KNOWLEDGE OF PROVENANCE AND ITS EFFECTS ON THE TRANSLATION PROCESS IN AN INTEGRATED TM/MT ENVIRONMENT

Carlos S. C. Teixeira

Universitat Rovira i Virgili

A research design submitted in partial fulfilment of the requirements for the Doctorate in Translation and Intercultural Studies

Doctoral program in Translation and Intercultural Studies
Universitat Rovira i Virgili, Spain

Supervisor:
Dr Anthony Pym, Universitat Rovira i Virgili

2012

Abstract

Over the last few years, machine translation (MT) has been increasingly used in professional translation settings and integrated into translation-memory (TM) systems. This TM/MT combination has turned pre-translation + post-editing into an even more attractive alternative in terms of translation speed – and thus of productivity and profitability – for all parties involved in the translation process. In some cases, source files are pre-translated using a combination of TM and customised MT before reaching the translators, who then become reviewers, or post-editors. But how does this actually affect productivity and quality, and how do translators feel when performing this new activity?

In order to look for answers to those questions, and drawing on a pilot experiment conducted during my Masters, I have devised a research design that compares two translation environments. The basic difference between the two is the availability to the translator of information on the provenance of the suggested translation for a particular segment (whether it comes from MT, TM, and at which match percentage). Data will be collected using screen recording, keystroke logging, eye tracking and retrospective interviews.

Keywords: translation technology, translation memory, machine translation, process research, speed, productivity, quality, job satisfaction, performance, provenance, trust.

Introduction

Until recently, machine translation (MT) and translation memories (TM) were seen as totally different approaches to using technology in translation. While the first approach was targeted – in recent years – mainly at end users interested in grasping the general idea of a text written in a language they could not understand (usually while browsing the Internet), the second was addressed to professionals in the translation industry, such as translators, translation agencies or translation departments in large companies.

However, this scenario has been changing at a rapid pace over the last few years, mainly due to quality improvements and the general availability of statistical machine-translation systems, based on large amounts of human-produced bitexts. This has allowed MT to be progressively integrated into TM tools in professional translation environments, bringing new possibilities as well as new challenges.

The potential productivity gains derived from integrating machine translation (MT) and translation memories (TM) are calling for new work methods in the translation market. As an example, some translation agencies pre-translate their source files using a combination of TM and customised MT before sending them out to translators, who then become reviewers, or post-editors. In this scenario, translators may review each segment without knowing its provenance, i.e. whether it comes from a translation memory (and at which match percentage) or from a machine-translation engine. Could this missing information have an impact on the way translators perform their tasks, compared to a more traditional environment, where translators would know where each translation suggestion comes from? In other words, how does the "knowledge of provenance" of translation suggestions affect translators' behaviour in environments that integrate TM and MT?

The results from a pilot experiment conducted during my Masters corroborate the "provenance" assumption and indicate that the information displayed about each suggested translation has an impact on performance. Although quality was not affected, certain types of translation suggestions, namely exact matches and high-percentage fuzzy matches, showed pronounced variations in speed and editing effort between the two environments.

Literature review

Historically, the focus of study of translation research had moved from the source text to the target text, in a myriad of product-centred approaches. Recent years have seen the rise of process-centred approaches, i.e. studies that focus on what happens while translators translate, in an attempt to address the new questions. Crucial for these studies are the strategies and technologies that allow researchers to collect real-time data from the actual translation process, which today typically takes place on a computer. Alongside more traditional data-gathering methods such as think-aloud protocols (TAPs) (Ericsson & Simon 1998; Jakobsen 2003), other methods such as keystroke logging (Jakobsen 2002, 2006), eye-tracking (O'Brien 2006, 2009) and screen recording allow us to identify where attention is being placed and even to measure cognitive load, e.g. through pupil dilation (see O'Brien 2006; Shreve & Angelone 2010). Several studies also combine those methods (see Alves 2003; Dimitrova 2005; Carl et al. 2011) for better confirmation of results.

