography. In particular, the section describing the work at the Institute of Exact Mechanics and Computing Techniques is almost identical to five other publications by members of the group (A 4, 8, 15, 18, 19). Some of the differences in detail are of interest to specialists, and Muxin's paper (A 15) in particular is somewhat more detailed than the others. His flow-chart for the translation of 'of' has an error which will lead to a correct translation of "... the result of experiments ..." but to an incorrect translation of "... the result of simple experiments. . ., " a weakness typical of ad-hoc programs based on intuition and the study of small samples. However, there is no reason to doubt that continuing effort by this group will have led to considerable progress by now.

The papers by Moloshnaja et al. (A 14) and by Kuznetsov et al. (A 11) are general surveys, again of some interest to specialists, but not particularly significant for this review.

In conclusion, it is worth giving special notice to the book of A. I. Kitov "Electronic Digital Machines" (A 6) which superficially describes Soviet work on automatic translation, but which is most noteworthy for a very good chapter on automatic programming, in which a system suggested by Ljapunov and elaborated by Janov (B 4, 5) is described in some detail probably for the first time.

For the sake of completeness, some interesting papers on information organization and retrieval and on automatic programming, subjects with which automatic translation has some kinship and which are of great importance in their own right, are listed in section B of the Bibliography.

The third section of the Bibliography (C) lists some books scheduled for publication in 1958. It is evident that the meetings described earlier in this review are having some effects. The Table of Contents of the first seven issues of the Bulletin of the Seminar on Problems of Machine Translation is given in part D of the Bibliography. This publication is also available in the Harvard and M.I.T. libraries.

---

8. Vladimir Dudintsev, Not By Bread Alone, (translated by Edith Bone), E. P. Dutton and Co., Inc., New York, 1957, p. 426.

BIBLIOGRAPHY

A. Soviet Works on Automatic Translation

1. [112]† N. D. Andreev, "Mashinnyj Perevod i Problema Jazyka-posrednika" (Machine Translation and the Problem of an Intermediary Language), Voprosy Jazykoznanija, No. 5, 1957, pp. 117-121.

2. [125] L. S. Barxudarov, G. V. Kolshanskij, "K Voprosu o Vozmozhnostjax Mashinnogo Perevoda" (On the Problem of the Potentialities of Machine Translation), Voprosy Jazykoznanija, No. 1, 1958, pp. 129-133.

3. [147] I. K. Bel'skaja, "Machine Translation of Languages," Research (London), 10, October 1957, pp. 383-389.

4. [107] I. K. Bel'skaja, L. M. Korolev, I. S. Muxin, D. Ju. Panov, S. N. Razumovskij, "Nekotorye Voprosy Avtomatizatsii Perevoda" (Certain Problems of Automatic Translation), Vestnik Akademii Nauk SSSR, 26, No. 12, 1956, pp. 24-33.

5. V. V. Ivanov, "Komitet po Prikladnoj Lingvistike" (Committee on Applied Linguistics), Voprosy Jazykoznanija, No. 3, 1958, pp. 136-137.

6. A. I. Kitov, Elektronnye Tsifrovye Mashiny (Electronic Digital Machines), Izdatel'stvo "Sovetskoe Radio," Moskva, 1956, Ch. 6, pp. 214-223.

7. [111] L. M. Korolev, "Kodirovanie i Svertyvanie Kodov" (Coding and Code Compression), Doklady, AN SSSR, 113, No. 4, 1957, pp. 746-
747.

8. L. Korolev, S. Rasoumovski, G. Zelenkevitch, Les Expériments de la Traduction Automatique de 1'Anglais en Russe à 1'Aide de la Calculatrice BESM, Académie des Sciences de 1'URSS, Moscou, 1956.

9. [128] O. S. Kulagina, "Ob Odnom Sposobe Opredelenija Lingvisticheskix Ponjatij" (On a Method for Defining Linguistic Concepts), Bulletin of the Seminar on Problems of Machine Translation, Moscow State Pedagogical Institute of Foreign Languages, No. 3, 1957, pp. 1-18.

10. [101] O. S. Kulagina, I. A. Mel'chuk, "Mashinnyi Perevod s Frantsuzkogo Jazyka Na Russkij" (Machine Translation from French to Russian), Voprosy Jazykoznanija, No. 5, 1956, pp. 111-121.

11. [100] P. S. Kuznetsov, A. A. Ljapunov, A. A. Reformatskij, "Osnovnye Problemy Mashinnogo Perevoda" (Fundamental Problems of Machine Translation), Voprosy Jazykoznanija, No. 5, 1956, pp. 107-111.

12. [115] I. A. Mel'chuk, "Soveshchanie po Voprosam Razrabotki i Postroenija Informa - tsionnyx Mashin" (Conference on the Problems of Development and Construction of Informa - tion Machines), Voprosy Jazykoznanija, No. 5, 1957, pp. 161-162.

13. [108] T. N. Moloshnaja, "Nekotorye Voprosy Sintaksisa v Svjazi s Mashinnym Perevodom s Anglijskogo Jazyka na Russkij" (Certain Questions of Syntax Connected with Machine Translation from English to Russian), Voprosy Jazykoznanija, No. 4, 1957, pp. 92-97.

14. [136] T. M. Moloshnaja, V. A. Purto, I. I. Revzin, V. Ju. Rozentsvejg, "Nekotorye Lingvisticheskie Voprosy Mashinnogo Perevoda" (Certain Linguistic Problems in Machine Translation), Voprosy Jazykoznanija, No. 1, 1957, pp. 107-113.

15. [69] I. S. Mukhin, An Experiment of the Machine Translation of Languages Carried out on the BESM, Academy of Sciences of the USSR, 1956 , (In English) (Also in Proc. I.E.E. 103, 1956, Part B, Supplements 1-3, pp. 463-472, with minor editorial modifications).

16. T. M. Nikolaeva, Analiz Russkogo Predlozhenija (Analysis of the Russian Sentence), Institut Tochnoj Mexaniki i Vychislitelnoj Texniki, Moskva, 1958.

17. [152] D. Ju. Panov, A. A. Ljapunov, I. S. Muxin, "Avtomatizatsija Perevoda s Odnogo Jazyka na Drugoj" (Automatization of Translation from One Language to Another), Session

† Numbers in brackets refer to the numbered abstracts which appear in each issue of MT.

of the Academy of Sciences of the USSR on the Scientific Problems of Automation of Produc - tion, October 15-20, 1956, - Plenary Sessions. Izdatel'stvo Akademii Nauk SSSR, Moskva, 1957, pp. 181-214.

