The effect of translation memory databases on productivity

MASARU YAMADA

Rikkyo University, Japan

Although research suggests the use of a TM (translation memory) can lead to an increase of 10% to 70%, any actual productivity increase must depends on the TM content. If the target renditions included in the TM database exhibit more free characteristics, this may adversely affect the translator's productivity. This paper examines how productivity is affected by different kinds of TM databases. A pilot experiment was undertaken to investigate the impact of two different versions of a TM database – free vs. literal TMs. All participants translated the same source text but used different TMs. The results show that in the higher fuzzy-match categories, translators using the less literal TM did not gain as much speed as was the case when using a more literal TM.

Key words: translation memory, productivity, localization.

Introduction

The role of the technical translator has changed as a direct result of translation memory (TM) technology. Translators are no longer focused on translating texts from scratch, but on recycling previously translated texts: the essence of the technology is "text re-use" (García 2009). As TM databases have changed into network-based systems, the translated texts are no longer locally managed by translators but rather centralized by translation bureaus or Language Service Providers (LSPs). In addition, due to the use of sections within localization projects, independent players such as quality-assurance checkers and client reviewers make extensive changes to the translated texts. By the time the texts are finalized, translators have lost control over their own translations.

Under these circumstances, translators using a TM provided by the LSP must deal with the imposed segments they have not generated themselves. This means more time checking and editing, thus adversely affecting productivity. Although previous studies have shown that the use of TM can lead to a 10 to 70% increase (Bowker 2005, Dragsted 2004, O'Brien 1998, Somers 2003), the actual productivity gain must depend on the TM content.

Target renditions exhibit more *free* characteristics when a product is adapted, customized, or highly localized, possibly due to reviewers' extensive modifications in the course of the localization process. If these texts are put into the TM database, it may have an impact on the translator's productivity. This present study will therefore examine how productivity is affected when different renditions are put into the TM database.

Relevant literature on TM productivity

The desire to increase productivity is one of the main reasons for using a TM, and this aspect has been investigated in empirical studies. According to O'Brien's experiment (1998: 119), anything from 10% to nearly 70% can be leveraged from the TM. Somers (2003: 42) states that "while on occasion a TM product might result in a 60% productivity increase […], 30% may be a more reasonable average expectation." Dragsted (2004: 210) has indicated that the average increase was 16 % for students and 2% for professionals. Bowker's pilot study (2005: 17)

From *Translation Research Projects 3*, ed. Anthony Pym, Tarragona: Intercultural Studies Group, 2011. pp. 63-73.
http://isg.urv.es/publicity/isg/publications/trp_3_2011/index.htm

shows that translators without using a TM could not finish a 387-word translation within the 40-minitue time frame, while participants using a TM completed the task.

Whatever the exact increase might be, it has been established that the use of TM increases productivity. However, none of these studies has taken different types of TM content into consideration. In Bowker's experiment (2005), two different versions of the TM (original TM vs. error-included TM) were prepared, but her main objective was to compare the quality of products. No difference in productivity was recorded. I therefore decided to undertake a pilot study with the aim of investigating the impact of different TM databases on translation productivity.

Pilot study

The pilot study was carried in March 2009 using eight student translators as participants. They came from the Translation and Interpretation program of the Monterey Institute of International Studies in the United States.

The reason for using student translators rather than professional translators was mainly convenience. As a visiting scholar at the Monterey Institute during the 2008 academic year, I had access to a group of 8 students who volunteered to join the pilot experiment. The students' ages ranged from the early 20s to the early 30s. All students were at Masters level: two subjects were second-year students and six were first-year students. These students had diverse backgrounds: some came directly after finishing their undergraduate degrees, others had a few years of work experience, and one student had professional translating experience. Their language background also varied: half of the students were Japanese native speakers while the rest were English native speakers. Despite this diversity, all the students were highly proficient in both English and Japanese, and we assume their translation skills to be at "near professional" level. This assumption is not without precedent: when Tirkkonen-Condit (1991) compares the translation behavior of professional and non-professional translators, the second-year students represent the "professional translators", and Bowker (2005) uses Masters students for her TM error propagation analysis.

Our subjects' TM skill level was, however, not at a proficient level. They were only novice or moderate users of TM. Some of them had completed a course on Translation Memories at the institute, others had not. To make sure that they were comfortable with the tool, I provided a training session and exercise lessons prior to the experiment. At the end of the trainings, I did not find major technical difficulties, nor did I see any when observing the actual screen recordings of their behaviors. Nevertheless, as Ribas (2007) points out, the translators' relative computer literacy may affect their translation performance in regard to the quality. This factor may thus be seen as a limitation of this experiment.