Machine translation has traditionally been a subject of study mainly within the computational linguistics community, possibly due to the technicalities involved. Translation scholars are more interested in knowing what those systems can offer than in the meanders of the underlying technology. Translation memory systems, on the other hand, for their immediate applicability and simpler conception, have become a frequent subject within translation circles. Tool manufacturers and the translation industry in general have published extensively on different types of tools, while academia has tried to keep up with the developments in a more conceptual way (cf. Pym 2012 for the different kinds of contributions that can be expected from either community).

Although the idea of automatic translation can be traced back to as early as the 17th century, machine translation in its more contemporary form dates back to the first half of the 20th century and has undergone several stages of development, with the current trend being towards data-driven statistical machine translation (SMT) and hybrid approaches (combining SMT with rules-based MT).[1] As stated before, MT in itself was not a topic in translation research, but revising and post-editing have deserved

[1] For an introduction to machine translation and its history, see Hutchins & Somers (1992). For a more recent overview of MT developments, see Way (2009).

serious attention e.g. in Krings (2001), Guerra Martínez (2003) and Mossop (2001, 2007). Several studies have tried to determine whether there is an actual increase in speed while post-editing machine translated segments when compared to translating from scratch. Allen (2003, 2005) conducted several studies on machine translation post-editing with specific tools and provides some guidelines for improving its results. Lee & Liao (2011: 142) "suggest various benefits for the use of MT, such as facilitating source text comprehension and reducing translation errors." This is related to effort and quality, respectively. Nevertheless, the authors' productivity assumptions that MT can save much time "from needing to type out words" (Lee & Liao 2011: 141), although plausible, deserve further empirical testing. A recent study with seven participants working in the English-to-Danish language combination (Carl et al. 2011) actually points in the opposite direction. García (2010) compares time and quality between translating "entirely from the source text" vs. "editing machine translation" and also finds that "time differences were not significant", although "the machine translation seeded passages were more favourably assessed" (2010: 7).

Much more recent than machine translation, CAT (computer-aided translation) tools appeared first in the form of terminology management systems in the mid-1980s and then as translation memory (TM) systems in the early 1990s. The first academic studies to deal with these translation technologies appeared before the turn of the millennium.[2] Later on, Dragsted (2004) investigated how the forced (usually sentence-based) segmentation in TM systems reflects and interferes with cognitive segmentation. Her study raises an important debate on an aspect of translation tools, also applicable to MT, that is usually taken for granted but can have an impact on all the metrics relevant for the field (including quality). More recently, Christensen & Schjoldager (2010) provide a good general overview of research on translation memories and suggest some paths for future research, namely "on how translators interact with TM technology and on how it influences translators' cognitive processes" (2010: 99). Yamada (2011) investigates how the type of content ('free translation' vs. 'literal translation') in a translation memory affects translation speed, and concludes that literal translations are more advantageous for higher fuzzy-match categories. Christensen (2011) focuses on

[2] Webb (1998) presents a comprehensive study of TM systems and summarises the state of affairs at the end of the 1990s, with a prophetic foresight for the following decade. For a very recent history and overview, with future prospects, see Zetzsche (2012).

studies that deal with "mental processes" in the interaction between translators and TM tools.

As far as the integration between TM and MT is concerned, O'Brien (2006) compares the performance of TM vs. MT when translators work in a 'traditional' TM system that integrates MT suggestions. One of her findings is that "cognitive load [and processing speed] for machine translation matches is close to fuzzy matches of between 80-90% value" (2006: 185). For fuzzy matches above 90%, including exact matches, TM processing is faster and requires a lighter cognitive load, whereas the opposite happens for fuzzy matches below 80%. Guerberof (2009) analyses time and quality when editing TM suggestions vs. MT suggestions in a post-editing environment (a TM was used during the pre-translation phase, reproducing an actual scenario that can be found in industry). Her findings suggest that "translators have higher productivity and quality when using machine-translated output than when processing fuzzy matches [at any percentage level] from translation memories" (2009: 11).

