

SYNTACTICAL ANALYSIS OF RUSSIAN MECHANICAL TRANSLATION

Abstract of notes to be presented for discussion at the informal meetings preceding the International Conference on Mechanical Translation on 20 October 1956 at the Massachusetts Institute of Technology, Cambridge, Mass.

by

PHILLIP AREY

The Ramo-Wooldridge Corporation
Los Angeles, California

The method outlined below is intended for the MT of scientific and technical Russian to English, although the adaptation of the method to other subject areas should not be ruled out. It will be apparent, too, that the MT of other languages might be handled the same way as Russian to English, though of course the coding and programming would have to be altered to fit the needs of the different languages.

To obviate the difficulties encountered in reading W/W MT, one needs a general purpose computer with a special program that can perform a contextual analysis of the source material. Thus in the target language, when needed, one can:

- 1 - transpose words
- 2 - generate additional words
- 3 - delete words
- 4 - select the most appropriate equivalent(s)

This, for the moment hypothetical, machine will have an idioglossary which can be in either paradigm or stem-affix form. For the sake of convenience in explaining the system a paradigmatic idioglossary will be used.

Each Russian word in the machine idioglossary will be coded. There may be several different codes for each Russian word. All Russian words will be coded for their part of speech, except that one may label any word with any part of speech even when that is grammatically incorrect.

There will be words which alone or in combination with other words in the idioglossary may be idioms. These words, in addition to having the regular English equivalent(s) in a separate dictionary, will have an idiom phrase equivalent. This latter equivalent will be used only after the machine has determined that the Russian word in question is being used in its idiomatic sense.

Finally, there may be still another code which will permit the word to be treated still differently as a result of other words in the sentence that would have to be recognized only as certain types of words or words belonging to a certain class rather than individual, specific words as in idioms.

Obviously the proposed coding scheme and the special computer program will be intrinsically complex. However, it will permit the realization of sentence-for-sentence MT, which eventually may rival human translations.