For the experiment, we prepared two different types of TM database for the same source text. All participants translated the same source text, but they used different TM databases. The first type was free-translation content (hereafter referred to as TM-F), which was based on authentic material used in an actual localization project. The pre-translation entailed a number of additions and deletions. The other type was more of literal-translation content (TM-L), for which I made modifications on the basis of the TM-F database. The TM-L content was not necessarily a *literal* translation: it was at the level of the current translation norm in the localization industry. Some examples of the differences between TM-F and TM-L are shown in Table 1.

In TM-F, for instance, the source word *application* in the first example was rendered as *program* in the target Japanese text. In the second example, the source word *current* was not translated. In the third example, the source phrase *basing on the features and tasks of your computer* was eliminated in the target text. Although some may claim that these features may be close to mistranslations in terms of formal correspondence, the TM-F content was, after all, authentic and was accepted in the market.

It is important to note the "match rate" of the TM. In principle, the TM functions as a database that stores previously translated content as paired source and target segments, and retrieves the translation segment for "recycling". The similarity level is indicated by the match rate, based on the syntactical structure of the source text. For instance, if a new sentence is said to be an "80 percent match" (fuzzy match) of an existing sentence, this represents the high resemblance and only a few corrections are required by the translator in the target text. If the new sentence is a "100 percent match" (exact match), this means that there is a high probability of no change at all in the target text. The source text used in our experiment was identical for both groups; therefore, the match rate for each sentence (or segment) was the same. However, because different target renditions were prepared for each type of the TM database, I expected different editing efforts to be required by translators.

Table 1: Sample sentences from TM-F and TM-L, with back-translation into English

ST	TM-F (Free)	TM-L (Literal)
Application Configuration Wizard	プログラム設定ウィザード 'Program Configuration Wizard'	アプリケーション設定ウィザード 'Application Configuration Wizard'
Obtaining information on the current status of components and tasks and statistics on them	コンポーネントとタスクのステータスと統計情報の取得 'The status of component and task, and statistics information –ACC obtaining. '	コンポーネントとタスクの現在のステータスおよび統計の情報の取得 'Information on component's and task's current status, and statistics –ACC obtaining.'
Its task is to help you configure the initial settings of the application basing on the features and tasks of your computer.	設定ウィザードはコンピュータへ適切な保護設定を行うお手伝いをします。 'Configuration Wizard helps to configure the protection settings properly to your computer'	アプリケーション設定ウィザードは、コンピュータの特徴やタスク基づき、アプリケーションの初期設定を行うお手伝いをします。 'Application Wizard helps to configure the initial settings of the application, basing on the features and tasks of your computer'

Hypothesis

Because the use of a TM forces the translator's cognitive segmentation into smaller linguistic chunks (cf. Dragsted 2004), the translator, if using TM-L, should be able to easily identify one-to-one correspondences between the source and pre-translated texts. I therefore hypothesized that TM-L would correlate with faster translation speeds than TM-F.

General experiment design

Translators were requested to translate a text of about 500 words from English into Japanese using the TM. The text was from an anti-virus software manual, a topic normally encountered in the localization industry. The translators were put into two sub-groups: TM-L and TM-F. All of them were asked to translate the same source text, using either the TM-L or the TM-F database. They were not notified of which TM database they would be using.

The experimental set-up was designed to reflect the translators' natural work environment. No time restriction was given for the task. The subjects were allowed to use their own computers and were permitted access to their usual reference materials, including the Internet, in addition to the TM provided.

All of the subjects' operations on their PC screens were recorded using BB Flashback. It recorded searches of electronic resources, cursor movements, clicks, and keystrokes as well as the translations. The recorded data were analyzed to trace the history of each translator's activities. BB Flashback was installed on each subject's computer and worked in the background so that it did not affect the subject's natural work environment.

The TM tool used for this experiment was SDL Trados 2007, the most common tool of this kind in Japan and the market leader in the world localization industry. Nearly 80% of translation service providers in Japan that use some kind of translation memories adopt SDL Trados (Japan

Translation Federation, 2008). According to Lagoudaki (2006: 21), the TM most used worldwide is also SDL Trados.

Because this experiment was a pilot study, the sample size was small. A total of 8 students was obviously not a high number, especially to assess the statistical significance of quantitative analysis.

