INSTITUTE OF EXACT MECHANICS AND COMPUTING
TECHNIQUE
ACADEMY OF SCIENCES USSR

ON SOME PROBLEMS OF MECHANICAL TRANSLATION

Moscow - 1963

The work was carried out by

L. N. Bykova, V. M. Zherebin and G. A. Tarasova,

workers of the Institute of Exact Mechanics and

Computing Technique, Academy of Sciences USSR.

Edited by I.S. Mukhin

ON SOME PROBLEMS OF MECHANICAL TRANSLATION

Researches in the field of MT have been carried on in the USSR from 1954. Since then the theory and practice of MT have passed a long and rather difficult way from the first attempts at the "word for word" translation up to the today's experiments making use of programming programs and automatic computers to provide formal description of the text.

At present there are many research groups in the USSR, whose activities are closely related to the problems of MT.

The work concerning MT is conducted on the following languages! Russian, English, Chinese, French, German, Japan, Hungarian, Arabic, Indonesian, Burmese, Romanian, Norwegian, Turkish, Armenian, Georgian, Ukrainian, Usbek etc.

The first experiment in MT (from English into Russian) was carried out in this country at the institute of Precision Mechanics and Computer Technology, Academy of Sciences, USSR on the "BESM" computer (Academician Lebedev's design) at the end of 1955.

In the summer of 1956 an experimental translation from French into Russian was performed with the use of a "Strela" computer at the Steklov Institute of Mathematics,Academy of Sciences, USSR. The MT algorithms used in those experiments were binary, each sentence being analysed on the word-by-word principle.

It should be mentioned that certain groups of rules of the French-Russian algorithm may be characterised as an Independent analysis.

The very early experiments in MT demonstrated the insufficiency from the MT point of view of the existent methods of text analysis and description of grammatical phenomena. In particular, it became evident that the compiling of the algorithms orientated to concrete texts providing numerous checks of concrete words was inexpedient.

An increase in the amount of text material under analysis in such algorithms resulted in either the number of concrete checks augmenting to an unacceptable figure or the algorithms becoming no longer applicable.

It was necessary to develop some generalized rules of description of lingual phenomena for making up compact and universal algorithms.

Furthermore the binary word-by-word algorithms used in the first experiments on MT owing to their specific features possessed some other disadvantages.

Thus inasmuch in compiling those algorithms provided the establishment of some correlation between linguistic units of the source and the target languages which was not systematic enough and those units were in many cases taken from different linguistic levels; the algorithms obtained as a result formed a complicated mosaic pattern. This rendered the algorithms cumbersome and difficult to test and programme.

The situation considerably improved in this sense with the Introduction of the configurational principle of analysis into the MT practice. As a typical example of an algorithm based on the configurational principle of analysis the binary (including elements of the independent analysis) English-Russian algorithm compiled by T.K.Moloshnaya (Slavonic Institute Academy of Sciences USSR) based on such a principle may be brought under consideration. In the Moloshnaya's sense the configuration is a typical context combination of classes or subclasses of words implied by concrete words which arranged in definite sequence and can have corresponding grammatical indicators. To detect the configurations and to speed up the analysis procedure in that algorithm the method of reduction of configurations (in a strictly fixed sequence) in the English sentence was applied which allowed the successive detection of the English configurations with the subsequent unfold of the equivalent Russian configurations [12] .

Configurational algorithms featuring the analysis independent of the target language characteristics have gained wide use.

These algorithms make extensive use of the sentence reduction technique.

Compiling MT algorithms with independent grammatical analysis not only reduces the required number of analyses and syntheses (multiple algorithms i.e. algorithms destined to translation from many into many-languages) but simplifies the algorithms themselves logically (binary as well as multiple ones). Thus the three constituent parts of the multiple MT algorithm - the independent analysis, the independent synthesis,and the rules of transition from the former to the latter are characterised by a more homogeneity of tasks and rules of their solution than the binary algorithm as & whole.

The most popular approaches to the main problems of MT maintained by the Soviet research groups are presented below in an order corresponding to that in which they are usually treated in the development of MT algorithms. Such specific problems as operator notation and input are treated separately.

I. Dictionaries for MT. For the application of an algorithm to a concrete text first of all the grammatical variables should be determined which are behind every wordform of the text under consideration. Therefore the real material operated by each algorithm is the concrete grammatical data drawn from the dictionary. Allowing for this every word kept in the MT dictionary is provided with a set of certain invariant grammatical indications. In the dictionaries designed for binary algorithms words are supplied with besides it's serial number also with the number of the target language equivalent .

