

What do professional translators think about post-editing?

Ana Guerberof Arenas, Universitat Rovira i Virgili, Tarragona

ABSTRACT

As part of a larger research project on productivity and quality in the post-editing of machine-translated and translation-memory outputs, 24 translators and three reviewers were asked to complete an on-line questionnaire to gather information about their professional experience but also to obtain data on their opinions about post-editing and machine translation. The participants were also debriefed after finalising the assignment to triangulate the data with the quantitative results and the questionnaire. The results show that translators have mixed experiences and feelings towards machine-translated output and post-editing, not necessarily because they are misinformed or reluctant to accept its inclusion in the localisation process but due to their previous experience with various degrees of output quality and to the characteristics of this type of projects. The translators were quite satisfied in general with the work they do as translators, but not necessarily with the payment they receive for the work done, although this was highly dependent on different customers and type of tasks.

KEYWORDS

Translation memory, machine translation, post-editing, review, professional translators, reviewers, experience, localisation, output, pricing, MT, TM.

1. Introduction

It is not very often that translators are asked their opinions about post-editing and machine translation in the localisation industry. This could be out of fear of an adverse or negative response or due to the fact that translators are often invisible in the localisation work-flow, and we feel this invisibility is increasing as process automation increases, and all aspects related to technology seem to acquire more relevance than the act of translating itself. There are many reasons for this and price can be one of them. We have carried out a questionnaire and debriefings with professional translators to know their opinions about this relatively new process. This was part of a larger research project with 24 translators and three reviewers to obtain productivity and quality data in the post-editing of machine translation (MT). The first part of the project involved the use of a web-based post-editing tool designed by CrossLang to post-edit and translate a text from English into Spanish. A trained Moses (Koehn *et al.* 2007) statistical-base engine was used to create the output with a BLEU score (Papineni *et al.* 2002) of 0.6 and a human evaluation score of 4.5 out of 5 points. The final output was then evaluated by three professional reviewers, who registered the errors using the LISA QA model criteria¹ and also filled in an on-line questionnaire. We are presenting in this article a summary of the second part of the project, the qualitative aspect, that

involved an on-line questionnaire and debriefings to gather information related to the translators' and reviewers' own working experience and opinions about post-editing and MT. This is a small group of translators; however, the results describe certain characteristics and problems that might be applicable to many translators.

2. Group profile

The group of professional translators had considerable experience with localisation tools and some experience in post-editing MT output, although the task represented a low percentage of their work and had not been performed for a very long period of time. Their experience in localisation ranged from more than two years to eight years or more (75 percent had more than six years' experience in localisation). Their experience in post-editing, however, was less extensive. Exactly 79.2 percent of the whole group had no experience or less than four years' experience in post-editing which was low in comparison to their general experience in localisation, tools and the subject. The reviewers, on the other hand, had more than eight years' experience in localisation; more than eight years' experience using tools; between two and eight years' experience in the subject matter; between two and six years' experience in post-editing and more than four years' experience in reviewing.

3. Results from the translators' questionnaire

SurveyMonkey was the tool used to publish the questionnaire. There were 11 questions that are presented here in tables containing the responses from the 24 translators. All quotations are gathered from the translators exactly as the participants inserted them in the on-line questionnaire.

Please estimate how often the following statements describe your revision procedure (you will need to select an option in each statement):						
Answer Options	Never	Rarely	Sometimes	Frequently	Always	Response Count
As I translate, I recheck my translation before going to the next segment.	1	1	2	5	15	24
Immediately after I finish the translation of one file, I go back and review all my translations.	1	4	7	7	5	24
After I finish the translation of all files assigned to me, I review the whole batch of files.	2	1	4	8	9	24
After I finish one day of work, I go back and review all work done in	6	5	7	4	2	24

that day.						
Other (please specify)						3

Table 1. Revision procedures

The data show that the revision procedure this group of translators tends to use most frequently is checking their translation before proceeding to the next segment and reviewing the whole batch of files assigned after finishing. This is in line with conclusions on revision styles drawn by Dimitrova (2005: 144): “a segment is often revised before going on to the next segment.” Reviewing at the end of one day’s work seems to be much less frequent.

If you have post-editing experience, which of the options below best represents your experience?		
Answer Options	Response Percent	Response Count
My productivity when post-editing has been constant over time.	45.0%	9
My productivity when post-editing has increased over time.	40.0%	8
My productivity when post-editing has decreased over time.	0.0%	0
I do not know.	15.0%	3

Table 2. Post-editing learning curve

We were interested in knowing how translators perceived their productivity in post-editing with growing experience. It seems that some of them might perceive an increase in productivity (40 percent) but others do not (45 percent). Two out of the three “I do not know” responses belong to translators that had declared previously not having experience in post-editing, so this is in keeping with that. None of the translators think that their productivity decreases over time. This response is interesting. Since post-editing can be a very repetitive task (correcting same type of errors over time) it is logical that translators feel that their productivity has remained constant or has increased. However, precisely because it is very repetitive, this could cause tiredness and potentially result in a decrease in productivity.

