INTEGRATING TRADOS TRANSLATOR'S WORKBENCH WITH MACHINE TRANSLATION

Mark Berry
MCB Systems
325 West Washington Street
San Diego, CA. 92103
619-298-8300
sales@mcbsys.com
www.mcbsys.com

The TRADOS Translator's Workbench for Windows is a translation memory program designed for making human translators more efficient and consistent in their work. It is not a machine translation program, which means it does not try to formulate sentences on its own. Rather, it stores translation work at the sentence level, and is able to recognize and insert a translation for any sentences that are repeated later. Since the Workbench uses human translations, there is very little post-editing compared to machine translation.

While translation memory requires very little post-editing, it does require new text to be translated from scratch. Machine translation provides very quick translations of new text, but the translations are often imperfect and require significant post-editing. A good approach to getting fast, accurate translation from the computer is to integrate the two systems.

Currently, the most effective strategy for integrating translation memory technology with machine translation is batch-oriented. It involves running segments not found in translation memory (rather than the document) through machine translation. Here's how it works:

1. Analyze the document with the Translator's Workbench, and export all segments which are not found in the translation memory.[1] The Workbench generates a file that resembles an exported translation memory, but with a copy of the source segment where the target segment should be. In concept, it looks like the following (you will find an explanation of codes on the next page):

    ```
    **

    <Author>MT!
    <USE>This is a text segment.
    <GER>This is a text segment.
    **
    ```

---

[1] If earlier documents were not translated with the Workbench, i.e. their segments do not occur in translation memory, you may want to use a program like WinAlign to generate segment pairs from the legacy documents. Those segments in the new text that already occurred in the legacy documents will not be passed to MT and need only be edited as human-translated sentences.

2. Run the exported file through the machine translation software. The job of machine translation software is to parse the target text, translate it, and overwrite the target text with its translation. The translated unit might look like this:

```
**

<Author>MT!
<USE>This is a text segment.
<GER>Dies ist ein Textsegment.
**
```

Explanation:

**	** starts each translation unit
<Author>MT!	<Author> is a tag which identifies the source of the translation, MT! means machine generated
<USE>This is a segment.	<USE> is the language label, U.S. English. The text is used as the source and index for translation memory.

3. The translated file is then imported back into the translation memory. It is merged with any translation units that have been previously translated by a human.

4. Translate (post-edit) the document using the Translator's Workbench. Based on the author tag, the Workbench alerts the translator when a translation has come from a machine so that extra attention can be given to post-editing MT-based text.

Advantages of Integration

- **No repetitive post-editing.** Once a segment has been post-edited, it is considered to be human-translated, and will not go through MT again.

- **User-friendly post-editing.** Post-editing of machine translation can be done with the source and target text on the screen at the same time. The Translator's Workbench automatically keeps corresponding source and target sentences together in the same window, so there is less window switching and cursor movement than post-editing in a word processor alone.

- **Easier handling of DTP tags.** Since most translation of text generated in high-end desktop publishing systems is not done in a WYSIWYG environment, translators must manually manipulate formatting tags. These tags may get misplaced or corrupted when they are touched by humans, but good machine translation software handles formatting fairly well. At the very least, using machine translation means that the formatting tags will automatically be placed in the target, and the human post-editor only has to adjust their position.

Certainly, if there is a large volume of translation, especially repetitive translation, then it is a good idea to use either MT or translation memory technology to save time and money. To take it a step further, integration of machine translation with the Translator's Workbench will help reduce turnaround time and DTP headaches in almost any translation situation.