Every word of the text fed into the computer is compared with the units kept in the dictionary. When analysing such text as those in the hieroglyphic Chinese where intervals between words do not exceed gaps between the hieroglyphs the look-up of words in the dictionary and the division of a sentence into words proceed simultaneously.

Every time dictionary search starts from the first hieroglyph of the phrase and goes on until the right hand end of the word is determined then the next hieroglyph is considered the first hieroglyph of the next word and so on [5]

In the MT dictionaries for Inflectional languages as a rule only the stem or other fixed part of the word is held in the computer memory. In such cases dictionary look-up operations furnish only part of the required information, the rest information being obtained through the morphological analysis.

In MT dictionaries for analytic languages such as Chinese words are stored in their full forms with the complete grammatical information required for the analysis. This permits to drop the morphological analysis as a separate stage.

In the former case the search must yield a partial coincidence with the fixed part of the word stored in the dictionary and in the latter case full coincidence with the dictionary entry should be achieved.

Big sizes of MT dictionaries (e.g. only mathematical dictionaries contains from 2 to 5 thousands of words) necessitates paying special attention to the problems of reducing the size of the dictionaries and speeding up the search procedure.

The reduction of the size of the dictionary is achieved by means of a "code compression" technique in which the character code is split into sections which is then summed up *(the* statistic information concerning the frequency of character combinations is desirable for using here).

To speed up the search in the dictionary a dichotomy look-up is also used with splitting the dictionary successively into two groups.

The most favorable type of dictionary organization is admittedly the one in which commonly used words and terms are held in separate sections of the dictionary (a term section may be replaced with another whenever necessary).

A considerable part of the work on dictionaries for MT purposes as well as ordinary word - frequency vocabularies may be assigned to a computer.

The semi-automatic method of making-up dictionaries is also widely used. Thus in compiling alphabetical and frequency vocabularies for a medical text of 20 000 word-uses (about 1000 sentences) (the Leningrad University) the information of word-forms containing general grammatical data is written down manually upon examination of the context [2].

Z.N. Zasorina a worker of the Leningrad University is at present using a computer in her work of making a Russian word - frequency vocabulary with a volume of the excess of 1 000 000 occurrences. Another frequency vocabulary of Russian is being compiled by the

Group of Machine Translation and Mathematical Linguistics of the Computing Center of the G.I.P.T.I. on the base of M.Gorky's works. In the latter case the proposed input data include an indication of the part of speech, grammatical connections of the word in the sentence and tables of flexions.

2. Morphological analysis. Morphological analysis deals with a single word taken separately for the purpose of obtaining some necessary information about it.

The part of morphological analysis takes in MT algorithms is fully conditioned by the source language structure, by the mode of expressing grammatical values in the language. Thus, morphological analysis makes the integral part of algorithms for those languages in which words are split into morphemes.

In binary algorithms (the French-Russian algorithm, the Steklov Institute of Mathematics) morphological analysis is used to provide the target language information about the word under analysis, which is heeded in the translation; in the independent analysis algorithms it provides the source language information, necessary for the further analysis of the text.

Morphological analysis supplements the dictionary information about the word under consideration; and in case the word is a homonym it may provide the basic information for the word, i.e. the "part-of-speech" indication.

Some algorithms morphologically determine the part-of-speech and some other characteristics for the words which are missing from the dictionary (the Independent Analysis Algorithm for the English Language, Institute of Precise Mechanics and Computing Technique).

By now the morphological analyses for MT algorithms from English, Hungarian, Georgian, German, Russian, Swedish and other languages have been developed in this country.

This creates practical possibilities for compiling a general scheme of morphological analysis for a number of languages; the typology of the languages concerned provides the theoretical basis for the scheme.

The general scheme of morphological analysis for the affixal languages has been compiled by I.A. Melchuk (Institute of Linguistics, Academy of Sciences, USSR).

The scheme covers all the morphological phenomena of the languages involved, both encountered in practice, and possible theoretically. At the same time it comprises all the logical possibilities for morphological analysis in the concrete algorithms.

With the general of analysis scheme available ess, each morphological phenomenon in a language is provided with the part for its proper analysis.

The general scheme for morphological analysis consists of five parts.