18. [79] D. Panov, Concerning the Problem of Machine Translation of Languages, The Academy of Sciences, USSR, Moscow, 1956.

19. [80] D. Ju. Panov, Avtomaticheskij Perevod (Automatic Translation), Nauchno-Populjarnaja Serija, Izdatel'stvo Akademii Nauk SSSR, Moskva, 1956. (A second, augmented edition has been issued in 1958.)

20. [120] S. N. Razumovskij, "K Voprosu ob Avtomatizatsii Programmirovanija Zadach Perevoda s Odnogo Jazyka na Drugoj" (On the Problem of Automatizing the Programming of Translation from One Language into Another), Doklady, AN SSSR, 113, No. 4, 1957, pp. 760-761.

21. [126] I. I. Revzin, Addendum to the Paper "Some Problems in the Formalization of Syntax," Bulletin of the Seminar on Problems of Machine Translation, Moscow State Pedagogical Institute of Foreign Languages, No. 3, 1957, pp. 20-29.

22. V. Ju. Rozentsvejg, "Raboty po Mashinnomu Perevodu s Inostrannyx Jazykov na Russkij i s Russkogo na Inostrannye v Sovetskom Sojuze" (Work in the Soviet Union on Machine Translation from Foreign Languages into Russian and from Russian into Foreign Languages), 4th International Congress of Slavists, Moscow, 1958.

23. A. V. Topchiev et al. (Eds.), Sessija Akademii Nauk SSSR po Nauchnym Problemam Avtomatizatsii Proizvodstva, 15-20 Oktjabrja, 1956 g. Plenarnye Zasedanija (Session of the Academy of Sciences of the USSR on the Scientific Problems of Automation of Production, October 15-20, 1956 - Plenary Sessions), Izdatel'stvo Akademii Nauk SSSR, Moskva, 1957.

24. [102] L. I. Zhirkov, "Granitsy Primenimosti Mashinnogo Perevoda" (Limits of Applicability of Machine Translation), Voprosy Jazykoznanija, No. 5, 1956, pp. 121-124.

B. Miscellaneous Bibliography

1. V. P. Cherenin and B. M. Rakov, Experimental Information Machine of the Institute of Scientific Information of the USSR Academy of Sciences, Institute of Scientific Information of the Academy of Sciences of the USSR, Moscow, 1955.

2. V. P. Cherenin, Nekotorye Problemy Dokumentatsii i Mexanizatsija Informatsionnyx Poiskov (Certain Problems of Documentation and the Mechanization of Information Searches), Institute of Scientific Information of the Academy of Sciences of the USSR, Moskva, 1956

3. L. I. Gutenmaxer, "Statisticheskie i Informatsionnye Mashiny Novogo Tipa" (Statistical and Information Machines of a New Type), Vestnik Akademii Nauk SSSR, 26, No. 10, 1956, pp. 12-21.

4. Ju. I. Janov, "O Matrichnyx Sxemax" (On Matrix Schemata), Doklady, AN SSSR, 113, No. 2, 1957, pp. 283-286:

5. Ju. I. Janov, "O Ravnosil'nosti i Preobrazovanijax Sxem Programm" (On Equivalence and Transformation of Program Schemata), Doklady, AN SSSR, 113, No. 1, 1957, pp. 39-42.

6. L. V. Kantorovich, "Ob Odnoj Matematicheskoj Simbolike, Udobnoj pri Provedenii Vychislenij na Mashinax" (On a Mathematical Symbolism, Convenient for Doing Calculations on Machines), Doklady, AN SSSR, 113, No. 4, 1957, pp. 738-741.

7. S. A. Lebedev, Ehlektronnye Vychislitel'nye Mashiny (Electronic Calculating Machines), Nauchno-populjarnaja Serija, Akademija Nauk SSSR, Moskva, 1956.

8. V. D. Moiseev, Avtomaticheskie Vychislitel'nye Mashiny i ix Primenenie na Zheleznodorozhnom Transporte (Automatic Calculating Machines, and Their Application to Railroad Transportation), Gosudarstvennoe Transportnoe Zheleznodorozhnoe Izdatel'stvo, Moskva, 1957.

9. B. A. Traxtenbrot, "Algoritmy i Mashinnoe Reshenie Zadach" (Algorithms and the Machine Solution of Problems), Matematika v Shkole, No. 4, 1956, pp. 3-10 and No. 5, 1956, pp. 5-14.

10. L. R. Zinder, "Ob Odnom Opyte Sodruzhestva Fonetikov s Inzhenerami Svjazi" (On a Trial Cooperation of Phoneticians with Communications Engineers), Voprosy Jazykoznanija, No. 5, 1957, pp. 111-116.

C. Forthcoming Publications †

1. N. A. Krinitskij and A. I. Kitov, Ehlektronnye Tsifrovye Mashiny i Programmirovanie (Electronic Digital Machines and Programming).

2. Ju. Ja. Bazilevskij (Ed.), Voprosy Teorii Matematicheskix Mashin (On the Theory of Mathematical Machines).

3. A. A. Ljapunov (Ed.). Problemy Kibernetiki (Problems of Cybernetics).

4. M. A. Kartsev, Arifmeticheskie Ustrojstva Ehlektronnyx Tsifrovyx Mashin (Arithmetic Units of Electronic Digital Machines).

5. M. R. Shura-Bura and others, Metod Standartnyx Podprogramm (The Method of Standard Subroutines).

6. Matematika v SSSR za 40 Let (Mathematics in the USSR over 40 years), Section 16: "Matematicheskie Mashiny" (Mathematical Machines) by A. A. Ljapunov.

7. N. A. Arxangel'skij and B. I. Zajtsev, Ehlektronnye Tsifrovye Mashiny (Electronic Digital Machines).

D. Table of Contents of the Bulletin of the Seminar on Problems of Machine Translation (mimeographed).

Bjulleten' Ob"edinenja po Problemam Mashinnogo Perevoda (Ministry of Higher Education, USSR - First Moscow State Pedagogical Institute of Foreign Languages).

Number 1 (Meeting of December 24, 1956)

V. Ju. Rozentsvejg, "Vstupitel'noe Slovo" (Introduction).

I. I. Revzin, "Nekotorie Voprosy Formalizatsii Sintaksisa" (Some Problems in the Formalization of Syntax).

Debate: I. Eh. S. Raxmatulova, A. A. Balaban, G. V. Kolshanskij, V. V. Ivanov, V. Ju. Rozentsvejg, L. S. Barxudarov, T. A. Amirova, A. V. Dogopol'skij, G. M. Strelkovskij, T. K. L'vova.

