

Machine Translation Post-Editing at TransPerfect – the ‘Human’ Side of the Process

Ana Zaretskaya



Ana Zaretskaya
azaretskaya@translations.com
TauYou-TransPerfect

Abstract

Machine translation post-editing (MTPE) is becoming a common practice in translation industry. Proper collaboration with linguists is key to successful MT implementation in the translation workflow. At TransPerfect, we apply different procedures to make MTPE convenient and beneficial for translators, which are outlined in this article.

Keywords:: machine translation, post-editing, translation industry, post-editing training, linguistic feedback, machine translation improvement.

Resum

La postedició de traduccions automàtiques (PETA) és cada cop més una pràctica comú dins la indústria de la traducció. Una col·laboració apropiada amb els traductors és clau per implantar la TA amb èxit dins el procés de traducció. En aquest article expliquem els diferents procediments que apliquem a TransPerfect per fer de la PETA una eina pràctica i beneficiosa per als professionals de la traducció.

Paraules clau: traducció automàtica

Resumen

La posesición de traducciones automáticas (PETA) está adquiriendo el carácter de práctica habitual en la industria de la traducción. La adecuada colaboración con los profesionales de la lengua es un elemento fundamental para la correcta implementación de la TA en el proceso de traducción. En este artículo se indican los diversos procedimientos que empleamos en TransPerfect para que la PETA les resulte práctica y ventajosa a los traductores.

Palabras clave: traducción automática, posesición, industria de la traducción, formación en posesición, observaciones de los profesionales de la lengua, mejora de la traducción automática

1. Introduction

Machine translation post-editing (MTPE) is an important part of the translation industry. The amount of post-edited texts is growing from one year to another. For big and small translation businesses, as well as governmental institutions like the European Commission, MTPE has become a common practice, and TransPerfect is no exception. TransPerfect¹ is one of the world's largest translation companies with over 4000 employees and over 90 offices all over the globe. We have thousands freelance linguists collaborating with us and hundreds of them are trained and certified in post-editing. Furthermore, we are currently expanding the pool of linguists who are keen on learning about machine translation (MT) and acquiring this new skill. Besides the post-editing task itself, linguists also collaborate with us on other MT-related tasks, such as MT evaluation and improvement.

There is substantial research on MTPE and many language service providers (LSPs) apply it to a significant amount of translation work they deliver. The level of acceptance of MT in the translators' community also seems to have raised in the past years and one can say that post-editing has become another regular type of services linguists provide, along with translation, proofreading, subtitling, copywriting, and others. Nevertheless, some linguists, especially those who have been in the industry for a very long time, might still be somewhat reluctant to do this type of work (O'Brien & Simard, 2014).

Our day-to-day communication with linguists shows that the negative attitude towards MT often has its roots in the lack of knowledge about what MT systems we use and how we do it. In many cases, linguists who are initially reluctant to post-editing have some misconceptions about our MT. The most common ones are the following:

- They believe that we use third-party MT systems that provide low quality translations and they are not useful and will only slow them down. In fact, we use our own proprietary technology.
- They believe that we offer lower rates for post-editing jobs only because they involve MT and the linguists will eventually lose money. Instead, we focus on maximizing their revenue as well.
- They believe that post-editing is similar to proofreading, i.e. correcting someone else's translation errors instead of producing one's own translation. However, post-editing is more complex than that, and the linguist performing it is the author of the translation.

We see the task of MTPE as a process rather similar to translation, during which the linguist is offered MT suggestions that they can use as a draft that they would barely need to correct (or even use them without any modification when the suggestion is completely correct), or delete them when they are not useful. This way, the linguist performing post-editing is in charge of the final translation.

¹<http://www.transperfect.com/> [accessed on 20.12.2017].

This perspective is different from many definitions of post-editing one can find in the literature, where the linguist's role consists of barely 'editing' the output of an MT system instead of actually producing their own translation. Consider some of the existing definitions of MTPE below.

MTPE is "usually understood as a human being (normally a translator) comparing a source text with the machine translation and making changes to it to make it acceptable for its intended purpose" (Koby, 2001: 1)

"..the process of improving a machine-generated translation with a minimum of manual labor" (TAUS, 2010)

"edit and correct machine translation output" (ISO 18587, 2017)

Machine translation post-editing has proven to be a recurrent solution across the industry for achieving higher translation productivity, and there are research studies showing that with the right processes in place MTPE indeed helps increase productivity (Federico et al., 2012; Läubli et al., 2013; Zampieri & Vela, 2014; Zhechev 2014). The first absolute condition of success is, of course, the quality of the MT, which is ensured through the deployment of cutting-edge technology, thorough quality evaluation and improvement.

Our MT specialists are constantly working on creating new and improving already existing MT engines. Those include customized client-specific engines of various types, such as hybrid (combination of statistical MT and automatic pre- and post-processing rules) and neural MT. These engines are created internally with high-quality training data and are continuously improved through applying linguistic feedback, machine learning techniques, and incremental re-training with updated linguistic data.