The first part is the main one. It contains general rules (the control algorithm).

This part determines the use of the algorithm as whole, finding out if the word gives a remainder after the dictionary search has been completed, then it selects the working table where the remainder is looked-up, retrieves the information and attaches it to the remainder then finishes the analysis or shows its further direction. The task of the first part *is* also the treatment of certain specific cases, when the suffix is not identified, or is identified wrong, because of the coincidence of the homonymic stems, homonymy or conjunction of the suffix with another one or with the beginning of the stem followed (e.g. with other-than-the first stem of a compound word).

The cases that have been solved are processed again.

The unsolved ones are stored.

The tables of suffixes with attached information make the following part of the scheme. The tables may be arranged in the scheme according to the part-of-speech indications, thus, the part-of-speech information of the word under analysis determines the choice of the working table. If the number of different suffixes in a language is sufficiently small, they all may be combined into a single table.

Certain combinations of suffixes in a word may serve a basis for solving homonymy problems. Such combinations are taken into account in the arrangement of the tables in the scheme. This arrangement determines the sequence in which the suffixes are processed. The experimental version of Georgian morphological analysis (Institute of Electronics, Automatics and Telemechanics AS, GSSR) uses the principle described above.

The next to the last part presents a set of rules taking account of the morphemic peculiarities and falls into some divisions according to the types of these peculiarities in the languages concerned.

The last part of the general scheme is a branch of the first one. It is designed for solving homonymy of the forms of the words and contains the algorithm for their processing. If the homonymy is to be solved syntactically, an indication on the number of the corresponding configuration table in the syntactical section of the algorithm is given.

The general scheme for morphological analysis is a step toward composing a morphological topology of languages. The scheme replaces the morphological description of each language concerned, for, being a general one it reveals the most essential morphological peculiarities of the language, showing the way of their treatment. At the same time, the general scheme for morphological analysis facilitates comparing algorithms for different languages and their programming. It may be used in any research on the automatic compiling morphological analysis algorithms.

The algorithm is written by way of a special symbolic language and in terms of operators. On the basis of the general scheme of morphological analysis the Russian Language algorithm was compiled, it was programmed for the "Kiev" Computer, Computing Center, Academy of Sciences USSR, and tested in April, 1960.

Morphological analysis may provide the sentence-structure determination for agglutinative languages, which is considered to be the final aim of the analysis; so it serves to be the main method for textual analysis in the corresponding algorithms.

In the Independent Analysis Algorithm of Japanese language (Institute of Precise Mechanics and Computing Technique), the sequences of affixes and auxiliary words are examined, and, as a result, not only the morphological information for the words is obtained, but their syntactical function is defined through which the source sentence structure is determined.

The dictionary search in this algorithm is effected after the analysis procedure has been completed. Morphological analysis is used similarly in the Tatar and Uzbek algorithms (Computing Center, State University. Tashkent).

Morphological analysis also serves to provide syntactical information in some researches concerned with affixial and fusional languages (e.g. Russian). Thus the algorithm analysing the sequences of Russian endings has been compiled (Mathematical Linguistics Laboratory, State University, Kiev).

To obtain the sentence structure information a research was made on the sequences of the last three letters of Russian words connected by prepositions and coordinative conjunctions; the dictionary of the combinations of affixes and auxiliaries has been compiled. The work has been carried out on the base of Russian text on mathematics (Institute of Precise Mechanics and Computing Technique).

3. Syntactic analysis. The syntactic analysis follows the morphological one and therefore it deals with sequences of informations about the words rather than with the words themselves).

However various approaches are, there are three main levels to be singled out in every algorithm:

1) Phrase-into-clauses splitting
2) Clauses-into-segments splitting
3) Determination of words relations within a segment and determination of segments relations within a clause; determination of clauses relations within a phrase.

The method is based on reviewing all syntactic relations possible within a phrase (certain grammatical restrictions are given in advance).

Various methods of finding out syntactic structures are used by various MT groups. The so-called valency method is applied in a number of algorithms being elaborated at the Leningrad University.

The valency implies a potential possibility for one linguistic element (here it is a word) to combine with others. Syntactic analysis program can be built on the principle of a successive valency processing, based on the categorial valency of words, characteristic of a given word category only. A corresponding valencies list, which contains all potential combinations of a definite type is here addressed. The computer compares the potential combination in the list with ones in the text and determines the further processing of a word. Such a principle is applied in the syntactic program of the Russian analysis by L.I. Zasorina and T.S.Tseitin (the Leningrad University).

