Computers and the Humanities 18 (1984) 39-46

Machine Translation in the U.S.S.R.

Yu. N. Marchuk

In discussing the state-of-art in machine translation (MT) in the U.S.S.R., three aspects may be considered: theoretical, practical and informational. The theoretical aspect includes the principal questions of linguistic theory associated with MT, the practical aspect is devoted to the implemented systems, and the informational aspect covers incorporation of MT into the operating informational service network. Any other assessment of MT without such an incorporation would not be expedient.

Theoretical Considerations

An increase in the quality of the modern MT systems is connected inseparably with the nature of the translation models in use. In the effort to improve these models, one of the main trends is simulation of translation in the hope of understanding and reproducing the actions of a human translator. The existing translation models involve all operations that are specific to human translation. Reciprocally, all actions of a human translator must, to some extent, be simulated in computer-aided translation.

Simulators of translator actions rarely build a general model, but rather confine themselves to a certain type of translation. Thus, Z.M. Shalyapina[1] decomposes written translation into a sequence of operations, part of which can be easily implemented by the computer (operations on the surface level), whereas another part is difficult for machine implementation or cannot be implemented at all at the present stage of development.

A.F. Shiryaev[2] describes a model of the simultaneous translation functional system based on theoretical and experimental studies of the simultaneous translator actions. He proposes that the simultaneous translation functional system be treated not in general but as specific, on the ground that simultaneous translation can be mastered normally by development and arrangement in other functional translation systems. The basic techniques of simultaneous translation are: timing, understanding of a source text, searching for and implementing translation options, and verifying and correcting them.

The leading role is assigned to the timing technique, represented by various levels of actual cognition, unconscious verification, conscious verification, etc. Yu. N. Marchuk does not think that simultaneous translation is absolutely unique as a form of translation and in his concept of the translator's actions (oriented mainly to the simultaneous interpretation), he does not stipulate the specific features of the latter but links up the interpreter's actions with a certain concept of linguistic understanding.[3] The increased interest in interpreter actions and their simulation corresponds to the existing trend in the world and reflects the importance of the "transfer stage," i.e., the translated correspondences proper in constructing modern systems of machine translation.

Yu. N. Marchuk is director of the All-Union Centre for Translation in Moscow.

Another important direction of the theory is a study of specific features of sublanguages in connection with the simulation of translation. After a period of experimental use of many MT systems, it has become apparent that the quality of translation can be improved if the specific features of sublanguages which aid automated analysis are reflected in dictionaries and in algorithms of analysis and synthesis. The theory of sublanguages or language subsystems has been first formulated by N.D. Andreev.[4] In a recently published book, L.L. Nelyubin[5] offers a theory concerning sublanguages from the viewpoint of machine analysis and translation. The sublanguage is described by four models: functional-communicative, statistical, informational and linguo-statistical. The translation system based on these models organizes and manages documents translated from English into Russian. Its computer dictionary was compiled especially for this purpose.

Problems of lexical analysis based on formal indices, even if they are not explicitly linked to MT, are of great importance to the latter since compilation of dictionaries is the most labor-consuming activity in any MT, whereas their completeness and adequacy to the formulated objective greatly improves the MT quality. First of all, in analysis of the vocabulary, attention is focused on a word's ambiguous nature (polysemy), whose resolution is rather important for translation, P. Ya. Serdobintsev[6] refutes the assertion of R.A. Budagov[7] that polysemic words represent about 80 percent in any glossary. He gives data certifying that out of 10,515 words analyzed in two volumes of the *Modern Russian Literary Dictionary,* there are 8657 monosemic words or 83.5 percent, and 1872 polysemic ones, or 16.5 percent. Thus the real picture is directly opposite to that obtained on the basis of intuition. Although the concept of a word ambiguity in MT is associated not with the traditional polysemy but rather with the existence of several translation options, nevertheless this data is undoubtedly of a great interest.

Modern MT is, however, not concerned primarily with literature but with the translation of scientific and technical texts. A complex of problems associated with this material is the correlation between the terminological and common-use vocabularies, which is under intensive studies. Of particular interest are two questions: how words of the common-use vocabularies become terms and how terms come into common use.[8] In most cases the authors draw conclusions that the meanings of a word are inseparable from its context.

