TRANSLATION MEMORY: A TEACHING CHALLENGE

Shih Chung-ling,
National Kaohsiung First University of Science and Technology, Taiwan
cshih@ccms.nkfust.edu.tw

Abstract
This article provides a brief introduction to Translation Memories and their use in the classroom. It defines the phenomenon, describes its development, and then discusses some of its uses, notably in the teaching of translation to would-be translators in order to familiarise them with working conditions in the market.

Key words: Translation Memories (general); English-Chinese; electronic tools in translation; teaching tools; alignment tools; term banks.

What is Translation Memory?
Although Machine Translation, usually abbreviated as MT, has been around in some form or other for more than 50 years, it still is not capable of producing fully automatic high-quality translation. Nevertheless, the Translation Memory (TM) systems developed in the 1990s rekindled professional translators’ interest in technology-enabled translation. These systems were developed by companies, mostly in the software industry, for translation for internal use, but were eventually put on the market and have enjoyed growing popularity. Many companies and virtually all international organisations as well as most professional translation agencies handling large quantities of translation use TM systems as aids.

The definition
TM stands for ‘Translation Memory’. The basic element in TM is the automation of the reuse and recycling of previously translated terms and sentences. A TM database system automatically stores translations and allows for their reuse. The database created is called ‘translation memory’. The translation unit saved in the translation memory consists of a source language segment and its target language equivalent.

TM overlaps with machine-assisted human translation and computer-assisted translation in that they all depend on computers to improve or speed up the translation process. (Hedden 2002)

In 1966, the Automatic Language Processing Advisory Committee reported that fully automatic, high-quality machine translation was impossible, and instead, the committee “recommended the development of computer-based aids for translators”. (Somers 2003b:14)

In the late 1970s, the first proposal was made for “what is now called translation memory, in which previous translations are stored in the computer and retrieved as a function of their similarity to the current text being translated”, a tool then known as “Repetitions Processing”. (Somers 2003b: 14) The 1980s saw the development of computational linguistic techniques and Alan Melby “proposed the integration of various tools into a translator’s workstation”, such as word-processing, terminology management, and automatic dictionary lookup. (Somers 2003b: 14)
In 1980, Martin Kay suggested that translators might “examine past and future fragments of text that contain similar material”, (1980: 19) and thus established the basic notion of TM in the first of a small number of studies on the topic in the 1980s. However, “it was not until the mid 1990s that TM systems became commercially available, and then in a short period of time they were quickly accepted.” (Somers 2003c: 33) Consequently, the mid-1990s saw the release of numerous TM products, such as Trados, Translation Manager (IBM), and Déjà Vu (Atril).

The TM concept

A TM system is based on the reuse or recycling of previously translated terms and sentences. In a TM system, translators can “consult a database of previous translations, usually on a sentence-by-sentence basis, looking for anything similar enough to the current sentence to be translated, and can then use the retrieved examples as a model”. (Somers, 2003c: 31) TM assists human translators by searching databanks for similar or exact previous translations. In principle, human translators thus need to translate these same words, phrases, or contextual segments only once. For areas in which there are large databanks and for which source texts and translations require regular updates or revisions, it is obvious that TM can reuse stored data and thus increase efficiency. As Bert Esselink suggests in reference to localisation, this helps “reduce time to market of localized versions drastically”. (Esselink 2003: 80)

It is not surprising then that TMs have been introduced in many workplaces, spanning from international organisations to small-sized professional translation agencies, since the 1990s. Once all of the 3,000 translators at the European Union organisations could access shared databases, these have constituted common and shared pools of reference.

In order to exemplify the usefulness of such networking, we posit that Translator A has completed the English-to-Chinese translation of the sentence

The above mentioned business should be managed in accordance with the terms and regulations concerned

and that this translation is immediately stored in the databank. Translator B is translating the sentence

The business should be managed following the terms and regulations concerned

and translator C is preparing to translate the sentence

The above mentioned affairs should be managed in accordance with the terms and regulations concerned.

Instead of laboriously working out their own renditions, translators B and C can retrieve translator A’s translation from the databank. What is more, if the context of their respective translations demand so, B and C can modify A’s translation and these renditions will then also be stored in the databank.

TM databanks also ensure that, e.g., specialist terminology is consistent with different translators. Such terminological consistency streamlines the final editing. Today, no large translation project completed by a team of translators can be done without TM tools.

The appropriate text for TM use

Since TM is suited particularly for the automatic recycling of previously translated terms and sentences, it is most appropriate for texts that repeat the same terms, phrases, and sentences. This is true in particular for technical texts
Specialised domains of technical texts include software (hardware), the automobile and pharmaceutical industries, medicine, and electronic software and hardware.

