



What we know and what we would like to know about post-editing

Dr Ana Guerberof Arenas
Universitat Rovira i Virgili/hiSoft International

Heidi Depraetere
Crosslang

Dr Sharon O'Brien
Dublin City University/CNGL

ABSTRACT

This article presents the view of three professionals and researchers in the translation field about the post-editing task with a view to examining what it is known about this relatively new task both in the localization industry and in the academic arena, as well as highlighting what should be investigated further in order to shed light on different aspects of post-editing and the use of machine translation.

Keywords: machine translation, post-editing, MT, translators, translation, post-editors, research

RESUM (*Allò que sabem i allò que ens agradaria saber sobre postedició*)

Aquest article presenta la visió de tres professionals i investigadors del camp de la traducció sobre la tasca de la postedició. L'objectiu és determinar què se sap sobre aquesta relativament nova tasca, tant per a la indústria de la localització com per a l'àmbit acadèmic. A més, es proposa posar de manifest aquells elements sobre els quals caldria fer més investigació per tal de posar el focus sobre diferents aspectes de la postedició i l'ús de traducció automàtica.

Paraules clau: traducció automàtica, postedició, MT, traducció, posteditors, recerca

RESUMEN (*Lo que sabemos y lo que nos gustaría saber sobre posesición*)

Este artículo presenta la visión de tres profesionales e investigadores del campo de la traducción sobre la tarea de la posesición. El objetivo es determinar qué es lo que se sabe sobre esta relativamente nueva tarea, tanto para la industria de la localización como para el ámbito académico. Además, se propone poner de manifiesto aquellos elementos sobre los que habría que realizar más investigación para arrojar luz sobre diferentes aspectos de la posesición y el uso de la traducción automática.

Palabras clave: traducción automática, posesición, MT, traducción, poseedores, investigación



Introduction

Heidi Depraetere has over 20 years' experience in the localization and language technology industries. She is a founder and director of Crosslang, a Gent-based consulting and systems-integration company dedicated to translation automation technology. CrossLang is a leader in independent translation automation solutions and works for a growing number of global companies. Dr Sharon O'Brien worked in the localization industry before becoming a full time professor at Dublin City University where she has done extensive research on controlled language and post-editing. She is a member of the Centre for Next Generation Localization. Dr Ana Guerberof has just finished her thesis on productivity and quality in the post-editing of machine translation and translation memory outputs. She also has 20 years' experience in translation and she is currently Director of European Operations at hiSoft International.

Questions & Answers

1. What are in your opinion the most interesting findings (in academic research as well as in the commercial environment) in post-editing in recent years? Why are these findings relevant?

Heidi: The fact that post-editing is now on the research agenda tends to confirm that machine translation (MT) is being adopted. The skill set required for post-editing and its impact on the translation output are two issues that need to be addressed and assessed in-context. However, it is important to develop a methodology that gauges the effects objectively. This is by no means an easy task as there are many factors and parameters to be taken into consideration.

Sharon: That MT + post-editing (PE) productivity can be equal to or greater than productivity for editing fuzzy matches, while still adhering to quality guidelines. This is relevant because there seemed to be a general assumption that MT+PE could never be as fast as translation memory (TM) editing and that, if it was as fast, the quality would be sub-standard.

Ana: Obviously, one relevant finding is that the task of post-editing within the localization workflow works as a translator's aid rather than hindrance if the raw MT output quality is high. This implies a perception change towards MT and its applicability. Moreover, MT output could be compared to a high fuzzy match in terms of editing speed or resulting quality if the engine generating this output undergoes proper training. This is not unusual now but when research on this area started, this was not an obvious conclusion. In my personal work, the most interesting finding was probably the inter-subject variability when looking at the productivity and the quality produced by the post-editors. This is relevant because it signals the difficulty we might experience when measuring time and number of errors in translators adding to the complexity of setting up prices and defining quality at the same time. Also, it makes observing top performers interesting in order to perceive what they are doing, that others are not, that is working so well. The human factor is so important in machine translation from training the engine to the actual post-editing and this is sometimes disregarded. Another interesting finding is the inter-segment variability when looking at cognitive effort and also number of edits and its relationship with time. Some segments might have few edits but require substantial amount of time, while others with more edits require less time.



2. Several studies (O'Brien 2006, Plitt and Masselot 2010, De Sutter and Depraetere 2012, Guerberof 2008 and 2012, and many reports from the commercial sector) report that post-editing MT output can reach productivity levels similar to those when editing high level fuzzy matches. Can you explain under which circumstances this is achievable and under which circumstances this is not be feasible?