A particular recent trend in the comparative study of linguistics is called contrastive linguistics, which always deals with the data of two languages, or of several languages studied in pairs. Its direct object is a set of arranged systems or subsystems. In this aspect, linguistics links closely with problems of translation in all their modifications.[9]

The semantic level is also being studied in connection with MT. The possibilities here, however, should not be overestimated. If a sentence has passed through all levels of analysis and achieved a unique semantic representation, then the synthesis will ensure its normal translation. However, this level is incapable of resolving all ambiguities; sometimes it can make a syntactically monosemic sentence ambiguous.[10] To resolve polysemy in machine translation, attempts are made continuously to employ a semantic dictionary of the combinatorial and glossary type in the formal model of the language.[11] A tendency to formalize the meaning of word chains longer than a single word or sentence brings us to a concept of developing the meaning of a total text. A problem here is that texts written in natural languages do not accept fixed semantics. The meaning of a word can be determined only in a contextual environment.

A word, as such, devoid of any potential context, is simply a sign or a name of some object, but it is meaningless since, even potentially, it is not an element of meaningful statement.

Each word can be included in a multitude of meaningful texts. No set of words is therefore meaningful in itself because for some words in it the remaining words may not form an appropriate context. Naturally, the meaning of the text can become clear to a reader only to the extent that he is familiar with the language, i.e., that he knows the potential context of each word. Thus, the question concerning the nature of a word's meaning can be confined to the nature of the meaningful statement.[12] Researchers describe the concept of a statement's meaning through the concept of role structures, which are understood as abstractions of a functionally integral situation, as a set of "roles" regardless of the particularities of the elements in the statement.

Everything mentioned above illustrates the multilateral policy in theoretical studies in the field of MT, which covers as before a wide range of problems.

Practical Considerations

The practical activity in MT is based on a number of MT systems with post-editing. In the U.S.S.R. Center for Translation of Scientific-Technical Documentation (UTC), the machine translation system called AMPAR (Automated Machine Translation from English into Russian) has been operating in the industrial environment for a number of years.[13] It is intended for translation with post-editing of texts covering radio-electronics, computer science, programming and a number of other technical fields.

The linguistic support of the AMPAR system, based on a special translation model using translation correspondencies, consists of two components: a dictionary and a grammar. The entire translation process is divided into 17 stages, each performing a specific operation: analysis, translation or text synthesis. Source text analysis (stages 1-7) covers morphological analysis and word-form matching against the dictionary, search in the text, analysis and translation of set expressions, resolution of all forms of homonymy, and syntactic analysis by parts of speech and by parts of the sentence. Translation per se (the set of "transfer" stages, 8-14) involves translation of unambiguous and ambiguous words using the contextual environment analysis. Synthesis of the target text (stages 15-16) is performed in two stages: syntactic synthesis, i.e., establishment of syntactic and morphological correspondences between the English text and the Russian one, and morphological synthesis. The entire translation process is completed with listing of the target text (stage 17). Under various options, listing can be page-by-page or a parallel display of the Russian and English text.

The dictionary component of the system represents a sophisticated interaction of a number of dictionaries. The source dictionary (over 25,000 words) is compiled according to subject fields. The English dictionaries used in the operation are three: common vocabulary, general technical vocabulary, and computer science and programming. The word combination dictionaries are also subdivided according to subject fields. The target Russian language dictionary contains about 35,000 entries. With the aid of tables arranged according to subject fields of the source dictionary, matching unambiguous English words are translated into their Russian equivalents. The ambiguous words are translated by means of specific algorithms that establish a particular translation by analyzing a word's content usage.[14] Updating available dictionaries and creating new ones is a routine process.

The translated text is submitted to the specialized editing board for post-editing, and then delivered to the user as typewritten copy, as line-printer listing, or on magnetic tape. To speed delivery, unedited text which can be understood by a specialist is in most cases provided as preliminary pilot information. The volume of literature being translated totals several hundred signatures per year.

The circle of users is constantly expanding. The translation of huge volumes of texts in the industrial environment helps to update and enlarge the system dictionary, and describing new sublanguages extends its subject fields. Practice has shown that the texts of a sublanguage can be satisfactorily covered by supplementing the system dictionary with 4000 to 5000 Russian and English lexical units, and the word combination dictionary with 5000 to 6000 dictionary entries. This job can be managed by eight scientific/research workers in three to four months.

