Comparing forum data post-editing performance using translation memory and machine translation output: a pilot study

Morado Vazquez, Lucia; Rodriguez Vazquez, Silvia; Bouillon, Pierrette

Comparing forum data post-editing performance using translation memory and machine translation output: a pilot study

Lucía Morado Vázquez† **Silvia Rodríguez Vázquez†** **Pierrette Bouillon**
Multilingual Information Processing Department (TIM/ISSCO)
Faculty of Translation and Interpreting (FTI) - University of Geneva
† Cod.eX Research Group
{lucia.morado; silvia.rodriguez; pierrette.bouillon}@unige.ch

Abstract

In this paper we present the results from a pilot study undertaken with translation students to compare community forum content post-editing performance based on suggestions from different translation systems. Output from both Translation Memory (TM) and Statistical Machine Translation (SMT) was presented to participants in the ACCEPT online post-editing environment, where they needed to perform a translation task with and without translation proposals. Observed data showed that post-edited MT output obtained higher results on each of the variables measured: the amount of time needed to complete the task, the participants' keystroke movements and the quality of the resulting translations.

1 Introduction and Background

The translation technologies landscape has been dramatically influenced by the current localisation industry demands. Quicker turnaround times are required for high translation volume rates at a low cost, but quality expectations remain the same (Van Genabith, 2012). Although Translation Memories (TMs) initially appeared to be a suitable solution to challenge those requirements, Machine Translation (MT) has emerged as an efficient alternative, both for translators and translation end users (O'Brien, 2012; Plitt and Masselot, 2010). Moreover, recent studies have suggested that a combination of both systems could result in a significant productivity gain (Guerberof, 2009; He, 2011). Nevertheless, efficiency in the use of such solutions may also vary depending on several contextual factors. For instance, light has already been shed on source text characteristics (complexity, ambiguity, style) or TM content types (Tatsumi and Roturier, 2010; Yamada, 2011) as variables that can have implications on the post-editing results.

The work presented in this paper is part of the Automated Community Content Editing PorTal (ACCEPT) European research project [1], where adequacy receives greater importance than other factors precisely due to the source text nature. ACCEPT aims at exploring the potential of using Statistical Machine Translation (SMT), complemented by pre-edition and post-edition modules, for translating community-generated content. Information exchange through specialized Web fora is becoming increasingly popular; however, up to now, the use of MT for community forum data has not proved successful due to its nature: short and concise messages, closer to oral language rather than to written discourse, for which a fast-delivered translation is required.

One relevant step to be taken in order to achieve ACCEPT's goal is to compare SMT with other translation systems, such as Rule-Based Machine Translation (RBMT) and TMs. In this study, we intended to determine if, in the case of forum user-generated content, it is preferable to work with SMT output, TM 80-95% fuzzy-match suggestions or to translate from scratch in the English-French language combination. We also aimed at observing participants' satisfaction regarding the post-editing environment and the type of task required. The experiment results showed that post-editing performance is higher when working with text of SMT provenance, and that participants have a positive attitude towards post-editing as a translation-related activity.

[1] http://www.accept.unige.ch/index.html

2 Experiment Design

Research efforts have been devoted to MT and TM fields with different purposes. In particular, they have recently focused on the combination of both paradigms for improving MT performance (quality-oriented studies) and estimating post-editing efforts to evaluate productivity rates or working environment conditions (translator-oriented studies). Instead of placing emphasis on MT and TM integration, such as in (He et al, 2010; Tatsumi and Roturier, 2010), our work focuses on the appropriateness of each system's set of proposals for a specific text genre. In this regard, we followed a similar approach to Guerberof (2009a), where the post-editing effort is measured on MT-based and TM-based translation proposals, considering post-editing speed, quality and post-editors' experience. In a later related study by the same author (2012), results illustrated that there was no significant difference between the segments' quality produced with the help of the two systems, nor between post-editing time spent on 85-94% TM fuzzy matches and the post-edition of MT suggestions. Our work differs in the variables measured, as well as in the source text context and the post-editing environment, specifically designed to be integrated in community fora. Results were obtained through an experimental methodology approach. Due to the limited available resources and time constraints, we decided to have translation students among our participants. A further study with ACCEPT's portal end users is envisaged in the near future.