If you have post-editing experience, which of the options below best describes your experience?		
Answer Options	Response Percent	Response Count
Experience has not affected my ability to spot MT errors - I correct them the same way as when I started.	30.0%	6
As I acquire more experience it is more difficult for me to detect MT errors, as I have become used to them.	0.0%	0
As I acquire more experience it is easier for me to detect MT errors, as I look for the same patterns.	55.0%	11
I do not know.	15.0%	3

Table 3. Post-editing proficiency

Most of the translators (55 percent) think that experience helps them to detect errors when post-editing. Note that 30 percent declared that they correct errors in the same way as when they started. None of the translators think that experience with post-editing might affect their ability to detect errors. This is quite a positive response from the translators, since it is often suggested that “growing accustomed” to the errors could result in a decline in the overall quality.

If you have post-editing experience, which of the options below best represents your experience? Reviewing means here to go over a human translation, identify and correct errors. Post-editing means here to go over MT out-put, identify and correct errors.		
Answer Options	Response Percent	Response Count
Post-editing, for me, requires the same effort as reviewing human translations.	30.0%	6
Post-editing, for me, requires more effort than reviewing human translations.	40.0%	8
Post-editing, for me, requires less effort than reviewing human translations.	20.0%	4
I do not know.	10.0%	2

Table 4. Post-editing effort

Interestingly four translators responded that post-editing required less effort than reviewing human translation (20 percent). These translators, except one, gained higher productivity with MT segments than with Fuzzy matches in this project. Still, 40 percent think there is more effort required in post-editing. Except for two translators, who in this project had higher productivity when editing Fuzzy matches, the participants showed higher productivity with MT matches. It is important to note that effort not only refers to productivity or time gained - it also implies *cognitive* effort (Krings 2001, O’Brien 2006). Translators might perceive a higher (cognitive) effort when post-editing and still be more productive than when editing human translations. Furthermore, the data that we obtained in this study do not represent all the experience these translators

have in post-editing or reviewing, and the MT output quality was high in our particular project.

How satisfied are you with the price per word you receive from your customers?							
Answer Options	Highly unsatisfied	Unsatisfied	Satisfied	Very satisfied	Highly satisfied	Rating Average	Response Count
I am	2	7	11	4	0	2.71	24
Comments							6

Table 5. Price satisfaction

The average rating for this question is 2.71 out of 5. Although this is above the median 2.5, there are still nine translators that were either “Unsatisfied” or “Highly unsatisfied.” The comments were varied. Translator 9 mentioned that due to the current economic recession the prices had gone down and that she makes less money now than when she started 15 years ago. Translator 17 commented that it depended on the task or the agency (some offer better prices than others, and this could depend, in turn, on their direct customers). Translator 20 mentioned that MT matches were poorly paid and that in her opinion it took less time to edit fuzzy matches. In this project, Translator 20 performed faster in MT but the difference with Fuzzy matches was not as high as with other translators (20.25 in Fuzzy match as opposed to 23.03 words per minute in MT match). Moreover, this is only one small project and Translator 20 has between four and six years’ experience in post-editing. Different translators have performed differently in very similar situations: one can be faster when post-editing than when reviewing and therefore a particular payment method might be better suited to that translator than to another. Pricing, however, does appear to be a problematic aspect in the view of this group of 24 translators.

How satisfied are you with the work that you do as a translator (not considering price now)?							
Answer Options	Highly unsatisfied	Unsatisfied	Satisfied	Very satisfied	Highly satisfied	Rating Average	Response Count
I am	0	1	4	13	6	4.00	24
Comments							2

Table 6. Job satisfaction

The translators give a 4 average rating for their satisfaction if price is not considered, and 19 are very satisfied or highly satisfied. So price is definitely a factor that causes some dissatisfaction at least in this particular group of translators. Only Translator 9 is clearly unsatisfied with the work done as a translator:

On the whole, I'm working on very tight schedules, with a bad organisation on behalf of many customers, in [sic] very short projects (or parts of them) that need

way too long to get ready before the actual translation, and I'm receiving no recognition for the good jobs done. Translation is no longer enjoyable to me.

This seems to be a good summary of reasons for being unsatisfied, and it suggests ways to improve the situation. These comments might be shelved as facts translators need to cope with if they work in localisation ("this is the way things are"), but we do believe a lot can be done on this front to make translators' work more satisfying (for example, in the quantity and quality of instructions). Of course, this is a small sample and a comment from only one translator, but the comments seem quite relevant: stress, too many tasks for too little compensation, and anonymity are problems to be dealt with. The other comment came from Translator 5, who enjoyed the challenges the job had to offer. Variety was an important point for this particular translator (variety is a sought after characteristic among translators as remarked in Lagoudaki 2008), who was "Highly Satisfied," and this might address the question of whether post-editing is an activity that can be done continuously throughout eight hours of work. It seems that alternated and new tasks might be a good strategy to keep translators interested and motivated.

