

## RUSSIAN-ENGLISH SYSTEM

GEORGETOWN UNIVERSITY

Michael Zarechnak

The Georgetown University Russian-English System is running on IBM 360/70 .CPU time for 2000 words @ 9 seconds. The texts translated include scientific, technological, and economic materials.

M.Zarechnak in close cooperation with the linguistic research staff. The linguistic statements are coded in symbolic language designed by Dr. A.Brown ('SLC'-Programming Language). Input/output is in Assembler language.

A dictionary entry contains a split or unsplit Russian stem, grammatical coding, lexical number, and English part. The clustered entries are recognized through special local operations when the calling signals occur within the sentence under processing.

Syntactic analysis is partly based on morphosyntactic markings and partly on semantic coding.

Users: Primarily scientists at ORNL. Users' comments essentially favorable.

The unddited translation is used primarily for information purposes, although in a few instances, the translations were post-edited when the user requested it.

The quality of the present translation is the same as it was in 1964. No linguistic improvements were inserted in the system although there are some linguistic programs ready to be inserted.

The semantic level will be added. Its underlying procedures are based on the semantic collocational and colligational distributional patterns as observed in the real corpora, with such generalization as these corpora would suggest. It is hoped that after large corpora will be described both semantically and analytically, then some theories might be developed and tested deductively for the improvement of the next MT cycle. Each sentence is scanned from the left to the right, and from right to left at least forty times, following a path of certain priority-based strategies. All these scannings in both directions are grouped into four levels: word recognition, syntagmatic, syntactic, and synthesis of English. Some parts of the synthesis are independent of the Russian input.

Size of the dictionary: 50,000 stems.

## MICHAEL ZARECHNAK

*Associate Professor of Linguistics*  
*School of Languages and Linguistics*  
*Georgetown University*

Born November 18, 1920, Czechoslovakia.

Education: PhD Harvard University in the field of Slavic Languages and Literature.

Experience related to the seminar: Teaching Russian to American students on introductory, intermediate, and advanced levels of proficiency.

Doctoral thesis: "Application of A.A. Kholodovich's theory of subclasses to Russian Temporal Nouns" (1967)

1956-64: While teaching at GU, participated in the Machine Translation Project at GU and had a significant role in the development of General Analysis Techniques (GAT), a system for computer translation from Russian to English in various scientific fields.

1964-66: Conducted research at Computer Concepts, Inc., in Silver Springs, Ma., in the field of automatic analysis of Russian, English, and German semantics, and automatic abstracting and indexing.

1966-67: Worked as programming specialist in Oak Ridge, Tenn., at the Union Carbide Computer Technology Center. Programmed in Cobol, PL/I on the IBM 7090, and IBM 360 computers. Also conducted research on Russian to English Machine Translation GAT-SLC field tested at Oak Ridge jointly by CTC and ORNL.

1967-68: Worked as research associate at GU on the GU MT Project in coordination with the MT Project of the University of Texas, under the general direction of Prof. W. Lehmann.

1968-Present: Have taught various computational courses at GU and a course on the theory of translation and its application. Worked also as a consultant for Union Carbide at Oak Ridge, updating the existing MT dictionary and doing semantic and syntactic research for Russian-English MT system used at ORNL by the scientists. Published articles in the field of MT.