The state-of-the-art in machine translation (MT) in the USSR is considered in three aspects: theoretical, practical, and informational. A survey is presented of theoretical studies in the field of linguistics for MT purposes. Then a description is given of several operational MT systems like AMPAR, NERPA, and FRAP. The question of MT product use in information services is discussed with a view of assessing its economic efficiency.

The machine output quality of modern machine translation systems is connected inseparably with the nature of translation models in use. One of the main trends in simulation of translation is research study and reproduction of actions performed by a human translator.

Existing translation models involve all operations that are specific to human translation. On the other hand, all actions of a human translator involve operations that can, to a certain extent, be simulated in computer-aided translation.

When simulating translator's actions, the authors rarely build a general model, but rather confine themselves to a certain type of translation. Thus, Z. M. Shalyapina [1] considers written translation as a continuity of operations, part of which can be easily implemented by the computer (operations on the surface level), and part of which is difficult for machine implementation. A. F. Shiryaev [2] offers description of a model of a simultaneous translation functional system based on theoretical and experimental studies of a simultaneous translator's actions. The simultaneous translation functional system is treated by the author not as a system in general, but as a specific system on the assumption that simultaneous translation can be mastered normally by means of development and rearrangement implemented in functional systems of other kinds of translation.

Basic techniques of simultaneous translation are: timing, understanding of a source text, search for and implementation of translation options, and their verification and correction. A prime 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 (geared mainly to 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's actions and simulation thereof corresponds to a similar trend elsewhere in the world today and reflects the importance of the 'transfer stage', i.e. the translation correspondences proper in constructing modern systems of machine translation.

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

Problems of lexical analysis based on formal indices, even if these are not explicitly tied up to MT, are of great importance to it since the compilation of dictionaries is the most labor-consuming part of any MT, whereas their completeness and adequacy to the formulated objective greatly facilitate the improvement of the MT quality. First of all, in vocabulary analysis, attention is focused on the ambiguous nature of a word (polysemy), whose resolution is rather important for translation. N. Ya. Serdobintsev [6] refutes the assertion of the outstanding Soviet linguist R. A. Budagov [7] that polysemantic words comprise about 80 per cent of any glossary. He gives data certifying that out of 10,515 words analysed in two volumes of the 'Modern Russian Literary Dictionary' there are 8,657 monosemantic words, or 83.5 per cent, and 1,872 polysemantic ones, or 16.5 per...
The linguistic support of the AMPAR system is based on a special translation model using translation correspondences and consists of two componentic: the dictionary and grammar ones. The entire translation process is divided into 17 stages, each performing a specific analysis, translation or text synthesis operation. 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, syntactical analysis by parts of speech and by parts of the sentence, Translation per se (the set of 'transfer' stages), stages 8—16, 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: syntactical synthesis, i.e. establishment of syntactical 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). Listing can be accomplished using various options: page-by-page listing, parallel listing of the Russian and English text, etc.

The dictionary component of the system represents a sophisticated interaction of a number of dictionaries. The source dictionary (over 25 thousand words) is subdivided according to subject division of the source dictionary. Ambiguous words are translated using specific algorithms that establish this or that translation of a word by analysing its context usage [14, 15]. Updating of available dictionaries and creation of new dictionaries is a routine process.

The translated text goes to post-editing and is then delivered to the customer as, per his request, a typewritten copy, the line printer listing, or on magnetic tape. The quality of translation is such that it can be understood by a specialist. This allows us to deliver unedited text to the customer in most cases as preliminary information to speed up its use. The volume of literature being translated totals several hundred signatures per year, and the number of new orders is constantly increasing.

Translation of huge volumes of texts under industrial conditions serves as a good updating source and a system dictionary enlargement source. Work is being done on describing new sublanguages to extend the subject fields of texts translated. Practice has shown that to satisfactorily cover the texts of a sublanguage, it is sufficient to supplement the system dictionary by 4—5 thousands of Russian and English lexical units and the word combination dictionary by 5—6 thousands of dictionary entries.
tionary entries. Eight scientific research workers can manage this job in 3—4 months. As compared to the qualitative enlargement of the dictionary files, routine work is carried out on qualitative improvement and upgrading of the system. AMPAR-2 is being developed on the basis of source dictionary which will be based on a wider use of semantic and word combination properties of a word. The system also provides for the perfection of syntactical analysis, thereby allowing translation quality to be enhanced. The idea used as the foundation for the AMPAR system support and implemented in a model of translation correspondences ensures perfection of the system without any dramatic changes of its framework and in such a manner that additions do not deteriorate system performance, as it sometimes occurs with systems in operation.