Results

The results of the pilot study are shown in Table 2. Contrary to my prediction, the overall difference in speed between TM-F and TM-L was not highly significant. The average production time shows TM-F 1:04.22 vs. TM-L 1:05.44, meaning that the overall production speed with TM-F was actually marginally faster.

This may partly be attributed to two factors. The first factor was that subject A1 in TM-F group recorded an exceptionally high productivity gain so that this translator’s speed contributed greatly to the overall average speed in the TM-F group. The WPM (words per minute) of A1 was 11.29, compared with a median of 7.48. The second cause was subject D1 in TM-L group, who was the slowest translator of all, achieving only 6.88 WPM. The data without these two translators would make the result TM-F 7.21 WPM and TM-L 8.13 WPM, which is in line with our expectations. Nevertheless, such intentional manipulation cannot be an option here unless there is a legitimate reason to do so.

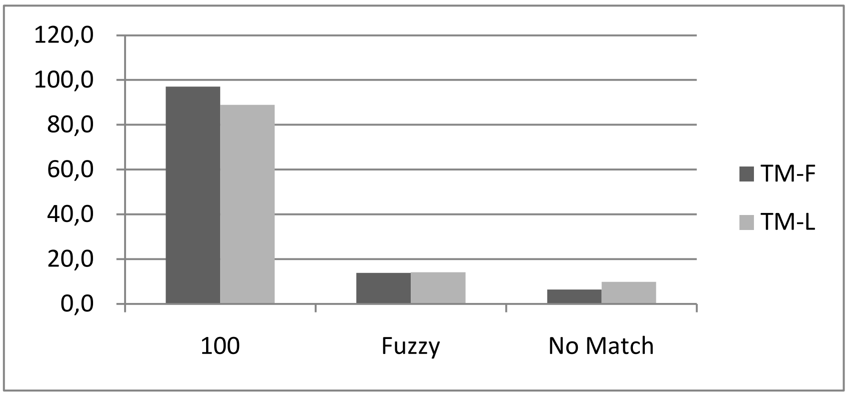
Table 2: Translation productivity: TM-F vs. TM-L

		Total Time	Mean Time	WPM	Mean WPM
TM-F	A1	0:45:10	1:04:22	11.29	8.23
	A2	1:12:42		7.02	
	B1	1:11:46		7.11	
	B2	1:07:48		7.52	
TM-L	C1	1:00:06	1:05:44	8.49	7.82
	D1	1:14:05		6.88	
	D2	1:08:31		7.44	
	D3	1:00:15		8.46	

Production speed for100% matches

Although the overall average data did not indicate any clear advantage of TM-L over TM-F, we found some differences in translation speed by subdividing detailed data into match-rate categories. Figure 1 illustrates the WPM for different match rates and a comparison between the two types of TM databases.

Figure 1: Speed comparison: 100%, fuzzy, and no match (WPM)



Comparing the speed at the 100% category, we see that TM-F was still faster than TM-L. Again, this was not in line with my prediction. However, detailed observation of the screen-recording data shows that this difference was mostly due to the translators' technical skill and how they handled the 100% matches (EM=exact matches). Some translators were familiar with short-cut key commands to semi-automatically skip the EM segments. Short-cut features should reduce or eliminate any time spent on the EM segments, and translators who took advantage of these functions normally paid little attention to these segments. If they were more cautious and not in the habit of using the short-cut key features, they took some time to check EM segments. Subject A1 in the TM-F group, the fastest translator of all, made the most use of this feature. That is probably why the overall processing speed with TM-F in the EM category was higher, and it had nothing to do with the influence of the content included in the TM database.

We have not yet seen any significant difference between the databases in other matched categories, other than the fact that translation speed was higher for the EM segments.

Fuzzy-match speed in detail

In order to analyze more closely the effect of two different types of TM, I measured each individual translator's speed for every 10% of the fuzzy-match ranges. Table 3 gives the mean speed of individul translator sorted by the match rate, and Figures 2 and 3 show their behavioral patterns.

