3-27-1981

BYU Linguistics Symposium A Suggestion Box Translator Aid

Alan K. Melby

Follow this and additional works at: https://scholarsarchive.byu.edu/dlls

BYU ScholarsArchive Citation

Available at: https://scholarsarchive.byu.edu/dlls/vol7/iss1/19
Computers and other electronic machines have already been shown to be useful tools in translation, provided that the computer system is appropriate to the type of translation. No existing computer system can produce fully-automatic high-quality translations. Computers are currently used only as tools to increase the effectiveness of human translators. For example, Canadian weather forecasts are translated from English to French daily by a computer system developed at the University of Montreal. The computer translates about 80 percent of the sentences of the forecasts and sends the rest to human translators. Two Provo firms are marketing machine-assisted translation systems designed primarily for technical translation. These systems produce a draft translation of an entire text and then present the text to a translator who corrects and revises it with a word processor.

A word processor does not translate, but it does allow a text to be revised and corrected without retyping the parts which are already acceptable. Even literary translation can be assisted by modern word processing equipment. The human translator does all the translating, but the translation is typed using a word processor instead of a typewriter. Then the translation can be revised without retyping the entire text. This saves time and avoids the introduction of new errors while correcting old ones. It would probably be correct to claim that computers can be useful in all types of translation.

However, a translation tool can be used inappropriately. The Canadian weather forecast system would not be useful for literary translation. A literary translator draws on a very large vocabulary (well beyond 20,000 words) while weather forecasts can be covered almost exhaustively by about 1,000 words. Literary translation is highly creative. Weather forecast translation is rather boring.

This paper will discuss a type of translation called "standard text" translation. Then various machine aids to translation will be considered, including a new type of translation aid called a "suggestion box" system, which might be appropriate for standard text translation.

I. STANDARD TEXT TRANSLATION

Typically, a translator specializes in one or more pairs of languages and translates a different text at each translation session. The source and target languages may remain fixed but the text varies endlessly. In fact, if the text is too repetitious, as in weather forecasts, the work becomes terribly boring.
In standard text translation, a relatively stable set of texts in one source language is translated into one target language after another. A well-known example of standard text translation is Bible translation.

The Bible has been translated into over 1,600 languages. There are approximately 4,000 languages in the world. However, the speakers of the 300 most common languages comprise over 90% of the world's population. Thus less than 10% of the world's languages cover more than 90% of the world's population. (Statistics obtained from the LDS Translation Division.)

Standard text translation therefore differs significantly from other types of translation in that standard text translation involves a standard text going into many target languages while most other types of translation involve many different texts going into standard languages. It should not be surprising that translation aids for standard text translation might be different from those appropriate to other types of translation.

Another example of standard text translation is the work of the Emerging Languages section of the LDS Church Translation Division. There is a standard package of materials which is translated in preparation for missionary work in a given language area. These materials are intended to accompany a Bible translation already done by some Bible Society and includes selections from the Book of Mormon, missionary pamphlets, organizational handbooks, etc. After these basic materials have been translated, the second phase of translation begins as needed and another relatively fixed set of texts is translated. To date there are approximately 40 languages into which the translation of even the basic texts has been completed. It is unclear how many translations must be completed before the LDS message can be said to be available in "every tongue." But the author feels that "every tongue" means at least 250 languages and perhaps 10 times that. In any case there is a mammoth task in standard text translation ahead of the LDS Church.

II. MACHINE AIDS FOR STANDARD TEXT TRANSLATION

The remainder of this paper will consider various ways in which machines might be useful in standard text translation. The discussion will consider some well-known aids: concordances, word processing, dictionary maintenance, and full-translation. Finally, a new aid called a "suggestion box system" will be proposed. A suggestion box system includes word processing and a new twist on dictionary lookup but does not attempt to do as much as a full-translation system.

A. Concordances

A concordance is a well-known tool for detailed study of a text. The meaning of a word or phrase can often be illuminated by studying all its occurrences in a text in their respective contexts. In standard text
translation, there are often words and phrases which have no established
equivalent in the target language. Choosing translations for such words
and phrases is a challenging and important task because a precedent is
often set which will endure for better or for worse. The author
proposes that for certain texts a concordance could be very useful in
studying the meaning of a term and selecting a translation for it.

