

CERTAIN FUNDAMENTAL PROBLEMS PROBLEMS ENCOUNTERED DURING MT RESEARCH ON THE  
RUSSIAN LANGUAGE

Notes for discussion at the Meetings of Oct.16-19 preceding the International  
Conference on Mechanical Translation on Oct.20,1956 at the  
Massachusetts Institute of Technology, Cambridge, Mass.

by

Dr. Lew R.Micklesen  
Assistant Professor of Russian  
Department of Far Eastern and Slavic Languages  
and Literature  
University of Washington  
Seattle

I. Russian Compounds and MT Linguistics:

Reifler's excellent theoretical discussion of the dissection of compounds with particular reference to the German language led me to an investigation of the compounds of Russian. As is generally assumed, Russian compounds are not formed with great facility, but there are definite productive types. In addition, the number of established compounds and the case with which their components may be dealt in the memory make some scheme for their dissection highly desirable.

The work on Russian compounds developed for these forms five significant features for MT purposes. These features are the following:

- 1) General configuration (number, shape, distribution of components)
- 2) Form classes and productive combinations of components (adjective substantive, adjective-adjective, etc. – these combinations are actually quite limited)
- 3) Memorization of components (yes or no)
- 4) Autonomy of components (free or bound or both; if both free and bound their bound or free status may be of importance in the solution of multiple meaning)

- 5) Dissection scheme (designed to secure maximum speed and reliability; it involves essentially memorization of international components and special combining forms, one matching procedure – left-to-right)

II. The status of the Definite and Indefinite Articles in Machine Translation From Languages Without to Those With Such Determiners:

MT work on Russian, in particular, has raised this apparent problem. There are two questions inherent in this problem: 1) Can the articles be supplied with complete accuracy? 2) If not, will any serious ambiguity arise through complete omission of the articles?

My answer to question No. 1 is no – if individual words, modifying phrases or clauses, or previous mention, or not of a given substantive were alone decisive criteria, prediction could possibly be made; but when the context carries implication then no accurate prediction is possible.

My answer to question No. 2 is also no – here context is helpful. Context informs the speaker or reader of a language which lacks articles of the definiteness or indefiniteness of a given substantive; any ambiguity that might develop is solved by use of other determiners.

III. A Simple Problem of Multiple Non-grammatical Meaning:

Multiple non-grammatical meaning is universally considered the besetting problem of MT. A very simple problem involving non-grammatical meaning is represented by the translation of Russian chemical nomenclature into English. A reasonable number of logical operations can convert any unsatisfactory result of a word-by-word translation into practically perfect translations.