Heidi: For obtaining these productivity levels it is vital that the MT output is customised. Low quality MT output is not realistic for achieving productivity gains and will only result in frustration and aversion to integrating MT in the translation production process. The quality level must be such that it is perceived as a genuine translation aid.

Sharon: This is only achievable if the raw MT output quality is high enough in the first place to allow for this kind of productivity. Of course, the quality will differ from segment to segment, so we have to talk about the overall quality of the sample here. I also think this is achievable when the post-editor has an open attitude to MT. A negative attitude may well impact on productivity.

Ana: As I said before the key is the output quality. What these studies have in common is the high quality of the trained engines (both SMT and RBMT engines). This means that an engine is specifically trained with a substantial amount of bilingual data. This facilitates the speed gained by the post-editors in the different languages. Also, in these particular cases I believe most of the post-editors were professional translators and this also might increase the speed. Still, in some cases, not all translators performed at the same level, as I mentioned before. This means that some translators might not avail of this type of productivity gain when using translation aids. Further, some translators complain about the low quality level of the MT output in certain projects they have worked with. This means that not everyone in the industry is using MT as would be desirable or advisable, but the same level of discounts are expected as those given for high quality output.

3. What research or commercial trend in MT or post-editing is likely to have most impact in the forthcoming years?

Heidi: For commercial use MT is probably best seen as a pipeline process. Producing high quality MT output is important for increasing translators' acceptance of the post-editing task. In order to achieve "post-edit friendly output", on top of the customised MT output, additional text engineering routines and processing tasks will be needed, resolving customer specific terminology, reflecting corporate language adaptation and resolving tag handling issues.

Sharon: I suspect that the automation of post-editing will have a major impact in the coming years.

Ana: Technology changes very quickly and it is difficult to predict. However, I find the research on confidence scores could potentially have a great impact. Basically, these scores aim at giving an indication from the MT engine of the ranks of different candidate translations, thus emulating a fuzzy match level in a translation memory. If this is refined it can speed up the process of discarding segments and preparing the post-editors for the task ahead, thus lowering the frustration level and I presume the cognitive effort. Also, speech-to-text technology can play a fundamental role as part of a localization process. It is not easy to envisage how right now, but using voice together with MT could speed up the process tremendously. I have not seen the application in localization or translation yet, but I wouldn't discard it. Commercially, I think the trend is to extend the use of MT to more languages and domains (verticals) making engine training more user friendly in a way that would allow language service providers (even individual translators) manage their own MT process, thus replacing the way in which translation memories are used at a fuzzy match level. We are



seeing this happening right now. Perhaps the two technologies will merge, so that the 100 per cent matches from a TM can be used throughout, and then, depending on the confidence score, MT or a high fuzzy match TM can be applied with some interaction on the part of the translator.

4. Very few universities offer post-editing as part of their Translation BA or Masters courses (although this might be on the increase). Do you think this should be an area covered in a Translation BA or Technical translation/Localization Masters course? What do you think are the most important aspects that a post-editing course should cover today?

Heidi: A course can be developed along the lines of a course that teaches students how to revise translations. Background knowledge of the different MT technologies may be useful as well as the post-editing context. It is important to include hands-on assignments that give students the taste of MT output, error types and correction techniques. As I see it, too much focus goes to levels of post-editing. Post-editing is likely to be one of two kinds, either the client gives a strict set of instructions as to what needs correcting or requires full human quality.

Sharon: Yes, I think that students should be aware of trends in the industry and MT + post-editing is an important trend these days. Important aspects are: How to gauge effort (and a fair price) vis-à-vis the quality expected; How to minimise the technical (keyboarding) effort; how to interact with MT technology.

Ana: It would be wise to include it at a later stage in their training, at least briefly with not only practical exercises but also with some guidelines so students can have their own criteria when dealing with this type of text. They would need to negotiate discounts, so they should have an idea of what they do and how they do it. Before the post-editing training, I think translators should follow a course on "editing" (reviewing) texts translated by humans. I have not seen a lot of universities that offer editing modules and it is so important and practical since translators will end up editing more than they will ever want. Also it is a good basis from which to build on post-editing. If the student has already a localization background, then the course should cover basic MT and post-editing concepts and definitions, error typology, quality definitions and expectations, post-editing guidelines, and practical exercises, for example.

5. In your experience, what is the attitude of young students/graduates towards MT and post-editing?

Sharon: They are generally open to it and willing to give it a try; some find that they like it; others find that they prefer traditional translation. As they are usually focussed on trying to get jobs, I think they are willing to be flexible about the use of technology.