The NERPA system has been recently put into experimental industrial operation. The number of translated texts is small. Efforts are currently being made 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. Software features the following [17]:

- division of the translation process into a number of stages;
- subdivision of processing files (dictionaries, schemes, tables) into subject field subfiles;
- use of a specialised programming language alongside with the Joint Computer System (ES) Assembler;
- use of a language support (Process Control Language) to specify input/output instructions for files handled and modes of handling;
- the opportunity to reorganise the system structure (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.

A modular structure concept has been employed in the software system complex whenever individual problems are solved by stage programs, each consisting of program modules. The modular structure concept also pertains to 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.

Great attention has been paid throughout system development and operation to questions involving the linguist’s efficiency in handling the system. As a result, the system’s linguist can:

- directly participate in creation and debugging of the programs (schemes) that implement the specific algorithms for processing compound word combinations, translation of ambiguous words, resolution of homonymy and analysis, i.e. participate in those stages which are most likely to change when system capabilities are expanded (a specialised language has been developed to simplify the process of programming and updating),
- obtain information about the words which are not available in the system dictionaries and about the types of errors that occur in the translation process;
- quickly localise an error when translating and determine its nature (selective printout of the system operation results is employed at any assigned section of the text providing highly detailed information with the accuracy reflecting functions of an individual scheme or module of a system);
- without hindering system operation, create various versions of the system. Each version can include new and/or modified schemes or a modified order of their operation;
- verify operation of the created versions using a wide spectrum of texts to select the most efficient version and to 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 in the process of their functioning and sets their operation
modes by means of instructions written in a special pro-
cess control language. This language contains a set of
directives. Each directive causes the module to perform
I this or that operation.

The NERPA program complex differs from that of
the AMPAR system in routine and information modules
developed additionally to take into account the specific
nature of translation from German into Russian.

The FRAP machine translation system (Automated
Machine Translation from French into Russian) opera-
tes using somewhat different principles [10, 18] whose
essence lies in the explicit use of the semantic level and
in producing semantic translation with validity
checks on a contextual level but not on that of transla-
ted correspondences. In the first version of the system
(1976—1980), the main attention was focused on deve-
lopment of the linguistic support: linguistic structures of
various levels — morphological, syntactical, semantic;
grammar and algorithmic complexes. The existing ver-
sion of the system proves the validity of the chosen
linguistic ideology that, at any given moment of trans-
lation, ensures availability of information on all levels
which have been built by that moment.

The software development immediately follows the
linguistic support.

The linguistic support is not adjusted to a particu-
lar subject matter. The main dictionaries are formed on
the basis of the unspecialised common vocabulary. The
terminological dictionary covers three subject matters
electronics, computer science, aviation and aircraft con-
struction.

The FRAP system has a modular structure. Its lar-
gest subdivision is the subdivision into dictionaries and
grammars or, to be more precise, into a dictionary com-
plex and a grammatical and algorithmic complex. Each
complex extends to the following levels:

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

The system operates in several modes. The first one
is auxiliary word-by-word translation. This mode enab-
les us to check the main system dictionaries, the source
French morphological and syntactical and the target
Russian morphological and grammatical dictionaries.

The second mode, the principal mode, includes the syn-
tactical component which references the semantic com-
ponent to verify the meaning of links and translated
equivalents. Translation itself is done through syntacti-
cal representation of individual sentences, which is why
the mode is called syntactical one. The third mode is
a semantic or textual-and-semantic one; it has been in-
stalled in the current system version but can be imple-
mented only in the next system version. This is trans-
lated through semantic representation, which may be
accompanied by conciseness and semantic editing of the
text content. Lastly, the fourth mode is an informatio-
nal one, which assumes selective dissemination of infor-
mation to the customer. The system must ensure trans-
lation of only those text extracts which meet the cus-
tomer's informational requirements.