Quality assessment continues to pose the most difficult problems to translation research. Several studies resort to human reviewers to assess translation quality based on predefined criteria (e.g. the LISA grid) and rely on individual subjectivity. Others use metrics originally intended for raw machine translation evaluation (BLEU, NIST, etc.). While still relying on subjective criteria, Huang (2011) stresses the need to shift from the concepts of "good" and "bad" translation to a distinction between "acceptable" and "unacceptable", and suggests a statistical approach. An alternative method that has also been used as an industry-relevant measure of quality is revision time, i.e. the longer it takes a professional reviewer to edit a translation up to delivery standards, the lower the quality of the translation. Although revision time provides a concrete measurement (words revised per hour), it still relies on the reviewers' subjectivity and specific usability or acceptability expectations.

Research question

The focus of my research will be on a particular aspect that distinguishes translation memory systems from other post-editing environments: TM systems show translators the 'provenance' and the 'relevance' of the translation suggestions coming from the memory, whereas most environments for post-editing MT display the 'best translation suggestion possible' without any indication of its origin or degree of confidence. The

presence or absence of provenance information might influence translators' performance and the failure to make this distinction might explain the discrepant results of some studies that compare translation times when (post-)editing MT and TM suggestions (such as O'Brien 2006 vs. Guerberof 2009). No research work in the field seems to have focused on this particular distinction.

My main research question may thus be summarised as: What are the differences (if any) in the translation process between a situation where translators know the provenance of the translation suggestions they are editing and a situation where this information is not available?

Hypotheses

In order to answer this question, I will compare two translation environments, which I describe in more detail below. In the first environment, translators do not know the provenance of translation suggestions, whereas in the second environment translators do have access to this information. These are my working hypotheses:

- Hypothesis 1 (H1): The overall *translation speed* is higher when *provenance information* is available.
- Hypothesis 2 (H2): The overall *amount of editing* is smaller when *provenance information* is available.
- Hypothesis 3 (H3): There is no significant difference in the *quality* level when *provenance information* is available.

Some definitions are necessary in order to operationalize the variables we want to test:

- *Translation speed* is measured as words per hour. There will be separate counts for the first rendition (drafting) and second rendition (self-revising).

- *Amount of editing* is measured as the percent ratio between the number of keyboard key presses + mouse clicks done by translators and the total number of characters in the resulting segment.

- *Provenance* of translation suggestions is indicated by showing their origin (TM or MT) and, in the case of TM, by highlighting the differences between the actual segment and the matching segment in the TM, as is usually the case in most TM systems.

- *Quality* is (inversely) measured as the time it takes a professional reviewer to edit the final translation in order to make it adhere to predefined acceptability criteria (see further details below in the 'Data collection' section).

Methodology

I have designed an experiment with ten experienced translators from English to Spanish, who will receive two similar source texts of around 500 words each, to be translated in the two different environments described below:

Environment B will present the source-text segments on the left-hand side of the screen and a pre-translated version of the source text (obtained through the pre-processing of the file with TM and MT) on the right-hand side. In this case, all no-matches will have been replaced by MT suggestions, and the whole text will be presented as a sequence of pre-translated segments. Translators will be able to edit the pre-translated segments as if they were revising a translated file and they will have no information on the provenance of each of the pre-translated segments (i.e. whether they come from a TM segment or an MT segment). Let us call this environment B (as in 'Blind') for mnemonic recall. This environment tries to reproduce as closely as possible the environment used in Guerberof (2009).

Environment V is similar to the previous one in that translators will also have access to the source-text segments on the left-hand side of the screen and an editing space on the right-hand side. However, the difference consists in that, instead of working by 'blindly' editing pre-translated segments, translators will be able to see where the default translation suggestion is coming from (either from the translation memory or from the MT engine). Additionally, in the case of TM suggestions, translators will see the highlighted differences between their actual source segment and the TM source segment. For mnemonic recall, let us call this environment V (as in 'Visual'). This environment tries to reproduce as closely as possible the environment described in O'Brien (2006).

Subjects

I plan to have ten English-to-Spanish translators who are native speakers of Iberian Spanish: five male and five female subjects. They will be selected by means of a questionnaire and will have 5+ years of professional experience working with translation-memory systems on technical or marketing texts. Formal training in translation will not be a prerequisite.

