On the Problem of Mechanical Translation †

D. Panov, The Academy of Sciences, Moscow, U.S.S.R.

HAVING STARTED WORK on mechanical translation, we arrived at the conclusion that both the lexical meaning and the morphological shape of the word can and should be utilized in analyzing the text, and that for purposes of translation it is impractical to omit the information which can be thus obtained.   The utilization of the lexical meanings of words as well as of their contexts may also affect problems of coding.   These questions are extremely important to automatic translation.

We based our work on the following principles:
1.   Maximum separation of the dictionary from the translation program.   This enables us to enlarge the dictionary easily without changing the program.
2.   Division of the translation program into two independent parts: analysis of the foreign language sentence and synthesis of the corresponding Russian sentence.   This enables us to utilize the same Russian synthesis program in translation from any language.
3.   Storing all the words in the dictionary in their basic form.   This enables us to design the program for synthesis of the Russian text according to the standard rules of Russian grammar.
4.   Storing in the dictionary all the constant grammatical properties of words.
5.   Determination of multiple meanings of the words from the context, whereas their variant grammatical characteristics are determined by analyzing the grammatical structure of the sentence.

These principles have proved quite reliable in the practice test to which they were subjected. Hence it seems to us that they constitute a reliable basis for the solution of the problem of MT.

The contents of the dictionary, for our experiments, were determined by an analysis of mathematical textual material, starting with Milne's "Numerical Solution of Differential Equations".   For the practical experiments, which were carried out on the BESM (the USSR Academy of Sciences' high-speed electronic computer), a dictionary of 952 English and 1,073 Russian words was compiled.

For a number of English words (121 words, in our case), the place-in-the-vocabulary indication is replaced by special digit indication to show that these words have multiple meaning. The proper Russian word is chosen in this case by utilizing a special program of automatic translation, which we call "the Polysemantic Dictionary".

If the spelling of the word in the text coincides exactly with that of a word in the dictionary, i. e.*,* their numerical codes coincide, this fact can easily be established by the operation of matching.   This is the principle used for finding words in the dictionary.

In order to find words in the dictionary which possess an affix (say, 's' or 'ing' or 'ed'), the machine must discard these endings after which it must repeat the search for the word with the discarded affix.

To determine the meaning of a polysemantic word, the words surrounding it in the given sentence are analyzed.   Both the semantic and the grammatical characteristics are established. The routines for determining the particular meaning of a polysemantic word are based on an elaborate analysis of a great body of concrete material and are placed together in a special part of the translation program called the "polysemantic dictionary".   Idiomatic expressions are also included in this part of the program.

It should be noted that the establishment of the most simple and general criteria for determining a particular meaning of a word (or group of words) is the result of substantial preliminary work by our linguists on actual texts.

If a word in the sentence to be translated is not found in the dictionary, it is stored unaltered in the memory of the machine.   When the translated sentence is printed out, such a word will be printed in Latin script.

Investigations in the area of the dictionary are fairly extensive. In our group they have been carried out by L.N. Korol'ev.

Of great importance is the space that a dictionary occupies in the memory. A method of "code compression" devised by L.N. Korol'ev

---

considerably reduces this space.

The automatic translation program is divided into two main parts — analysis and synthesis.

In the first part, the form of the English words, their place in the sentence, and the grammatical information given in the dictionary are analyzed with a view to the determination of both the grammatical form of the corresponding Russian words and their place in the Russian sentence. The resulting information is recorded by means of indices, thereby permitting passage to the second part of the program -- "Synthesis of the Russian Sentence". Here, Russian words, taken from the dictionary in their basic form, acquire grammatical form in accordance with the indices obtained from the analysis.

Both English and Russian grammar is presented as a series of special schemes for the basic parts of speech: verbs, nouns, adjectives, numerals, etc. The working basis of each scheme is dichotomic analysis, i.e., a system of "checking" for the presence or absence of a certain grammatical (morphological or syntactical) characteristic of the analyzed word. In checking, only two answers are possible, either positive or negative. Each of these answers admits either a final conclusion and the development of the corresponding grammatical indices for the given word, or the continuation of the check for the presence of the next characteristic until a definitive answer is obtained together with an indication of which grammatical indices must be developed for the given word.

Different parts of the program are ordered in a sequence which ensures the development of the indices necessary to carry out further operations.

Starting with the input of the English sentence into the machine, the entire translation process has been carried out automatically with no human intervention whatsoever. To make the machine translate in the manner just described, an enormous amount of preliminary research work by philologists was required especially by I.K. Belskaya, our philologist-in-chief, and by the mathematicians I. S. Mukhin, L.N. Korol'ev, S.N. Razumovskii, G.P. Zelenkevich, and, in the early stages, N.P. Trifonov.

S.N. Razumovskii has been studying translation schemes and programs and their logical structure. He has developed a system of symbols that makes possible the recording of the details of the above mentioned schemes in an appropriate manner.

Our opinion is that the principles according to which machine translation of languages should be organized have been sufficiently clarified by now and that the time is ripe to undertake work on a large scale. We have started research work in automatic translation from German, Chinese, and Japanese into Russian.

In our discussions of machine translation from Chinese and Japanese, we thought that great difficulties would be presented by the input in these languages. However, this problem, apparently, will be solved easily by using the Chinese telegraph code.

The work on German is being carried out under the direction of Belskaya by G. P. Zelenkevich and E. A. Khodzinskaya; Chinese by A. A, Zvonov and V. A. Voronin; and Japanese by M. B. Efimov.

We also plan soon to take up the problem of translation from one foreign language into another. For this we intend to use Russian as the "inter-language".