A translation task was designed using real data from the forum of Symantec, one of our research partners. The twelve participants who took part in the experiment had to perform a post-editing and translation task from English into French, as well as to answer a demographic questionnaire and a task-specific questionnaire.

2.1 Environment

The experiment took place in one of the computer rooms, equipped with Windows 7, of the Faculty of Translation and Interpreting at the University of Geneva. The task was carried out using the online post-editing environment included in the ACCEPT portal.[2] Participants were allowed to use Internet and any online resources[3] that they might consider appropriate to complete the task; a glossary[4] containing a terminology list was also provided to all participants.

It is worth highlighting that this portal was initially conceived for measuring post-editing efforts, not to work with translation proposals from TM or to carry out traditional translation activities, as we did in our study. It must be also stated that translation proposals from TM are normally displayed in Computer Assisted Translation (CAT) tools, highlighting the fuzzy match to clearly indicate the difference between the TM hit and the sentence to be translated. In our case, participants were working blind: they were informed that suggestions may come from TM or MT, but the origin was not specified in any of the segments. Particularly, regarding TM origin, they did not have any information about the fuzzy match, i.e., percentage of text leveraged or differences found between the hit and the current segment. Research on the impact of using these two different working environments (online post-editing portal vs. CAT tool) has already been initiated by Teixeira (2011). Repeating the same study in a TM-based software scenario (where differences between fuzzy matches and the current segment could be highlighted) would help us to extract more conclusive results about the differences between TM and MT output.

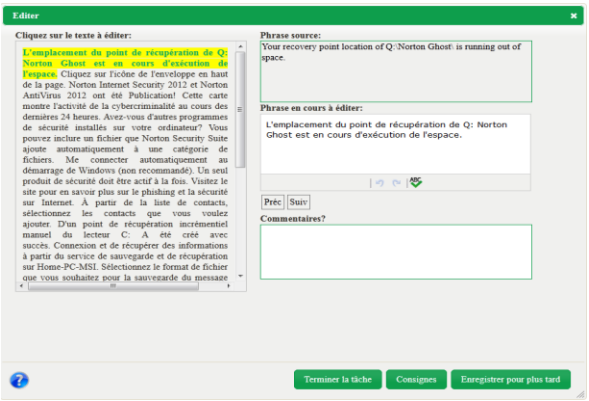

Figure 1. ACCEPT Post-editing portal

The portal supports JSON format files. Two different template files were created for the two groups, including the source text with its corresponding translation suggestions. In terms of ecological validity, we tried to replicate as many aspects as possible of a post-editing and translation

[2] http://www.accept-portal.eu/AcceptPortal/Account

[3] Video data gathered shows that no online MT system was used by the participants. Google.ch/com was the most consulted webpage (37 times), followed by terminology search-related pages: Termium© Plus (16), Microsoft Language Portal (11), Iate (8), and WordReference (4). The official Symantec Norton webpage was visited seven times.

[4] By the time of the experiment, the terminology function was not directly integrated in the ACCEPT post-editing environment.

task, i.e., we gave specific instructions to the translators, we asked them to work in a familiar and controlled environment and they were provided with resources to carry out the task, namely, a terminology data base and access to other resources via the Internet. Non-invasive recording methods were used in order not to interfere with the task.

2.2 Independent Variable

The independent variable that we manipulated in our experiment was the absence or presence of a translation proposal in the editing environment, as well as its origin. Three different instances of the independent variable were used: (a) translation proposal from a MT system; (b) translation proposal from a TM; and (c) no translation proposal offered. The document that participants had to post-edit and translate was divided into three sections, each of them containing one of the three instances aforementioned.