I. I. Revzin, "Zakljuchitel'noe Slovo" (Conclusion).

Number 2

T. N. Moloshnaja, "Soobshchenie o Sostavlenii Grammaticheskix Pravil dlja Mashinnogo Perevoda s Anglijskogo Jazyka" (Report on the Compilation of Grammatical Rules for Machine Translation from English).

Debate: I. I. Revzin, G. V. Kolshanskij, G. M. Strelkovskij, M. A. Balaban, L. S. Barxudarov, A. B. Dolgopol'skij, I. S. Raxmankulova, V. Ju. Rozentsvejg, T. N. Moloshnaja.

Number 3 (Meeting of February 21, 1957)

O.S. Kulagina, "Ob Odnom Sposobe Opredelenija Lingvisticheskix Ponjatij" (On a Method for Defining Linguistic Concepts).

Debate: G. V. Kolshanskij.

I.I. Revzin, "Dobavlenie k Dokladu 'Nekotorye Voprosy Formalizatsii Sintaksisa' " (Additions to the Paper 'Some Problems in the Formalization of Syntax' ").

Number 4 (Meeting of April 18, 1957)

I. A. Mel'chuk, "O Mashinnom Perevode s Vengerskogo Jazyka na Russkij" (On Machine Translation from the Hungarian Language to Russian).

G. V. Kolshanskij, "K Voprosu o Vozmozhnostjax Mashinnogo Perevoda" (On the Question of the Possibilities of Machine Translation) .

Debate: I. I. Revzin, G. M. Strelkovskij, L. S. Barxudarov, V. Ju. Rozentsvejg.

Number 5 (Devoted to papers presented at a seminar on Mathematical Methods in Linguistics, Philological Faculty, Moscow University)

Ot redaktsii (From the editorial office).

"Dnevnik Seminara 'Nekotorye Primenenija Matematicheskix Metodov v Jazykoznanii' " (Diary of the Seminar "Some Applications of Mathematical Methods to Linguistics").

V. V. Ivanov, "Jazykoznanie i Matematika (Linguistics and Mathematics),

V. A. Uspenskij, "K Opredeleniju Padezha po A. N. Kolmogorovu" (Towards a Definition of the Case According to Kolmogorov).

R. L. Dobrushin, "Ehlementarnaja Grammaticheskaja Kategorija" (Elementary Grammatical Category).

† Gostexizdat; Gosudarstvennoe Izdatel'stvo Texniko-Teoreticheskoj Literatury.

V. A. Uspenskij, "K Opredeleniju Chasti Rechi v Teoretikomnozhestvennoj Sisteme Jazyka" (Towards the Definition of Parts of Speech in a Set-theoretical Theory of Language).

P. S. Kuznetsov, "Ob Osnovnyx Polozhenijax Fonologii" (On the Basic Principles of Phonology).

I. I. Revzin, "Po Povodu Opredelenija Fonemy, Dannogo Prof. P. S. Kuznetsovym" (Apropos of the Definition of the Phoneme given by Prof. P. S. Kuznetsov).

V. V. Ivanov, "Kod i Soobshchenie" (Code and Communication).

V. V. Ivanov, "n-Mernoe Prostranstvo Jazyka" (n-Dimensional Language Space).

V. V. Ivanov, "O Nekotoryx Ponjatijax Sravnitel'no-istoricheskogo Jazykoznanija" (On Some Concepts of Comparative-Historical Linguistics).

V. V. Ivanov, "Ponjatie Nejtralizatsii v Morfologii i Leksike" (The Concept of Neutralization in Morphology and Lexicography).

S. K. Shaumjan, "Ponjatye Fonemy v Svete Simvolicheskoj Logiki" (The Concept of the Phoneme in the Light of Symbolic Logic).

Number 6 (Meeting of September 25, 1957)

V. Ju. Rozentsvejg, "Vstupitel'noe Slovo" (Introduction).

V. V. Ivanov, "Voprosy Matematicheskoj i Prikladnoj Lingvistiki na VIII Mezhdunarodnom Lingvisticheskom Kongresse v Oslo" (Questions of Mathematical and Applied Linguistics at the VIII International Congress of Linguistics in Oslo).

I. I. Revzin, "Soobshchenie o Tematike Rabot Ob"edinenija" (Report on the Subject Matter of the Work of the Seminar).

A. A. Reformatskij, "K Voprosu o Mashinnom Perevode Ustnoj Rechi" (On the Question of Machine Translation of Speech).

I. A. Mel'chuk, "Nekotorye Vyvody Obshchego Xaraktera v Svjazi s Mashinnym Perevodom s Vengerskogo Jazyka" (Several Conclusions of a General Character in Relation to Machine Translation from Hungarian).

A. V. Superanskaja, "Protsess Transkribirovanija Sobstvennyx Imen i Vozmozhnost" ego Avtomatizatsii" (The Process of Transcribing Proper Names and the Possibility of Its Automatization).

"Materiali k Spetskursu i Spetsseminaram po Mashinnomu Perevodu" (Materials for the Special Course and Special Seminars on Machine Translation).

V. Osval'd i R. Louson, "Slovar' dlja Mashinnogo Perevoda" (V. Oswald and R. Lawson, A Dictionary for Machine Translation), (Translated from the English).

Number 7

I. I. Revzin, "O Nekotoryx Aspektax Sovremennyx Teoreticheskyx Issledovanij v Oblasti Mashinnogo Perevoda" (On Several Aspects of Contemporary Theoretical Investigations in the Field of Machine Translation).

I. A. Mel'chuk, "Statistika i Zavisimost' Roda Frantsuzskix Sushchestvitel'nyx ot ix Okenchanija" (Statistics and the Dependence of the Gender of French Nouns on their Endings).

A. L. Shumilina i Z. M. Volotskaja "Nekotorye Statisticheskie Dannye o Sushchestvitel'nyx Russkogo Jazyka (na Materiale Matematicheskix Tekstov)" (Some Statistical Data on Russian Nouns (Based on Material in Mathematical Texts)).

N. D. Andreev, "O Rabote Leningradskogo Seminara po Mashinnomu Perevodu" (On the Work of the Leningrad Seminar on Machine Translation).

Lju Jun-tsjuan', "Razrabotka Voprosov Mashinnogo Perevoda v K. N. R." (Work on Questions of Machine Translation in the Chinese People's Republic).

Review:

N. Chomsky, "Syntactic Structures," s'-Gravenhage, Mouton, 1957, Janua Linguarum, No. 4 (E. V. Paducheva).

Materiali k Spetskursu po Mashinnomu Perevodu v 1-m MGPIIJa. - Charl'z F. Xoket (Kornel'skij Universitet) - "Perevod Cherez Neposredstvennye Sostavljajushchie" (perevod s anglijskogo) (Materials for the Special Course on Machine Translation in the 1st MGPIIJa. - Charles F. Hockett, Cornell University). "Translation through Immediate Constituents" (Translation from the English).