Aside from the quantitative enlargement of the dictionary files, routine work is carried out on qualitative improvement and upgrading of the system. In AMPAR-2, which is being created now, the entries of the source dictionary will be more widely based on the semantic and word combination properties of each word. With the syntactical analysis perfected, the translation quality will be enhanced. The foundation for the AMPAR system support, implemented in a model of translation correspondences, ensures perfection of the system without any dramatic changes of its framework and additions that do not deteriorate the system performance, as sometimes occurs in running systems.[15]

German-Russian Translation

The same linguistic and programming principles used by AMPAR are employed in the NERPA translation system (Automated Machine Translation from German into Russian). The specific features of the German grammar have been incorporated into algorithms of this system.[16] In particular, morphological analysis based on a relatively broad system of inflections plays a considerable role in the system. At the same time homonymy of stems is avoided through a special index system based on a broad homonymy of inflections.

The NERPA system features a morphological (word-formative) analysis of words that have not been found in the dictionary. These are then synthesized in Russian in the form of an artificial word with a regular suffix attached to the stem available in the dictionary (искание-search, покрывание — coverage). The NERPA system differs from the AMPAR system mainly in the analysis of composite words so typical of the German language. The general principle is disintegration of the composite words into the component stems followed by their synthesis as a Russian word combination. For example, "Informationsverarbeitung" is translated as "обраготок информации" (data processing). This composite word analysis significantly reduces the volume of the German and Russian dictionaries. In contrast to the 50,000-60,000 words which would be required if sophisticated analysis were not available, the dictionary consists of only 10,000-15,000 words. Since the number of composite words in German texts is practically unlimited, the composite analysis stage is rather important for the system and represents its characteristic feature.[16]

Compared to the AMPAR system, the NERPA system widely uses the semantic and syntactical codes which are required, partly to distinguish the syntactical homonymies but mostly to accommodate the increased number of semantic classes. For instance, nouns may fall into classes denoting space, animals, organizations, artifacts, quality, processes, etc. The differentiation of semantic classes facilitates selection of a Russian equivalent for the ambiguous German words. Thus, the German word "Seite" will be translated as "страница" (page) if accompanied by the words with semantics of a number, or otherwise as "сторона" (side). "Ausstellung" is translated as "установка " (installation) if the neighboring word has the semantics of the artifact; otherwise it is translated as "составление" (putting together).

When recently put into experimental industrial operation, the NERPA system translated a small number of texts. Presently, efforts are being exerted to enlarge the dictionary, update the files and expand the subject fields. The main engineering field of system application is programming and computer science.

Both the AMPAR and NERPA systems have unified software featuring the following:

- division of the translation process into a number of stages;
- subdivision of processing files (dictionaries, schemes, tables) into subject field subfiles;
- use of a specialized programming language alongside the joint computer system assembler;
- use of a language support (process control language) to specify input-output instructions for files handled and modes of handling;
- capabilities of the system structure reorganization (creation of various versions to select the most efficient system version);

- capabilities of obtaining the results of system operation at any stage in the form convenient for analysis in the verification mode.

While the software system complex is being created, the modular structure concept has been employed whenever individual problems are solved by the stage programs, each consisting of program modules. The modular structure concept also governs the information files (dictionaries, tables, etc.). Since modules are relatively independent, it is possible to modify programs and information files in a comparatively simple manner by developing and including new modules or changing the sequence of their operation.

Throughout system development and operation, great attention has been paid to the questions involving human efficiency in handling the system. As a result the system's linguist can:

- directly participate in creating and debugging the programs (schemes) which implement the specific algorithms for such operations as processing compound word combinations, translating ambiguous words or resolving homonymy, and analysis; in other words, participate in the stages most likely to change when system capabilities are expanded (a specialized language has been developed to simplify programming and updating);
- obtain information about words not available in the system dictionaries and about typical errors in the translation process;
- quickly localize translation error and determine its nature (selective printout of the system operation results at any assigned section of the text provides highly detailed information with an accuracy reflecting functions of an individual scheme operand);
- without hindering system operation, create various versions of the system, each of which can include new and/or modified schemes or a modified order of their operation;
- select from a wide spectrum of texts the most efficient version and include it into the work file as a work version;
- trace the state of information files.

In both the AMPAR and NERPA systems, the operator communicates with the work and service routines as they function and sets their operation modes through a special process control language which contains a set of directives, each of which causes the module to perform a particular operation. To take into account the specific nature of translation from German into Russian, the NERPA program complex differs from that of the AMPAR system in its routines and additional information modules.