2.3 Dependent Variables Measured

Three dependent variables (time, quality and keystroke information) were measured. The answers to the task-specific questionnaire also provided us with an insight of the translators' thoughts on the task and their own experience.

The ACCEPT portal itself measured the time spent on modifying each segment, and provided us with keystroke information too. However, these data were related to the time spent on the post-editing window and did not take into account time spent elsewhere, i.e., when a translator tried to find something on the Internet, or when they added a comment into the corresponding window (see Figure 1). We therefore put in place an additional recording system that captured all movements on the participants' screens; the program used was BB Flashback recorder.[5]

The ACCEPT portal keeps information about editing time and keystroke movements in a XLIFF (XML Localisation Interchange File Format) file that can be extracted and analysed. These data were included in the tag count by means of user-defined values (i.e., keys and editing time).

The quality measurement was obtained using the LISA QA model (LISA, 2007); an external professional translator and reviewer followed this model to blindly examine each of the documents that were divided into three sections (each of them corresponding to the type of translation proposals that the translators had received). This resulted in three values per document and allowed us to observe different quality rates per section.

2.4 Participants

A call for participation was announced for translation students of the Master in Translation registered in the modules 'Localisation and Project Management' and 'Computer Assisted Translation'. We recruited students with the language combination English into French (and only those who were native French speakers). They were paid for their time. Twelve participants agreed to take part in the experiments (five of them were in their first year of the MA, and the other seven were in their second year). The mean age among participants was 25 years old and they were all female students. They had experience with CAT tools[6], but they had almost[7] no previous experience with post-editing. We randomly distributed the participants in two groups (A and B).

2.5 Data & Tools Used

The text participants had to translate was extracted from the real user forum of one of our research partners in the ACCEPT project. The corpus contained 566,028 segments (with more than 8.8 million words).

The TM used to find matches within the forum content was created by aligning the English and French texts from Symantec's official documentation. It included English-French parallel data coming from product manuals, marketing content, knowledge base content and website content (ACCEPT, 2012). The obtained TM was made of more than 1.6 million translation units (with more than 14.7 million words in the English variant), but had to be divided into three different TMX files in order to be processed by the TM system. The TMX files were manually checked to confirm that source and target elements were mutual translations.

We used a commercial CAT tool (SDL Trados Studio 2011[8]) to compare the data from the forum with our translation memories and try to find

[5] http://www.bbsoftware.co.uk/bbflashback.aspx

[6] A mean of 3 was obtained in the question about CAT tools from the demographic questionnaire, where 6 was "I use them in all my translation activities" and 1 "I have never tried them". A similar result ($\bar{x} = 2.91$) was obtained from the TM question.

[7] A mean of 1.5 was obtained in the question from the demographic questionnaire, where 1 was "I have never done this activity" and 6 "I work mainly doing post-editing".

[8] http://www.sdl.com/products/sdl-trados-studio/

fuzzy matches between them. In the pre-analysis done by the tool the following fuzzy matches were found: 58,665 sentences (in the 50-54% range); 39,705 (55-59%); 37,151 (60-64%); 51,047 (65-69%); 14,672 (70-74%); 13,789 (75-79%); 8,260 (80-84%); 3,131 (85-89%); 4,509 (90-94%); 10,513 (95-99) and 61,922 (100%).

From this pool of sentences with fuzzy matches, we first selected those with at least five words and a fuzzy match between 80 and 95 per cent. A review of the literature suggests that using 80-90% fuzzy matches can be broadly comparable to MT post-editing in terms of translating effort (O'Brien, 2006). Hence, this range has been adopted as well by other researchers in the domain (Guerberof, 2009b; Morado Vázquez, 2012). A set of 4,630 sentences fulfilled those requirements, but after eliminating duplicates we ended up with a corpus of 1,413 sentences. We reduced this number to 181 by choosing sentences that had between 10 and 15 words length.[9] A final sieve process was manually done to obtain our final 36 sentences; too similar sentences or sentences without verbs were eliminated.