A different version of the valency analysis was used by B.M.Leikina in her English analysis algorithm. This syntactic analysis is mainly oriented to the individual valencies (syntactical singularities of one or more elements rather than of a whole class).

The correspondence between the so-called categorial and individual valencies depends on the way of classification of linguistic elements.

One may split large word classes into small ones with homogenous valencies, thus enlarging the total number of classes or, on the contrary, may single out only large classes with a sufficient number of subclasses, showing the structural singularities of certain class elements. To carry out rationally a syntactic analysis, it seems necessary to find an optimal correspondence between classes and subclasses of linguistic elements.

The difference between valency rules which show all possible syntactic relations of a given element to others, and so-called positional rules is not significant. The latter characterises the position of a given word in respect to others. Certain word classes which can intervene between a given word and the one related to it, are omitted in the analysis. It is provided by special dropping rules (in the algorithms worked out at the Institute of Precise Mechanics and Computing Technique AS USSR).

It seems convenient to represent rules for syntactic analysis in the form of a syntactic configurations list (word classes combinations with a "slide" classification used, where some grammatical and semantic information may partly cross), as it was done in the English-Russian algorithm by T.N.Moloshnaya and the syntactic analysis algorithm by I.A.Melchuk.

The standard configuration list represents a peculiar "syntactic dictionary". There exists a proper algorithm - the instruction how to handle this dictionary:

As the algorithm Is designed to handle definite places in the standard table, it is good for any language, provided it is supplied with configurational table filled in properly.

Peculiarities of a language which are not written in a strict form of configurations are registered with the help of separate rules processing a certain class of singularities. This adds some elasticity to the system. Actually at the Steklov Mathematical Institute AS USSR the attempts are being made to automatize the compilation of the syntactic analysis algorithm (for the Russian texts).

For source phrases a tree of dependences is established with the aid of arrows in the direction from the governed member to the governing one. Afterwards a list of configurations (the analyzer) is compiled to analyze a given text. Two ways of solution are possible here:

a) machine receives one version of the analysis comparing it with what was given and works out error-correcting rules.
b) machine receives all possible versions of analysis and chooses correct ones, following the tables where types of relations are fixed.

Results of syntactic analysis can be represented in the form of a tree of dependences for a given sentence, with numbered "branches" and with "the top" in the predicate. I.A.Melchuk represents the structure of the sentence with the help of 31 syntagmas. Another way is to use the grammar of immediate constituents (some analysis algorithms of the Institute of Precise Mechanics and Computing Technique AS SSSR).

4. Synthesis . Interest in the problem of synthesizing a text may be considered a peculiarity of Soviet research in M.T. However, many papers do not present algorithms as such, but mere parts of a future algo-

rithm or pure linguistic research on the synthesis problem. In this outline only two approaches to the Russian phrase synthesis will be mentioned.

A model of Russian phrase synthesis on the base of semantic notation was proposed by N.N.Leontjeva(IMSIPFL) The algorithm is designed to generate Russian phrases upon the base of "semantic multipliers", it includes the following parts:

1) access routine for the dictionary
2) generation of possible versions
3) routine for matching versions
4) routine of forms synthesis

The algorithm is designed for a special dictionary (nearer 130 words). Being rather complete, this algorithm is oriented to a definite type of automatic process on the whole and a definite type of the analysis.

T.N. Nicolayeva (Institute of Precise Mechanics and Computing Technique AS USSR) proposed a Russian synthesis where phrases are built only in one variant with a fixed word order (lexico-semantic aspect was not taken into account).

This version of synthesis is independent, that is, not designed for an intermediate language. It includes a dictionary and grammatical rules.

Grammatical rules of automatic translation into Russian with the independent synthesis consist of 3 main parts:

1) phrase building, or syntactic synthesis
2) providing morphological information within an output phrase.
3) form building, or morphological synthesis

For the synthesis rules to work properly, all the elements of an input phrase should be supplied with a series of necessary indications. For every significant word it is showed to which of 23 syntactic relations provided in the algorithm it belongs.

At the step of establishing correspondences, indications to syntactic relations and lexical units of the output text are given.