French-Russian Translation

The FRAP machine translation system (Automated Machine Translation from French into Russian) operates on somewhat different principles,[17] namely the explicit use of the semantic level and semantic translation, with validity checks only on the contextual level but not on the level of translated correspondences. In the first version of the system (1976-1980), attention was mainly focused on development of the linguistic support: linguistic structures of various levels, such as morphological, syntactical, semantic, grammatical and algorithmic. The existing version of the system validates the chosen linguistic ideology which, at any given moment of translation, ensures availability of information on all levels built by that moment. The software development immediately follows the linguistic support, which is not adjusted to a particular subject matter, and for which the main dictionaries are based on unspecialized common vocabulary. The terminological dictionary covers three subject matters: electronics, computer science, and aviation/aircraft construction.

The FRAP system has a modular structure. On the highest level it is subdivided into dictionaries and grammars, or, to be more precise, into a dictionary complex and grammatical and algorithmic complex, each of which extends to the following levels:

- analysis: graphemic, morphological, syntactical and semantic;
- translation of significant lexemes, relational words, syntactic links, grammatical classes, pronouns;
- synthesis: semantic, syntactical and morphological.

The system operates in several modes. The first, auxiliary word-by-word translation, enables us to check the main system dictionaries, the source French morphological and syntactic and target Russian morphological and grammatical dictionaries.

The second one, the principal mode, includes the syntactic component which references the semantic component to verify the meaning of links and translated equivalents. Translation as such will be performed through syntactic representation of individual sentences; that is why the mode is called syntactic. The third mode is semantic, or textual and semantic; it is unavailable in the current system version but can be implemented in the next system version. This is translation through the semantic representation, which may be accompanied by compression, that is, semantic editing of the text contents. Finally, the fourth mode is informational, providing selective dissemination of information to the users. The system must ensure translation of only those text extracts which meet the users' informational requirements.

In the current FRAP version, an interface has been established between the syntactical and semantic components. Thus, the sentence is described in terms of two representations: syntactic and semantic ones. These representations interact as follows:

a) realization in the semantic representation of those dictionary-covered word meanings which correspond to the given syntactic representation;

b) rejection, on the basis of the semantic representation, of some doubtful links in the syntactic representation.

Great attention has been paid to improving software to achieve the flexibility required for adequate simulation of structure transformations in machine translation. Since programming and debugging the entire cycle of syntactic analysis using the PL/1 language proved to be too labor-consuming and a practically unconvergent process, a decision has been made to change over to a more dynamic language in which program development and debugging can be performed by linguists themselves. This language is a variant of the standard statement language developed for the AMPAR system.

In the FRAP system, four machine data representations are employed as follows:

- the pre-syntactical level representation uses information in the simplest form;
- the most consistent and system-organized representation for the syntactical stage is phrase-oriented;
- the text-oriented representation for the semantic stage slightly differs from the previous one in that it has a larger depth due to semantic information;
- the representation for synthesis is phrase-oriented and word-form-oriented and can be reduced to the second and third representations.

It is assumed that the FRAP system will allow more convenient detection of translation units and thus will improve the quality of translation.

In the U.S.S.R. Translation Center, machine translation systems are also worked out with an automated dictionary designed to assist a human translator and editor. At the moment, this dictionary contains English, German, French and Hungarian lexical files and is oriented to computer science and aviation fields.

On an order from the Kazakh Academy of Sciences, the Chikment Pedagogical Institute provides lexical industrial translation of British and American texts on chemistry and polymers.[18] The initial stage of the system was the creation of the automated dictionary of word forms and turns of speech. This dictionary is oriented to a limited class of documents. The main criterion for selecting lexical items for the dictionary is a systematic approach (their place in the terminological system) and their frequency. This translation, designated by its authors as word-by-word, turn-by-turn, has for the past several years completely satisfied its users.

Materials are also published on the development of a microcomputer-translator. Although this problem is somewhat related to the automated dictionary, real industrial machine translation in the immediate future will invariably and most closely be connected with the automated dictionary, since the main problems of such translation are lexical ones. In essence, we refer in both cases to computer-assisted translation;

the only difference is that the machine translation system takes upon itself a larger part of man' s work (at least as it is planned), whereas in the case of the automated dictionary, the computer plays a purely auxiliary role.