The original 36 sentences from the English forum were isolated and randomly distributed in the JSON file. The total number of words of our text was 403, which is slightly higher than what O'Brien (2009) identifies as a manageable text size for this type of experimental research (200-300 words). The next step was to add the translation proposals. We created two different final files to each one of the groups. The text to be translated was the same for both, but the translation proposals distribution differed (see Table 1). The text was divided into three sections (S) with the same number of segments (S1 contained 139 words; S2, 130; and S3, 134). In the final file for Group A, we included translation proposals from MT in S1, translation proposals from the TM in S2, and we left S3 without translation proposals. In the final file for Group B, we included translation proposals from TM in S1, translation proposals from the MT in S2, and we left S3 again with no translation proposal. Due to the small number of participants, we decided to create only two groups with the above mentioned data distribution.

	Group A	**Group B**
1st Section	MT	TM
2nd Section	TM	MT
3rd Section	Ø	Ø

Table 1. Distribution of data by groups

The MT proposal set was obtained using one of the MT project's translation engines,[10] previously trained in earlier stages of the ACCEPT project with the same documents included in our TM, plus the WMT12 [11] releases of europarl and news-commentary (ACCEPT, 2012). This MT system received a Bleu score of 36.14 in previous evaluation on a different test-set (*ibid*).

3 Data Analysis

In this section we present the results from the data analysis. First, the three dependent variables measured are explained separately (quality, time and keystroke information). Since groups A and B obtained similar results in the three parameters observed, findings of both data sets are reported together per type-of-segment sections: translation proposal from MT (hereafter MT), translation proposal from TM (hereafter TM) and translation from scratch (hereafter Ø). Finally, we present the results from the task-specific questionnaire.

3.1 Quality

The LISA QA Model used by the reviewer is designed on an error basis. The initial "perfect quality" score is 100, but each of the errors found by the reviewer deduct from that score, depending on their nature and severity. The MT sections clearly obtained a higher quality rate ($\bar{x}$ = -14.17, *sd* = 95.01). The sections with TM output obtained lower quality results ($\bar{x}$ = -161.67, *sd* = 231.98) than MT output and translations without any proposal ($\bar{x}$ = -68.33, *sd* = 71.71).
Figures 2, 4, 5 and 6 contain boxplots showing the mean (thick horizontal line), the first and third quartiles (top and bottom edges of the box), and the outliers (circle data points). In general terms, the quality level was higher when a translation proposal from MT was suggested. On the other hand, translating from scratch obtained higher results than translating with a proposal from the TM.

[9] This range represents an 18% of the total number of sentences in the original forum corpus. We had previously analysed the sentences with less than 10 words (mainly courtesy formulas and error messages) and we considered that it would be extremely artificial to translate them without their context. On the contrary, sentences of a 10-15 word range were mostly well formed and semantically rich enough to be translated individually. The sample contained only 60 sentences of 16-20 words; 39 of 21-25 words, 13 of 26-30 words, and 4 sentences between the 31 and 44 words.

[10] http://accept.statmt.org/demo/

[11] http://www.statmt.org/wmt12

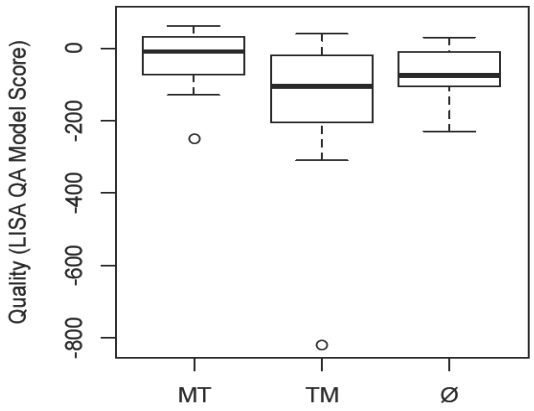

Figure 2. Quality Results

We performed an ANOVA test and we did not found significant results ($F_{(2,32)} = 53.67$, $p > .0.5$). We believe this may be due to the subjectivity of the only reviewer involved in the study. We envisage contrasting these initial scores with a second reviewer's evaluation to improve the significance of the results and achieve an interrater reliability.