Order of Subject and Object in Scientific Russian When Other Differentia Are Lacking

D. C. Hays, The Rand Corporation, Santa Monica, California

The order of subject and object is an adequate criterion for distinguishing between them when other grammatical properties are ambiguous.

HARPER[1] AND LEHISTE[2] have discussed the order of subjects and predicates in Russian scientific text. Lehiste concludes that "form and function" should be used to distinguish the subject from the predicate of a Russian sentence; although her conclusion may be accepted (subject to assumptions about the value of maintaining customary English order in the output), her dictum must be converted into programmable instructions.

To a certain extent, the most economical method of distinguishing subject from predicate is obvious and straightforward. Verbs, short-form adjectives and participles, and other potential "fillers of the predicate slot" are marked in the glossary and can be identified when they occur in text. Inasmuch as some glossary entries are marked (in effect) "possibly predicate," some difficulties are involved in finding the predicate, but we wish to pass over these to a specific problem of detail.

The formal characteristics by which a subject can be recognized are, roughly, part of speech, gender, number, person, and case. The subject and predicate of a sentence are, in fact, two of its members of specifiable parts of speech, agreeing in number and either person or gender, while the subject must be of specified case, i.e., nominative. Unfortunately, for example, two nouns in a sentence may be equally good candidates for the role of subject; this is true because the nominative and accusative cases are not always formally distinct. Thus, if two neuter nouns, each nominative or accusative, respectively precede and follow a third-person, singular, non-past verb (which takes an accusative object), the choice between these nouns must be made on grounds other than morphology.

Word order and semantic agreement immediately come to mind. Semantic agreement would require thoughtful, expensive research. The hypothesis that subjects precede their predicates whenever the latter contains a noun that could be mistaken (morphologically) for the subject can be tested rapidly and inexpensively by reference to a body of data already collected at The RAND Corporation.

Method

A large volume of Russian physics text has been keypunched into IBM cards, referred to a glossary, and analyzed by translators[3]; the structure of each sentence has been determined in accordance with a dependency theory, and each dependency relation punched into a card. For a sample of 22, 000 occurrences (running words) of text[4], a special report has been prepared (by machine processes), showing all dependents of every occurrence in the sample; the listing is ordered by the grammatical type of the governor.

Since subject and object are regarded as dependents of the main predicate element in our theory, it is simple to scan the section of this report that is devoted to verbs and their depend ents, noting the textual location of every verb with two dependents, of which either could be

1. K. E. Harper, "A Preliminary Study of Russian," in W. N. Locke and A. D. Booth, Machine Translation of Language, New York, Wiley, 1955.

2. Ilse Lehiste, "Order of Subject and Predicate in Scientific Russian," MT, 4, 1957, 66-67

3. H. P. Edmundson and D. G. Hays, "Research Methodology for Machine Translation," MT, 5, 1958, 8-15.

4. H. P. Edmundson, K. E. Harper, D. G. Hays, and A. K. Koutsoudas, Studies in Machine Translation - - 9: Bibliography of Russian Scientific Articles, The Rand Corporation, Research Memorandum RM-2069, October 16, 1958. (Corpus 2 was used in the present study.)

Table 1

INSTANCES OF MORPHOLOGICALLY INDISTINGUISHABLE SUBJECT AND OBJECT IN A SAMPLE OF RUSSIAN PHYSICS TEXT

Subject	Verb	Object	Order	Frequency
Nom./acc. sing. noun	3rd person sing., non-past	Nom./acc. sing. noun	SVO	27
do.	do.	do.	VOS	4
Nom./acc. sing. pronoun	do.	do.	SVO	1
Non-Cyrillic	do.	do.	SVO	6
Nom./acc. sing. noun	do.	Non-Cyrillic	SVO	3
Nom./acc. plural noun*	3rd person plural non-past	Nom./acc. plural noun	SVO	7
do.	do.	Non-Cyrillic	SVO	1
Conjunction of proper names	do.	Nom./acc. plural noun	SVO	1
Nom./acc. sing masc. noun	Masc. sing past	Nom./acc. sing. masc. noun	SVO	1
Name (indeclinable, masc.)	do.	do.	SVO	1
Nom./acc. sing. neut. noun	Neuter. sing. past	Nom./acc. sing. neut. noun	SVO	2
Conjunction of proper names	Plural past	Nom./acc. plural noun	SVO	1
do.	do.	Non-Cyrillic	SVO	1
		Subtotal	SVO	52
		Subtotal	VOS	4
	Total			56

* Three subjects are in apposition with conjunctions of Non-Cyrillic occurrences.

subject. All doubtful cases were noted as well. A 3x5 card was prepared for each such occurrence, and the cards (about 100 in number) were sorted into textual order.

Examination of all 100 occurrences required only about 3 hours. Doubtful cases were resolved, situations in which a modifier of either noun distinguished its case were recognized and discarded, subject and object were differentiated by careful human judgment, and their order was noted on each card.

Results

Just 56 instances of true ambiguity were found in 22, 000 occurrences.[5] They are summarized in Table 1. The subject precedes the verb 52 times; the object follows the verb 56 times. When both object and subject follow the verb, the object precedes the subject 4 times.

The 4 sequences V-O-S are:

Обращает внимание наличие (The presence [of..] calls attention [to..])

Имеет место состояние (a state that occurs)

Имеет место правило (a rule occurs)

Имеет место уменьшение (a decrease occurs)

Note that the verb-object pair might be regarded as idiomatic on grounds other than those of the present study; neither is translated literally.

Conclusions

On the basis of a preliminary study of the 56 relevant instances in 22, 000 running words of text, we conclude that: If two nouns in a sentence cannot be distinguished as subject and object of a transitive verb by their morphological properties, and if one precedes the verb while the other follows, the first noun is the subject. This rule, together with adequate coverage of idioms, appears entirely effective. The study should be repeated on a larger sample of text, however.

---

5. If an adjectival modifier forms an unambiguous noun phrase with either subject or object, or if negation of the verb calls for a genitive object, the instance is irrelevant to the present study.

The author is indebted to Kenneth E. Harper for guidance in the course of this study.

German Sentence Recognition †

G. H. Matthews and Syrell Rogovin, Massachusetts Institute of Technology, Cambridge, Massachusetts

A computer program is described which assigns one or more distinct immediate constituent analyses to every German sentence, thus indicating which of all possible sentences any given sequence of words may represent, and revealing all the information implicitly or explicitly contained in each of these sentences, that can be used in the choice of their translations.

