Towards Intelligent Post-Editing Interfaces

Sharon O'Brien, PhD & Joss Moorkens, PhD

Dublin City University

Sharon.Obrien@dcu.ie
Joss.Moorkens@dcu.ie

1. The New MT Paradigm

The increasing adoption of Machine Translation (MT) as a type of computer-aided translation technology is arguably as disruptive and revolutionary as the disruption caused by the introduction of translation memory technology in the 1990s. Although MT has been in use by some organisations for a long time, for example the Pan-American Health Association (Vasconcellos 1983), recent breakthroughs in data-driven computational techniques have resulted in a proliferation of MT engines that can produce a level of quality that is either acceptable to end users who wish to get a gist of the information contained in the source content (known as MT for assimilation) or that can be edited (or "post-edited") by professional translators to produce publishable-level quality (known as MT for dissemination).

MT for dissemination purposes is being used increasingly, especially in the IT localisation sector. It has been reported repeatedly at localisation industry conferences, for example, that multinationals such as Dell, Symantec, Adobe and Autodesk, to name just four, have integrated MT into their localisation processes, and a survey of almost 1000 language service providers in 2010 found that 42% offer post-edited MT to customers (DePalma and Hegde 2010). It is clear that MT is not being used for every translation domain, nor will it be in the foreseeable future, but its increase is notable and has important ramifications for those working in these domains, not least professional translators who are now expected to include "post-editing" as a service.

While post-editing used to be viewed as a service or task separate to "computer-aided translation" – i.e. translation done using translation memories (TMs) as an aid – the boundary has become increasingly blurred between these two activities. Where TMs are regularly deployed, there is usually a client-service provider agreement that fuzzy matches below a specific value (e.g. 75% similarity) are not of much use and are therefore considered to need full human translation, and are paid for as such. Such sentences are usually categorised as "No Match". These days, where source language words fall into the "No Match" category, they can be seamlessly submitted to an MT engine such that the

translator now engages in post-editing rather than full human translation for those source language words. The resulting situation is that translators switch between fuzzy match editing and MT post-editing as they move from one segment to the next within their TM tools.

Having established that post-editing is on the increase in professional translation, we will now consider why post-editing is sometimes viewed negatively, after which we will focus specifically on the technology environment in which post-editing takes place and report on findings from a survey on this topic.

2. Why translators dislike post-editing

As a result of this new MT/TM paradigm, many issues have emerged, which we discuss briefly here. First, we acknowledge that post-editing is not often viewed favourably by translators because:

- it involves "revision", which not all translators like as a task
- it involves revision of errors that appear on the surface to be caused by a lack of intelligence (informally, we call these "stupid errors")
- it is more cognitively demanding compared with human translation (or TM-aided translation)

We shall tackle each one of these points in turn. Professional translators make their livings primarily by translating and revising their own work ("self-revision"). Eventually, they may be asked to also review and revise others' work and it would seem that some translators enjoy this task, while others do not; some are good at it, and others are not, as highlighted in a review by Mossop (2007) . This applies whether the translation is generated by a human translator or a machine translation engine.

MT output may contain errors that a qualified human translator should not make, for example, mixing a masculine noun with a feminine adjective, or a singular noun with a plural modifier. These are "basic" linguistic errors that should not appear in human translation, but do appear in MT output. On the surface, these errors seem stupid to the human eye. Why doesn't the machine known that *voiture*, for example, is a feminine noun? When one appreciates the complex processing being carried out by the MT program to try to parse and make sense of language and transfer it to another language, one might not suggest so readily that the machine is "stupid". Human translation errors tend to come in a different form. For example, it's not uncommon for human translation to be inconsistent, whereas MT is generally consistent; it's also not uncommon for human translation to contain typos, but MT output does not suffer from this. The differences in the nature of errors in the two types of MT output leads to a dislike of the post-editing task.

