

# Post-Editing Free Machine Translation: From a Language Vendor's Perspective

**Luciana Cecilia Ramos**

Ocean Translations S.R.L.  
1716 San Lorenzo Street – 7<sup>th</sup> Floor  
2000 Rosario, Argentina

[luciana.ramos@oceantranslations.com](mailto:luciana.ramos@oceantranslations.com)

## Abstract

This paper presents a language vendor's perspective on the actual implementation of machine translation solutions in the translation/localization process. This lecture will be delivered at AMTA-2010 Conference, and a short video will accompany lecturer's speech.

## 1 Introduction

Who dares deny humanity is in constant change? Moreover, changes have always driven evolution and evolution is only possible as long as innovations are embraced by users, consumers... people at large. After many years of development and research, machine translation has taken the floor and it's worth listening to what it has to say, or taking advantage of what it has to offer - Why should we keep giving it the cold shoulder? In this presentation, we will look into MT output versus human translations, what the role of post-editors should be, what is quality in terms of MT and how can it be assessed. Experiences as language vendors [limited to the English into Spanish, English into Portuguese and Portuguese into Spanish language pairs] will be discussed, as well as a perspective based on figures and timelines, both from the client's and vendor's side. When lectured in front of the audience, this presentation will also include a visual tour depicting an experience using a free MT engine (Google Translator Toolkit) and the conclusion supported by lecturer's experience.

## 2 Why Choosing Machine Translation as Part of the Localization/Translation Process

When there is a need, it is (almost) always urgent; resources tend to be scarce at that time; and results have to surface soon so as to have the possibility of going on. Although this statement may not be true for every business, I dare say most of us have gone through such experience sometime... and especially in the translation arena. Unfortunately, I cannot claim Machine Translation has come to change the world for the better, but, definitely, It does contribute a lot to the growth of the language industry. Whatever the situation is, there is no reason why we should let innovation pass unheeded. Then why should we deny MT is worth using? If so, we will be denying all the hard work technology experts and researchers have been doing for more than fifty years: in 1952, the first MT conference was held at MIT (Massachusetts Institute of Technology). By that time, full automation of good quality translation was considered as virtually impossible and human intervention either before or after computer processes was thought to be essential (Hutchins, 2001). Since then, much has been done not only on the developers and researchers' side, but also from the translators' community, and it is time to expect for the massive use of Machine Translation solutions. Nowadays, MT technology and practices are varied. Some software developers or content creators (such as Microsoft, IBM, just to name a few) have built their own systems and outlined their own guidelines to implement it. The general public has Google Translator Toolkit for free; and

many language and technology companies have developed their own customizable engines (such as Asia Online's Statistical Machine Translation platform, Welocalize's GlobalSight Technology, Systran's Hybrid Machine Translation Engine, among others) to cater for a wide array of content types and particular needs. Real-life case studies have shown the combination of human post-editors and customized machine translation is powerful and effective. From our experience as a language company, all assignments overtaken have proved successful for all sides (clients, translators and ourselves), considering time spans have been shortened, costs reduced, quality controlled, workflows maintained; satisfaction has been achieved at both ends, end users and actual post-editors. By and large, and in controlled-enough scenarios, machine translation helps clients reach foreign-language markets in shorter times and to a larger scale. But, what is a controlled scenario? And what does it take? It encompasses a set of rules and procedures regulating and sensing MT post-editing; it takes, above all, experienced professionals in the technology and language areas, so that reasonable and adequate quality goals are set, procedures are delineated accordingly, and pre and post-editing phases are included in the loop. Under such circumstances, MT can offer an effective shortcut to multilingual communication.

### **3 Tailoring Content for MT and Tailoring MT to Content**

Once the decision of using MT is made, a linguist or translation expert is essential, either to make content suitable for MT or to depict the best MT-based approach based on content and intended use. As a start point, content type and language used in the source text should be analyzed to choose between statistical, rule-based or hybrid machine translation. Text intention is also relevant for making the decision, even on whether machine translation is feasible or what type of post-editing will be needed. If source material is not yet created, considering the inclusion of Controlled Language (CL) at the authoring phase would help render less erratic output. Testing actual output; timing post-editing of that output versus traditional translation/editing; surveying readers or real users of that output; setting goals of source text and assessing to which extent such goals are attained