THIS PAPER describes a routine that is based upon a theory of language which recognizes in each sentence of a given language an immediate constituent structure. Prior work on German sentence recognition [1, 2, 3] has been based on a linear view of language. Oswald and Fletcher, for example, "... found that the elements of the language in question and their functional relationships to each other could be treated most efficiently in terms of traditional descriptive grammar." [4] This theory of language that neither explains nor accounts for any features of language other than its linear structure has led them and other investigators to develop routines which merely rearrange lexical items and translate them individually into the output language.

Our general method of translation is based on the following assumptions: each sentence of a language has one or more discoverable constituent structures; there is a finite and manageable number of constructions that make up any given sentence; and these constructions, except when two or more share a single common element, are discrete from one another. Our attack on the problem of recognition has been to take one construction at a time, and develop a routine for finding its limits in any sentence, discovering that it is this construction, and finding its function in the larger construction of which it is a part. In such a program, then, difficulties do not arise from the length of a sentence, nor from the number or kinds of relationships, both syntactic and special, of its constructions; the constructions of the sentence are recognized one at a time, from the most inclusive to the least, and from the beginning of the sentence to the end. We feel that the most efficient program and the best output text can be attained by working from the outset from grammars of the two languages involved. These grammars are adapted for the computer from the type suggested by Noam Chomsky in Syntactic Structures. [5] Each grammar is a series of ordered, completely unambiguous rules, some of which are obligatory, and some, optional. Every sentence in the language is thus the result of all the applicable obligatory rules plus none or more of the applicable optional ones. Then, when the computer, as a final step in the translation process, is given the grammar of English and directs which optional rules of the grammar are to be chosen, the sentence so designated will be generated. Preceding this there is a routine which will translate lexical items, the syntactic functions of which will have been defined in the preceding step, the recognition routine. In the recognition routine the input sentence is sent through a program which ascertains those rules of gram-

† This work was supported in part by the U.S. Army (Signal Corps), the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U.S.Navy (Office of Naval Research); and in part by the National Science Foundation.

1. Oswald and Fletcher, "Proposals for the Mechanical Resolution of German Syntax Patterns", Modern Language Forum, Vol. XXXVI, no. 3-4 (1951).

2. Booth, Cleave and Brandwood, Mechanical Resolution of Linguistic Problems. (London, 1958) pp. 125-277.

3. Leonard Brandwood, "Some Problems in the Mechanical Translation of German", MT, Vol.V, no. 2, pp. 60-66.

4. Oswald and Fletcher, op. cit. pp. 2-3.

5. Noam Chomsky, Syntactic Structures, 'S-Gravenhage, The Hague, 1957.

mar which must have been applied in order to produce that particular sentence. The middle step, therefore, is not merely the translation of lexical items but also a translation of rules. In order to discover what rules of the grammar of the input language produced the sentence to be translated, we need, of course, a grammar of that language as well. We can thus outline the process of translation in three steps, "recognition of the structure of the incoming text in terms of a structural specifier; transfer of this specifier into a structural specifier in the other language; and construction to order of the output text specified."[6]

The authors believe that this system has advantages over those previously proposed. One feature which may appear to be a drawback is the fact that in addition to lexical translation, the detailed grammars described in the last paragraph must also be drawn up. The project is thus of necessity long range, the goal being to develop a program which will translate most effectively, rather than as effectively as possible after a short amount of time devoted to basic research. Furthermore, by basing the program on the theory that sentences are generated and thus have a traceable history, we can produce a superior output text.

It may be noted that the initial research required for our program may entail more work than that necessary for word-for-word programs but the generation of English sentences as a result of translation from any language at all will remain the same. Similarly, the recognition of German sentences will also remain constant as the first step for translation into any language. Thus two of the three sections of the program are not uniquely adapted to a particular pair of languages. If, however, the process of recognition, translation, and construction were integrated in a translation routine, the entire program would have to be unique for each pair of languages and no part of the program could be used in any other program. It is certainly reasonable to assume that we will eventually want to translate material to and from several languages. . It is therefore practical to develop a program which is not completely unique but one that has parts that can be used repeatedly, just as, within the program itself, we will want to build sections which can be used at several points in the program.

For the foregoing reasons the M. I. T. mechanical translation group has chosen to design the kind of translation program de scribed by Yngve.[6] The first step in such a program is a recognition routine; the one which we have designed is one type to come out of the approach we use. This does not preclude the possibility of others, some of which are already under investigation.

Problems of Recognition

Recognizing a sentence involves the discovering of the possible phrase structures that can be assigned to the sentence, as well as the particular morphemes used. Complicated as the generation of sentences in a natural language is, the recognition of those sentences is even more complex. The recognition process must take into account generation rules which delete, rearrange, expand, and reclassify constituents in the sentence. Further, recognition does not necessarily end when a single structure for a given sentence has been discovered, for a sentence in isolation may represent several structures, any one of which might be the "correct" one in the larger context from which the sentence was taken. The program described in this paper attempts to discover all possible structures for each sentence but obviously cannot decide which is the correct one.[7] Problems of multiple meaning have been discussed in several publications with various methods of solution proposed.[8, 9, 10, 11, 12, 13] One possible way, is that of looking at the context of one or two words before and after the word in question, but this is extremely time consuming. If it is possible to recognize the constituent structure, however,

6. V. H. Yngve, "A Framework for Syntactic Translation", MT, Vol. IV, no. 3, pp. 59-65.

7. Robert B. Lees, "Review of Noam Chomskys Syntactic Structures", Language, Vol. 33, p. 406 (1957).

8. Abraham Kaplan, "An Experimental Study of Ambiguity", MT, Vol. II, no. 2, pp. 39-46.

9. A. Koutsoudas and R. Korfhage, "Mechanical Translation and the Problem of Multiple Meaning", MT, Vol. III, no. 2, pp. 46-51.

10. Roderick Gould, "Multiple Correspondence", MT. Vol. VI, no. 1/2, pp. 14-27.

11. M. M. Masterman, "Thesaurus in Syntax and Semantics", MT, Vol. IV, no. 1/2, pp. 35-44.

12. Kenneth E. Harper, "Semantic Ambiguity", MT, Vol. IV, no. 3, pp. 68-69.

13. Kenneth E. Harper, "Contextual Analysis", MT, Vol. IV, no, 3, pp. 70-75.

then phonemically identical forms which belong to different form classes, such as gut and Gut will automatically be differentiated. However, wherever two phonemically identical forms belong to the same form class, such as Band = volume, Band = ribbon, and Band = bond, it is best to put off the solution until after the constituent structure has been determined, for it will then clearly designate just what the context is, and thus replace the ad hoc definition of context, which is used in the above cited papers.