Various studies indicate that the percentage of repetition in these domains ranges from 20% to 70%. Thus, for example, Andrei Gerasimov reports that in his work for Volvo Cars International, about 60% of the sentences in a brochure on the Volvo S/V80 were sentences that he had already translated for the Volvo S/V40, S/V60, and S/V70 brochures (www1), which means that a translator with access to a TM system does not have to repeat the translation of at least 60% of the new brochure.

**Distinctive functions of a TM system**

**Automated retrieval of translation suggestions**

For each new document to be translated, a TM system can automatically search and retrieve segments already stored in the translation memory. This illustrates the ability of a TM system to “match the sentences to be translated against the database”. (Somers 2003c: 37) The translators determine whether they want to be presented with only exact matches or with so-called “fuzzy matches” in which a specified percentage of the words and phrases match (and where few professional users working between Indo-European languages go below a 90% match). When there is an exact match, that is, between the source text segment and a corresponding translation segment, the TM system offers to paste it directly into the target text so that translators have the option of rejecting it, modifying it, or using it as it stands. “Fuzzy matches” are presented to translators for their reference or editing. The TM system highlights the differences in the fuzzy matches so that users can easily identify and edit their tentative translations. For example, if sentence (a) below is to be translated, the TM system would suggest sentence (b) as the fuzzy match source-language sentence retrieved from the TM databank, and sentence (c) as the translation of the retrieved sentence.

(a) Movie length is limited by capacity of memory card or 5, 15, 30 second clips.
(b) Movie length is limited by capacity of external memory card.
(c) 影片長度受限於外接記憶卡的容量.

**Terminology management**

A TM system enables translators to create term bases, add or delete terms, searches, and filters. During the interactive translation initiated by a TM system, users can search local or remote term bases and then insert the required translation of terms in the document they are translating. In a multi-user environment, the term base server system allows users to connect to remote or on-line term bases.

**Repetition analysis and time saving**

A TM system performs repetition analyses to calculate how many 100% matches, fuzzy-matches, and internal repetitions are encountered by means of a comparison of a source-text document with a translation memory. Such analy-
ses offer translators, agencies, as well as project managers some reference for estimates of costs, workload, and translation deadlines.

The use of a TM system can save 30% to 60% of the total time and cost of translation, depending on the type of text and the memory database. Lynn E. Webb (1998-1999) summarises the savings of companies, translation agencies, and companies with in-house translation departments after they employed TM systems, as follows:

**Table 1**

<table>
<thead>
<tr>
<th>Translation agency savings and profits after using TM tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without TM</td>
</tr>
<tr>
<td>Number of words in project</td>
</tr>
<tr>
<td>Turnaround, in days</td>
</tr>
<tr>
<td>Total translation cost (Per project)</td>
</tr>
<tr>
<td>Cost per word</td>
</tr>
<tr>
<td>Total profit</td>
</tr>
</tbody>
</table>

(Webb 1998-1999: 35)

The Table indicates that the savings in terms of time and money for each translation project ranges from USD 3,360 to USD 5,215, and that the savings on individual words ranges from USD 0.08 to USD 0.12, at one of the companies studied by Webb (1998-1999).

**Improvement in translation quality**

The use of the same terminological databank leads to terminological consistency when multiple translators are collaborating on a large translation project. And, terminological consistency is one factor that affects translation quality.

**Acceleration of localisation**

TM systems can be used for localisation when a new product with instructions written in several foreign languages is released. TM systems, such as Trados, automatically convert dates, numbers, and measurements based on the user settings across countries. Thus, TM use shortens the turnaround time of localisation and the time to market.

**Pedagogical foci in TM training**

TM technology has been introduced at several universities, such as, the University of Leeds, the Monterey Institute of International Studies, and the National Kaohsiung First University of Science and Technology in Taiwan (see Aula. int 2005). In addition to a theoretical introduction to the concept and purpose of TM development, TM translator training might incorporate components such as the following.

**Translation memory buildup**

Teachers can introduce students to at least three ways of creating a TM databank. The first method is the simplest one that TM systems support: each
translated sentence is automatically added and saved. When the same (or the next) translator works on a text that has been translated before, the latter can be loaded into the database and the newly translated texts are then added. This method is open to any user. The following example is illustrated with a (low) 65% match:

**Figure 1**

*Trados window showing 65% match with previously translated sentences*

In the figure, the Trados window shows the partial match of the sentence to be translated compared with the previously translated sentence.

The second method exploits the alignment function, which makes it possible to match source-text and target-text segments that are “sentences or other more or less easily distinguishable text portions, such as titles” (Somers 2003c: 34) into bilingual pairs, segment by segment. Some language pairs, e.g., English and Chinese, are so different in their syntactical structures and grammar that it is very difficult to align them accurately. However, some TM systems allow users to redo or modify incorrect alignments.