How adequate is the standard payment of fuzzy matches in Translation Memories in relation to the productivity you obtain with them? Standard means here that you receive approximately 20 to 30% for 95-99% fuzzy match, repetitions and 100% matches (if required); 60-66% of the word rate for reviewing 75-94% matches; and full rate for 0-74% matches.							
Answer Options	Highly unfair	Unfair	Adequate	Advantageous	Highly lucrative	Rating Average	Response Count
I think it is	1	5	17	1	0	2.75	24
Comments							8

Table 7. TM pricing

The rating average is 2.75 out of 5, similar to the rating for pricing. This is logical, since currently most localisation projects involve translation memories, thus Fuzzy-match payment. A total of 17 translators found the payment adequate, although there were numerous comments stating that it really depended on the quality of the translation memory, the language combination, the text, and the type of project. Translator 17 mentioned that "I rarely get paid for 99% fuzzy matches, repetitions and 100% matches." Translator 13 thought that the payment was "Highly unfair" because:

I pay for the tools to obtain productivity, pay to learn and certify myself on their use, pay to gain experience in the field and pay for faster computers, not the clients, the benefit should be all mine.

This is an understandable point, although one could argue that customers also pay for all of these, and they often populate the translation memory

with the contributions of many translators and not just one, as well as perhaps running numerous quality verification checks; therefore, it seems logical that they also want to benefit from the use of tools. There were five translators that thought it was “Unfair.” Four of these five were not satisfied in the previous question about pricing. Translator 8 thought the pricing was “Unfair” in this particular case, although she was satisfied with the pricing in general. She mentioned, “most of the time, fuzzy matches need as much work as no matches.” Translator 5 thought that the payment was “advantageous.” Incidentally, she was also “Highly satisfied” with her work.

How do you revise fuzzy matches when working in SDL Trados or similar tool? (You need to select one option per row.) After downloading a segment...						
Answer Options	Never	Rarely	Sometimes	Frequently	Always	Response Count
I read the Source, then correct the Target segment.	0	3	2	10	9	24
I read the Target, then the Source segment, then I make the changes.	4	6	8	5	1	24
I look at the changes marked by the tool, then I correct the Target segment.	2	6	3	10	3	24
I read the Target, then I look at changes marked by the tool, then I correct the Target.	5	8	5	4	2	24
Other (please specify)						2

Table 8. Fuzzy match revision

Regarding the methodology for revising Fuzzy match segments it seems that the most common practice in the translators’ opinions is to download the segment, read the source and then correct the target while also looking at the changes marked by the translation tool. Obviously, the table shows that the translators report a combination of methods, but it appears to be less common to just focus on the target texts. Another interesting point is that there is quite a spread in the option “I look at the changes marked by the tool, then I correct the Target segment,” as we would have imagined that almost all translators would select “Always” in this option. However, it seems from these responses that reading the Source first and then the Target is more frequent.

How adequate is the payment of proposed matches in Machine Translation in relation to the productivity you obtain with them?							
Answer Options	Highly unfair	Unfair	Fair	Advantageous	Highly lucrative	Rating Average	Response Count
I think it is	1	8	10	0	0	2.47	19
Please, explain how you have been paid to post-edit so far.							19

Table 9. MT pricing

For this question, the translators were asked to explain how they had been paid to post-edit so far. The rating average for this question is 2.47, lower than the other two questions that refer to payment (general pricing and Fuzzy match payment). There are 19 responses to these questions because, as we saw above, six translators declared not to have had experience in post-editing MT output. However, one of these six translators (Translator 22) responded “Unfair” to this question, because “I don’t like the idea of Machine Translation.” Of the 18 translators with experience, ten think the payment is “Fair.” Translator 13 thinks that the payment is “Highly unfair” (as the fuzzy match payment) but she was “Satisfied” with payment in general. Perhaps she carries out other types of work where she is satisfied and post-editing is only a small part of her work (1-25 percent according to her response). Seven translators think that it is “Unfair.” They are normally paid “per word but at a higher rate than editing human translations,” “the same rate as when translating with CAT tools,” it is “dependent on the customer,” “70 percent of the word rate” and “50 percent of the word rate” or “the same rate as proofreading.” Translator 15 commented:

It depends on the client/project. Many times quality expected from postediting is the same as from human translation (this goes against the idea of postediting, btw). Some clients will ask me to take lots of things into account (terminology, style, etc.) when postediting so in the long run it is not cost effective to me.

The ten translators that think the payment is “Fair” commented that they were paid something “between the no match rate and revision rate;” “70 percent of the word rate” (Translator 8 said that it was fair because she doubles her productivity in this type of projects); “full rate;” “per hour with an agreed productivity rate reflecting the real time the task takes;” “based on the quality of the MT output;” or they were paid a rate “corresponding to a high fuzzy match” (it was then fair when the MT output was “good” but not so when the MT output was “poor”).

What would be, in your opinion, the ideal payment method for post-editing MT output?		
Answer Options	Response Percent	Response Count
Per word	41.7%	10
Per hour (with an agreed productivity rate)	50.0%	12
Other	8.3%	2
If you chose other, please specify...		2

Table 10. Ideal payment method for MT output

There is no clear answer on the ideal payment method. The two translators that selected “Other” summarised this dichotomy. Translator 15 says that per hour should be more appropriate if all requirements and expected quality are defined at the start of the project. Translator 22 mentions that it depends on the MT engine and output, and she finds there is no “universal solution”: both per word and per hour can work together but each project should be treated individually.