Table 3: Individual translator's WPM at detail match rate

	TM-F				TM-L			
	A1	A2	B1	B2	C1	D1	D2	D3
99-90	23.49	15.01	5.79	9.54	18.21	8.68	30.00	23.29
89-80	34.41	11.22	11.33	11.97	15.78	7.85	12.77	19.15
79-70	19.66	15.31	11.12	16.51	18.07	7.79	15.26	25.22
below 70	12.49	8.05	11.65	11.93	11.75	8.89	7.24	13.52
0.00	4.45	7.25	8.40	5.52	12.34	13.33	5.72	7.78

Figure 2: WPM change at detailed match rate for TM-F

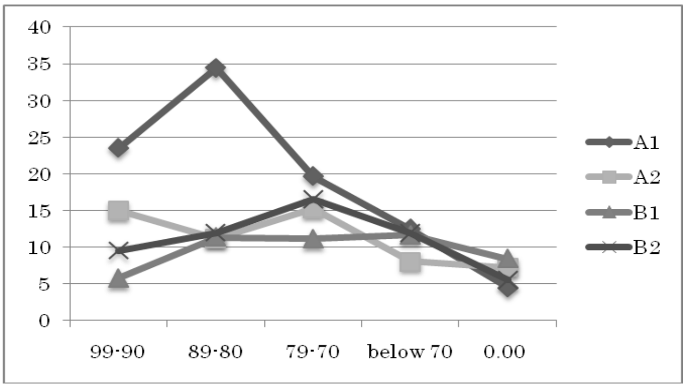
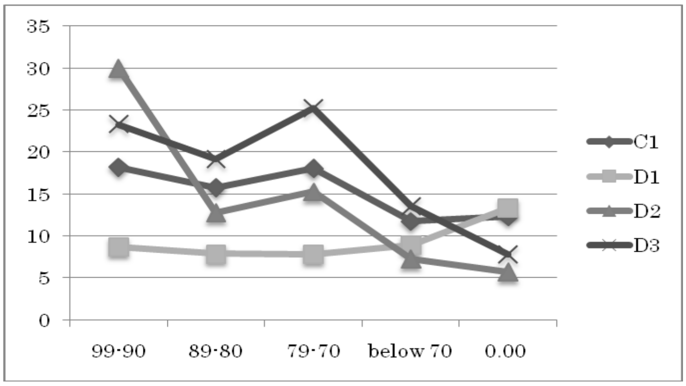


Figure 3: WPM change at detailed match rate for TM-L



In the graph for TM-F (Figure 2) we see that the production speed with subject A1 was much higher than the rest of the translators. A1's speed began at 23.49 WPM in the 99-90% match range, increased to 34.41 WPM at 89-80% category, then constantly dropped almost proportionally to the decreasing match rate, until it reached no-match with 4.45 WPM. A1's dynamic range (the difference between the highest and the lowest WPM) was approximately 30 WPM.

Conversely, subjects A2, B1, and B2 exhibited much narrower ranges of speed change through all fuzzy matches. Their overall trend appeared "flatter" than A1's. For instance, A2's fastest peak came around 79-70% with 15.31 WPM, but the slowest speed at no-match showed 7.25 WPM. A2's range difference was only about 8 WPM; the trend curve did not show as much dynamic movement as A1's. A similar tendency was also observed with B1 and B2.

This result suggests that the subjects in the TM-F group did not gain the same benefits from each fuzzy match segment proposed by the TM. It is normally expected that translation speed at a high match rate is faster than with lower matches, but this prediction was not applicable to the case of TM-F. The only exception was noticed with translator A1, whose speed curve changed almost proportionally to the match rate. However, even in the case of A1, the processing speed at the 99-90% category was significantly slow. This implies that free TM content may have reduced the translator's segmentation recognition speed in higher match categories.

In contrast, the overall trend with TM-L (Figure 3) showed more consistency and a wider range of speed leverage. The production speed increased almost in step with the increasing match rate. An exception was found with subject D1, whose curve contradictedly went up as the match rate decreased. D1's result was something that should not be observed in professional use of a TM. Perhaps D1 did not follow any proposed translations presented by the TM. This was also evident from this translator's post-experiment comments, which I requested participants to submit after the experiment. Subject D1 stated "TM was of no use for me".

Other than this exceptional case, however, translators C1, D2, and D3 indicated very similar characteristics: the speed at 99-90% match was the highest or near highest, and then dcreased toward no match, almost in proportion to the fuzzy-match rate.

The dynamic range in the case of TM-L was also wider. Translator D2 gave 30.00 WPM at 99-90% match category and 5.72 WPM at no match. The difference was over 24 WPM. D3's range was also over 17 WPM, although C1's trend fell within the range of a little over 6 WPM. Nevertheless, C1 still recorded a higher translation speed than the average for TM-F.

From these observations, we can conclude that the different types of the TM database seemed to have been affecting a productivity increase in fuzzy matches. Especially in the higher fuzzy-match categories, translators using TM-F did not gain as much productivity leverage as they did in the TM-L group. Hence, the overall dynamic range in TM-F was narrower than that in TM-F.