Ana: I have only been exposed to Masters students in Spain and I have observed a variety of attitudes. Some students are already using MT as freelance translators (sometimes without their customers knowing), others are positively surprised by the results when applied in a simulated working environment (they might have used MT only to find out mistakes and not really to look at how it can be useful), others are sceptical about the use of MT in translation (thinking that a machine will never offer the same quality as a human translator). Over the years, the use of MT in translation has been "normalized" and I can perceive that students are more familiar with its use and application. However, when we discuss quality it is always controversial and interesting points of view emerge, those who believe that quality is only linguistic quality and more particularly their own personal view of quality, and those that are



more focused on what the customer wants. A lot of them realize that it is necessary to learn about post-editing but they would prefer to translate without it. After all, this is the reason why they enrol in translation (in general).

6. Do you think that the current pricing model (applicable to translation memories) would be applicable to MT and post-editing or is this model likely to change?

Heidi: MT output fits in with traditional fuzzy match categories, so traditional pricing models can be applied. Currently research is undertaken to establish correlations between MT matches and TM matches.

Sharon: I think it could be applicable, but it could also change to a more time-based or effort-based model. We will have to wait and see what the market decides.

Ana: The TM model is being applied somewhat. MT is presupposed to give a certain level of match and a discount is applied. Sometimes with little evidence, which is worrying, but then again I have not seen much evidence in the application of TM matches either, apart from the fact that it has been applied over the years. I think the model is likely to change as it becomes more difficult to know which segments are being used and which ones are discarded, and to which extent the segment is used. Some companies apply the pricing, for example, after the job has been completed according to the level of usage/effort, but this does not appear to be totally transparent to translators and I have heard some complaints. Transparency and clarity are key factors in whatever method is applied. And above all, and this might seem obvious, translators cannot make less money than they do when translating on their own, because prices have remained almost stagnant for over 10 years, and individual translators do not have much room for discount anymore. If we consider that translators have a lot of inter-subject variability, setting up prices is more difficult than it seems. It is true that variability already existed in translation and that the industry has agreed standard throughput figures. However, now we are not only calculating a price based in a daily throughput (and then a translator can go slower or faster for the same amount of words) or even a gain according to a match with a human translation, we are assuming a percentage of benefit for all translators in a series of segments (of variable quality), and this factor adds to the complexity of setting the prices.

7. In the new Trados Suite 2011, SDL MT engine is available to translators for free. This means that post-editing can be included in the localization process almost by default like an on-line glossary or terminology look up. How do you think technology will support the post-editing task in the near future?

Heidi: Seamless integration of MT into translation editors is a key requirement for establishing MT as a productivity aid just like TM and therefore does not require additional post-editing tools. It is important that translators can continue working in their preferred translation editing environment.

Sharon: I think that editors will become more sophisticated so that post-editing will be less tedious. I think also that editors will have built-in preference filters so that individual translators can decide where their thresholds lie in terms of quality (e.g. do they want to see all fuzzy matches first, followed by MT, or are they willing to look at MT proposals if there is no very high fuzzy/exact match; will they take provenance data into account or not (who created the match, how long ago etc.); I also believe that editors will make use of confidence scores coming from the MT system and that these will be used in the preference filters by translators; finally, I believe that technology will allow for real-time re-training of MT engines, as machine learning becomes more sophisticated.



Ana: As I mentioned before, I think technology will merge MT and TMs and terminology lookups as Trados 2011 or even further, and the translator will interact and choose the best option. The proprietary issue is something that needs serious consideration and debate because even if translators sign non-disclosure agreements, it is not very transparent what they do while working if they have free access to these tools, they will use MT to speed up the work. It would be desirable to see more tools supporting post-editing tasks specifically rather than just offering the MT proposal.

8. If we consider post-editing, what are the areas where you see more research is needed from a translation studies point of view? And from a commercial point of view (if different)?

Heidi: More research focusing on identifying recurring error patterns so automatic post-editing routines can be applied before the post-editor steps in.

Sharon: How to better support the PE process through better user interfaces and data filtering (see 7 above).

Ana: It would be interesting to know more about where translators employ higher brain power when post-editing and where they employ less. This would be good for MT developers and for companies. Analysing how a top performer (meaning optimum time vs. quality) operates to see what suggestions or advice might be given to other translators. From a commercial point of view, more information on productivity is needed at a larger scale into more languages, and its correlation with the initial quality of the engine. For example, set correlations between MT quality automatic metrics and the time gain by translators. This will make it easier to define discounts for different languages according to the real utilization of the engine.

9. Why do you think research in post-editing is useful and who benefits from research in this area?

Heidi: As mentioned before research in post-editing confirms the uptake of MT in commercial environments. This is probably one of the areas where industry and academia are closely working together since objective measurements are needed not only in terms of potential productivity gains but also in terms of impact on quality.