The third claim is that post-editing is cognitively demanding, or is at least perceived to be so. If we consider the human translation (HT) task, a very

simplistic view of it is that the translator reads the ST and translates in order to create the TT. With MT, the translator also has to read the ST, then read the MT output and try to create a new TT from the intermediate machine representation. Depending on the quality of the MT output, this may require one simple interaction (e.g. addition of a plural marker), or it may require a significant re-structuring or re-wording of the sentence. Having to deal with the ST, interim TT and final TT is, we believe, the reason why translators *perceive* post-editing to be more cognitively demanding. We do not yet have enough research on this topic, but some research measuring the cognitive load of post-editing has indicated that it is similar to processing fuzzy matches in the 80-90% range (O'Brien 2006) or 85-94% range (Guerberof 2012). Of course, these results are dependent on the *quality* of the raw MT output; in both of the studies mentioned above, the quality could be categorised generally as being reasonably good, on average. Some (as yet unpublished) research is now emerging that suggests that post-editors report that they are less productive when working with MT compared with translation from scratch, but the throughput data demonstrate clearly that this is not the case. Thus, the post-editing task is disliked because it is perceived to make people slower and make them work harder when in fact the reality is not necessarily so.

There are other challenges we could mention here, such as productivity expectations, potential impact on quality and economic issues, but space does not allow for a full discussion on these issues. Instead, we wish to focus on the technological environment in which post-editing takes place.

3. Lack of support from technology

One of the issues that emerges in post-editing, but which has received very little attention to date, is the technology used for editing MT output. More specifically, the text editors or TM tool editors used during post-editing have not been designed with this specific task in mind and, we argue, could better support the post-editing task.

Several research tools have been developed to test whether post-editing-specific functionality can be integrated into a text editor. The TransType project, for example, aimed to suggest completions of a segment from an MT engine to the user (Langlais, LaPalme, and Loranger 2002). The Matecat tool, currently in limited production use, adds MT quality estimation, and incremental retraining of the MT output based on post-edits (Cettolo, Bertoldi, and Federico 2013). In a project related to Matecat, the Casmacat tool will attempt to add novel functionality, such as intelligent auto-completion and confidence measures, on a web-based platform (Alabau et al. 2013). Other tools that could be mentioned in this context include Caitra (Koehn 2009) and PET (Aziz, de Sousa, and Specia 2012). For industry-led research, Crosslang's tool has been used in some experiments (Guerberof 2012). An example of a functioning TM tool that has been adapted for research and industrial interests is iOmegaT, which is designated as an "instrumented" version of the open-source OmegaT tool. The

instrumentation involves detailed logging of the post-editing and TM-translation processes (Moran and Lewis 2011).

We take the view that the integration between TM and MT is probably one of the better environments in which to engage with MT. Even if TM matches above circa 70% are not available in a TM, it still makes sense for the post-edited MT to be saved into a TM as this immediately removes one of the significant issues translators face with MT – fixing the same errors in the same sentences more than once. For this reason, we set out to understand how existing CAT tools could be improved in order to better support the post-editing process.

3.1 Survey and Interview of Translators

We designed a survey of translators in order to gather opinions on editor features that might better support the post-editing task. Based on previous post-editing research, we had some initial ideas about features that might be useful and asked the translators for feedback on these ideas. Ideally, the surveyed translators would have had a chance to experience those features and provide experience-based feedback on the features. However, these features did not exist in tools at the time of the survey and the idea was to elicit input *prior to feature implementation*, something that is unusual in translation tool design (Lagoudaki 2008). Moreover, our project planned a future implementation of these features, which would then also be evaluated in full user evaluations once the features had been implemented.[1]

The survey was divided into five segments: (1) biographical details, (2) current working methods, (3) concepts regarding the ideal user interface (UI), (4) presentation of TM and MT segments and (5) intelligent functions for combining MT and TM matches. The survey was administered online and a link was sent to several localisation service providers asking them to promote the survey via their translator network. We deliberately chose to limit the focus to this potential set of respondents because it was clear that MT implementation and post-editing was most advanced in the IT localisation domain and it followed that we would be able to elicit opinions from translators who had some experience in post-editing. Of course, experience of post-editing was not absolutely necessary for participation in the survey. Following the survey, we also selected a number of responders for interviews in order to gain a more detailed understanding of their responses. The selection criteria involved (1) whether or not the respondent had experience as a post-editor and (2) we aimed to cover a range of target languages. An interim report on the survey findings is available in Moorkens and O'Brien (2013), so we will report only in summary on the main findings here.

