[*Mechanical Translation*, vol.3, no.2, November 1956; pp. 33,41]

Machine Translation Development at the University of Washington

Erwin Reifler, Far Eastern Department, University of Washington, Seattle

MACHINE TRANSLATION development at the University of Washington is a joint enterprise of the Department of Far Eastern & Slavic Languages & Literature and the Electrical Engineering Department.

MT research at our University began in November 1949. We realized very early the importance of a close cooperation between linguist and engineer and the advantages of working jointly for a definite project with well defined linguistic and engineering conditions and limitations. The result was the planning of an MT Pilot Model by Dr. Thomas M. Stout, then of our Electrical Engineering Department, and its construction under the supervision of Prof. Hill.

During my research, I developed linguistic solutions for the identification by machine of grammatical categories, of both predictable and unpredictable compound words whose constituents occur in the machine memory, and for the automatic recognition and transfer to the output of words which, both graphically and in meaning, are shared by the two languages concerned in the machine translation process. It is for the purpose of testing the fundamental engineering feasibility of these linguistic solutions that the pilot model was planned.

Along with these researches went a steady development of an adequate terminology by the linguists and engineers of our group working in close cooperation.

At present, I am continuing research in all categories of words which can be omitted from the machine memory without any loss in the intelligibility and accuracy of the output text. I am also studying the problem of how to deal with proper and geographical names, which are also members of the general vocabulary of a language but should be left untranslated.

My research has been supported by two grants from the Rockefeller Foundation.

While my research, though primarily based on German language material, took into consideration the identical or analogous phenomena of a variety of languages, Dr. Micklesen directed his investigation primarily toward the Russian language and particularly toward the application of my results to Russian.

Supported by two grants from the Graduate School of our University, Dr. Micklesen carried out two studies. In one he investigated the process of compounding in the Russian language and elaborated proposals for the economical dissection of compounds by machine. The other developed into an exhaustive analysis of MT form classes of the Russian language, the prerequisite for the mechanical determination of intended grammatical and non-grammatical meaning. He also worked out a complete tabulation of all subclasses of Russian paradigmatic form classes and determined the number of distinctive forms in each paradigmatic set. These classes are purely formal, representing the most economical (structural) breakdown into Stems and endings.

Dr. Micklesen has also been very much interested in the theoretical aspects of the linguistic problems of MT. As a structural linguist, he has been especially concerned with fitting the results of MT research into the general framework of present-day linguistic thought. He recently contributed a chapter entitled FORM CLASSES—STRUCTURAL LINGUISTICS AND MECHANICAL TRANSLATION to "For Roman Jacobson" (Mouton & Co, The Hague, 1956).

Professor Hill has given much of his time to the study of the engineering aspects of a program for machine translation using a high capacity store. The recent development of large-capacity, rapid-access storage systems permits adopting a point of view different from that previously employed. It is no longer necessary to reduce the number of entries by dissection of stems and endings or by the use of "ideoglossaries". In fact, the vocabulary can be expanded to include idiomatic sequences as well as single words.

From the machine standpoint even a whole string of words which for reasons of source-target semantics has to be handled as an entity can be entered in the store and given an idiomatic translation. Such strings of words are the longest representatives of what we call "semantic units". Furthermore, punctuation marks and even the graphically very distinctive space

Continued on page 41

REIFLER *from page 33*

between words can be considered as letters of an extended alphabet and as part of a "semantic unit". This extension of the concepts of alphabet and word provides additional graphic and semantic distinctiveness which greatly improves the translation product.

Based on these points of view a program for machine translation has been devised which 1) provides for the translation of words and word sequences, 2) permits the dissection of compounds, and 3) permits the handling of prefixes and certain types of suffixes. Each unit of input is compared serially with the entries of the store to find the longest possible memory equivalent that matches an initial portion. This is accomplished by a logical ordering of the store to place any memory equivalent that is an initial portion of a longer one behind the longer one. Each entry consists of the memory equivalent of a "semantic unit" of the source language, its target language equivalent or equivalents, the control symbols for operating the machine, and the editing symbols intended to help the reader of the output text. In a more advanced machine the editing symbols become logical tags used in a computer to edit the information extracted from the memory and thus to supply a better translation product.

Since May 15 of this year our group has been working on a project for machine translation from Russian scientific texts into English by means of the photoscopic memory device being developed for the Air Force by the International Telemeter Corporation of Los Angeles. The project is based on a contract of the University of Washington with the International Telemeter Corporation. The term of the contract is one year.