Do you like using MT as part of the localisation process?							
Answer Options	Strongly dislike it	Dislike it	Am Indifferent	Like it a little	Like it very much	Rating Average	Response Count
I...	5	4	8	5	2	2.79	24
Please, tell us why							20

Table 11. Predisposition to MT

The average is 2.79 out of 5. Five translators “Strongly dislike it” coincidentally three out of these five translators declared not to post-edit, so it seems natural that translators who dislike the task will not accept this type of work. During the project, however, these translators did show productivity increases when working with MT, but of course this does not mean that they were actually “enjoying” it. Four said that they “Dislike it” (Translators 4, 5, 17 and 18). Again, Translator 17 had declared that she did not post-edit, so this seems quite natural. Translator 5 dislikes MT because she believes her productivity goes down when using it. In this particular project, Translator 5 did increase her productivity, but of course this is not applicable to all of her post-editing projects and we know that in this particular case the quality of the output was high. Translator 4 commented, “Some segments are disastrous. I think this will improve with time.” Eight were indifferent (Translators 2, 3, 6, 7, 13, 14, 15 and 23) although two of these had no experience in post-editing (Translators 3 and 6). Translator 14 made an interesting comment:

I think MT is marked by current business trends. As a translator, I evaluate each task proposed in term of time and rate, and if I agree on the job proposed, I accept it. MT is a new tool and as a professional I should be acquainted with it so as not to be out of date.

Translator 23 mentioned, "I have to review the translation given by the machine the same way I review my own translation, so it is fine with me." There were seven translators (Translators 1, 8, 9, 10, 11, 12 and 19) that either "Like it a little" or "Like it very much," and they gave the following reasons: they can leverage content; it is dynamic and "physically advantageous" as it avoids having to type continuously; it helps consistency and the translation of repetitions (especially if the translation memories used to train the engine are well maintained); it increases productivity; it helps accuracy; and it is useful especially in texts with similar patterns.

3.1 Reviewers' opinions

There were only three reviewers on this project but we thought it was still important to know their opinions about the work they do, since they are not only reviewers but also professional translators.

When they were asked how they review their own work, the reviewers gave different responses. They review after finishing one segment, a file or a batch of files. They hardly ever review their work after they have completed a day's work. On the other hand, to review others, the most common method was to read the source, then check the proposal from the tool, and then implement corrections in the target (for Reviewers 1 and 3) and read the target, then the source and then implement corrections (for Reviewer 2). If they were dealing with fuzzy matches, all three reviewers read the source text and then implement the changes in the target text; two reviewers check the changes marked by the tool (Reviewer 2 does not tend to do this). All of them use either the LISA QA form or other proprietary forms (from their customers). They do not use J2450² and they always use some kind of form to report on the quality of the translations.

Reviewer 2 is satisfied with the price paid for reviewing but Reviewers 1 and 3 are not, because the price is low considering that they also have to fill in the QA forms and that they have a greater responsibility for the translation quality. Coincidentally, Reviewer 2 was "Very satisfied" with the work she does as a reviewer, while Reviewers 1 and 3 were just "Satisfied" if the price was not considered. Perhaps, the price received is influential for Reviewers 1 and 3.

Reviewers 1 and 3 think the price paid for fuzzy matches is adequate, although Reviewer 1 mentioned that they are paid according to a fuzzy match rate despite the fact that the TM is not the only reference they have to consult, and this might make the payment unprofitable if there is a high number of reference material they need to check for each segment.

Reviewer 2 finds the price paid “Unfair” due to file formats and tagging in files.

Reviewer 3 finds the payment of proposed MT match segments “Fair,” but she mentions that this depends on the quality of the output, so she states that it is not always “Fair.” Reviewers 1 and 2 found it “Unfair.” Reviewer 1 mentioned that this is due to incorrect terminology and excessive tagging. Reviewers 2 and 3 think that it is better to get paid per word for the post-editing task, while Reviewer 1 thinks it is better per hour with a productivity rate agreed upon.

Reviewers 1 and 3 dislike working with MT output. Reviewer 1 dislikes it:

[...] unless it is a very simply structured document without specific vocabulary and without tags, it normally takes more time to post-edit it than translating from scratch, and you are discounted a significant percent of the fee.

Reviewer 3 dislikes it mainly because:

[...] we cannot predict how useful it is going to be in the end. Sometimes it makes things much [more] difficult than translating without it. I cannot trust in [sic] the MT output, and the most time-demanding task in the process, that is, consulting the reference material, glossaries, etc. is still part of the job. In few occasions [sic], we have been told the MT is trustworthy and it actually was.

Reviewer 2 is indifferent.