However, an equally important component of successful post-editing is continuous work and regular communication with linguists. In this article we present an overview of all the procedures we have in place at TransPerfect that help ensure that our linguists are well prepared for the post-editing task and are satisfied with all sides of the process. These procedures include a training and certification process, collection of linguistic feedback to improve MT quality, and regular open discussions and Questions and Answers sessions.

2. Post-editing training

MT post-editing is a relatively new type of service and many linguists are not completely aware of what it entails. That is why training is essential and there are several companies and organizations who provide training on MTPE. The most well known is probably the online training courses created by TAUS². At TransPerfect, completing our training on MTPE is an absolute requirement for becoming certified for post-editing services. Some of the main goals of the training are outlined below.

² <https://www.taus.net/academy/taus-post-editing-course> [accessed on 20.11.2017].

- 1/ Improve the general perception of MT in the linguistic community.
One of the objectives of the training is to explain our approach to MT, why and how we utilize it. We see MT as an additional productivity tool rather than a technology that is aimed at replacing human linguists. Human expertise is key for achieving high-quality translation.
- 2/ Manage the linguists' expectations regarding the MT quality and the task itself.
Many linguists expect MTPE to be similar to proofreading human translations. However, both the quantity and the nature of MT errors are different from those you can find in a human translation. Consequently, the effort and time necessary to post-edit MT texts is not the same as the one spent on proofreading a text delivered by a human translator.
- 3/ Explain what is expected from the linguists in terms of final translation quality.
As many other LSPs, we distinguish two types of MTPE quality, which are Light and Full post-editing. Light PE consists of only correcting major MT errors in order to make the translated text *understandable*. The quality of the translation after Light PE is lower than the one expected in a regular translation project, it can sound too literal or unnatural and contain minor objective errors. In contrast, Full PE quality requirements are higher and the translation after Full PE cannot contain any errors or stylistic defects. Our PE training includes a detailed description of the two post-editing quality levels with examples, which is one of the most important and difficult aspects of PE work.
- 4/ Provide tips and recommendations on how to be more efficient when post-editing.
There are a number of tricks one can apply to make the most of post-editing and save time. This topic is covered in detail in Section 4 of this article.
- 5/ Explain to the linguists how to provide valuable feedback to help improve MT quality.
Our dedicated MT team works together with linguists in order to improve the performance of the MT engines. In other words, linguists advise us what has to be improved, so the quality of MT increases over time. In the following section, we explain how exactly we collect this information and what makes good and useful linguistic feedback.

3. Linguistic feedback

There are different methods of improving the performance of MT systems. For instance, post-edited translations are analyzed using statistics and machine learning techniques to extract the most frequent and relevant changes made by linguists. These changes are then applied automatically. However, the most efficient way of improving MT quality

is the linguistic feedback. We have a special process in place where we work together with linguists to improve MT systems to obtain more useful MT output. Linguists are direct users of the MT systems and they are the ones who know best where output can be improved and they can help us do that. Our linguists' role consists of more than just editing MT output, through constant collaboration with the MT team some of them become real MT experts. In order to be useful for MT improvement, linguistic feedback has to comply with several essential criteria.

3.1. Feedback should be efficient

We ask linguists to provide only *frequent* or *systematic* errors they find in the MT output instead of simply listing all the errors. The goal is to reduce the amount of manual work needed to edit MT output in the future. The improvements we perform based on this feedback must have the biggest possible impact on the quality. In other words, if a certain error occurred only once and probably will never occur again, reporting it will not have a big impact on the quality of the MT. The two examples below are useful to improve MT quality.

- (1) Punctuation: if the whole sentence is inside parentheses, the period should be inside the closing parenthesis.

Source (French)	MT (English)	Correct translation (English)
(...on extrapolera à la dose supérieure).	(... it will be extrapolated to the highest dose).	(... it will be extrapolated to the highest dose.)

- (2) Terminology

Source (French)	MT (English)	Correct translation (English)
pipette électronique	pipette email	electronic pipette

In contrast, the example below is not very significant as this error is rather uncommon. If this term were a frequent one, only then would it make sense to report this error.

(3)

Source (German)	MT (English)	Correct translation (English)
Bedruckung Lot	Printing lot	Lot printing

We have a standard MT feedback form that linguists are recommended to use to report the MT issues they want to fix. We advise to have it open during post-editing and fill in recurring errors while working on a file, this way there is no need to go

back to the post-edited file and try to remember the errors that were already corrected.

3.2. Feedback should be specific

It is important to make it clear for the MT team what exactly was incorrect in the segment produced by the MT engine. Usually, feedback is provided in a three-column format, where the first column corresponds to the source phrase or term, the second column to the MT suggestion that was wrong, and the third column shows the correct translation. However, simply copying the entire segments is not enough. Consider the example below.