The third method is to import compiled translation memory databanks into the trainee’s workstation. Despite the fact that there are many different TM products, most use a common interchange format, which means that “TM databases using one product can be imported into another.” (Somers 2003c: 34)

### The use of TM-driven concordance

Although the results would be more reliable if training institutions could use huge and varied databanks, instructors may rarely have access to more than the bilingual concordances embedded in TM systems. These, however, often allow teachers to carry out tasks in class along the lines of the following.

#### Evaluating different translators’ choice of words

Concordance can be used for the identification of different word choices made by translators. For example, we developed three small TM databanks of the American novel *Gone with the Wind* using three English-to-Chinese translations. We then used the Concordance function to search for the word ‘defeat’ and obtained three translation sentences [(b), (c), and (d)] from the three concordance outputs. [(a) is the source language sentence.]

(a) The source text: “The army, driven back into Virginia, went into winter quarters on the Rapidan - a tired, depleted army since the defeat at Gettysburg - and as the Christmas season approached, Ashley came home
When students compared these translations, they found that the translator of (d) had the most concise style, because he/she translated ‘defeat’ as ‘大敗’ (two Chinese characters) instead of (c)’s four-character ‘吃了敗仗’, and (b)’s nine-character ‘吃了那麼一個大敗仗’. On the other hand, some Chinese readers may favour the second translation ‘吃了敗仗’ (c), because its style is more vivid.

Identification of Chinese courtesy terms of address

We have also used the TM-driven Concordance to identify the predominance of courteous expressions in Chinese translations of product instructions and users’ guidelines. For example, we developed a TM databank that consists of bilingual excerpts from a guideline for flight security, which is a technical, informative text-type. We searched for sentences that contained the word ‘please’. We found 14 sentences with ‘請’ in the Chinese translation, whereas the English source sentences only used ‘please’ two times and ‘ask for’ one time. A few instances are listed on the next page, the underlined spaces representing the absence of ‘please’.
| Table 2  |
|-----------------|-----------------|
| **The concordance output of the use of 请/please in English and Chinese in guidelines** |
| 1 | 疏散後，請協助其他乘客沿著緊急升降梯下降。  
After evacuating, please assist other passengers coming down the emergency slide.  
Created on: 2005/9/18, 15:29  
Created by: ALIGN!  
Changed on: |
| 2 | 請協助其他旅客直接下機。  
Please direct other passengers to evacuate away from the aircraft immediately.  
Created on: 2005/9/18, 15:29  
Created by: ALIGN!  
Changed on: |
| 3 | 請將行李放置在頭頂上的行李放置艙內或是座位下方。  
Please stow the baggage you bring with you in the overhead bins or under the seat in front of you.  
Created on: 2005/9/18, 15:29  
Created by: ALIGN!  
Changed on: |
| 4 | 為確保閣下與其他旅客的飛行安全與順利，請與我們配合，謹慎遵守基本法則和指引。  
To ensure a safe and pleasant journey for you and your fellow passengers, we ask for your cooperation in carefully observing some basic regulations and guidelines.  
Created on: 2005/9/18, 15:29  
Created by: ALIGN!  
Changed on: |

This illustrates cultural differences to students in a very concrete fashion. In the case in hand, the difference shows that Chinese business culture includes courteous terms of address to customers more often than does the American source text.

**Term-base buildup**

Teachers can also use the TM system to teach students to use the terminology management systems to develop term bases in different subjects. Term bases mainly save words and phrases, whereas databanks save texts. Sophisticated term bases allow translators to design and store a variety of entries, including a lexical item's grammatical features and sample sentences, and extralinguistic information about a lexical item, such as definitions and background data. In addition, trainee translators may learn the principles behind such systems for compiling term bases for customers regarding the same subject. This may also be used for demonstrating to students how different companies use the same specialist terms in different ways.

**Cost-effectiveness assessment**

Furthermore, TM systems can be used to show prospective translators how to calculate the cost of their work and how much they should charge for texts in which there is much repetition. Statistical techniques supported by the TM system can analyse the similarity of new source texts to the corpus already stored. Thus, TM systems allow users to establish how many translated segments correspond to the source-text segments. This, in turn, permits translators to assess immediately the amount of repetition of terms in the documents. This is obviously much easier in translations between Indo-European languages, such as English and French, for which words and sentence structures are more alike.
Web page translation with TM as an aid

Students can also be taught how to use a TM system to translate web pages. Some TM systems, e.g., Trados TagEditor, allow users to translate web pages without typesetting, because the TM tool separates the layout and formatting information from the translatable text and protects the original graphs of web pages. The interface distinguishes between translatable text and non-translatable items, which means that the structure and formatting of the page remains the same during the interactive translation. In addition, translators can immediately view the translated portions on the web page.

