TRANSLATION SUPPORT

Alan MELBY

Department of Linguistics
Brigham Young University
Provo, Utah 84602 USA

Summary

Seeking harmony between humans and computers in translation means seeking the appropriate role for each. This paper will discuss three areas of the total translation process: (1) the actual translation, (2) the total document flow from author to published translation, and (3) terminological research. In each of these areas, we will propose roles for humans and computers that will maximize the support they lend to translation.

This Paper and the Conference Theme

The theme of this conference, "Harmonizing Human Being and Computers in Translation", is very encouraging. It signals an end to the battle between human translators and machine transla- tion. Some machine translation zealots have, in the past, predicted that human translators would be completely replaced by computers by now. And some insecure human translators have claimed that there is no place for machine translation. The author has been trying to overcome this gap for years (Melby, 1985).

When the title of this paper "Translation Support" was assigned to me, I at first wondered what was supporting what. Did it refer to how humans can support machine translation or how computers can support human translation (using Martin Kay's classic distinction between Human-Assisted Machine Translation and Machine-Assisted Human Translation)? When the organizing committee replied that it could address either option, I decided to address both in the sense of discussing how, in a harmonious relationship, each partner supports the other.

Can humans and computers work together in roles that are not harmonious? To use an extreme example, imagine a strange, useless word processor which has a command to insert a word and a command to delete a word, but no way to type in the letters of a word. The human types the INSERT command, and the computer inserts the word that the computer thinks should be used at that spot. The human is in "complete control" because he or she can delete the word if it is not the desired one and ask the computer to insert another one, repeating this cycle (indefinitely?) until the computer inserts an acceptable word. Imagine further that the word processing program, after inserting a word, asks the human after which word to split the line. That is, the simple word wrap operation, which all commercial word processors handle automatically, is assigned to the human, while the complex word selection operation, which no commercial word processors handle

automatically, is assigned to the computer. Obviously, such a
"chimpanzee" approach will not produce harmony. Although the
choices are not as clear cut, the assignment of roles between
human and computer is just as important in translation.

Selecting Human or Machine Translation

A basic decision for each text is whether to assign the draft
translation to a human or to a computer. There are clearcut
cases where machine translation is appropriate (such as Canadian
weather bulletins). There are also clearcut cases where human
translation is appropriate (such as literary translation).
Clearly, the role of the human should not always be restricted to
post-editing.

Boogaard (1988) lists an number of factors which should be
considered when making the decision between human and machine
translation. These include document length, subject matter,
complexity, machine-readability, style, need for quality, need
for speed, and number of translators involved. And, of course, a
decision cannot be made based on a single document. If there is
a flow, over time, of many similar documents then machine trans-
lation may be called for, while a variety of documents or an
uneven flow points toward human translation. As Boogaard points
out, the ultimate criterion is whether the human translator can
"enjoy and benefit from the computer's participation in the
translation process", or in other words, whether the partnership
has been harmonized.

Preserving Machine Readability during Document Flow

The path of a document from author to published translation can
be long and painful. Computers can best support the document
flow when the text is in machine readable form. In practice,
however, the machine readability of a text is often lost some-
where along the path and a human has to re-key the document.
This is not the proper role for a human: it is not only costly
and error prone but irritating. The most common reason for
losing the machine readability of a text is an incompatibility
between two hardware or software systems. For example, the text
may arrive on a 3 1/2 inch diskette and the machines in the
office only have 5 1/4 inch drives. Or it may arrive on a
9-track magnetic tape, and the office has no tape drive. Or the
character codes used to represent accented characters are
incompatible with the character codes on the target machine.
Those responsible for translation in an organization would do
well to study the different forms in which texts to be translated
arrive and prepare conversion methods in advance.

Even if the text remains in machine-readable form, the format
codes are often lost going from one system to another. This is a
particularly difficult problem since each major word processor
has its own set of format codes. One possible solution which
deserves further study is the use of SGML (the ISO standard

generalized markup language). Other contenders for standard ways of representing format codes are IBM's DCA and the Postscript markup language. And, of course, format codes may be lost during the machine translation step. It is very important that all machine translation vendors find solutions to the problem of carrying at least the basic format of a document all the way through to the translated output. Solving this problem alone may contribute as much to overall productivity as improving machine translation algorithms.

It might even be useful to have a utility which produces a skeleton document (based on a source text) with format codes but blanks instead of text, to be filled in by a human translator, preserving the format.

Sharing New Terminology

In all but the boring and repetitive translation tasks, there is a continual need to find translation equivalents for new terms. Since a new term normally occurs in a number of documents, it is likely that a human translator or terminologist has already struggled with the term and come up with a solution. This makes it very important to find ways of sharing new terms and their translation equivalents.

One way to share new terms is through post-editor feedback. For example, assume an organization which uses machine translation with human post-editors. What happens when a human post-editor corrects an inappropriate translation of a technical term? How does that correction get back into the master dictionary files of the machine translation system? If these questions are not answered with procedures that work smoothly, then the same error will have to be corrected over and over again. Cooperative roles are needed.

Another possibility for increasing harmony and mutual support in translation is to develop computer programs that help pre-editors normalize a text. (This type of component is built into the DLT machine translation system.) A pre-edit program could check not only spelling but syntax. If the text does not conform to the formal language built into the program, the human editor would be given the task of rewriting the offending sentences until they do conform, producing a cleaner machine translation. Progress in translation may depend on reevaluating the role of humans and computers.

Melby, Alan, 1985. "Machine Translation vs. Translator [Tools]: A False Dichotomy," in Humans and Machines. Ablex, edited by Stephanie Williams (from a paper presented in 1982).

Boogaard, Jill, 1988. "Deciding on the Level of Computer Participation in Translation," in Technology as Translation Strategy, State University of New York at Binghamton, edited by Muriel Vasconcellos.