Among papers devoted to micro-translators, it is important to mention that in research being carried out in the Minsk Institute of Foreign Languages, extralinguistic and linguistic components are being developed. These are the heart of the data bank of the microcomputer translator, performing translation of conversational cliches. Four thousand parallel pairs of English and Russian colloquial cliches have been translated. Representing the stereotyped situations of communication in towns and cities, the Data Bank includes the following blocks:

a) Russian-English and English-Russian colloquial cliche dictionaries;

b) bilingual microdictionaries serving individual cliches;

c) bilingual subject-field dictionaries serving colloquial cliches for different situations of one subject field;

d) the common bilingual dictionary serving all subject fields, situations and cliches.[19]

Summary

Such is the development of operating MT systems which have been described in publications from 1980 up to the present moment. The informational aspect of machine translation, apparent after several years of experimental and industrial operation of the system, is closely interconnected with the economic aspect of MT. Only in those cases where large, sufficiently homogenous chunks of text have a specific form, and a large number of texts to be translated are input into the system, does the economic efficiency of this new kind of informational product appear. In the economic-technological equation, a major factor is the quality of translation and the volume of post-editing (or inter-editing, or pre-editing). The experience of the U.S.S.R. Translation Center shows that various texts require different degrees of editing. Some users are satisfied with a rough, practically unedited machine product. Others require rather deep post-editing, which practically nullifies the advantages of MT and equates it with manual translation.[20] The Center must maintain a permanent contingent of supernumerary editors responsible for editing the machine product.

The differentiated use of MT is possible with an economic benefit in the integrated scientific-technical information systems where all forms of service, including machine translation as pilot information, diverse kinds of MT, which differ in the required editing depth, are defined, such texts which are not translated by the machine are assigned, etc. It is clear, however, that many such services have long-term traditions and are not going to give way to any other product in the scientific and technical information system. As an example, there are special services which carry out the abstracting activity, and the replacement of a traditional abstract by the pilot machine translation may hardly be smooth for either the customer or the abstracting services. These and other questions of including MT in the traditional network providing information for scientific and technical development support are about to be solved.

Within its technological scheme of translation processing, the U.S.S.R. Translation Center, as a leading organization in the scientific and technical field in the U.S.S.R. and in the international information services *Interinform-perevod,* is now creating a stock of machine translations, some of which are edited and others which have been ordered by the user in the unedited form. This special stock is less expensive, for in contrast to the usual translation stock, which is stored in hard copies, it is stored on magnetic media. Materials from it can be delivered in any form on orders of the customers. The U.S.S.R. Translation Center has already gained experience in using it for informational service to users.

Summarizing everything said, it is possible to note that MT in the U.S.S.R. is perhaps developing rather slowly but in an undeviating manner because other ways are unavailable to overcome increasing language barriers in the country, as well as all over the world. Even though the scientific foundations of MT must be reevaluated, they nevertheless are continuously enriched as a result of our MT development and the contribution of related sciences, primarily linguistics.[21] New operating MT systems are being introduced, microcomputer-translators are appearing on the market, automated dictionaries are put into practice, and in-depth studies of the scientific-engineering style of speech are being conducted.[22] We are at the stage where the greatest stimulus for MT development can become the practical activity.