B. Word Processing

Word processing is an effective aid to many types of translation. The
major benefits are the ability to revise a translation without retyping
it and the possibility of photo-typesetting of the text without a manual
typesetting step. However, applying word processing to standard text
translation is particularly difficult because many target languages are
involved, each with a different alphabet. Word processing equipment is
available which will handle several Roman alphabets. Of course, non-Roman
alphabets such as Cyrillic and Arabic present another level of
difficulty. A word processor can certainly be designed for a particular
non-Roman alphabet, but a problem arises when the same word processor is
expected to be useful for many different alphabets. Problems include
the changing of key caps, arrangement of the keyboard, letters which are
too intricate to be represented in a small matrix of dots, and letters
whose form is changed by surrounding letters and diacriticals. It may
also be difficult to obtain high-quality printed output in many
languages on the same word processing system. Of course, if the target
language is non-alphabetic (like Chinese) then using word processing
becomes much more difficult still. If a translation is done on location
in a less developed area, it may also be difficult to obtain service and
stable power for a word processing system. Nevertheless, word
processing is advancing, becoming more flexible and less expensive and,
in the author's opinion, will become an important machine aid in
standard text translation.

C. Dictionary Maintenance

A word processing system can be extended to include a dictionary look-up
capability. In technical translation, a word processor could be linked
to a terminology bank contributed to by other technical translators of
the same subject matter. In standard text translation, a similar
possibility exists. If several translators are translating the same
text into the same target language they need some system to keep their
use of terminology consistent with each other. Whether the dictionary
is consulted directly through a computer terminal or whether the
computer is used to maintain current dictionaries which are consulted on
paper or microfiche, the computer can help maintain consistent
terminology without the publishing delays associated with traditional
otypeset dictionaries.

D. Full Translation

A considerable distance beyond word processing and dictionary
maintenance is a full-translation system which includes a complete
dictionary and grammar and produces complete translations ready for
post-editing. A full-translation system involves an enormous investment
in linguistic analysis and programming, and it is questionable whether
such an investment is justified in standard text translation.

E. A Suggestion Box Aid

So far, this paper has defined standard text translation and considered
several ways in which machines might be useful in the translation of
standard texts. It was proposed that since the source text is rather
stable, it would be worthwhile to produce a concordance of it as a
reference work for the translator. It was also proposed that it would
be helpful to have machines for word processing, including dictionary
lookup and maintenance. It was then proposed that the development of a
full-translation system might not be appropriate in standard text
translation.

Assuming that a full-translation system is too ambitious and costly for
standard text translation, one may ask whether there is a machine aid
which goes beyond word processing and dictionary maintenance without
requiring a large investment in development before being useful. The
rest of this paper discusses one such intermediate aid. This aid will
be called a "suggestion box system." In the suggestion box approach, it
is assumed that the source text is available in machine readable form.
Any text important enough to be translated into several languages will
probably already be in machine readable form and at any rate it need
only be done once. The suggestion box system reads the source text a
segment at a time. As each segment is read into the computer, the
program identifies the words and looks them up in the suggestion
text. Then the program presents to the translator the segment of
source text and, to the side, suggestions, i.e., suggested translations
for selected words and phrases. The screen of the computer terminal is
divided into several areas. One area contains a segment of text to be
translated and a second area of the screen (called the "suggestion box")
contains the computer's suggestions. A third area is a working area
where the translator enters the translation, and a fourth area accepts
translator commands which maintain the computer's dictionary of
potential suggestions. As each new segment of text is presented to the
translator it is accompanied by a list of suggestions in the "suggestion
box." The translator examines the source text and the suggestions. Bad
suggestions are ignored and good suggestions are incorporated into the
translation.
1. An Example of Using the System

For example, consider the following segment of text:

"My dear brothers and sisters, the stake president has asked us to pay our fast offerings."