4. Debriefings

After the translators and reviewers completed the questionnaire, we asked them if they would accept to take part in a one-on-one discussion about the assignment that would be recorded. The objective of the one-on-one interview was to add qualitative data that would help explain the quantitative data collected. The data would be gathered immediately after finishing the assignment so the ideas could be fresh in their minds. We decided to use an informal semi-structured interview with a flexible structure. The questions aimed at eliciting a conversation with the participants and listening to what they might have omitted during the assignment or we might have not asked in the questionnaire, and to see where they would place particular emphasis when giving an opinion or where they had found difficulties in the assignment or even in the translation process or the localisation industry. We offer here a summary of the translators’ and reviewers’ debriefings and not the full interview transcripts.

During the interviews, the translators made several references to machine translation. Some of them had had positive experiences, others had not.

The majority had mixed feelings about machine translation. The term “mixed” here means that translators thought that on some occasions the experience had been good and on others poor; it does not mean that they were unsure or had doubts about the MT output quality in a given project. Three translators made reference to their positive experiences because, although in some cases, especially with long sentences, the task was complex, if MT output was used correctly it could be “very dynamic,” “it is faster,” “less monotonous,” “interesting.” Eleven reported having had mixed experiences in the past. For example, they found that on some occasions the terminology was perfect, but the sentence structure was very poor, that sometimes the whole sentence had to be “reshuffled” but in other cases the result speeded up their work, or that in some cases the results were “terrible” but in others, as in this project, the results were good; that sometimes “you don’t have to intervene really but other times you go crazy,” or that sometimes MT is better, sometimes it is worse, so that MT becomes difficult to quantify.

Translator 15 made an interesting comment: “In general, machine translation does not need to be perfect but only understandable.” In his view, MT was more beneficial to him financially if the quality requested was “understandable” and where he did not “have to worry too much about the style,” that is, if customers used MT for material that was not highly visible. On the other hand, if the quality expected was very high, then, he felt he had to make many changes (style and terminology) and it became unprofitable. In other words, it seems that Translator 15 preferred to use MT for “fast post-editing” rather than for “full post-editing” (Allen 2003), and that customers should be more flexible in their style and terminology requests if they are using MT in their localisation process.

Translator 20 made a relevant remark regarding tags in the documents: “You work with the variable tags and you always have to touch the segments, you always have to change the order of something.” In our experience in the commercial sector, translators often complain that with a heavily tagged document it is easier to work from the source text and not from a proposed text where tags need to be rearranged completely in each segment. Translator 20 also commented in the questionnaire, “I find the process unfairly paid at times, and I miss the creative feeling, even if it is software manuals, that translating from scratch brings.”

Translator 23 summarised well the feelings that some translators expressed during the debriefings, particularly in relation to the varying quality of MT. She also expressed an interesting opinion on this type of study:

[...] a lot of importance is placed on time employed in doing the job and I think this sometimes goes against the translator because there are sentences that are easier than others, or depending on the translator's experience he would go faster or slower [...] I think that machine translation should be considered from the linguistic point of view almost exclusively.

Although doing research or simply measuring the use of MT from a time point of view is insufficient, we cannot negate the fact that the use of MT is directly related to speed (reaching markets more quickly), volume (more content in more languages) and saving costs. We agree that time on its own only tells part of the story (as we clearly saw in the quantitative analysis of this project) but time is nevertheless an essential part of this story. As Translator 23 remarks, analysing MT is a very complex topic and many factors are involved such as quality of the output, experience, training, purpose of the post-editing job or even quality of the "translator."

There were four translators that gave clearly negative feedback about MT (Translators 14, 17, 20 and 22). They had also responded in the questionnaire that they were "Indifferent," "Dislike" or "Strongly Dislike" using MT as part of the localisation process. There were several reasons: "projects are full of instructions and a lot of glossaries to follow," "technical aspects that present obstacles," proposals are "so bad" that the translator had to return the assignment, the rate is lower than the effort required, or segments have to be completely redone. Translator 20 openly said that she was not a "fan of machine translation":

I thought I was working for little money for the time I had to invest in that type of translation and then also it is a personal preference because I don't like revising in general. I prefer a process of creating from zero, to translate... There might be people that prefer to revise, I don't know... There was this customer that I'm thinking of right now that pays us the same rate as *Trados fuzzies* [English in original] but really I can't tell you if the effort is equal, lower, higher but I have the impression that I have to stop more and I don't trust it as much.

It is interesting to see that Translator 20 does not like revising and prefers to create something from zero, and this is regardless of her productivity when doing so or the rate. Also, she does not know but she feels her effort is higher when post-editing than when editing TM fuzzy matches, although there is evidence (O'Brien 2006, Guerberof 2008) that MT correlates well with high ranges of TM fuzzy matches in terms of "time" effort. We could hypothesise that, for translators, if the cognitive effort is higher with certain MT segments (Krings 2001, O'Brien 2006), their perception of the whole post-editing exercise is that it takes longer, although this might not actually be the case in terms of temporal effort. Finally, Translator 20 stated that she trusted a Fuzzy match segments because it comes from a human translation but did not trust something

that came from a machine. We have already noted that some translators might trust fuzzy match segments more readily (Guerberof 2008).

