[*Proceedings of the National Symposium on Machine Translation*, UCLA February 1960]

Session 3: CURRENT RESEARCH

CURRENT RESEARCH AT THE UNIVERSITY OF WASHINGTON

Erwin Reifler
University of Washington

MT research at the University of Washington has, it appears to me, hitherto differed mainly in three important respects from that of other research groups in this country and abroad. These are:

1. The apparent emphasis on lexicography rather than on the structural linguistic aspects of the MT problem.

2. The apparent emphasis on the language and vocabulary of many fields of science rather than of one field or sub-field of science.

3. The apparent emphasis on and preference for the "free-form approach" rather than what we may call the "stems and endings approach". This means the reception into the MT-operational lexicon not only of the undissected customary lexical forms such as the nominative singular of nouns, the present infinitive of verbs, etc., but even of all relevant forms of the paradigms of all eligible semantic units of the source language [1].

The result of this "free-form approach" has been an MT-operational lexicon containing almost 170, 000 Russian-English entries which we supplied to our sponsors, the Rome Air Development Center, Air Force Research and Development Command, by June 1959. The Russian part of these nearly 170, 000 entries represents only about 14,000 Russian semantic units. The details are found in our first Comprehensive report [2].

The decision to take such a course has by no means been an arbitrary one, but was based on the following facts:

1. Our sponsors did not require us, during the initial phases of our project, to undertake structural analytic researches, but explicitly charged us to compile a Russian-English lexicon for MT purposes. In the course of the ensuing lexicographical studies we soon realized that in certain types of cases of higher frequency it is possible to solve grammatical and/or non-grammatical problems [3] by lexicography alone--that is, without the necessity of logical procedures and logical machine operations. This approach I shall outline

in my second paper entitled "The Solution of Machine Translation Linguistic Problems through Lexicography". We decided therefore to try to achieve an optimum of lexicography in which we would solve as many problems of grammatical and non-grammatical ambiguity as possible, leaving the unsolved problems to be dealt with by logical programs based on considerations of environment and structural hierarchy.

2. As for our apparent emphasis on the language and vocabulary of many fields of science rather than one field or sub-field of science we have to stress the following two points:

(a) The first is that we were privileged and fortunate to be asked by our sponsors to do our lexicographical and, later, linguistic work in consideration of a translation system which includes the truly ingenious photoscopic disc designed by Dr. Gilbert King. This permanent memory device has, or soon will have, a storage of such capacity that the reduction of the size of the MT glossary has lost its initial significance. If specialized glossaries, idio-glossaries, for one field or sub-field of science will still be used in MT in the future, it will not be because of any limitations in the storage capacity of the permanent memory device.

(b) The second point is the fact that in scientific publications it is not the scientific and technical terms which present a serious semantic problem, although they are frequently shared by more than one field of science and then sometimes require different translations. The important point to stress here is the fact that in scientific publications it is the non-scientific and non-technical general-language vocabulary which constitutes the major semantic problem in our MT research. And this general-language vocabulary is shared by all fields and sub-fields of science and presents in all of them the same semantic problems. We decided, therefore, to compile the general-language vocabulary current in modern Russian

scientific texts, and to work out procedures for the pinpointing of intended meaning in the case of scientific and technical terms belonging to more than one field of science. This problem will be presented in detail in the paper "An Experiment in the Automatic Selection or Rejection of Technical Terms" by Dr. Lew R. Micklesen who until recently was on the staff of the University of Washington MT project and now has joined Dr. King's MT research staff at the IBM Research Center [4] .

These two reasons, namely, the availability of the photoscopic disc and the fact that the general-language vocabulary is shared by all fields of science, made it unnecessary for us to limit ourselves in our MT research to one single field or sub-field of science.

3. The reasons for our apparent preference of the "free-form approach" instead of the linguistically and engineering-wise really very attractive "stems-and-endings approach" we have frequently stated very emphatically. Let us repeat them here today. The "stems-and-endings approach" means automatic dissection of input forms, automatic identification of the constituent stem and ending or endings, and the automatic determination of the intended meaning of the dissected free-form in consideration of the information coded with the stem and the ending. All this requires, of course, quite a number of machine operations. Moreover, the dissection into stem and ending means the destruction or loss of the meaning or meanings the free-form carries, and this lost semantic information has to be synthesized subsequently by the machine. All this is necessary as long as the storage capacity of the permanent memory is limited, and it may, for all I know, actually become the established practice of the MT of the future even when permanent memories with very large storage capacities are available. But since we at the University of Washington had the good fortune to be able to work in consideration of the photoscopic disc, we felt that we should not increase our, in any case, already stupendous non-grammatical meaning problem by destroying the meaning or meanings of free-forms through dissection into stem and ending.

Even during the time of our concentration on lexicography,

some members of the University of Washington MT project had begun structural analytic researches for the purpose of elaborating logical procedures for the solution of all those problems lexicography alone cannot solve. The preliminary results have been published in our first comprehensive report [5]. These researches were continued until after the completion of the MT-operational Russian-English lexicon, and they will be reported in detail in our forthcoming second comprehensive report.

Apart from these investigations of lexicographical and structural linguistic problems of MT carried on at the University of Washington, I should also mention Dr. Robert E. Wall's research into various engineering aspects of MT, partly already reported in our previous report [6]. For his latest research results I have to refer you again to our forthcoming report. On one of these researches Dr. Wall will report before this conference in his paper entitled "System Design of a Computer for Russian-English Translation". Some research has furthermore also been carried on at our University concerning the problem of an electronic print reader [7]. I may also mention here that on October 28, 1959, the Electrical Engineering Department of our University was awarded a special contract by the Air Force for research which may become important for the further development of MT. The project will be concerned with the study of the quality of machine translations. In the determination of the quality, the concepts of probability and information theory will be considered.

REFERENCES

[1] Erwin Reifler, Outline of the Project, Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English, the University of Washington, Seattle, June, 1958.

[2] Lew R. Micklesen, Procedural Report, 5. 7: Total Count of Semantic Units in the University of Washington MT-Operational Lexicon, Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English.

[3] Erwin Reifler, op. cit., 3.3: Meaning.

[4] Cf. Udo K. Niehaus, Automatic Pinpointing of Intended Meanings, Linguistic and Engineering Studies. . . , Engineering Analysis, §§ 8.

[5] Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English.

[6] Ib., Engineering Analysis, §§ 2, 3, 5, 6 and 7.

[7] Philip M. Pahl and David L. Johnson, Pattern Recognition in an Electronic Reader, The Trend in Engineering at the University of Washington, July 1959.