All ten translators will translate both texts in both environments. Five translators will start working in B and the other five in V, in order to account for potential differences related to the order of the tasks.

Materials

Both source texts will be excerpts from the same section in a software manual. I will provide translators with a translation memory consisting of segments with five different levels of similarity with the original text: 'exact matches'; 'fuzzy matches' in the ranges of 70%-79%, 80%-89%, 90%-99%; and 'no matches.'

Data collection

The main methods for collecting data will be screen recording, keystroke logging and eye tracking using Tobii Studio 3.0 software and a Tobii X120 eye tracker. We also considered using Translog for keystroke logging, but since the tool has its own interface and does not allow for integration with other applications, it would not be possible to reproduce the two environments we want to test (cf. Jakobsen 2002, 2003 on the use of Translog for keystroke logging). The two translation environments will be created in SDL Trados Studio 2009 Freelance.

Think-aloud protocols will not be used as they are known to slow down the translation process (cf. Jakobsen 2003) and we are trying to measure translation speed in a natural(istic) environment. Retrospective interviews can provide the information we need on translators' feelings and satisfaction, without interfering with the translation process. Eye tracking was chosen as an additional data-collection method in order to provide a more precise way of identifying which segment is being focused on at a given point in time.

For assessing quality, all texts will be reviewed by two professional reviewers, who will receive instructions on what kind of errors to correct in order to make the revised text 'acceptable'. Although still subjective, this method has been considered as one of the best options available for quality assessment and reflects a common practice in the industry. The quality level of a translation will be (inversely) measured as the time spent for its revision.

Expected results

We can imagine several possible combinations of results for the main experiment. If all three hypotheses are supported, we can conclude that the optimal setting for the integration of MT and TM (as far as translation speed, quality and effort are concerned) should make provenance information available to translators. If no significant difference is detected from the testing of H1 and H2, we can conclude that the provenance information is irrelevant to speed and effort (although it might still be relevant to quality). If H1 is supported and H2 is rejected (or vice-versa), we will have shown that the variation in the amount of editing is not a sufficient explanation for the variation in speed. If both H1 and H2 are rejected (the provenance information actually slows down the translation process and increases the typing effort), then we could use the information gathered in the interviews to look for possible reasons for this phenomenon (e.g. the additional information clutters the translator's workspace). If H1 is supported and H3 is rejected, our conclusion will depend on what environment produces the best quality. For example, if it is B, then we can conclude that the gain in speed happens at the expense of quality.

Expected benefits

The answer I am ultimately trying to find is which strategy is more effective: either the pre-translation of files, with further review at the end, or a 'live' translation with knowledge of the provenance of suggestions. A deeper understanding of both processes, as envisaged in this doctoral research, can be beneficial for all parties involved in the translation scene, including independent translators, translation agencies, translation-tool developers and, ultimately, translation customers, as the results can contribute to devise optimal workflows and best practices.