At the step of the synthesis, this text is really built in the form of a sequence of words. According to the point of view which is now widespread the analysis rules can be built as a reverse of the synthesis rules, but there does not exist the homonymy problem which is essential in the analysis. A special interest can be presented by usage of synthesis rules for automatic postediting (i.e., for translations from a source language into the same one with the fixed synthesis rules). Texts actually received in experimental translations need the hand post-editing.

5. Interlanguage correspondences. Compiling multiple algorithms of independent analysis and algorithms of independent synthesis implies the existence of a special group of rules operating the transition from the information furnished by the analysis to the initial data for synthesis. The work of establishing interlanguage correspondences presents in fact one of the methods of typological comparison of languages.

The composition of groups of correspondence rules is determined primarily by the necessity of translation of meanings (grammatical or lexical) of the source language by means of units of the target language belonging to different linguistic levels, ambiguity of lexical and grammatical translation (i.e. polysemantics of the translation of configurations) information, and by insufficiency of the information furnished by the analysis for the synthesis.

In drawing up rules for translation from the analysis to the synthesis great attention should be paid to the standardisation of kinds of notation for the results of the analysis and source data for synthesis and to the unification of the methods of analysis and synthesis. Naturally the establishment of grammatical correspondences is achieved in the simplest way when both rules *of* analysis and synthesis are based on similar principles. Thus it does not cost much labour to solve the case of finding correspondences between the languages in which both analysis and synthesis are based on the same principle of binary configuration.

When constructing a correspondence system account-should be taken of the close relationship between the problems of the establishment of the lexical and grammatical correspondences [3].

At present the work is being completed concerning the construction of the English-Russian and Chinese-Russian.

6. Intermediate language. The task of the establishing of the an interlanguage correspondences system may be combined with the problem of construction of an intermediate language, especially for the purposes of MT, the introduction of which **is** one of logical consequences of application of multiple algo - rithms for independent analysis and synthesis. The idea of the intermediate language lays in the possibility of translation from many to many languages following the scheme: the source language - the intermediate language - the target language.

One of the possible ways of constructing an intermediate language consist in utilizing a set of units, obtained while finding correspondences between many pairs of languages (each unit presents a bundle of correspondences between equivalent words of various languages). An intermediate language, designed on the semantic principle forms a restricted set of elementary meanings combinations of which are used to describe meanings of words of a natural language. The constructing of the system of main meanings requires a semantic classification of words of natural languages.

The English-Russian MT program making use of an intermediate language is being devised in the MT Experimental Laboratory of the Leningrad University, The Group of Mathematical Linguistics of the Mathematical Institute of the Siberian Division of the AS USSR, undertakes experimental translation from German into Russian with the use of the semantic intermediate language which contains objects of two kinds: elements including kinds, actions, and properties and sense relations [3].

7. Translation by way of semantic analysis. The most adequate transfer of the sense, expressed by the source language sentence, by means of the target language is considered to be the main task of any linguistic translation, so the strict description of the sense expressed linguistically is one of the main MT problems. The work at the translation on the basis of semantic analysis has been undertaken at the MT Laboratory, 1-st Moscow Pedagogical Institute of Foreign Languages.

Since a sentence is assumed to be the minimal semantic unit expressed linguistically, for it presents the most adequate form of expression of any sense, the main task of the translation process consists in obtaining a semantic notation for the whole sentence.

This semantic notation must be free of the peculiarities of its linguistic expression. Furthermore, two word sequences having the same sense in respect to the corresponding situation, must be brought to the unique semantic notation and those different in their sense - to the different semantic notation, the degree and the characteristics of the dissimilarity taken into account. The semantic notation must reflect the sense of the whole sentence as well as the sense of each word.

The sentences having the identical sense were compared so the elemental sense units and relations called "semantic multipliers" were defined(e.g."Time","Element", "Reception", "Action", "Possession" and so on).

Then the dictionary was compiled, in which the sense of each word was written in terms of combinations of semantic multipliers. Each combination not always renders the sense of the word to the full, but it must represent those aspects of the sense which may be used to eliminate the ambiguity of neighbouring words or account for the participation of the word in different semantic operations. The sense of concrete words may be fixed by indicating their serial number in the sentence.

In the semantic notation of the whole sentence two types of relations between semantic multipliers are distinguished - the one placed and the two-placed predication, the attributive connection and the connection between the members of bilateral relation, the head members of these relations singled out. All this makes the relations between semantic factors similar to the syntactical relations of words, i.e. configurations.