Problems with TM

In the above, I have outlined only a few methods that can be used in order to familiarise and train translator trainees in TMs. However, this should also be balanced by pointing out the weaknesses of the systems.

Whereas a TM databank is useful for recycling and thus lightens the workload of translators, the retrieval does not distinguish between adequate and erroneous translations that are saved in a TM databank. This means that the individual user will have to critically review any previously translated segment that is displayed for approval or rejection – and may have to reject otherwise adequate renditions that are merely inadequate in the new context. It should be stressed to students that, in principle if not in practice, one method of improving the results stored in databanks is to have all data reviewed for accuracy and appropriateness before it is stored. From a long-term perspective, revisers might check the translations stored in the databanks, in some kind of cooperation with engineers. One possible short cut is to align source-language and target-language texts and use the aligned segments for developing a TM databank. But this should only be done provided the target-language segments are checked for adequacy and accuracy. However, at the present stage of development, I do not believe we can get any further than this.

Advantages of TM training

Learning TM strengthens students’ confidence about their future careers in the market place, with translation agencies, international globalised companies, and localisation industries. The pedagogical advantages of TM training at universities are summarised in Table 3 and are adapted from Kathleen Graves (2003: 83).
The pedagogical advantages of TM use.

<table>
<thead>
<tr>
<th>Areas</th>
<th>Pedagogical Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>1. The important financial advantages of TM tools.</td>
</tr>
<tr>
<td></td>
<td>2. The lexical, syntactic, and stylistic differences between different text-types, the individual choices made by translators, and appropriate translation models by means of TM-driven Concordance.</td>
</tr>
<tr>
<td></td>
<td>3. Real-time data recycling and shared information.</td>
</tr>
<tr>
<td>Awareness</td>
<td>1. The use of TM tools to translate technical texts in an efficient and cost-effective way.</td>
</tr>
<tr>
<td></td>
<td>2. The use of TM in translation.</td>
</tr>
<tr>
<td></td>
<td>3. The ability of TM tools to manage the budget, the process, and to assure quality and terminological consistency in translation projects.</td>
</tr>
<tr>
<td>Skills</td>
<td>1. How to establish TM databanks and customer-tailored term bases.</td>
</tr>
<tr>
<td></td>
<td>2. How to use TM tools to translate web pages.</td>
</tr>
<tr>
<td></td>
<td>3. How to use exact or fuzzy matching functions in TM systems to access previously translated sentences.</td>
</tr>
<tr>
<td></td>
<td>4. How to use TM-driven concordances.</td>
</tr>
<tr>
<td></td>
<td>5. How to use fuzzy or exact searches to retrieve terms from term bases.</td>
</tr>
<tr>
<td></td>
<td>6. How to use the alignment tool to compile parallel bilingual corpora.</td>
</tr>
<tr>
<td>Attitude</td>
<td>1. Increase student confidence in translation.</td>
</tr>
<tr>
<td></td>
<td>2. Improve the quality of students’ translations and develop the students’ self-confidence.</td>
</tr>
</tbody>
</table>

The use of TM-driven Concordance enables the trainee translators to learn about lexical and syntactic variations in several subject fields, stylistic differences in textual types, and appropriate translation models. This knowledge increases the competence of student translators. In addition, an awareness of the relevance of TM use in the market place prevents students from being isolated from the professional world and enables them to face the market-place with more self-confidence. The use of TM tools to translate is a valuable asset to trainee translators, and learning terminology management and TM databank building enables them to master skills that they will need for their professional work.

Conclusion

TM systems are used extensively in professional translation and translator training institutions have to introduce students to them. Disregarding the problem that in every country TM will meet with specific problems in terms of the size of databanks available and specific translation problems depending on the language pairs involved, teaching staffs, at present, are hampered by the small size of the corpora available for training purposes.

It would be a great advantage if specialised and professional TM databanks, e.g., those used by globalised companies for the translations of their equipment and manuals, were accessible – or at least cheap - to training institutions. This would be a great advantage for all would-be translators, as they could use such material to become familiar with the challenges and work procedures in the daily lives of professional translators in today’s world. One could imagine that this might also be beneficial to those who offered such access to translation schools,
since this would minimise the in-house training that they are otherwise obliged to give to newcomers.

No matter whether such chances of hands-on experience are offered or not, translation training at universities must keep up with the rapid changes in the industry. The integration of TM components into current translation education is imperative; the “resistance to change can represent a major impasse”. (O’Brien 1998: 121) TM training not only benefits future translators, but also enables industry to recruit qualified translators.

Works cited

Websites