In general, translators show some knowledge of how machine translation works and its error typology. For example, "some of the errors had to do with problems with structure, order, that tend to be typical of machine translation;" "MT is useful in some cases and not in others." In short, from the debriefings it was obvious that translators had experience of this task but also that they had a professional outlook on the topic rather than an emotional one. For example, Translator 14 presented an interesting view of the current situation for translators and MT:

I think that apart from the fact that you might like it or not, that you feel comfortable or not, these are the trends in the current market and we have to get familiar and up to date because it is what is being used at this moment. So, many times, one prefers other types of work but if you are not up to date and learn new tools and up to date with machine translation, the current market now, you are left out. This is, I think, a reality.

In conclusion, the group was highly familiar with machine translation and their attitudes were open and flexible. This does not mean, however, that they liked using MT. This signals a change with respect to previous views on how translators perceive MT where translators are seen as very reluctant to adopt MT as a working tool. In a 2010 survey about Post-editing, TAUS mentions translators' resistance as one of the main pains in post-editing management. Later, in their post-editing report (TAUS 2010b), they try to explain this resistance by suggesting that post-editing requires a higher cognitive load than translation and therefore it would be understandable for translators to show some kind of resistance. They also explain that for translators, dealing with MT is "similar to the emergence of TM tools in 1990" (TAUS 2010b: 15) and would be like dealing with a poor TM tool. Evidently, the opinions in the survey are from companies engaging in this type of activity and how they perceive translators' attitudes; this was not a survey designed to gather information from translators. The results might have been completely different if translators had been asked. Also, we are not entirely sure - and translators in this project do not seem to be either - that dealing with poor TMs is the same as post-editing MT. In our project, the translators seem to have had a very practical and open attitude towards MT, although some did not like working with it for different reasons. Tatsumi (2010: 185) has already commented on this in her doctoral thesis: "the answers to our questionnaire suggest that a flexible and down-to-earth attitude towards PE [post-editing] is the trend," and this was also the case in this study. Also, Lagoudaki (2008) in a survey conducted about the value of MT for the professional translators concludes, "machine translation appeared to be well received amongst translators who were familiar with it" (*ibid*: 265)

and also “translators also seem to be coming to terms with machine translation as an alternative means of translation production” (*ibid*: 268). We asked the translators which aspects of the questionnaire they found difficult to answer, if any. The main issue translators had when completing the questionnaire was answering the questions related to rates, since “sometimes you are more satisfied than other times,” “it is relative because not all the jobs in machine translation take the same amount of time.” In brief, translators reported that the question was too general, given that satisfaction with rates depends on customers, subject matter, nature of task (if it involves machine translation, for example), actual rates and the financial situation in a given country. Translator 17 explained some of the complexities regarding rates paid:

[...] it is very difficult to generalise on this topic [...] I work for several translation agencies and sometimes for publishing houses with different rates [...] in Argentina there is a range of prices that is huge and in general it depends on how many intermediaries there are along the way. [...] sometimes there are a lot of intermediaries, and what arrives to the freelancer is a low rate...

Another difficult question was the one referring to how they review their translations. Some translators found this difficult to define as it depended on the peculiarities of the projects, and others were not aware of the revision method used. Translator 18 explained “Sometimes all [the translation process] is so automatic that I don’t reflect on the actual process.” We saw, however, that this translator did have problems in the whole process (high number of errors, not understanding instructions and not following the glossary).

Some translators that had not performed post-editing tasks previously found those questions difficult to answer, understandably so (in fact, they were asked to answer them only if they had had experience). Translator 24 mentioned that he found it difficult to say if more effort was required to revise human translations or post-edited material (he responded that more effort was required to post-edit).

There were three questions in the debriefings directly related to the segments they had post-edited: one was related to their awareness of having MT segments, another to the differences in the segments in terms of type of edits, and the third to difficulties found in certain segments during the task. Thirteen translators realised that MT output was involved in the task, either when they received the instructions or when they started working on the assignment. Before starting the task, five translators knew or strongly suspected that there would be MT output. They imagined this when they received the emails to participate in the project, despite the fact that it was not mentioned, because the email said there would be Fuzzy matches. Although Translators 12 and 15 were among this group of translators and they performed very well in terms of

errors, the other three translators had an average or poor performance, so we cannot suggest that knowing a priori the exact nature of the task was an advantage or that it resulted in better quality results. Still, with the instructions and during the task, they all became fully aware, mainly because of the “changed structures,” “expressions where it was clear they had not been translated by a person,” the “very literal translation,” or “word order” and the fact that the instructions mentioned to correct only errors that they were certain about and “not [to correct] style issues.”

Translator 17 was not aware that the project involved MT, as she explains:

Really, the perception that I had was not that it was machine translation but that it was a translation memory. Except some segments that clearly, well I imagined they were modified to see the correction made or if the error was noticed.

This is due to the overall high quality of the MT output. Notwithstanding, some segments were poor. Translator 17 was not fully aware that she was dealing with MT and thought the segments were seeded with errors. Still she had a low number of errors (58 in total, aggregated value from three Reviewers) and she was in Speed group 3 (the fastest group). Fifteen translators perceived differences between the segments; six clearly stated that there were Fuzzy and MT matches. Translator 11 even thought that he was able to identify the type of match:

[...] I noticed the difference was mainly similar to any fuzzy segment in any translation memory that shows, for example, a segment with a 95% match. Everything is the same but there is one word or one section of the document that does not coincide with it, precisely because the update was not done with respect to the new source.