This segment contains several phrases which might have been previously entered into a suggestion lexicon. The common greeting "my dear brothers and sisters" might be stored with its French equivalent "mes chers frères et soeurs". The phrase "the stake president" might be stored with its standard equivalent "le président de pieu". And the phrase "fast offerings" has a standard translation of "dons du jeûne".

When the segment of English text is presented to the translator, the computer automatically scans the English segment for words and phrases that are in the suggestion lexicon. For the segment being considered, the computer might find suggestions for the three phrases mentioned above. The following information would be presented to the translator:

<table>
<thead>
<tr>
<th>Source text</th>
<th>&quot;Suggestion Box&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>My dear brothers and sisters,</td>
<td>#1:mfs Mes chers frères et soeurs</td>
</tr>
<tr>
<td>the stake president has asked us</td>
<td>#2:pp le président de pieu</td>
</tr>
<tr>
<td>to pay our fast offerings.</td>
<td>#3:oj dons du jeûne</td>
</tr>
</tbody>
</table>

This display includes the English source text in a column on the left with suggested translations for three common phrases. The abbreviations are mnemonics consisting of the first letters of key words in a suggestion.

Assuming that all three suggestions are acceptable, the translator might enter the following for a translation:

:mfs, #2 nous a demandé de payer les :oj.

The example shows that either the suggestion name or the suggestion mnemonic can be used as abbreviations.

This abbreviated line would then be immediately expanded by the suggestion box system to a full translation:

"Mes chers frères et soeurs, le président de pieu nous a demandé de payer les offrandes du jeûne."
If the translation were from French to English, the translator might enter the line:

: MbS, : sp has asked us to pay our : fo.

and see it expanded by the computer into:

"My dear brothers and sisters, the stake president has asked us to pay our fast offerings."

2. Advantages and Disadvantages of Possible Extensions

Clearly, this system saves keystrokes for the translator. On the other hand, the translator must scan the suggestions and decide which ones to use.

The author's hope is that after a period of adjustment, a translator could feel comfortable using a suggestion box system and be more productive using it than using just a word processor. A suggestion box approach could also increase consistency in the use of terminology. Of course, when a suggestion box system is first used on a new target language it will be only a word processor. No suggestions will appear on the screen until they are entered by a translator into the suggestion dictionary. So the system is at least as helpful as a word processor, and it is potentially much more useful. Since the translator has some control over the suggestion dictionary, the translator enters only those words and phrases which have consistent equivalents and which become boring for human translators to write out in full each time they occur.

Repeated, consistent retrieval of well-defined words and phrases is something computers are very good at. They do not get bored or tired. The suggestion box approach does not at first expect the computer to handle the difficult aspects of translation, as a full-translation system does. So a suggestion box system can be used with no development time beyond that needed to set up word processing capabilities in the target language at hand.

A suggestion box system need not stop with suggestions which are presented on the screen exactly as they appear in the suggestion dictionary. Each suggestion is based on a word or phrase appearing in a segment of source text. The source segment and the abbreviated translation can be examined by the computer to produce a guess as to appropriate inflections for the target language suggestion. The guess may be right or wrong. If it is right, the translator can save more time because the suggestion will not need to be edited for inflectional suffixes after it is inserted into the translation. A suggestion box system could perhaps even be extended to the point where the system suggests translations for whole clauses or even some entire sentences. An important point is that the system is useful even without a large investment in machinery and programming and yet can be expanded as resources are available.
Of course, there is also a disadvantage of the suggestion box approach compared to simple word processing. Some translator time is required to evaluate the suggestions, and it is conceivable that more time would be spent evaluating the suggestions than would be saved by using them. The effectiveness of the system can be maximized by entering into the suggestion dictionary mostly phrases rather than single words. It seems clear, for example, that it would be easier and faster to enter "#6" into a translation than to enter the standard translation of "The Church of Jesus Christ of Latter-day Saints."

The suggestion box approach allows for considerable flexibility. With the exception of certain entries flagged as standard translations approved by some official committee, the translator could modify the suggestion lexicon at any time and immediately see the results as the modified suggestion lexicon is consulted on subsequent segments of text.