with the use of human translated text versus MT-ed/MT post-edited text are other mission critical tasks that require the involvement of a language expert. Linguists and developers working together will render the best outcome in terms of time and cost efficiencies, considering the translator's input as the most useful device to filter errors out, thus avoiding the need of repetitive corrections or enhancing the engine. The translator's eye may even be able to predict MT errors or inaccuracies based on a pattern, and the course of the engine can be changed according to his/her feedback, creating new rules or settings different parameters. Issues related to consistency, choice of words and tone can also be addressed at an initial phase when both resources communicate effectively and proactively. Based on the target audience for the materials and/or the media throughout which target output will be disseminated, the post-editing phase may consist of one or two rounds or reviewing. Materials meant for publishing will definitely require a more polished language, which is not unattainable by the use of machine translation, although sometimes time and money savings will not be so apparent in such cases. However, it does offer a great possibility when resources are scarce, considering part of the job can be done mechanically. Many post-editors would be willing to work even with a poor quality output if they get paid closer to what they would get if they were translating. Although time spans will not be much shorter, there will be less hassle in the placement phase and still some reductions in production time, since not all content will be recreated. Post-editors may be led to accept what the computer offers, as long as the same message is conveyed within the target language and in concordance with the language grammar rules. Ideally, glossaries and style guides should be prepared in advance for the sake of consistency and to save precious time during the MT/MT post-editing process itself. Also, a smooth implementation will depend on resources skills for team-working and communication. Teamwork is a *must* for positive results to be accomplished, and results have to be assessed in the long run, since the initial phase tends to be much slower than the rest, and although it is customary for any assignment irrespective of the use of MT, when it relates to MT translators may tend to be reluctant to accept it. Initial investments of time, money and

resources involvement should pay off, therefore, ideal scenarios for successful stories of commercial use of MT should involve large scale projects.

#### 4 The Role of Editors and Post-Editors

As discussed above, post-editors, along with MT engine developers, play a key role in the MT utilization. The role of the human translation editor and MT post-editor should be clearly differentiated, and if expected results are to be obtained by the predicted deadline and within the set up budget then post-editors should be trained and be provided with guidelines to meet this end. Post-editors have to be able to focus on adjusting the MT output so that it reflects and conveys the meaning of the source text as accurately as possible (Krings, 2001). On the contrary, human translation editors tend to strive to disguise the fact that the text has been translated (Senez, 1998).

#### 5 Quality in the MT Sphere

Quality is always an issue by itself, it does not matter over which sphere we are hovering... but when it comes to MT, it seems to gain considerable weight. It is the main reason raised by those less keen on MT against welcoming it to the translation/localization industry. To begin with, translation customers have to clearly define quality goals and whether raw MT output or MT post-editing output will be taken into account. According to a study carried out by Rebecca Fiederer and Sharon O'Brien in 2009, in which human translation and MT post-editing outputs were compared, output is assessed in terms of three axes – *Accuracy, Clarity and Style*. In this study, results show a better performance by human translation only in the third category, *Style*, but, based on the scale used by authors<sup>1</sup>, both results fall under the same definition, 3, which rates style as 'Most of the language is appropriate. Mostly natural and idiomatic; flows fairly well'. In such study, results reflect HT and MT render quite similar levels of clarity and accuracy<sup>2</sup>. If raw MT is judged against ZT (Zero Translation), quality moves down to a lower position in the priority

---

<sup>1</sup> Evaluators had to number quality criteria within a scale of 1 to 4, and results for Style were 3.23 for Human Translation; 3.03 for MT.

<sup>2</sup> Clarity: HT, 3.44; MT, 3.43. Accuracy: HT, 3.32; MT, 3.47.

scale, being indexes such as time to market, market penetration, general information, etc. those of major importance.

#### 6 Google Translator Toolkit – Free Machine Translation Engine

Both for the individual or commercial user, this product from Google offers a free, easy to access and use global engine for machine translation. A user-friendly interface will guide users from a general experience to a more customized application, with the addition of property Translation Memories and glossaries. As language service providers, we (Ocean Translations S.R.L.) have been using this engine as the initial translation step for several clients (only when specifically requested) and results are encouraging. Our linguists, in their role of post-editors, are overall satisfied with the quality they produce and the fees they get for it. They claim it takes them one hour to post-edit 500-800 words (versus 800-1000 words edited when translation was done by another translator). A sustained flow of work (related mainly to the lengthy nature of such requests) is also reported as one of the advantages of working as post-editors. As an average, around 650 words are post-edited per hour, and an eight-hour working day is possible under post-editing working conditions, with minimal breaks in between, such as when working with human translation. When Internet connection is stable and acceptable, the server response times are reasonable and adequate. They have also reported being happy with the fact of not having to face the "blank page" (compared to their translation experience), and—even though there are some variations, which are usually offset by some great quality instances, usually related to short phrases or single-word phrases— output is easy to cope with. As an illustrative fact, we have recently completed a job of more than 90,000 words post-edited and revised (English into Spanish) in 9 calendar days (from client's confirmation to delivery) involving 1 project manager, 4 post-editors (one designated as a lead post-editor) and 2 reviewers, and estimated cost savings for the client were approximately 40%. As requested by one of our main and most valued translation clients, the Department of Reproductive Health and Research (RHR) at WHO Headquarters, we have carried out