Operation of the Routine

The routine itself is divided into several parts - initialization, dictionary search, determination of the kind of sentence that is being recognized (i.e. is it a question, declarative sentence, if-then construction, etc.), delimiting subordinate constructions and removing them from the main clause, establishing the limits and possible functions of the several noun phrases in the sentence, and determining what verb forms are present and what their governance relationships are. Finally the actual functions of the noun phrases are determined. After this operation has been performed on the main clause, the process is repeated for each dependent construction and indications are inserted concerning the use each construction has either in the main clause or in another dependent construction.

Initialization

Initialization involves bringing the sentence letter-by-letter into the workspace. ('Workspace' is the designation in the M.I. T. programming language, [14] for an expansible register in which strings of symbols are manipulated.) Each symbol is tested to see whether it is a space between words (space is treated as an orthographic symbol), in which case the sequence between it and the last such space is placed at the beginning of the workspace so that at the end of the initialization process the words are in reverse order. Each character is also tested to see whether it is a terminal punctuation mark, in which case the input part of the routine has been completed. Thus the unit of translation is a complete sentence. It is probable that in a connected text information gleaned from one sentence might be useful in recognizing the structure of following sentences. Such information would be useful in choosing among several possible phrase structures or meanings. However, to date we have not incorporated this information in our program.

Search

Following the initialization words are looked up in the order in which they appear in the workspace, i.e. from the end of the sentence to the beginning. The dictionary is divided into two separate parts; the first is a list of separable prefixes in which the last word of the sentence is first looked up. A typical entry in this part of the dictionary AUF // SW1 SEP 3. This is a rule in the programming language used at M.I. T. for expressing linguistic facts in a manner that can be interpreted by a computer. This rule means that if the last word in the sentence is auf it will be found, a note will be made that of the set of alternative rules designated by SW1 the particular rule that will be chosen is rule SEP, and the next rule to be applied is rule 3. This first part of the dictionary contains an entry for every separable prefix. Later SW1 SEP will cause the finite verb of the sentence to be looked up in conjunction with the separable prefix. When wieder appears as the last word in the sentence, it may present an ambiguity, e.g. Er kommt wieder, can be either "He is coming again, " or "He is coming back," if wieder is an adverb in the first sentence and a separable prefix in the second. In cases like this, two interpretations will be offered. All other words in the sentence, as well as the last one if it is not found in the separable prefix list, are looked up in the main dictionary. The entries in this dictionary have the effect of adding grammatical information in the form of subscripts to the word that is looked up. The specific form of the entry depends mostly on the form classes to which the entry word belongs, and partly on the particular word itself. Every possible German lexical item which one would want to translate is included in the dictionary. This is feasible because storage space in the form of tapes is essentially unlimited. Our program has been written so that the dictionary must contain an entry for every form to be translated. However, if it should prove to be more efficient, a sub-routine could be added which would remove endings from a stem. The dictionary would then need to contain only one entry for each morpheme. However, due to the productiveness of compounds in German, especially in scientific literature, it would be well to have a sub-routine which would indicate and look up separately their

14. V. H. Yngve, "A Programming Language for Mechanical Translation", MT, Vol. V, no. 1, pp. 25-42.

constituents.[15] This, of course, should not be done in cases where just one of two or more possible interpretations is correct, such as Literaturkunde, or where the meanings of the compound is not the same as the sum of its constituents such as Hochzeit. It would also be well to give two interpretations to ambiguous compounds such as Bluterzeugung. Some typical entries in the lexicon are:

```
BUCH      = 1/.1, CASE -GEN, PN 3S,
            GEND NEUT, CNG 1 5 9
LIEST     = 1/VRB, CASE ACC, PN 3S,
            FORM FIN, TYPE MAIN,
            TENSE PRES
DASS      = Y4 + SB1 + 1/CON -SUB
DEN       = 1/.15, CASE ACC DAT,
            GEND MASC PLUR, CNG 6 11
GEHENDE   = 1/.25, CASE NOM ACC,
            PN 3S 3P, CNG 1 2 3 4 5 7 8,
            FORM PRES-ADJ + Y1
IN        = 1/.20, CASE ACC DAT,
            CNG 5 6 7 8 9 10 11 12
SCHWEREN  = 1/.5, PN 3S 3P,
            CNG -1 2 3 5 7
```

In each of the above subscripts, the first symbol of a set between commas names a class and the following symbols of the set are the members of the class to which the lexical item may belong. The subscripts attached to BUCH give us the following information: . 1 means the word is a noun (numerical subscripts will be discussed later); CASE -GEN means the word may be any case except genitive; PN 3S indicates its person-number qualification is third singular; GEND NEUT shows it is in the neuter gender; (plural is also regarded as a gender); CNG stands for a coding which combines case, number, and gender in a two-dimensional scheme which shows number-gender horizontally in the order, neuter, masculine, feminine, plural and case vertically in this order: nominative, accusative, dative, genitive. Numbering begins at the upper left and moves horizontally.

For entry LIEST, CASE ACC means that the verb takes an object in the accusative case. FORM includes finite, infinitive, past participle, participle with an adjectival ending. TYPE is main, auxiliary, passive, modal or future.

If a word is not found in the lexicon, the sentence is automatically printed out and that word is letter-spaced. This would happen most often in the case of proper names. An alternative procedure would be to have a pre-routine which would merely look up all the words of the text in the lexicon, printing out those which are not found. Then, when entries for these forms had been made in the dictionary, the recognition program could proceed.

15. Erwin Reifler, "Mechanical Determination of the Constituents of German Substantive Compounds", MT, Vol. II, no. 1, pp. 6-14.

The Process of Recognition

Following the placement of subscripts on the lexical items, we come to the main portion of the routine. In effect it does the following: Considering the beginning of the sentence to be the left and the end to be the right, the program scans from the left looking for the finite verb. Arriving at the right, the scanner then proceeds in the other direction to locate dependent constructions, each of which is removed from the main clause, whereupon a marker is left in its place. Once more at the left, the scanner reverses its direction and moves along locating and classifying the phrases which remain.

Location of Finite Verb

We shall now examine the process of recognition in more detail: When all the forms in the German sentence have been looked up in the dictionary, their order in the workspace is reversed, so that they are now in the order of the original sentence. Then the finite verb of the main clause is located, placed at the end of the sentence, and its original position is marked. This is done in order to connect the verb stem with a possible separable prefix. The finite verb form of the main clause is moved so that all clauses, dependent and independent, may be treated alike by the rules which follow. We now come to the previously discussed set of rules, SW1. If rule SEP has been indicated, the last two elements in the workspace, i.e. the separable prefix and the finite verb, will be looked up again in the dictionary, and a different set of grammatical information will be assigned to it.