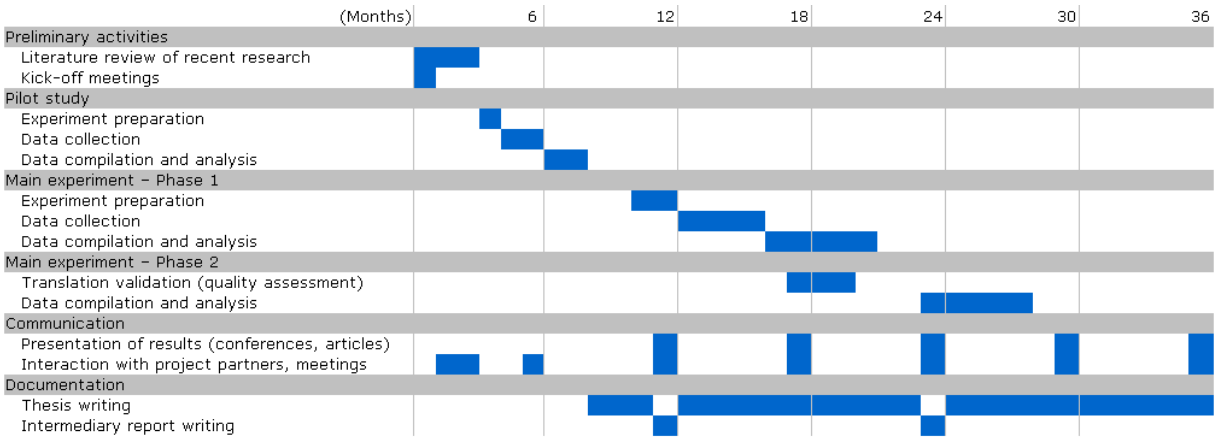
Besides the potential impacts on earnings (and savings), the search for optimal processes can increase the volume of text that can be processed. We must remember that in the European Union, as well as in many international organizations, large amounts of text remain untranslated due to time or budget constraints. Even more important, it is our concern to try to optimise the translation process in order to increase job satisfaction among translation professionals. Finally, I hope the results will also be of intellectual importance, as we are trying to demonstrate that the impact of technology is not just in what it does, but also in what the stakeholders know about what it does.

Transfer of results

This research design is part of a larger, three-year project under the Seventh Framework Programme (FP7) of the European Commission's Marie Curie Actions. In addition to the yearly reports and the final doctoral thesis, it is my intention to present papers regularly at conferences and submit a series of articles to top academic journals on Translation Studies, such as *Meta*, *Target*, *Machine Translation* and *Perspectives*. A tentative schedule for the transfer of results is shown below under the Communication group. Partial results from a related pilot study were published in Teixeira (2011).

Work schedule

Below is a preliminary time plan for the 36-month period over which I intend to conduct my doctoral research.



Cost estimation

For the fulfilment of all the activities planned for this doctoral research, I estimate the following costs:

Item	Cost (Euros)
Equipment	
Portable computer (PC with Windows)	800
Software license (SDL Trados Studio 2009 Freelance)	800
Conferences (includes fees, travelling, accommodation), 3/year	7200
Eye tracker (Tobii X120)	25000
Transportation*	400
Payment of translators†	1000
Payment of reviewers‡	600
TOTAL	35500

* I am anticipating that most of my subjects will be located in Barcelona. I am considering having to meet each of them twice, 20 euros round-trip.
† Since we want to study experienced translators, it is reasonable to assume that they will not accept to waste several hours of work without being paid. I will offer a payment of 100 euros for each translator, which corresponds to approximately 25 € per hour.
‡ Two professional reviewers will be hired. I am estimating 30 minutes to review each text (20 in total) at a rate of 30 euro/hour.

References

Allen, Jeffrey. 2003. "Post-editing." In: Harold Somers (ed.) *Computers and Translation: A Translator's Guide*. Amsterdam & Philadelphia: Benjamins. 297-317.

Allen, Jeffrey. 2005. "What is Post-editing?" *Translation Automation Newsletter* 4: 3-5.

Alves, Fabio (ed.). 2003. *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam & Philadelphia: John Benjamins.

Carl, M.; Dragsted, B.; Elming, J.; Hardt, D. & Jakobsen, A. L. 2011. "The Process of Post-Editing: a Pilot Study." In: *Proceedings of the 8th International NLPCS Workshop – Special theme: Human-Machine Interaction in Translation*. Copenhagen Studies in Language 41. Copenhagen: Samfundslitteratur. 131–142. Available at: http://www.mt-archive.info/NLPSC-2011-Carl-1.pdf

Christensen, Tina Paulsen. 2011. "Studies on the Mental Processes in Translation Memory-assisted Translation – the State of the Art." In: *trans-kom* 4 [2]: 137–160.

Christensen, Tina Paulsen & Schjoldager, Anne. 2010. "Translation-memory (TM) research: what do we know and how do we know it?" *Hermes – Journal of Language and Communication Studies*: 89-101.

Dimitrova, Birgitta Englund. 2005. *Expertise and Explicitation in the Translation Process*. Amsterdam: Benjamins.