The others perceived differences in terms of quality of the segments: some were very good, and others were poor (although they specified that poor quality was rare). Some translators seemed to imply that those segments that were “very good” “belong to a human,” as Translator 22 explained:

There were some that were very good, you could say that they belong to a human, but others were obviously from a machine.

Others mentioned that if this was MT output then it was “very good.” It is interesting to see that some translators assumed that the proposed texts that contained fewer errors were human translations when in fact many MT segments were not changed by translators because they were of acceptable quality. A similar experience reported in He *et al.* (2010) where post-editors mistake MT outputs for TM outputs. We have no way of knowing how translators perceived each individual segment and whether the segments they thought were human were indeed human and not MT.

The majority of translators in this group were familiar with certain type of MT errors (word order, wrong structure) and they appeared to be able to identify these segments very quickly but when it came to segments that flowed well they might have assumed that these were human translations. The other four translators did not think there was a clear difference between the segments, although they might have thought that some were better than others in terms of linguistic quality.

Seven translators made reference to the overall quality of the segments: some mentioned that it was “pretty good,” “fairly acceptable,” “very good quality,” “fairly good,” “a high percentage of what was already translated was better than what I expected, really”; another mentioned that, “segments that were longer and more complex you had to almost completely change them, they were not the majority,” or that “I had to read again the source and then rewrite, reformulate practically the whole sentence... and there were also many segments that were perfect.”

The translators were asked how they felt during the task. Ten stated that they had liked the task because “it was interesting,” “because it isn’t what I normally do,” “it was something enjoyable,” “I liked it quite a lot,” “I was positively surprised” (especially with regard to their perceived quality of the MT). Translator 4 was particularly pleased for the following reasons:

To be honest, very good. It was a pleasure. I’m very used to this type of translations, all that is software, etcetera. And to be honest the task was very good, mainly because we had all the material. If there was anything to consult, any terminology to consult and it was not in the glossary, I took the Microsoft terminology databases as terminological reference but we had all the tools to be able to do it, and the instructions were very clear.

The translators seemed to be pleased that the task was short, uncomplicated and at the same time it was outside their normal routine. They also found it interesting because they felt they were involved in a research project that involved acquiring knowledge about the profession. Seven translators were quite neutral in their comments. Some felt it was another job, “like a normal project, like the ones that are normally done for example with Trados,” “it is very similar to what I do as a professional,” “I felt comfortable because it was very similar.” Others felt they had not experienced any particular problem; and still others thought that although they had to pay more attention: “it was not that horrible.” Finally, Translators 21 and 22 did not like that the task because they found it either tedious or they did not like working with machine translation. In their own words:

A translation of this type, I found it tedious, it takes a long time and in the end you end up retranslating almost everything in the end, at least in my case. I haven’t done it a lot, but the times I have done it, this was my experience, no, no, I wouldn’t like to do this daily and with large projects. Yes, I think it was tedious. (Translator 21)

A bit uncomfortable, because I insist, I don't like working with *machine translation* [English in original]. Even if this one I could tell it was of very good quality, but no, it isn't something that I like. (Translator 22)

Translator 22 also mentioned in the questionnaire, "One never can trust 100% [sic] on a Machine Translation." It is interesting that these two translators also had quite a high number of errors (Translator 21 had 80 errors in total and Translator 22 had 98) and that they had fewer errors in Fuzzy matches than in the other two categories: MT and No match, and they had no post-editing experience. It would be interesting to know whether not liking the task had any influence on their performance, but of course, we would need to test them doing a task they liked doing, for comparison purposes, and that was beyond the scope of this project. What we can say is that other translators that *did* like the task had a similar number of errors or equal speed. Therefore, in this particular project, we cannot establish any correlation in this respect.

5. Conclusions

All translators except one were satisfied with the work they do, but not necessarily with the payment they receive for different tasks, although this was highly dependent on different customers. The payment for Fuzzy and MT matches might be inadequate if the quality of the translation memory or MT output is poor and the translator has to invest more time in fixing those segments than if they did the translation from scratch, while they are paid only a fraction of the word rate for translation. It was not clear if they prefer payment per hour or per word, but translators did indicate the need for a payment related to the quality of the MT output or TM, or to the nature of the task requested. The methodology for reviewing (texts and fuzzy matches) tends to be to open the segment, read the source, apply changes to the target and check the tool to see the changes marked. Before handing back the files, these translators would recheck the batch of files received. Reviewing after a days' work or after finishing a file is less frequent. There were several problematic issues that they signalled in the translation process: the excessive number of instructions to complete small tasks, terminology maintenance, excessive reference material, and tagging in documents that force translators to rearrange every single segment regardless of the level of fuzzy match or quality of the MT output.