AUF-STEIGT = 1/VRB, etc.

The following are the rules for determining the finite verb: 1) In sentences containing a single clause, the finite verb is the first verb in the sentence which can be finite. 2) In complex sentences where the dependent clause precedes the finite verb of the main clause, we require that the dependent clause be followed by a comma and that each such relative clause which does not begin the sentence be preceded by a comma. Assuming that these requirements are met, we choose as the finite verb of the main clause the first finite verb-form of the sentence which is not within a dependent clause. 3) Sentences

which fall into neither of the above categories (e.g., with final dependent clauses), can be treated under the first rule.

Dependent Constructions

The next part of the routine establishes the limits of the dependent constructions — subordinate clauses, relative clauses, and participial phrases — and places them at the beginning of the workspace in the same order in which they occurred in the sentence. In establishing the limits of these constructions, those which are nested within other dependent constructions are, for the time being, ignored and are automatically moved to the beginning of the workspace with the constructions in which they are embedded. The general method of discovering these limits is to work from the end of the sentence and to place a right parenthesis, so to speak, at the end of each such construction and a left parenthesis at the beginning. Whenever the number of lefts equals the number of rights, the leftmost and the rightmost are the limits and every thing between them is moved to the beginning of the workspace. This process is repeated until the beginning of the sentence is reached. Whenever a dependent construction is moved to the left of the workspace, an indication of it is inserted in its original position, and it is separated from other constructions by special marks.

The criteria applied in placing these parentheses are: a right is placed 1) after each sequence of a finite verb plus a punctuation mark and 2) after each participial form with an adjectival ending. A left is placed 1) before a subordinate conjunction and any punctuation that precedes it, 2) in the case of a participial construction, between any constituent of a prepositional or noun phrase and a word which could not be a constituent of the same phrase, and 3) in the case of relative clauses, before an unambiguous relative pronoun or before a sequence of comma (or comma plus preposition) plus a definite article which is in turn followed by a word which could not be part of the same construction as the article. In the case of transitive participles the program recognizes the fact that the noun preceding the participle is part of the participial construction. Thus, in ein Leben spendendes Weib, the left parenthesis is placed after ein.

Identification of Phrases

At this point, the main clause of the sentence is at the end of the workspace and a mark has been placed at its beginning. The next part of the program is designed to delimit the several noun phrases and prepositional phrases and to establish their possible functions.

Since the dictionary entries attach code numbers to prepositions and all constituents of noun phrases — prepositions, articles, numerals, adjectives, and nouns, numbered from highest to lowest, respectively, — the program accomplishes the first of these operations by scanning the workspace comparing the numbers and whereever there is a sequence of one number followed by a higher number, an equal number which is not the adjective number, or by no number at all, that point is regarded as the end of the phrase, the grammatical information previously attached to each element by the dictionary is compared in order to find the possible functions of this construction in any German sentence.

DER/. 15, CASE -ACC, GEND -NEUT,
CNG 2 7 8 11

GUTE/. 5, CASE NOM ACC, PN 3S,
CNG 1 2 3 4 5 7 8

MANN/. 1, CASE -GEN, PN 3S, GEND MASC,
CNG 2 6 10

The grammatical information associated with the words of this noun phrase is compared by an automatic process akin to taking a logical product. The results of this are indicated at the beginning of the phrase on a marker Y4, Y4/. 1, CASE NOM, PN 3S, GEND MASC, CNG 2 This process is repeated for all phrases in the clause, and the markers then represent the gram matical meaning of each of them. In the case of a prepositional phrase, the grammatical information attached to the preposition is compared with that of the elements of the noun phrase to discover its function in the sentence.

Following this, the verbal elements of the clause are considered. The purpose of this portion of the routine is to recognize what verbal elements occur, what their relationship to each other is, and to place an indicator at the end of the clause to represent the grammatical meaning of each of these forms. In the case of ambiguous verb sequences such as Das Kind wird vergessen, if selection rules allow the noun phrase to be both a subject or an object of the main verb in its active voice, the program will first designate the sequence as both passive and future and in a later part of the program it will provide two constituent analyses, one passive and one future, each of which is represented by the sequence of words in the sentence.

Assignment of Syntactic Functions

The program next assigns syntactic functions to the several noun phrases. In general, the criteria for choosing which of the noun phrases is the subject are the same as those outlined by Oswald and Fletcher[16] and by Brandwood.[17]

The program is here divided into three sections, one to treat each of three types of sentences, — passive sentences, active sentences which take accusative objects, and all others. In passive sentences, if the main verb takes an accusative object, the first possible nominative that agrees with the finite verb in person and number is regarded as the subject. In other passive sentences the first noun phrase which is of the case that the main verb would take as its object in the active voice is marked as the subject. In active clauses in which the main verb does not take an accusative object, the first nominative that agrees in person and number with the finite verb is marked as subject; if there is no such nominative noun phrase, the first dative noun phrase is marked subject. In active clauses with verbs that take an accusative, if there is an unambiguous nominative it is designated as subject; otherwise the first possible nominative noun phrase that agrees with the finite verb is designated the subject. (By a very simple addition to the program, a sentence which has two noun phrases, both of which fulfill all the grammatical qualifications for subject and object, could be printed out twice with a different assignment of subject and object in each case). The object of all active clauses is the first noun phrase that has the case required by the main verb and has not been designated subject. Noun phrases that can be either genitive or dative and which follow another noun phrase are designated genitive; other such noun phrases are designated dative.

The recognition of the main clause of the sentence is now complete. The workspace now contains the dependent constructions in the same order in which they occurred in the original German sentence but separated from the main clause and placed, with indication of their limits, in front of the main clause. However, dependent constructions that are embedded within other dependent constructions are not so separated. Following the string of dependent constructions is the main clause with one change in order, i. e. the finite verb has been placed at the end of the clause and combined with a possible preceding separable prefix.

In addition to the original words of the main clause, each with its respective grammatical information, there are also several markers, each indicating the syntactic function of the following noun phrase or the preceding verbal elements. There is also a marker which shows the original position of the finite verb, and there are indicators in the original positions of each of the dependent constructions.