To pass from a semantic notation of separate words to that of the sentence as a whole, some information about word dependences in the sentence is required. Thus, determination of syntactic relationship between the words in the source sentence presents the preliminary stage of the semantic analysis. The special syntactic analysis algorithm has been worked out by J.S.Martemjanov, worker of the Laboratory.

In the process of semantic analysis the words which are the connecting members of bilateral relations (i.e., nouns, verbs, prepositions, conjunctions et cet.) are defined first since they form notations for another members of the relations, and the bilateral relations are filled in.

The words are united in their semantic notation according to the dictionary information about their form and some semantic characteristics.

As a result of the work of the algorithm the semantic notation for a number of Russian sentences has been obtained, which is irrespective of the concrete ways of its linguistic expression. This notation is continous and not divided into separate words. It allows free varying of translating units and their different combinations in the synthesis.

The main task of the synthesis is to express the source semantic notation by means of the target language. Thus, single semantic notation in the dictionary for synthesis is referred to by the total of the synonymous words and forms of expression of this sense in the target language. The choice of the ultimate translating version is conditioned by the requirements of semantic transfer of the source sentence, and by the peculiarities of the target language, i.e. the potentialities of its linguistic means, the restrictions imposed on by linguistic expression and general stylistic considerations.

The translation by use of semantic analysis gives the opportunity to obtain automatically a set of versions of one and the same sense which are expressed by all the ways admissible linguistically; on the other hand, the resulting semantic notation is economic, it eliminates tautology and synonymy of semantic expression, which are brought out by distribution of certain separate words in different environments. The work at the semantic analysis on the lines of its formalization may serve to model some basic semantic notions, which are necessary for MT as well, e.g. the notion of the word meaning , with the neighbouring notions - synonymy, multiple meaning, the notion of context, the sense identity in the process of translation.

The semantic notation expressed in semantic multipliers allows to outline the semantic limits of the word and its inner semantic structure.

As a result of the semantic analysis described above the universal list of elemental semantic units may be compiled, the dictionaries of different languages, presented in semantic factors, as well as algorithms for the transfer to the semantic notation of the sentence and algorithms for the transfer of the resulting semantic notation to the words and sentences in some other languages.

8. Operator notation. For successful solution of such difficult problems as the working out of algorithms and their programming great attention should be paid to the elaboration of the standard wordy and symbolic notation. Such a strictly formalized system is required for the better operation of algorithms and creating more suitable ways for their programming. To create such a special language having necessary properties of laconicity, flexibility, simplicity and accuracy the possibility of dividing algorithms into a number of repeating elementary operations (or their successions) may be utilized, A repeating part of an algorithm having the standard program of realisation is named an operator [10]. The notion of the operator in the MT algorithms of analysis was firstly introduced by O.S. Kulagina (who offered 17 standard operators) and then developed by I.A.Melchuk. Kulagina as well as Melchuk distinguishes the operators for the better programming (Kulagina's operators and Melchuk's program operators) and the operators for abbreviating and systematizing algorithms (respectively simple rules and algorithmic operators).

For example, the following operators are chosen: simple check (the operator executes if the meaning of the given indication of the object in question meets a certain condition or not), simple record (record of the value to one indication in the given information oт in the given set of intermediate information), remembrance operator (the operator brings the information operated to a certain place from where this information *may* be taken for following operations) and ect. Each operator has its own name, an individual number and the number of the operator to which should pass after this operator is fulfilled. The further development of the operator notation is connected with the introduction of the notion of quasioperator by I.A. Melchuk. A quasioperator is a part of an algorithm bigger than an operator which is standard in general but may differ in details depending upon concrete situations. The algorithm may be at first constructed with quasioperators corresponding to the outline of a cense task and then unfolded following the peculiarities of the situation. [10].

Standardisation of the notation language of algorithms is tightly connected with the tasks of the automatisation of programming.

The elaboration of the operator notation for algorithms has great significance in cybernetics sense generally because the description of translation operations in groups of elementary operators and reduction of the translating programs create the conditions for the more concrete patterning of the human translation process [6].

Basing on the description of language in terms of Kulagina operators (Steklov institute of Mathematics) at present experiments are carried on for the application of computers for compiling algorithms.

9. Input problems. Texts designed for automatic analysis are actually put into the machine mainly by hand. Coding is made letter for letter, and the coded words are separated by the gap code. The coded text is put into the machine in groups, each of which contains

so many bits how many bits are contained in a place.

