REGULAR USE OF MACHINE TRANSLATION OF RUSSIAN AT OAK RIDGE NATIONAL LABORATORY

Martha W. Gerrard

Oak Ridge National Laboratory*

Oak Ridge, Tenn. 37830

Abstract

User reaction has been favorable to routine computer translation of Russian scientific articles at Oak Ridge National Laboratory. Speed is the chief advantage of the machine translation, illegible copy being one of the greatest problems. Costs are comparable with those of human translation. Training of key punchers is not difficult. The machine dictionary is updated frequently, and fields other than the original chemistry of the dictionary are being included.

User Reaction

The ORNL program is aimed at machine translation, not machine-aided translation. We see no prospects of eliminating either pre- or postediting in the immediate future. Our earlier work was usually post-edited, and we have eliminated some of the necessary post-editing by judicious pre-editing.

A report, "User's Evaluation of Machine Translation," prepared by Bozena Dostert, was issued in August 1973. Dr. Dostert's study was based on 10 years of use of the Georgetown machine translation system at Oak Ridge National Laboratory and at the Euratom Research Center in Italy. The results of the study indicated that 92% of the persons responding to the questions judged machine translation to be "good" or "acceptable," i.e., to be generally informative and readable. Learning to read "machinese" seemed not to present any particular problems. Ninety-six percent of these users have recommended or would recommend machine translation to their colleagues. Eight-seven percent even expressed a preference for machine translation over human.

*Operated by Union Carbide Corporation, Nuclear Division, for the Energy Research and Development Administration

Since this evaluation was made, the program in use at ORNL, formerly operated on the IBM 7090 computer, has been converted for operation on the 360. During this conversion our regular use of machine translation was suspended, but recently we have started using it again on a regular basis. Again, we are finding a generally favorable reaction. When the reaction is unfavorable, we usually can elicit favorable comment after we explain the limitations of machine translation. For example, a user was rather disturbed because a Russian word (I forget what it is) was translated "descendents" instead of "progeny." When I explained to him that meanings are selected for the dictionary which are most generally applicable rather than using meanings that apply to a specific field and that we felt that "descendents" would be meaningful even if not specific for his field, he sent us another article to translate.

Before the conversion from the 7090 to the 360 some 75% of our research scientists' needs for Russian translation were met with computer output. We expect to be operating at that high level again very soon. The 25% that we do not translate on the computer, except for requests from a few die-hards, are articles that are too badly printed or copied to be readily legible and articles with a very high proportion of mathematical equations and symbols. We encourage our users to send us the best available copy. Even so, a key puncher may not be able to distinguish among the Russian letters и, и, и, ш, and п. A human translator, because of knowledge of the language, can usually decide, maybe with a hand lens, which letter is present. But the key punchers know no Russian and can only guess. Parenthetically, I might mention that not knowing Russian typography sometimes has its advantages. A key puncher does not tend, as I do, to type "sh" for ш, but uses the "w" as required by our system.

Articles with a high proportion of mathematical expressions and symbols are not too well suited for computer translation because the material that cannot be key punched must be inserted by hand on the finished copy, a time-consuming procedure. A recent development in our program has, however, facilitated such insertion. Formerly the key puncher typed in the word(s) "long equation" or "symbol" whenever something occurred that could not be key punched. Now, a means is provided for leaving a space so that the omitted material can be written or pasted in conveniently. With

this development we are less reluctant to use the machine for translating mathematical articles than we were formerly.

Advantages of Machine Translation

Of course, the greatest advantage of a machine translation is speed. A 10-page article can be translated on the computer in minutes. Key punching requires 3 to 4 hours, and pre- and/or post-editing maybe half an hour. Thus a requester could have the translation back the day after he asks for it.

Costs

A recent estimate indicates that costs of key punching, computer operation, and the small amount of pre- and/or post-editing that we do are comparable with those of human translation.

We do a minimum of pre-editing. I usually go through an article and mark, the first time it occurs, letters or words that are not to be translated. The key puncher then punches so that "Vitamin A" is translated as such rather than as "Vitamin and". Manual post-editing consists in indicating, at the first appearance, the meaning of a work that is not in the dictionary. The time for this is really chargeable to research rather than to routine translation because we then code such words and enter them in the dictionary. We have a program for post-editing which can be used to change the meaning of a word that is obviously wrong in the context; for example, we can instruct the machine to change the word "floor" as a translation of the Russian "pol" to "sex" in a biological article.

Training of Key-Punchers

Training of key punchers is not difficult, even though, naturally, some of the letters look alike, e.g., ь, ы, Ь, and Ь. However, I have been pleasantly surprised at how quickly the operators learn to distinguish these letters.

Updating

We have a program developed by Fred Hutton for updating the dictionary readily. Formerly we had to have around 2000 words to enter before the adding of new words became economic. Now we can add a few or large number whenever we have a list coded.

The original dictionary was primarily for chemistry, with some physics and nuclear energy terms. We are expanding the dictionary with nuclear energy terms, obviously because of our particular interest, and other energy terms and are adding biology and other fields. We are working on a means for indicating that a word has one meaning in the field of chemistry and another in biology, for example.

Future

We are able to take on a few customers from outside the Laboratory if arrangements can be made for transfer of funds. We are now charging $3.00 per hundred words for this service, which is about what it costs us. We are currently paying $3.00 per hundred for human translation. We request feedback from our customers on meaning of words or possible misinterpretations of grammatical construction and are using this feedback to improve our system.

Persons interested in our services are invited to call on us for more information.

Acknowledgment

Oak Ridge National Laboratory uses the Georgetown MT system. The system was brought to ORNL by Dr. François Kertesz and its use and further development were supervisied by him until his recent retirement. The project is now monitored by the Office of Language Services, a division of the ORNL library, and we have been fortunate in having as consultants Drs. Anthony Brown and Michael Zarechnak from the original project and Mr. Fred Hutton of the Computing Technology Center in Oak Ridge. The maturity of the program is indicated by its now being supervised by the library and used for routine translations.

Office of Language Services
Oak Ridge National Laboratory
P. O. Box X
Oak Ridge, TN. 37830