3.2 Time

As explained in section 2, two data collections systems were put into place in order to obtain the amount of time spent by participants on each section. The ACCEPT Portal recorded the editing time when participants were working in the editing window. The video from BBFlashback recorder allowed us to capture the total time spent in the process as a whole. The results from both systems differ considerably and they are thus presented separately.

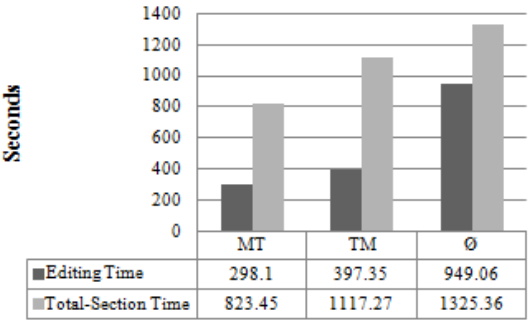

Figure 3. Editing Window Time vs. Total-Section Time (arithmetical mean)

Editing Window Time

By taking into account sections with similar proposals together, we can observe several differences between them: participants spent less time on the MT section ($\bar{x}$ = 298.10 seconds, sd = 123.61; $\bar{x}$ = 2.21 seconds to process each word), followed by TM ($\bar{x}$ = 397.35 seconds, sd = 156.02; $\bar{x}$ = 2.95 seconds per word). Participants spent considerably more time in the Ø section ($\bar{x}$ = 949.06 seconds, sd = 204.84; $\bar{x}$ = 7.08 seconds per word). This section was 3.1 times slower than MT and 2.3 times slower than TM. The ANOVA test indicated that the results were statistically significant $F_{(2,32)} = 53.67$, $p < .05$.

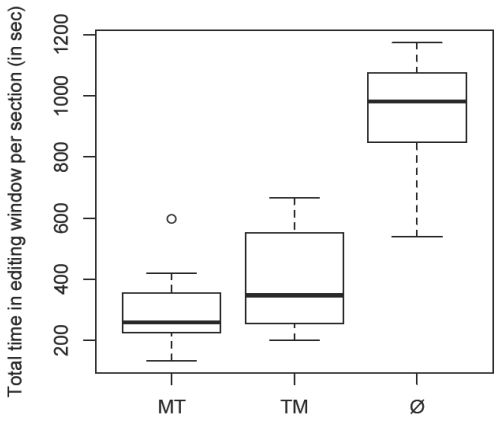

Figure 4. Editing Window Time

Total Section Time

Contrary to what we observed in the editing window time analysis, the difference between sections is not so significant if we consider the total time spent per section. However, the same pattern is repeated overall: the fastest section is MT ($\bar{x}$ = 823.45 seconds, sd = 273.66; $\bar{x}$ = 6.12 seconds per word), followed by TM ($\bar{x}$ = 1117.27 seconds, sd = 873.58; $\bar{x}$ = 8.30 seconds per word). The section with no proposal is again the slowest one ($\bar{x}$ = 1325 seconds, sd = 330.80; $\bar{x}$ = 9.88 seconds per word). In this case, the ANOVA test was not as conclusive as in the editing window time analysis: $F_{(2,32)} = 3.03$, $p = 0.06$.