Problems of automatic coding of texts, that is, of the data input without a preliminary transfer to punch cards or tapes, is closely connected with that of building up reading devices. In its turn, their construction is not separated from a broad problem of recognizing visual images.

At present in the USSR a number of organizations develop great investigations, including more universal devices for recognizing images up to the very contour of animals and portraits (Institute of Automation and Telemechanics of the AS USSR).

The research work is carried out in a number of directions. There can be enumerated some of them: 1) fragment methods, using the presence or the absence of a black field of representations in the definite regions of a general field; 2) topologioal principle (review of the possessing transformational stability); 3) method of recognizing a sign by linear elements of its form and their place, size and correlation (geometrical approach); 4) correlational method of a recognizing a print (a coefficient of correlation between a given sign and a standard of a concrete print is counted).

The method of teaching the machine to recognize signs without a given standard is applied; on this case machines can work in the self-teaching regime.

The statistical and probability approach is widely applied (some average probability characteristics of a representation are used in distinction from the determined ones).

The greater part of experimental devices include several methods mentioned above [3].

The realization of the translation from one language into another is a many-sided problem requiring for its solution a union of various specialists, linguists and mathematicians first and foremost.

The role of mathematics in MT is exclusively important, especially in the rigorous and precise description of linguistic phenomena. Here it is necessary to mention the set-theoretic conception of the language, proposed by O.S.Kulagina, papers by A.N.Kolmogorov, R.L.Dobrushin, S.Y. Phitialov, J.S.Tseitin and others, who have created the theoretical foundation for the description of word classes and grammatical categories on the base of mathematical methods applied to the linguistic analysis.

At the same time, with the introduction and elaboration of automatic analysis methods and realization of experiments on the automatic programming, favourable conditions have arisen to divide rationally the labour between mathematicians and linguists , working at the compilation of MT algorithms.

10. Literature.

1. Машинный перевод. Вып.2, ИТМ и ВТ АН СССР, М.,1961.
2. Материалы к отчетно-координационному совещанию по автоматическому переводу.I МГПИИЯ, М.,1963.
3. Тезисы докладов на конференции по обработке информации, машинному переводу и автоматическому чтению текста. Институт научной информации АН СССР, М.,1961.
4. Питания прикладной лингвистики,Черновцы, 1960.
5. В.М.Жеребин. Некоторые вопросы машинного перевода с китайского языка. ИТМ и ВТ АН СССР, М.,1962.
6. В.В.Иванов Некоторые вопросы машинного перевода в СССР. Доклады на Конференции по обработке информации,машинному переводу и автоматическому чтению текста. М., 1961.
7. О.С.Кулагина, И.А.Мельчук Машинный перевод с французского языка на русский. "Вопросы языкознания", М.,1956,N° 5, стр.III-121.
8. О.С.Кулагина Операторное описание алгоритмов перевода. Сб."Машинный перевод и прикладная лингвистика",М.,1959, N° 2(9) и 3(10).
9. Б.М.Лейкина Некоторые аспекты характеристики валентностей. Доклады на Конференции по обработке информации, машинному переводу и автоматическому чтению текста, М.,1961.

10 И.А.Мельчук Морфологический анализ при машинном переводе. "Проблемы кибернетики", вып.6,М.,1961, стр.207-277.

11. И.А.Мельчук Об алгоритме синтаксического анализа языковых текстов. Сб."Машинный перевод и прикладная лингвистика", N° 7, М.,1962.

12. Т.И.Молошная Алгоритм перевода с английского языка на русский, "Проблемы кибернетики", вып.3, М., 1960, стр.209-272.

13. И.С.Мухин Опыты машинного перевода на электронной вычислительной машине БЭСМ, АН СССР, М., 1956.

14. Т.М.Николаева Некоторые лингвистические вопросы машинного перевода с русского языка и на русский язык.. Дисс., М., 1961, стр.219-259.

15. Г.А.Тарасова, Л.Н.Выкова О структуре алгоритма независимого анализа английского языка. ИТМ и ВТ АН СССР, М., 1962.

16. Г.С.Цейтин, Л.Н.Засорина О выделении конфигураций в русском предложении. Доклады на Конференции по машинному переводу и автоматическому чтению текста. Изд-во АН СССР, М.,1961.