A natural extension of the system would be to have the computer print out the suggestion lexicon in the form of a glossary. Thus the system also becomes a dictionary maintenance tool.

If careful quality control is needed, then, after a document is translated, the system could automatically produce a bilingual concordance of the source and target texts. A bilingual concordance can be an effective tool in evaluating consistency of terminology in a translation.

3. Origin of the Suggestion Box Approach

The idea for the suggestion box approach to translation aids was developed in discussions of possible translation aids for Emerging Languages. Special thanks are due to Jill Peterson, Jared Burt and several others at the BYU Language Research Center. The suggestion box approach is based on three other translation aids: The least lookup aid (simple dictionary access for a word specified by the translator), the extended lookup aid, and the expansion code aid. The extended lookup aid was proposed several years ago by Eldon Lytle and others at the BYU Language Research Center. The idea of an "extended lookup" was for a computer to consult an extensive dictionary and provide a word for word translation of a text as a reference for a translator. This approach, as described, may include too many suggestions to be effective.

The idea of including a short code in the translation (e.g. #3) which the computer then expands into a word or phrase can not only save time but can also reduce the tedium that can be a part of translation. The idea of expanding codes came from a translator aid called an expansion code system. The expansion code system was implemented at BYU by Olivia Rojas, Steve Richardson, and others. In this approach, the translator enters short codes which are looked up in an expansion dictionary and expanded into full words or phrases. For example BYC might expand into Bishop's Youth Council.
In the expansion code approach, the computer does not look up any words or phrases on its own initiative. The translator instructs the system to look up an entry by including an expansion code in the translation. The expansion code approach is certainly effective and useful, but it is inherently limited to the expansion of certain codes. A suggestion box system, on the other hand, is more extensible. A simple suggestion box system is very similar to an expansion code system in that the codes entered by the translator refer to fixed entries in a dictionary. However, an expansion code system could be extended by including some of the processing involved in a full-translation system. Then the system might suggest the translation of some entire clauses, if the clause matched some pre-defined format. The suggestion box approach differs from an expansion code system or a least lookup system in that the suggestions are retrieved automatically instead of upon the specific request of the translator. The suggestion box approach also differs from extended lookup. In extended lookup all the words are looked up and no provision is made for incorporating them into the translation except by typing them out. A suggestion box system differs from a full-translation system in two ways. First, it is useful from the beginning while a full-translation system is not useful until it has an extensive dictionary and grammar. And second, it can easily avoid the most unusual and difficult constructions by simply not providing translations for them. In a full-translation system, everything needs to be handled or else the translation may not be good enough to post-edit.

In summary, the suggestion box approach combines features of the expansion code approach and the extended lookup approach and allows for expanded capability at any time, yet providing useful aid immediately.

4. Future of the System

In the author's opinion, a suggestion box translation aid represents a good division of labor between machine and human -- letting the computer handle the repetitious aspects, thus freeing the human to spend more energy on the creative aspects.

A suggestion box system could even be used by a translator-secretary team -- even if they are separated by a great distance. The first text would be translated by traditional manual means and the secretary could enter the translation and enter suggestions in the dictionary for recurring phrases. Then the next document to translate could be sent with suggestions printed out, derived from the translators first document. The translator could then write out the translation and refer to good suggestions by code and request further suggestion entries for the next document. This could save translation time and text entry time, and increase consistency of spelling and wording in fixed expressions.

An experimental suggestion box translation aid is now being programmed by Jared Burt. The author invites comments concerning the suggestion box approach and invites all interested parties to develop
their own variation of the system. In the next few years, the author hopes to implement suggestion box translation aids on several computers and develop the systems to the point where they can be tested by professional translators doing serious translation. Then, a future paper will evaluate the success or failure of the suggestion box approach to translation. The suggestion box approach is one new idea which may or may not turn out to be useful. But the author is certain that other new ideas will appear and computers will be more and more used in translation, as they are in so many human activities.

SELECTED REFERENCES

Meta, an international journal for translators, Vol. 26, No. 1, March 1981 (special issue on machine aids to translation).


Automated Language Processing Systems, Inc., Provo, Utah

Weidner Communications, Inc., Provo, Utah