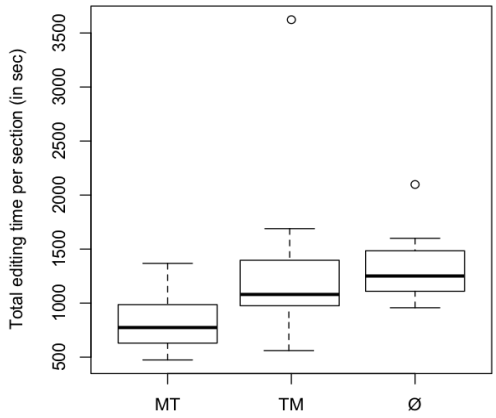

Figure 5. Total-Section Time

3.3 Keystroke Information

The ACCEPT Portal also captured keystroke information from all participants. The portal only

recorded the keys typed by our participants, and not their mouse clicks. These data can help us to understand how many keys were necessary to post-edit and/or translate each segment.

Having into account sections with the same proposals together, the section with no translation proposals typed the highest number of keys ($\bar{x}$ = 1219.33 keys, *sd* = 254.05; $\bar{x}$ = 9.09 keys per word); on the other hand, the lowest amount of keys typed was observed in the section with MT proposals ($\bar{x}$ = 159.83 keys, *sd* = 99.76; $\bar{x}$ = 1.18 keys per word), followed by the section with TM proposals ($\bar{x}$ = 258 keys, *sd* = 66.87; $\bar{x}$ = 1.91 keys per word). The higher number of keys typed in the Ø section (7.6 times higher than MT and 4.7 than TM) can be explained by the fact that in the Ø section translators had to write the whole target text rather than partially modifying it. The analysis of the variance showed in this case that the results were statistically significant $F_{(2,32)} = 150.7$, $p < .05$.

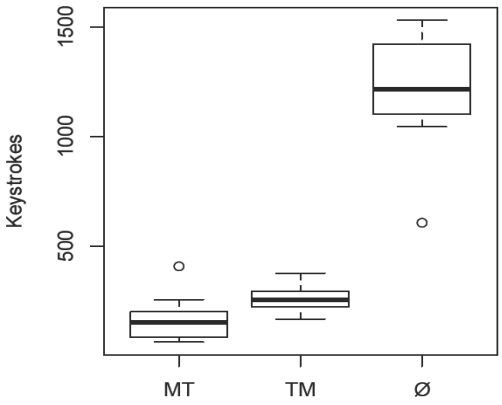

Figure 6. Keystroke Information

3.4 Insights from the Task-Specific Questionnaire

Questions from the task-specific questionnaire were designed to gather information about the translator's task experience. From the analysis of the answers, we collected the following information: participants were not familiar with the topic[12] of the text nor with Symantec[13] products; they found the task quite difficult[14] from a linguistic point of view. The doubts that they experienced the most were terminology-related (*N*= 9 participants), linguistic-related (*N*= 5), technical-related (*N*= 3), experiment-related (*N*= 2), and tool-related (*N*= 1). On the other hand, the translation interface was easy[15] for them to use. They also declared that the translation proposals had helped[16] them to perform the task.

We asked participants if they had considered post-editing as a career option and to justify their answer. Responses were all positive: a third of the participants used the term "interesting"; four participants mentioned the idea of post-editing as a side job, three of whom declared that they would not considered it as a full job (PB8 said "I see post-editing as complementary to translating. I would find it very tiresome to do only post-editing"; PB2 declared "Not as a full time job, I like the translation process itself, but part-time yes, why not"; and PA5 stated "Not as a single activity").