Sharon: When there are prejudices and assumptions about a specific technology, I think the only way to address those is through high-quality empirical research. When there are enough results to demonstrate that a misconception is just that – a misconception – then the community will be ready to move on and tackle the new phase. Although technology is sometimes seen as a 'necessary evil' in translation, research benefits all members of the community, the translators and the clients.

Ana: I think research in this area is very useful and everyone in the community benefits: Translators, translator trainers, translation agencies, MT developers, and translation buyers. For example, something as simple as identifying post-editing as a task where translators' "behave" differently can cause agencies to be very specific about what is needed and create specially designed guidelines and influence the way tool developers design features in their own tools. Knowing that translators can increase productivity under certain circumstances (quality of the raw output) helps implement the use of MT across the industry, knowing when it works well and when it can be discarded. All is done through empirical research or some kind of "measured" experimentation, preferably in an industrial setting.



10. What steps would you advise translators to take if they want to become post-editors?

Heidi: To look at MT as progress and keep an open mind – high quality MT matches do not crucially differ from the traditional fuzzy matches. It won't be long before MT will be accepted as a routine aid just like TM came to be accepted in the eighties.

Sharon: Becoming a post-editor should not be too difficult in itself. It requires a bit of practice, but it is not rocket science! What translators need to do is to ensure that they know about the benefits and limitations of the technology and to use that knowledge to their advantage. So, rather than blindly dismissing MT, they should engage with it and educate themselves so that they know more than their clients and can even advise their clients on its use.

Ana: I would advise them to first learn a little bit about how an engine is trained and customized (this will help to come to the task with realistic expectations), then to make sure they know something about the particular project assigned to them (level of quality of the raw MT output is important to know); to make sure they have the right set of instructions to know what errors to correct and clear guidelines on expected quality, also that they measure productivity when using MT over a period of time to see how beneficial it is for them. If it is not beneficial or the benefit is not as high as the discount offered, they should speak out and let the customer know with concrete figures. Discounts will not change if one translator complains but it might if more than one complains about a particularly poor MT output. Communication is a very important aspect especially now that there is little knowledge and MT has so many variables (maybe the output is okay but the terminology is outdated, for example).

Conclusions

MT and post-editing increases the productivity of translators and functions as a translation aid in a similar way as a translation memory as research has shown, but this is only achievable if the quality of the initial raw output is high. To achieve this, MT engines need to be trained and customized specifically and carefully. The full implementation of MT in the localization process is a reality for many companies and its implementation will only increase with time. Important developments in the future are: post-editing automation, tool development to include specific features and filters for post-editing, and terminology handling with custom-built domain engines and tag handling. In other words: full integration of MT and TM technologies through powerful interfaces. Although research has been a key element in this area so far, more is needed in areas such as: error patterns to apply automatic post-editing, user interfaces and data filtering to better aid the post-editing task, more data on productivity across different settings and correlation testing between automatic metrics and PE effort. Pricing seems to be applied in a similar way to translation memories, but there is a possible trend towards setting the pricing according to the effort applied in a given project (percentage of segments used). All agree that post-editing should be a topic taught at university because even though it is not an extremely difficult task, it does require an open-minded attitude, certain knowledge and practice of how MT works: both benefits and limitations. With this knowledge translators will be able to use the technology to their own advantage and be able to communicate needs and shortcomings to the relevant players in the industry.

References

De Sutter, N. Depraetere, I. 2012. "Post-edited translation quality, edit distance and fluency scores: report on a case study". Presentation in Journée d'études Traduction et qualité Méthodologies en matière d'assurance qualité. Université Lille 3. Sciences



humaines et sociales. Lille. Available from http://stl.recherche.univ-lille3.fr/colloques/20112012/DeSutter&Depraetere_2012_02_03.pdf Accessed May 2012.

- Guerberof, A. 2008. Productivity and Quality in Machine Translation and Translation Memory outputs. Masters Dissertation. Tarragona. Universitat Rovira i Virgili. Also available from https://docs.google.com/open?id=0B_cVNsnGLfvMjc5MDc3ZWYtZjRiZC00NWMzLTlhMTEtNTEzOWlwODY0MDc1
- Guerberof, A. 2012. Productivity and Quality in Machine Translation and Translation Memory outputs. PhD Thesis. Tarragona. Universitat Rovira i Virgili.
- O'Brien, S. 2006b. "Eye-tracking and Translation Memory Matches" Perspectives: Studies in Translatology. 14 (3): 185-205
- Plitt, M. Masselot, F. 2010. A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. The Prague Bulletin of Mathematical Linguistics. Prague: 7-16. Available from <http://ufal.mff.cuni.cz/pbml/93/art-plitt-masselot.pdf> Accessed June 2012.