a quick analysis of shifting from the current model (CAT Tool-based) to Google Translation Toolkit for the localization of the Reproductive Health Library, for which updates are generated every year. Our report to them clearly reflects our overall experience in relation to this engine, and it is encouraging –although challenging, of course– to see how happy both parties can be with this solution, bringing time and money savings to one side, and a better utilization of resources and contracts for larger volume to the other. The report, intended to gain some enhancements and adjustments from the Google and WHO collaborative efforts, was as follows:

### What We Did

- We have created 5 files from the existing TM in order to validate it. Size problem solved.
- We noted that for validation, apart from size restrictions, there are language combination restrictions. In fact, our TM is set for ES\_EM, which means target locale is International Spanish, but Google Translate only accepts ES-ES as a valid language configuration for Spanish. We have changed locales.
- Users have created customized Gmail accounts.
- We have saved XML files as HTML using Dreamweaver 8.
- We changed glossaries extensions as CSV, saving Word glossaries as XLS and then renaming them as CSV.
- We processed a couple of files - everything seems to run smoothly.
- We have received the heads-up that the project will start in January 2011 under this new "translation" modality.

### Our Insights

- High speed Internet connection is required.
- Interface is user-friendly and includes basic editing features.
- TM and glossaries can be easily shared among groups or uploaded individually once files are prepared appropriately.

- Task assignment should be done in the regular fashion.
- Matching type is specified only when it is an exact match. If fuzzy matches are found, only differentiators are identified, but not matching %.
- Machine-translation renderings are inserted by default - if exact match exists, it has to be chosen by clicking the **Use Translation** button. No much of a problem, but maybe if setting could be changed so that existing translation in TM is inserted at first, it would be great.
- Once finished, files can be saved in local disks with the same file extension as uploaded. BUT, HTML does not convert back as XML correctly. Options are: Google Translate is changed to support XML files OR WHO provides us with HTML or other supported file format.
- No spell-check feature available (Mozilla has its own; it can be set and activated, so Mozilla should be used).
- Glossary interface does not seem to pop-up when a term is found, so it seems that linguists have to click on the Glossary tab for every segment to check terminology. It does show if matches are found - It would be great if pop-up windows appear for every matching.
- Dictionary feature is interesting and useful!
- Repetitions are not self-propagated.
- Wordcount is not broken-down according to different match types (which is necessary for planning and critical for invoicing).
- Translators will have to be compensated for processing exact matches, since each file will have to be uploaded, opened and processed (selecting Use translation for each line).
- As a general conclusion, we liked it, we found it agile and we expect very good results!

## 7 Conclusions

As in any environment, fear to change has to be overcome and collaborative alliances are the best

tool for this. Machine Translation is no longer "the future" and it is here to stay. After many years of research and development, it has become a useful tool available to the population at large, which will yield the best results when used by the best trained and most qualified professionals: translators, terminologists and software developers. Any translation or localization endeavor has to be duly discussed and a pros and cons analysis respective of using MT should be performed by a team of language and technology professionals. In a nutshell, adding Machine Translation to the production cycle is an undeniable advantage, especially for large volume contents. Leveraging CAT and MT technologies with human post-editors, the translation industry may experience a significant increase in its figures and it does not have to mean human professionals of the language industry will lose their jobs - on the contrary, more and longer working opportunities may arise, as we have experienced, as translation becomes an easier and more flexible service to buy.

## **Acknowledgments**

Particular thanks should be given to my partners, Betty Galiano and Federico Carroli, my company's QA Manager, Sabrina Maceda, Metin Gulmezoglu, Asa Cuzin-Kihl, from the Department of Reproductive Health, WHO, and Michael from Google for their valuable collaboration they have made to this paper.

## **References**

- Rebecca Fiederer and Sharon O'Brien. 2009. *Quality and Machine Translation: A realistic objective. The Journal of Specialised Translation, Issue 11*. Dublin City University.
- Dorothy Senez. 1998. "Post-editing Service for Machine Translation Users at the European Commission". In *Proceedings of Translating and the Computer 20*. London.
- Hans P. Krings. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Geoffrey S. Koby (ed/translator). Kent, Ohio: The Kent State University Press.
- John Hutchins. 2001. *Machine Translation over Fifty Years. Histoire, Epistemologie, Langage*, Tome XXII, fasc. 1, p.7-31.