NOTES

1. Z.M. Shalyapina. Automatic Translation as a Model of the Human Translation Activity. Int. Forum Inf. Doc. 1980, vol. 5, No. 2, pp. 18-23.
2. A.F. Shiryaev. O psikholingvisticheskikh osobennostyakh funktsionalnoi sistemy sinkhronnogo perevoda. V kn.: Perevod kak lingvisticheskaya problema. Sb. st. (Redkol.: Garbovskii) N.K. (otvetstvennyi redaktor) i dr.-M., Izd-vo Mosk. un-ta,-120 s.
3. Yu. N. Marchuk. Blok-skhema deistvii perevodchika. V kn. Urovni teksta i metody ego lingvisticheskogo analiza. AN SSSR-VTSP, M., s. 21-31.
4. N.D. Andreev. Statistiko-kombinatornye metody v teoreticheskom y prikladnom yazykoznanii. L, Nauka, 1967, 248 s.
5. L.L. Nelyubin. Perevod i prikladnaya lingvistika. M. Vysshaya shkola, 1983, 207 s.
6. N. Ya. Serdobintsev. Semanticheskaya struktura slova i ego konnotatsii. V kn. Teorya slova i funktsionirovaniya slovarnych edinits. Saratov, 1981. s. 4-25.
7. R.A. Budagov. Chelovek i ego yazyk. M., Izd-vo MGU, 1974, 261 s.
8. See, for example, V.L. Tyulina. Izuchenie vzaimodeistvya terminologicheskoi i obshcheupotrebitelnoy leksiki kak lingvoperevodcheskaya problema. Tezisy dokladov i soobshchenii zonalnoi konferentsii "Problemy perevoda nauchno-technicheskoi literatury i prepodavaniya inostrannych yazykov v tekhnicheskich vuzach". Yaroslavl, 1983, str. 123-125.
9. V.P. Neroznak, S.L. Nikogosov, Kontrastivnaya lingvistika i problemy perevoda. See ditto, s. 223-224.
10. N.N. Leonteva, S.L. Nikogosov. Sistema FRAP i problema otsenki kachestva avtomaticheskogo perevoda. V kn. Machinnyi perevod i prikladnaya lingvistika. Vyp. 20., M, MGPIIYa im. M. Toreza, 1980, str. 57-78.
11. Yu. D. Apresyan. Tolkovo-kombinatornyi slovar v formalnoi modeli yazyka i razreshenie mnogoznachnosti. V kn. Aktualnye voprosy prakticheskoi realizatsii sistem avtomaticheskogo perevoda. M., MGU; 1982, str. 108-128.
12. A.G. Dorfman, V.M. Sergeev. Morfogenez i skrytaya smyslovaya struktura tekstov. V kn. Voprosy kibernetiki. Logika rassuzhdenii i ee modelirovanie. M., AN SSSR, 1983 str. 137-147.
13. Yu. N. Marchuk. Problemy mashinnogo perevoda. M., Nauka, 1983, 233 s.
14. Kontekstologicheskii slovar dlya mashinnogo perevoda mnogoznachnykh slov s angliiskogo yazyka na Russkii. Sostavitel Yu. N. Marchuk. M., VTsP, 1976, part I-264, part II-256 s. Yu. N. Marchuk. The Contextological Dictionary: Use in Programmed Language Teaching. Computers and the Humanities. New York, 1979, v. 13. No. 4, pp. 277-281.
15. The AMPAR system standard statement language is described in Ju. N. Marcuk, B.D. Tihomirov, V.I. Scerbinin. Ein System zur maschinellen Übersetzung aus dem Englischen ins Russische. In: Automatische Sprachübersetzung. Darmstadt, 1982, WissenschaftlicheBuchgesellschaft, 1982, SS 31-336. It has been slightly modified since then.
16. Yu. N. Marchuk, A.N. Vlasov. Nekotorye printsipy avtomatizatsii perevoda s nemetskogo yazyka na russkii. Zeitschrift "Fremsprachen", Berlin, 1980, No, 2, str. 91-99.
17. E.G. Sokolova, I.M. Kudryasheva. Komponent sintaksicheskogo analiza v sisteme FRAP. Sb. Aktualnye voprosy prakticheskoi realizatsii sistem avtomaticheskogo perevoda M., MGU, 1982, part I str. 167-183. See also N.N. Leonteva, et al., fn. 10.
18. P.A. Abdullaeva, L.P. Prikhodko. Printsipy otbora leksiki v AS dlya MP patentov. Materialy seminara "Statisticheskaya optimizatsiya prepodavaniya yazykov i inzhenernaya lingvistika". Chikment, 1980, str. 38-39.
19. A.V. Zubov, O.A. Nekhay, L.I. Rtibis. Ekstralingvisticheskie i lingvisticheskie komponenty banka dannych mikrokompyutera-perevodchika. V kn. Problemy vnutrenney dinamiki rechevykh norm. Sb. statei. Minsk, 1982, s. 191-198.
20. B.D. Tikhomirov. Osobennosti ispolzovaniya mashinnogo perevoda v sisteme nauchno-tekhnicheskoi informatsii. Materialy seminara "Statisticheskaya optimizatsiya prepodavaniya yazykov i inzhenernaya lingvistika", Chimkent, 1980, str. 132-134.
21. R.G. Kotov, Yu. N. Marchuk, L.L. Nelyubin. Machinnyi perevod v nachale 80-tykh godov. Voprosy yazykoznaniya, M., Nauka, 1983, No. 1, str. 31-38.
22. R.G. Piotrovskiy, N.P. Rakhubo, M.S. Hazhinskaya. Sistemnoe issledovanie leksiki nauchnogo teksta. Kishenev, Shtiintsa. 1981, 156 s.