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 thesis, Copenhagen Business School: Samfundslitteratur.

Ericsson, K. Anders & Simon, Herbert A. 1998. "How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking." In: *Mind, Culture, and Activity* 5:3. 178–186.

García, Ignacio. 2010. "Is machine translation ready yet?" *Target* 22 (1): 7-21.

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

Guerra Martínez, Lorena. 2003. *Human translation versus machine translation and full post-editing of raw machine translation output*. Master's dissertation. Dublin: Dublin City University.

Huang, Harry J. 2011. "Intermediality and Human vs. Machine Translation." In: *CLCWeb: Comparative Literature and Culture* 13: 3. Thematic issue New Perspectives on Material Culture and Intermedial Practice. Ed. Steven Tötösy de Zepetnek, Asunción López-Varela Azcárate, Haun Saussy, and Jan Mieszkowski. Available at: http://docs.lib.purdue.edu/clcweb/vol13/iss3/10

Hutchins, W. John & Somers, Harold L. 1992. *An introduction to machine translation*. London: Academic Press.

Jakobsen, Arnt Lykke. 2002. "Translation drafting by professional translators and by translation students." Empirical Translation Studies: Process and Product. *Copenhagen Studies in Language* 27: 191-204.

Jakobsen, Arnt Lykke. 2003. "Effects of think aloud on translation speed, revision, and segmentation." In: Fabio Alves (ed.) *Triangulating Translation. Perspectives in process oriented research*. 69-95.

Jakobsen, Arnt Lykke. 2006. "Research methods in translation – Translog." In: Sullivan, K.P.H. & Lindgren, E. (eds.). *Computer Keystroke Logging and Writing: Methods and Applications*. Oxford: Elsevier. 95-105.

Krings, Hans. 2001. *Repairing Texts* [edited by G.S. Koby, translated from German by G.S. Koby, G.M. Shreve, K. Mischerikow and S. Litzer]. Kent, Ohio: Kent State University Press. (1994 post-doctoral thesis)

Lee, Jason & Liao, Posen. 2011. "A Comparative Study of Human Translation and Machine Translation with Post-editing." In: *Compilation and Translation Review* 4:2 (September 2011). 105-149.

Mossop, Brian. 2001. *Revising and Editing for Translators.* Manchester: St Jerome.

Mossop, Brian. 2007. "Empirical studies of revision: what we know and need to know." *The Journal of Specialised Translation* 8, unpaginated.

O'Brien, Sharon. 2006. "Eye-tracking and translation memory matches." *Perspectives: Studies in Translatology* 14/3: 185-205.

O'Brien, Sharon. 2009. "Eye tracking in translation-process research: methodological challenges and solutions." In: *Copenhagen Studies in Language* 38. Copenhagen: Samfundslitteratur. 251-266.

Pym, Anthony. 2012. "Democratizing translation technologies – the role of humanistic research." In: Valeria Cannavina and Anna Fellet (eds.) *Language and Translation Automation Conference*. Rome: The Big Wave. 14-29.

Shreve, Gregory M. & Angelone, Erik. 2010. *Translation and Cognition*. Amsterdam & Philadelphia: John Benjamins.

Teixeira, Carlos S. C. 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. Copenhagen: Samfunds-litteratur. 107-118. Available at: http://www.mt-archive.info/NLPCS-2011-Teixeira.pdf

Way, Andy. 2009. "A critique of Statistical Machine Translation." In: *Linguistica Antverpiensia* 8/2009. Special issue on Evaluation of Translation Technology. 17-41.

Webb, Lynn E. 1998. *Advantages and Disadvantages of Translation Memory*. Master's dissertation. Monterey: Monterey Institute of International Studies.

Yamada, Masaru. 2011. “The effect of translation memory databases on productivity.” In: Anthony Pym (ed.) *Translation research projects 3*. Tarragona: Intercultural Studies Group. 63-73.

Zetzsche, Jost. 2012. “Translation technology comes full circle.” In: *MultiLingual*. April 2012. 50-51.