As we were interested in calibrating the participants' general attitude towards machine translation, we asked them what their general opinion was concerning MT. From their answers, we can observe that most of the participants (*N*= 10) had a generally positive attitude towards MT, using the following adjectives: useful (*N*= 4), helpful (*N*= 3), great (*N*= 1), and practical (*N*= 1). Half of the participants also specified that MT works better when using repetitive (*N*= 3), technical (*N*= 2) and simple (*N*= 1) texts. Two participants said that there is still work to do on the MT area to obtain good results (PA5 stated "I think it's great, in particular for specific domains. There's still research / work to do though" and PA1 stated "I think it can be really useful when it is done in a way that allows the translator to focus more on the language. But that requires a lot of work for MT to really be helpful"). Finally, two of the participants declared that human translation was better than MT (PB4 stated "(…) if you are a translator used to translate from this language, you'll be quicker and more efficient than the machine"; and PA6 "(…) [MT] it is not as good as human translation: we need to check every sentence and modify or completely re-translate it").

4 Discussion and Conclusions

Results from the translation quality analysis demonstrated that using MT proposals with community forum content helped translators to

[12] A mean of 2.6 was obtained, a 6 points scale was used where 1 stood for "completely unfamiliar" and 6 "very familiar".

[13] A mean 2.1 was obtained following the same scale system.

[14] A mean of 2.9 was obtained, a 6 points scale was used where 1 stood for "very difficult" and 6 "very easy".

[15] A mean of 1.3 was obtained, a 6 points scale was used where 1 stood for "Easy to use" and 6 "Difficult to use".

[16] A mean of 2.3 was obtained, a 6 points scale was used where 1 stood for "Absolutely yes" and 6 "Not at all, I would have preferred working from scratch".

obtain a higher quality level compared to TM suggestions. Translating from scratch obtained better results than translating with the aid of TM proposals. The latter contradicts a previous study using technical documentation, where better results with TM than translating from scratch were observed (Morado Vázquez, 2012). We can thus hypothesise that the traditional use of TM might not be the correct strategy to follow when working with user forum content.

In terms of time (both editing window time and total section time), participants spent more time in the section without translation proposals, and the best results were obtained again in the MT section. The TM section was carried out more slowly than the MT section, but faster than the section without translation proposals. In this case, these results correlate with previous studies (*ibid*). Interestingly, translation from scratch received worse results than MT in all the variables measured, which matches the results obtained by Plitt and Masselot (2010).

Taking an overall look at the results, we could also state that, in our particular context, working with proposals is faster than translating forum data from scratch; however, taking a shorter amount of time does not always correlate with quality rates. Therefore, if time is the main constraint in a specific translation task (and the quality is not as important), translation proposals (preferably from MT output rather than TM) should be presented to the translators. Our participants also stated that translation proposals helped during their task.

We have also observed a general positive attitude towards post-editing and machine translation in our participants, such as in (Guerberof, 2013), which contradicts the general assumption that translators tend to reject these approaches (Arevalillo Doval, 2012, p. 181; Yuste Rodrigo, 2013). Nonetheless, it is also worth mentioning that our participants might be more technology-oriented (they are at least taking part in translation technologies-related courses) than other translators who may be reluctant to include translation technology aids in their working routine. It would be interesting to study if the same attitude results are found among the whole group of Master students, as not necessarily all of them take part in the translation technology courses. Our participants considered post-editing as a career option, though a third of them would not considered it a single activity but just a side job. This leads us to think that post-editing is not yet seen as an activity that can be carried out as a full-time job.

To sum up, our findings indicate that using a context specific MT system has a greater impact in the translation activity (in terms of time, quality and keystroke effort) than using any other translation alternatives, which in our case were TM output and traditional translation without proposals.

5 Limitations and Future Work

Although we did find significant results in almost all of the variables measured, which lead us to think that, within the ACCEPT project, working with MT output would be beneficial for translators, the main limitation of this pilot study is its lack of external validity, which is not one of the aims of our research, as it was carried out under specific and controlled circumstances: type of data used (forum specific content), type of participants (students), and a particular controlled environment (ACCEPT post-editing portal). Consequently, we cannot predict that same results would be obtained if other data, subjects or environment are to be considered. However, in order to extract more conclusive results we intend to continue this line of research through two main paths: a) we are planning to repeat the experiments using regular forum users (which are the target audience of our online tool; b) we are also considering repeating the experiments using a traditional CAT tool environment, where differences in the fuzzy matches from TM could be highlighted.

