

COPING WITH MACHINE TRANSLATION

J. Richard Ruffino  
XEROX Corporation  
Rochester, New York  
U.S.A.

Perhaps the title of this paper seems a bit strange, in fact somewhat "trendy" as we say in the United States. "Coping" is certainly a popular Twentieth Century word, having gained widespread usage in the last two decades as a modern technical society has called upon all of us to learn to "cope". The title of this presentation has a dual significance because machine translation is part of this technological onslaught, but the dictionary definition of coping as well is applicable to our subject. "Coping" is defined as a struggle to contend on equal terms or with some success. Certainly, we are now "coping" with machine translation, yes struggling, but still meeting with some success.

Computers are now a part of our everyday life. From pocket calculators to banking to airline reservations to industry and government where computers write our paychecks and track our taxes, we have all been touched, and deeply, by this powerful tool. So why not translation? Why not language? As more of a humanist than a scientist I often feel a twinge of triumph over computer science's failure to meet this challenge completely. Language is not fully empirical. The thought processes of man in speech and communication in general defy the powerful machine, and often defy its understanding of the logic in sentences that even a child can understand and interpret better than the computer.

Lest we fool ourselves, we must admit that the machine's logic is only a product of man's creation, so its deficiencies lie in our own lack of understanding of our own thought processes. Without a doubt, the key to true machine translation lies in the fields of cognitive psychology and anthropological linguistics. These fields of study are still too young to be able to provide the information we need to program a machine to translate as the human translator. I seriously do not believe that this will ever be accomplished, at least not in our lifetimes.

Now that I have been what might seem a bit negative in my opinions of this technology and fatalistic in my predictions for its total success, you may be wondering why I choose to continue working in this field. Realistically, I admit that machine translation is the "too early weaned child", and feel that it will never completely reach maturity. Linguistics as a science cannot alone supply the foundations to this project. Yet, as it stands today, is a tool for the translator, and becomes a more and more valuable tool each day. Machine-assisted translation, and I deliberately add the word "assisted" at this point, elevates the role of the translator because the function of translation is merged with computer technology. To the translator is presented a new, multifaceted and creative challenge, for he or she must learn to analyze and comprehend machine logic, to manipulate this logic in order to derive the fullest benefits from the machine. Those translators who fear for their jobs have, I believe, unfounded fears. The human factor in machine-assisted translation is not replaceable, and never will be. The mystique of language and the thought processes that language involves preclude the elimination of the human element from any translation activity.

In Systran, as well as other machine translation systems, we hear the word "analysis" used as standard terminology. I take some objection to the use of this word in this context, because only the human mind can truly "analyze". Computers can indeed "compute", and be programmed to follow logical sequences, but can never analyze. Among those who work closely with machine translation, there is often a tendency to personify the computer in its capabilities to deal with language. This personification no doubt stems from the computer's role in such a truly human function, that is, language.

I have made an attempt to put the field of machine translation in perspective as to its limitations due to our own limitations, but by coping, some success can be met in the field of machine-assisted translation. The translator is the one who must learn to cope, but it is the responsibility of the organization, public or private, to provide the necessary tools to help contend with the deficiencies of machine translation. At the XEROX Corporation, it was never assumed that Systran or any other computer-assisted translation system would take the place of the translator. Again, I stress the importance of the word "assisted" because this idea of

of machine translation as an aid to the translator has been the focus within our industrial environment. Systran has always been considered as a tool within our production environment, a tool for gaining speed, accuracy and productivity.

XEROX developed a very unique approach to the translation problem. With the goal of using Systran to translate service and training documentation into five languages, French, Spanish, Portuguese, Italian and German, XEROX investigated the three areas which could be influenced to increase translator productivity in the post-edit cycle.

First, there is a possibility of controlling the input, that is, the English material to be translated. Next, the software of the system itself could be improved through research and enhancements, and this too could help bring about a more productive system. Finally, the post-editor himself or herself could be given certain tools to help speed the correction cycle, but this in and of itself would not improve translation quality, but merely facilitate the post-edit, such as the use of a terminal for revision of the side-by-side translation.

It was decided to attack the problem of productivity of C.A.T. (Computer Assisted Translation) at XEROX by controlling the input. Multinational Customized English (MCE) was born. Basically, this system involves a limited vocabulary, in many ways unique to XEROX, and a set of writing rules which encourage a clear, concise English and a minimization of ambiguities. Except in the field of Natural Language Processing, little work has been done in the area of controlling English input for the purposes of machine interpretation. It is a well known fact that it is very difficult to use one's own language in a restricted or limited fashion, however, it was determined that at XEROX, this would be the best approach for capturing the greatest productivity from C.A.T. This choice is a natural one when we consider the benefits of such a controlled English input. Besides the increase in translation quality, RICE provides a uniformity in terminology and serves as an aid to the manual translators from those countries who would not be included in the C.A.T. system. In addition, MCE's uniformity and unambiguous style facilitate the use of this documentation by those whose first language is not English.

MCE's writing rules have been generated from two sources. The

Systran staff made suggestions as to those rules they knew would benefit translation, and through experience with Systran we have established additional guidelines for our technical authors.

The key to quality in using the Systran System lies not only with a control of English input, but also with the quality and comprehensiveness of the Systran dictionary. It is necessary to be constantly sampling materials both before and after the actual translation jobs are completed, in order to be able to update the dictionaries.

As we have seen, it is a combination of two major factors, writer education for a controlled input, and accurate, comprehensive dictionaries, which have made Systran a productive tool for the XEROX Corporation. We have experienced a five-to-one gain in translation time for most texts and in some cases this gain has been higher. We have been able to "cope" with machine translation, a not yet perfected technology, through our dictionary maintenance and Multinational Customized English. The limitations of any machine translation system are obvious, but "some success" can be achieved with the proper approach.