The program now turns its attention to the dependent constructions. Starting at the leftmost construction it goes through the routine described above and then places that construction in the main clause in the position of the first indicator that follows it. In the recognition of a dependent construction, constructions which are in turn dependent on it are treated according to the general rule, i. e. they are placed at the beginning of the workspace and indicators are put in their places in the sentence. Thus, if the leftmost dependent construction is always taken as the next to be recognized, and upon having been recognized is placed in the position of the first indicator which follows it, all of the dependent constructions will be returned to the same place from which they were taken. In the case of participial constructions it is necessary to insert a coded symbol to function in the routine as a subject and, in the case of past participial constructions, one to function as a finite verb — auxiliary after intransitive participles and passive after transitive participles — so that the rules will apply correctly. These symbols are removed when the recognition of the construction has been completed.

The foregoing is a description of an actual program which is written in the M.I.T. programming language, a language that is being adapted for an IBM 704 computer. The authors do not claim that this program can recognize all German sentences. There are orthographic restrictions as well as grammatical ones which must be observed in order that a sentence be recognizable by this routine. An example of the former is the fact that adjectives in a series must not be separated by commas. Grammatical difficulties arise with such sentences as: "Gesprochen werden können die Worte eines Satzes. .. " or "Gehen können wir nicht. " In both cases, our program would fail to find the finite verb. These limitations on the usefulness of the routine are, however, far from disheartening. Inspecting the program one readily finds the appropriate points at which to build in a

16. op. cit., pp. 10-13.

17. Booth, Cleave and Brandwood, op. cit. pp. 161-182.

sub-routine to recognize constructions that are not at present included. The limitations do not represent an inherent weakness in the system. Rather they exemplify the results of optional transformations which we have not yet treated.

Bibliography

M. V. Sofronov 170
The General Principles of Machine Translation from the Chinese Language
Voprosy Yazykoznaniya, vol. III, no. 2, 1958, Moscow, pp. 116-121

In this article it is assumed that as a first step, the sequence of morphemes represented by the original text has already been combined into words. A recognition routine for the analysis of the Chinese sentence is then outlined. Some rules are given for translation into Russian of various types of Chinese syntagmas (defined in this article as sentence fragments in which the relationship between words is determined only by their order), and for the combining of syntagmas into a Russian sentence. The procedure is illustrated by an example.

The types of grammatical structures considered are of interest in the translation of technical and scientific literature but do not include many structures found in literary works.

D. Lieberman

A. F. Parker-Rhodes 171
The Use of Statistics in Language Research
Mechanical Translation, vol. 5, no. 2, pp. 67-73

The literature concerning the application of statistics to linguistic problems and in particular to mechanical translation is reviewed. The conclusion is that much of the work done is of little direct use for mechanical translation, and that some of it is based on a misapprehension of what statistical techniques can in fact do. Statistical methods can play a useful part in the development of mechanical translation procedures once these have been well established, but have little to contribute at the present stage of the work.

Author

L. N. Korolev 172
Method of Selecting the Required Words from the Dictionary
U. S. Soviet Publication Research Service 487-D, Jan. 16, 1959, pp. 28-30. English translation of an article in Vychislitel'naya Tekhniki, Moscow, 1958, pp. 116-118

Several approaches to machine storage and search are discussed very briefly.

D. Lieberman

O. S. Kulagina 173
Machine Translation from French
Izvestiya Vysshikh Uchebnykh Zavedenii, Matematika, no. 5, 1958, pp. 46-51

A resume of an article with the same title by Kulagina and Mel'chuk (Voprosy Yazykoznaniya, no. 5, 1956) is given. The mechanization of translation algorithm programming through the use of standard subroutines is discussed, and a list of 17 standard subroutines based on the French-Russian translation algorithm is given.

D. Lieberman

Leonard Brandwood 174
Some Problems in the Mechanical Translation of German
Mechanical Translation, vol. 5, no. 2, pp. 60-66

The problems discussed are those of syntactical ambiguity and multimeaning in translating relative pronouns from German to English. The former, which is of concern for the English word order, arises from the coexistence in German of homomorphous inflections and variable word order, the latter from this combined with gender dissimilarities in the two languages. Some statistics are given of the frequency with which such ambiguities were encountered in scientific texts, and some possible solutions or partial solutions discussed.

Author

S. M. Newman, R. W. Swanson and K. C. Knowlton 175
A Notation System for Transliterating Technical and Scientific Texts for Use in Data Processing Systems, Patent Office Research and Development Reports, no. 15, 1959

This report describes a notation system by which textual material, particularly the contents of scientific and technical texts, can be represented. The system is intended to be used for input to a machine program which will rewrite the text in a standardized format, which can be used conveniently for both information retrieval studies and linguistic analysis programs.

J. R. Applegate

O. S. Kulagina 176
The Operator Description of Translation Algorithms, Mashinnyi Perevod i Prikladnaya Lingvistika, 2(9), Moscow, 1959

The need for a special form for writing translation algorithms is indicated, and some of the necessary features which such a form must possess are discussed. An attempt to develop a set of standard subroutines, or operators, based on an existing French-Russian translation algorithm, is described.

D. Lieberman

T. M. Nikolaeva 177
Soviet Developments in Machine Translation: Russian Sentence Analysis
Mechanical Translation, vol. 5, no. 2, pp. 51-59

The principles of Russian sentence analysis for machine translation from Russian are set forth, describing the various ways of obtaining grammatical and lexical information for every word in a sentence being analyzed. The syntactic analysis of a Russian sentence is described.

D. Lieberman

T. M. Nikolaeva 178
Russian Sentence Analysis
Institut Tochnoi Mekhaniki i Vychislitel'noi Tekhniki, Moscow, 1958
See preceding abstract

William S. Cooper 179
The Storage Problem
Mechanical Translation, vol. 5, no. 2, pp. 74-83

The bulkiness of linguistic reference data, contrasted with the limited capacity of existing random-access memory units, has aroused interest in means of conserving storage space. A dictionary, for example, can be considerably compressed, yet at the same time virtually all of its usefulness can be retained. Various approaches to compression are described and evaluated. One of them is singled out for extensive treatment. This approach allows considerable compression of the "argument" part of each dictionary entry, yet it introduces no chance of lookup error, provided the item to be looked up is indeed in the dictionary.

Author

Reuben A. Brower, Ed. 180
On Translation
Harvard University Press, Cambridge, Mass., 1959

In this collection of articles on the problem of translating from one language to another, two approaches are evident. On the one hand, there are essays by translators of a wide variety of material — from the Bible to Kafka — based on their personal experiences in translating. On the other hand, there are more theoretical discussions of the problem. The final section, a critical bibliography, although by no means complete, provides a useful list of works, by authors from Cicero to contemporary workers in the field of mechanical translation, for those who are interested in the development of current points of view on the subject.

J. R. Applegate