In the current FRAP version, interface between the
syntactical and semantic components has been worked
out. Thus, a sentence is described in terms of two repre-
sentations: a syntactical one and a semantic one. These
representations interact as follows:
(a) realisation in the semantic representation only
of those dictionary-covered word meanings which cor-
respond to the given syntactical representation;
(b) rejection of certain doubtful links found in the
syntactical representation on the basis of semantic re-
presentation.

Much attention is being given to perfecting the soft-
ware so as to achieve the flexibility required for ade-
quate simulation of operations concerning structure trans-
formations in the course of machine translation. Prog-
gramming and debugging of an entire cycle of syntacti-
cal analysis using the PL/1 language proved to be high-
ly labor-consuming and practically non-agreeing proces-
ses. A decision has been made to change over to a more
dynamic programming language in which program deve-
lopment and debugging can be performed by the lingu-
ists themselves. This language is a variant of the stan-
dard statement language developed for the AMPAR
system.

In the FRAP system, four different machine data re-
presentations are employed as follows:
— pre-syntactical level representation which uses
information in the simplest form;
— most consistent and system-organised representa-
tion for the syntactical stage, which is phrase-oriented;
— text-oriented representation for the semantic sta-
ge. It slightly differs from the previous one in that it
has a larger depth due to semantic information;
— representation for synthesis. It is phrase-oriented
and word-form oriented and can be reduced to the se-
cond and third representations.

It is thought that a more convenient detection of
translation units will be achieved in the FRAP system
and that this will make for improved translation quality.

In the All-Union Centre for Translation, machine
translation systems are being worked out in parallel
with work on an automated dictionary designed to as-
sist a human translator and editor. At the moment, this
dictionary contains English, German, French, and Hun-
garian lexical files and is oriented to computer science
and aviation fields.

The Chimkent Pedagogical Institute provides lexical
industrial translation of British and American texts on
chemistry and polymers [19]. The initial stage of the
system was the creation of an automated dictionary of
word forms and turns of speech. This dictionary is orien-
ted to a limited class of documents. The main criterion
for selecting lexical items for the dictionary is a syste-
matic approach (their place in the terminological sys-
tem) and their frequency. The authors call the transla-
tion a word-by-word one; it has completely satisfied cus-
tomers' needs for several years now.

A number of publications have also appeared on the
development of a microcomputer-translator. The prob-
lems here relate more to the automated dictionary;
however, one may assert that practical industrial machine
translation will inevitably be connected with the auto-
mated dictionary in years to come, since the main prob-
lems of such translation are lexical ones. It is essential-
ally in either case a matter of computer-assisted transla-
tion, the only difference being that the machine transla-
tion system assumes a larger part of man's work (at
least as planned), whereas the computer plays a purely auxiliary role in the case of the automated dictionary. Among chain translators, research carried out in the Minsk Institute of Foreign Languages may be noted. Extralinguistic and linguistic components are the heart of the data bank of the microcomputer-translator, which translates conversational cliches from Russian to English and vice versa. 4,000 pairs of English and Russian unedited by the customer. This is a special stock. Its use is less expensive. In contrast to the usual translation stock, where translations are stored in the form of hard copies, this stock is stored on magnetic media. Materials from this stock can be delivered in any form on orders of the customers. The All-Union Centre for Translation has already acquired experience in using this stock for informational servicing of customers.

Summarising, one may say that as a result of MT’s development being aimed at language barriers elimination, its scientific foundations are being re-evaluated. They are being continuously enriched by MT’s own development and by the contribution of related sciences, primarily, of linguistics [22]. New operating MT systems are being introduced, microcomputer-translators are coming out, automated dictionaries are being put into practice, and in-depth studies of scientific-and-engineering speech style are being conducted [23]. We are now at the stage where practice can give MT the greatest stimulus.

REFERENCES

8. Tsyulina, V. P. Studies into interaction between terms and commonly used words as a linguistic and translation problem. In: Summaries of papers presented at a local conference on 'Problems in translating scientific and technical literature and teaching foreign languages at higher education institutes of technology'. Yaroslav, 1983, 123—125.


[All references except 1, 15 and 17 are in Russian]