The overall differences between TM-F and TM-F are shown in Figure 4, where subject A1 has been excluded from caclucation. As mentioned above, subject A1's processing speeds for fuzzy-match categories were much higher than the remainder of the participants in the same group. Further investigation of this translator's processing is needed. Given this, however, Figure 4 still provides us with an overview of the productivity difference between TM-F and TM-L.

Figure 4: Average speed for fuzzy/no-match categories

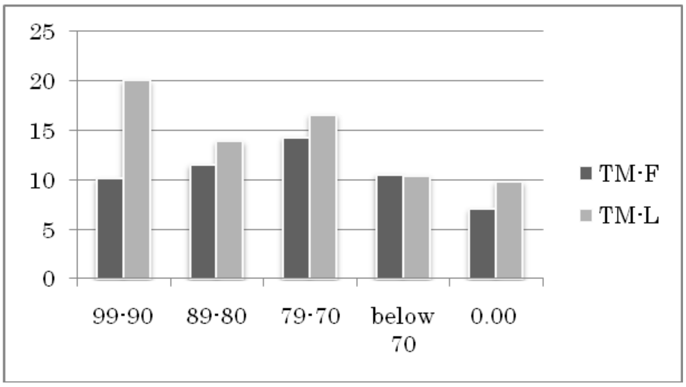


Figure 3 shows that the production speed with TM-L is equal to or higher than with TM-F in all categories. Especially in the 99-90% match category, the speed for TM-F was significantly lower, at approximately half that of TM-L.

Concluding remarks

In sum, the TM-L production speed for fuzzy match segments exhibited faster WPM than did work with TM-F. That is, if a TM content is highly customized or localized as in TM-F, it may reduce productivity.

The reason for the reduced speed has not been analyzed in this paper. It may be related to the translator's focus range or translation unit. Under the TM-F condition, where the target renditions contained many deletions and additions, translators require more effort to recognize one-to-one correspondence between the source and the target text. Because the use of TM restricts the translator's segmentation range to a sentence or smaller unit, chunk-level recognition would be more difficult when using the *free* translation content (TM-F). This in turn may imply that translators using a TM are actually working on a sub-segment unit rather than an entire sentence or the discourse level.

In terms of efficiency in the actual practice of localization, if *free* segments are put into the TM database, there is a chance that this may adversely affect the translator's performance. The freer the renditions in the TM, the less effective the localizability may be. In order to improve the efficiency, it is necessary to review both the project workflow and the TM database, because the TM databases, like translators themselves, are no longer isolated from the project: they are part of the localization team.

References

Bowker, Lynne. 2005. "Productivity vs Quality: A pilot study on the impact of translation memory systems". *Localisation Focus* 4(1): 13-20.

Dragsted, Barbara. 2004. *Segmentation in Translation and Translation Memory Systems: An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process.* Doctoral dissertation, Copenhagen Business School: Samfundslitteratur.

García, Ignacio. 2009. "Beyond Translation Memory: Computers and the professional translator". *The Journal of Specialised Translation* 12: 199-214.

Guerberof, Ana. 2009. "Productivity and quality in the post-editing of outputs from translation memories and machine translation". *Localisation Focus* 7(1): 11-21.

Japan Translation Federation. 2009. *Honyaku Hakusho 2009* [Translation White Paper 2009]. Tokyo: Japan Translation Federation.

Lagoudaki, Elina. 2006. *Translation Memories Survey 2006: Users' perceptions around TM use*. http://www.atril.com/docs/tmsurvey.pdf. Visited May 2010.

O'Brien, Sharon. 1998. "Practical experience of computer-aided translation tools in the localization industry". Lynne Bowker, Michael Cronin, Dorothy Kenny and J. Pearson (eds) *Unity in Diversity?: Current Trends in Translation Studies*. Manchester: St. Jerome. 115-22.

Ribas, Carlota. 2007. *Translation Memories as vehicles for error propagation. A pilot study*. Minor Dissertation. Tarragona: Intercultural Studies Group, Universitat Rovira i Virgili.

Somers, Harold. 2003. "Translation memory systems". Harold Somers (ed.) *Computers and Translation: A Translator's Guide*. Amsterdam and Philadelphia: Benjamins. 31-47.

Tirkkonen-Condit, Sonja. 1990. "Professional vs. Non-Professional Translation: A Think-Aloud Protocol Study". M.A.K. Halliday, J. Gibbons, and H. Nicholas (eds) *Learning, Keeping and Using Language*. Amsterdam: John Benjamins. 381-394.