This group of translators was in general quite familiar with machine translation and post-editing, but not all of them were performing these tasks on a regular basis. They could identify clear MT segments and knew what to change in those cases. Although some did not like doing post-editing, mainly because the quality of certain MT segments was poor or

the instructions too cumbersome to follow, or they did not like to review, the overall attitude was nevertheless flexible and practical. The translators that dislike post-editing would in general not perform the task, and those that post-edit find that experience helps them spot errors and that in some cases it increases their productivity. Post-editors do not feel that they grow accustomed to MT errors or that their productivity decreases over time. Most find that post-editing requires either similar or more effort than editing human translation, and this could refer not just to a higher cognitive effort for this task (not necessarily a temporal effort), but also to the fact that each translator might have different experiences with previous post-editing jobs and might also perform differently because of their own personal characteristics. Also, many were aware that post-editing will be a necessary task in the future of localisation and that outputs will improve over time. From this group of professional translators we can see that those doing post-editing are well-informed about the process and the current shortcomings. We do not find a negative attitude towards working with MT (although the majority of translators might dislike it) but rather problems with how the task is paid or organised. The reviewers, like the translators, were not satisfied in general with the payment they received for the task of reviewing, and they thought that for Fuzzy and MT matches it depended on many factors but it was sometimes unfair. They either dislike or were indifferent to working with MT, mainly because it was highly dependent on the quality of the MT output, the tagging in the text, and the quality of the terminology proposed.

Acknowledgments

The research project this article stems from was funded by the European Association of Machine Translation (EAMT) and the Centre for Next Generation Localization (CNLG). I would also like to thank Crosslang, MicroStrategy and hiSoft Technology International Limited for their key contribution to the project.

I'm especially grateful to Dr. Anthony Pym and Dr. Sharon O'Brien for their continuous support on the completion of the doctoral thesis that this project is part of.

Bibliography

- **Allen, Jeff** (2003). "Post-editing." Harold Somers (ed.) (2003). *Computers and Translation: A Translator's Guide*. Amsterdam/Philadelphia: Benjamins, 297-317.
- **Englund Dimitrova, Brigitta** (2005). *Expertise and Explicitation in the Translation Process*. Amsterdam/ Philadelphia: Benjamins.
- **Guerberof Arenas, Ana** (2008). *Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation*. Masters Dissertation. Universitat Rovira i Virgili. https://docs.google.com/open?id=0B_cVNsnGLfvMjc5MDc3ZWYtZjRiZC00NWZlTlhmTEtNTEzOWIwODY0MDc1 (consulted 15.05.2012).
- **He, Yifan, Yanjun Ma, Johann Roturier, Andy Way and Josef van Genabith** (2010). "Improving the post-editing experience using translation recommendation: A user study." *Proceedings of the 9th Annual AMTA Conference*. Denver, 247-256. <http://doras.dcu.ie/15803/> (consulted 15.05.2012)
- **Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin and Evan Herbst** (2007). "Moses: Open source toolkit for statistical machine translation." *Proceedings of the 45th Annual Meeting of the ACL on Interactive Demo and Poster Sessions*. Association for Computational Linguistics. Prague, 177-180.
- **Krings, Hans** (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. G. S. Koby (ed.). Ohio: Kent State University Press.
- **Lagoudaki, Elina** (2008). "The value of Machine Translation for the professional translator." 8th AMTA Conference. Hawaii, 262-269. http://www.amtaweb.org/papers/3.04_Lagoudaki.pdf (consulted 20.06.2012).
- **O'Brien, Sharon** (2006). "Eye-tracking and Translation Memory matches." *Perspectives: Studies in Translatology* 14 (3), 185-205.
- **Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu** (2002). "BLEU: A method for automatic evaluation of machine translation." *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Philadelphia, 311-318. <http://acl.ldc.upenn.edu/P/P02/P02-1040.pdf> (consulted 20.06.2012).
- **Tatsumi, Midori** (2010). *Post-Editing Machine Translated Text in A Commercial Setting: Observation and Statistical Analysis*. PhD Thesis. Dublin City University. <http://doras.dcu.ie/16062/> (consulted 06.03.2012).
- **TAUS**. (2010a). "TAUS Research. Results post-editing survey." <http://www.translationautomation.com/articles/> (consulted 20.06.2012).
- — (2010b). "Postediting in Practice. A TAUS Report." <http://www.translationautomation.com/articles/> (consulted 20.06.2012).

Biography

Ana Guerberof Arenas has over 20 years of experience in the localisation industry and she is currently Director of European Operations at Pactera. Ana holds a PhD in Translation and Intercultural Studies from Universitat Rovira i Virgili on Productivity and quality in the post-editing of outputs from translation memories and machine translation.



ana.guerberof@gmail.com

Notes

¹ The Localization Industry Standards Association (LISA) was an association dedicated to the creation and implementation of standards for the localisation industry. They were also responsible for the creation of the LISA QA Model for the evaluation of localised project quality. Although LISA closed in 2011, the review form or QA model originally created by them is still widely used in the localisation industry.

² J2450 is a quality metric created by the Society of Automotive Engineers (SAE) for language translation of service information to be used by automotive companies to compare quality of translation deliverables. Nowadays it is also used in other domains of the localisation industry to assess linguistic quality.