We should also state that the original 36 sentences that formed our text were selected based on their match with the TM (made of texts from official manuals, which were written in a more formal style than regular forum content). They do not, therefore, represent an entirely randomly selected sentence set from the forum data, where oral language is frequently used. In order to better study the use of MT with significant community-generated forum content, we propose to repeat the study using only randomly selected sentences and working with only two variables (MT output versus translating from scratch).

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 288769 (ACCEPT). The following individuals and companies have also donated

their time and experience to this research: Johann Roturier, Veronique Bohn, and the translation master's students of the FTI, UNIGE.

References

ACCEPT. 2012. Baseline machine translation systems. Project Deliverable D 4.1., available: http://www.accept.unige.ch/Products.htm [accessed 16 April 2013].

Arevalillo Doval, Juan José. 2012. La traducció automàtica a les empreses de traducció. *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, (10):179‑184.

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

Guerberof, Ana. 2009b. Productivity and quality in MT post-editing. In *MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT*, August 29, Ottawa.

Guerberof, Ana. 2012. *Productivity and quality in the post-editing of outputs from translation memories and machine translation*. (PhD), Universitat Rovira I Virgili, Tarragona.

Guerberof, Ana. 2013. What do professional translators think about post-editing? *The Journal of Specialised Translation*, (19):75‑95.

He, Yifan. 2011. *The Integration of Machine Translation and Translation Memory*. (PhD), Dublin City University, Dublin.

He, Yifan, Yanjun Ma, Johann Roturier, Andy Way, and Josef van Genabith. 2010. Improving the post-editing experience using translation recommendation: a user study. In *AMTA 2010: The ninth conference of the association for Machine Translation in the Americas, Proceedings*, Denver 247‑256.

LISA. 2007. LISA QA Model 3.1 User's Manual.

Morado Vázquez, Lucía. 2012. *An empirical study on the influence of translation suggestions' provenance metadata*. (PhD) University of Limerick, Limerick.

O'Brien, Sharon. 2006. Eye-tracking and translation memory matches. *Perspectives: Studies in Translatology*, 14(3), 185-204.

O'Brien, Sharon. 2009. Eye tracking in translation process research: methodological challenges and solutions. In: Mees I, Alves, F. Göpeferich S (eds) Methodology, technology and innovation in translation process research- a tribute to Arnt Lykke Jackobsen, Copenhagen Studies in Language 38. Samfundslitteratur, Conpenhagen, 251-266.

O'Brien, Sharon. 2012. Translation as human-computer interaction. *Translation Spaces*, *1*(1):101–122.

Plitt, Mirko and François Masselot. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical* Linguistics, 93(1), 7-16.

Tatsumi, Masuru, and Johann Roturier. 2010. Source Text Characteristics and Technical and Temporal Post-Editing Effort: What is Their Relationship? In *Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC 10)* Denver 43‑51.

Teixeira, Carlos. 2011. Knowledge of provenance and its effects on translation performance in an integrated TM/MT environment. In *Proceedings of the 8th International NLPCS Workshop–Special theme: Human-Machine Interaction in Translation. Copenhagen Studies in Language* (41):107–118.

Van Genabith, Josef. 2012. Centre for Next Generation Localisation Annual Report 2011, available: http://www.cngl.ie/drupal/sites/default/files/Annual_Report_2011.pdf [accessed 16 April 2013]

Yamada, Masuru. 2011. The effect of translation memory databases on productivity. *Translation research projects*, (3):63–70.

Yuste Rodrigo, Elia. 2013. La posedición en el flujo de producción de contenido multilingüe: tendencias, actantes e implicaciones tecnológicas. *Tradumàtica: tecnologies de la traducció*, (10):157‑165.