[1] This feature implementation is ongoing at the time of writing and we hope to report on the user evaluations in the near future.

403 people answered the survey, but only 231 completed it fully and the results presented here are based on the 231 complete responses. Survey participants ranged in age from 20 to 50+ and featured a wide spread across the age categories. Among responders, there was a reasonable spread of professional categories from freelance (31%), freelance with an agency (34%), employees for translation or localisation companies (26%), or translation company owners (9%).

By far the most commonly used tool was SDL Trados, followed by MS Word and then Wordfast. Approximately half of the respondents reported being dissatisfied with their current tools, with performance issues figuring strongly including problems with bugs, slow response times, poor layout, compatibility issues and lack of customisability of the UI. Having reported this, it should also be noted that some responders were entirely happy with the tools they use, pointing to features such as auto-propagation, integrated quality assurance checks and concordance as being particularly helpful.

What was most surprising about the survey responses was that most negative comments pertained to TM tools in general. We had not expected to find such high levels of dissatisfaction with current TM tools because they have been used in the market since the 1990s and could reasonably be expected at this point to have overcome stability and performance issues at a minimum. The responses on TM tools in general somewhat diluted the focus on features for post-editing support and highlighted that there are still some very basic problems with technology support for translators.

When the participants were asked what improvements their current editors might need in order to support MT, two topics were noteworthy: (1) better display of provenance data and (2) dynamic changes of the MT proposal, based on edits. Provenance data refers to meta-data displayed in TM interfaces that provides contextual information about the translation suggestions, for example, date of creation, product or project to which the segment belongs, creator (human translator in the case of TM matches or engine name in the case of MT matches). Translators clearly see such metadata as being important to the decision-making tasks in editing. The second suggestion pertains to the issue mentioned earlier in section 2 - that translators dislike having to fix the same MT error repeatedly. If the editing interface could identify recurring phrases and automatically (or semi-automatically) implement the revisions to each phrase after the first correction, this would reduce the post-editor's frustration. Clearly, issues may arise if a tool automatically implements a change without notifying the translator, so we suggest that controls can be implemented to ensure that the translators still has oversight of this process.

The topic of dynamic changes to raw MT output based on post-edits also transcends the unique local project environment in which a translator is working. Ideally, the tool would not just propagate changes within a text, but the MT engine would also "learn" from the post-editors' changes. This so-called "machine learning" has become a dominant focus in the MT research arena. In

the data-driven MT paradigm, machine learning can take place by retraining the MT engine on the updated translation memory (this is another reason in favour of carrying out post-editing in a TM environment since it reduces the need for post-translation alignment of source and target texts). However, the retraining step can take a significant amount of time (sometimes even days). The ambitious objective is for the machine to learn from the post-editor edits on-the-fly so that the post-editor can immediately gain benefit from the MT engine's improved output. This ambition is the focus of attention in the R&D domain, but it has not yet been realised.

The nature of the errors produced by the MT system, as mentioned above, are such that the revisions are deemed to be quite tedious by post-editors and current tools do not help eliminate the problem. In our survey we suggested some features to the respondents to see if they thought these features might be useful. Typical examples include:

- changing the capitalisation of a word that has become capitalised in MT or has lost its capitalisation
- re-applying punctuation to the TT, which has been deleted in the MT process
- moving formatting tags (in tagged environments such as XML) to the correct location in the TT, as tags are either stripped during the MT process or all placed at the start or end of the segment
- adding or deleting spaces that have been inserted or deleted during the MT process
- changing singular to plural or vice versa
- changing the gender or case inflections
- moving the position of words within a phrase (e.g. from Subject Verb Object structure to Verb Subject Object)

The survey participants said that they would all like to see features that would mitigate these tedious tasks. 80% of participants reported that their productivity is normally improved by using keyboard shortcuts, and this was the preferred method of triggering a function to make an automated change. However, many were not convinced that a machine could do this efficiently. A general suspicion of machine translation was notable throughout the survey results, with 56% suggesting that machine translation was "still problematic", and comments that said that machine translation was "still in baby shoes" or that MT quality is "just horrible".