(4)

Source (English)	MT (German)	Correct translation (German)
Unique reflector system for indirect light emission to avoid	Einzigartiges reflektor System indirekte Licht Emissionen	Einzigartiges Reflektorsystem indirekte Beleuchtung

This example is not specific enough: the MT team has to understand which phrase or term was wrong and how exactly it should be corrected. A good way of reporting issues in this segment would therefore be the following:

(5)

Source (English)	MT (German)	Correct translation (German)
reflector system	reflektor System	Reflektorsystem
light emission	Licht Emissionen	Beleuchtung

3.3. Feedback should be objective

Feedback should be supplied for objective errors only, and not stylistic or preferential changes. In the example below, the change reported is preferential.

(6)

Source (Spanish)	MT (English)	Correct translation (English)
¿Tiene preguntas acerca de sus pagos?	Do you have questions about your payments?	Do you have questions regarding your payments?

4. Efficient Post-Editing

Even though MTPE is similar to translation and also has some similarities with proofreading, it is still a different linguistic activity and has its nuances. Below we provide some recommendations we give to linguists on how to be more efficient when doing MTPE.

- 1/ Quickly assess the general quality of MT before starting to work on the project. If it is a new project you are starting to work on, it is useful to first see if the quality of the MT is useful and swiftly evaluate how much effort will be needed. If it's a project you are working on regularly, it is useful to confirm that the quality of the MT is on the same level. Any irregular or significant quality issues in the MT output have to be flagged immediately to be fixed.
- 2/ 2-second rule
Often you will find MT segments of borderline quality. It might be difficult to decide if these segments are useful and if you can correct the 'bad' parts, or if it will be faster to delete the MT suggestion and re-translate the segment from scratch. If you want to be efficient, you should not spend more than 2 or 3 seconds on this decision. Thinking for too long about every segment makes you lose time – in this case, it is better to re-translate the segment from scratch.
- 3/ Find & replace
'Find-and-replace' is a powerful tool when it comes to post-editing. You can find many repetitive mistakes, for example, in terminology or formatting – these can easily be handled with 'find-and-replace'.
- 4/ Pay special attention to proper names
Proper names are important as they often get translated by the MT system, which 'considers' them common nouns (e.g. *Alvaro Cadena* can be translated as *Alvaro Chain*). Pay special attention to proper names in order not to miss such cases.
- 5/ Pay special attention to the tone of voice (in languages like Spanish, Portuguese, and Slavic languages)
When translating from English into languages with formal and informal pronouns and verb forms, be careful and make sure the forms you are using are the ones required and are consistent. E.g. the use of *tú* and *usted* in Spanish, as well as the verb forms. Since English does not have this distinction, it can be handled incorrectly or inconsistently by the MT system
- 6/ Use shortcuts
Some shortcuts are particularly useful for post-editing, such as the shortcut for deleting entire target segment or cut and paste.
- 7/ Read through the whole file once finished and fix style and natural flow
Style is specifically important in full post-editing, especially if you are working with a phrase-based MT system. These systems consider words and combinations of words instead of entire sentences and often provide

translations that are too literal and sound unnatural. It is recommended to read through the translated text after you finish – this way you can better spot any stylistic problems.

5. Conclusions

The translation industry is changing and it affects all the stakeholders including translators. There is no doubt that MT will play a significant role in the near future. Even though it existed for many decades, MT has found a way into the translators' business and everyday work only recently. For translators, it is a relatively new type of technology and post-editing is a new task, which they have to learn and get used to. At TransPerfect, we invest time and resources to make this learning process as smooth as possible, and the post-editing process as beneficial as possible for the linguists. We believe that collaboration and dialogue are key to the successful implementation of MT in the translation workflow; in this way, all the parties involved can benefit from the process: our clients, our linguists, and us.

References

- Federico, M., Cattelan, A. & Trombetti, M. (2012), "Measuring user productivity in machine translation enhanced computer assisted translation", *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA)*.
- ISO 18587 (2017), *International Standard. Translation services — Post-editing of machine translation output — Requirements*.
- Koby, G. S. (2001), "Editor's introduction", in Krings, H. P. *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*, Kent State University Press, pp. 1-23.
- Läubli, S., Fishel, M., Massey, G., Ehrensberger-Dow, M., & Volk, M. (2013), "Assessing post-editing efficiency in a realistic translation environment", *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice*, pp. 83-91.
- O'Brien, S. & Simard, M. (2014). "Introduction to special issue on post-editing." *Machine Translation* 28(3), 159-164.
- TAUS (2010), *Post-editing in Practice Report*, March 2010.
- Zampieri, M & Vela, M (2014). "Quantifying the influence of MT output in the translators' performance: A case study in technical translation", *Proceedings of the EAACL Workshop on Humans and Computer-assisted Translation (HaCat)*, pp. 93-98, May.
- Zhechev, V (2014). "Analysing the post-editing of machine translation at Autodesk". in O'Brien, S., Winther Balling, L., Carl, M, Simard, M., & Specia, L. (eds.), *Post-editing of Machine Translation: Processes and Applications*, pp. 2-24. Cambridge Scholars Publishing, Newcastle upon Tyne.