4. Summary

Post-editing of MT is still rather disliked as a task, and for good reason. However, as MT is increasingly adopted, at least in some domains, it is

worthwhile investigating how the task can be made less demanding for those who engage in it. Our view is that technological improvements over time will remove some of the more annoying aspects of post-editing, such as fixing errors repeatedly and fixing annoyingly basic errors. A preliminary survey gathered opinions from translators working in the IT domain on CAT editors in general and on proposed PE-support features in particular. Although some appear to be quite content with the tools that they use, there was found to be a high level of dissatisfaction with TM tools in general, even without taking MT and post-editing into account. It is somewhat lamentable that after so many decades of use of these tools that this is the case. The next step in our research project is to implement some of our suggested features and to test their impact on post-editors to see if the features do indeed ease the task, as is our objective.

Bibliography

Aziz, W. / de Sousa, S. C. M. / Specia, L. (2012). "PET: a tool for post-editing and assessing machine translation", in: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidi (Eds.) European Language Resources Association, Paris, France.

Alabau, V. / Bonk, R. / Buck, C. / Carl, M. / Casacuberta, F. / García-Martínez, M. / González, J. / Koehn, P. / Leiva, L. / Mesa-Lao, B. / Ortiz, D. / Saint-Amand, H. / Sanchis, G. / Tsoukala, C. (2013). "Advanced Computer Aided Translation with a Web-Based Workbench" in: *Proceedings of Machine Translation Summit XIV Workshop on Post-editing Technology and Practice*. O'Brien, S., Simard, M., Specia, L. (Eds.) Nice, September 2, 2013, p. 55-62.

Cettolo, M. / Bertoldi, N. / Federico, M. (2013). "Cache-based Online Adaptation for Machine Translation Enhanced Computer Assisted Translation", in: *Proceedings of Machine Translation Summit XIV*. Khalil Sima'an, Mikel Forcada, Grasmick, D., Heidi Depraetere, Andrew Way (Eds.)

De Almeida, G. (2013). *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages.* PhD Dissertation. Dublin City University.

DePalma, D. A. / Hegde, V. (2010). *The market for MT post-editing*. Common Sense Advisory, Boston, MA.

Guerberof, A. (2012). *Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation*. PhD Dissertation. Universitat Rovira i Virgili, Tarragona.

Koehn, P. (2009). "A process study of computer aided translation", in: *Machine Translation* 23, 4, p. 241-263.

Lagoudaki, E. (2008). *Expanding the Possibilities of Translation Memory Systems: From the Translator's Wishlist to the Developer's Design*. PhD thesis. Imperial College, London.

Langlais, P. / Lapalme, G. / Loranger, M. (2002). "Transtype: Development-evaluation cycles to boost translator's productivity", in: *Machine Translation (Special Issue on Embedded Machine Translation Systems)*, 15, 2, p. 77-98.

Moorkens, J. / O'Brien, S. (2013). "User Attitudes to the Post-Editing Interface", in: *Proceedings of MT Summit XIV Workshop on Post-editing Technology and Practice*. O'Brien, S., Simard, M., Specia, L. (Eds.) Nice, September 2, 2013, p. 19–25.

Moran, J. / Lewis, D. (2011). "Unobtrusive methods for low-cost manual evaluation of machine translation", in: *Proceedings of Tralogy 2011*, Centre National de la Recherche Scientifique, Paris, France, 3-4 March 2011.

Mossop, B. (2007). "Empirical studies of revision: what we know and need to know", in: *Journal of Specialised Translation*, 8, p. 5-20.

O'Brien, S. (2006). "Eye-tracking and translation memory matches", in: *Perspectives: Studies in Translatology*, 14(3), p. 185–205.

Vasconcellos, M. (1983). "Management of the machine translation environment: interaction of functions at the Pan American Health Organization", in: *Proceedings of Translating and the Computer 5*. Lawson, V. (Ed.) London, 10-11 November 1983.