**Optimising the Machine Translation Post-editing Workflow**

Anna Zaretskaya

TransPerfect

As most large LSPs today, TransPerfect offers a variety of services based on machine translation (MT), including raw MT for casual low-cost translation, and different levels of MT post-editing (MTPE). The volume of translations performed with MTPE in the company has been growing since 2016 and continues to grow to this date (Figure 1, the numbers on the Y axis have been omitted as commercially sensitive information), which means tens of millions of words post-edited each month. In order to implement MT at such a large scale, the process has to be as easy as possible for the users (Project Managers and translators), with minimal or no additional steps in the
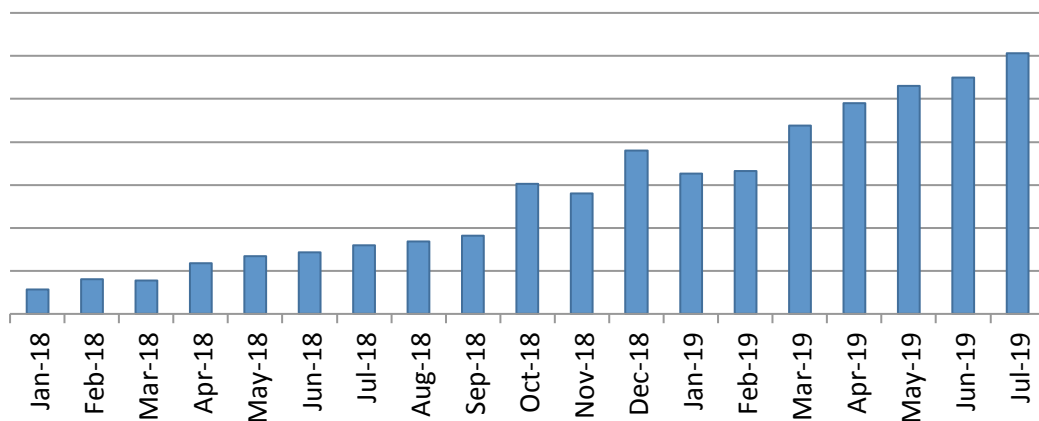
workflow.

Figure 1. Volume of MT post-editing TransPerfect

In our case, MT is integrated in our translation management system, which makes it very easy to make the switch from purely human translation workflow to the post-editing workflow (Figure 2). In this article we will share the methods we used to optimise the workflows when implementing MT, covering both the technical aspects and the processes involved.
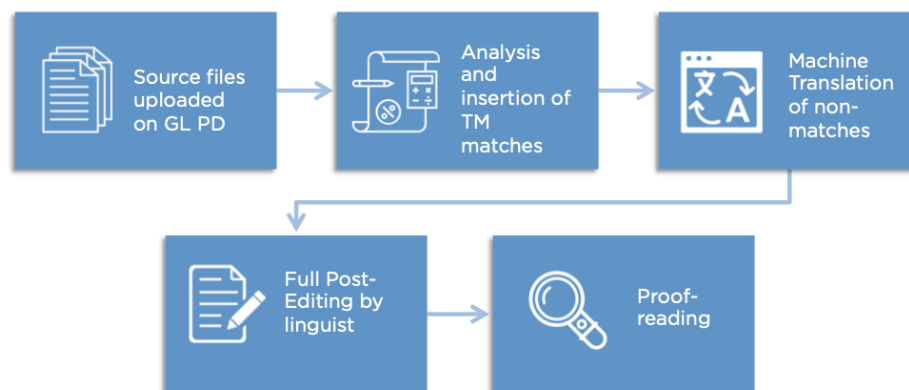

Figure 2. Standard MTPE workflow.

1. Machine Translation Systems

TransPerfect's MT systems use state-of-the-art neural technology and include a variety of off-the-shelf trained MT engines that we can choose from for any given project, including both generic systems and domain-specific ones. We also have access to a number of third-party MT services via APIs. Our approach consists in choosing the best possible solution and we are not limited only to proprietary systems.

Apart from generic and domain-specific MT engines we recommend customising the systems with the data specific to the content that will be translated. A typical use case is customisation for a specific client, where a generic model is *incrementally trained* using the client's Translation Memory. Engine customisation can yield improvement in the output quality of more than 20% compared to the baseline, but this is conditioned by the amount of data and the quality of this data. We have conducted several case studies, in which we explored how MT quality increases when adding more volume to the customisation data set (results of one of of these studies are summarized in Figure 3). We have found that, on average, a noticeable difference in quality is observed with incremental training with additional 50 000 translation units (TUs) in the data set, but it can differ based on the initial quality of the engine and the quality of the data. This number seems to be independent of the initial training data size of the base model, but rather to depend on the initial quality, however, this has to be confirmed by conducting more experiments.
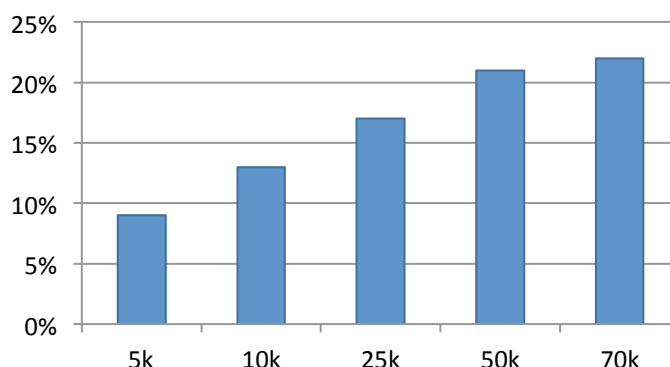

Figure 3. Case study: improvement over baseline MT quality with incremental training depending on the size of the training data set (in TUs).

After our MT systems are deployed in production they are improved over time. First of all, we collect feedback from the linguists who work on post-editing. They use a simple web interface to report frequent MT errors, which are then fixed using glossary enforcement techniques or pre- and post-processing rules. In addition, all the edits to the MT output are registered and used for engine improvement.

2. Linguist Experts

Providing training and support to our linguists is at least as important as the MT technology itself. We consider MTPE as a separate service, along with translation, subtitling, interpreting, and others, and we have established a special certification process for this service. In this process, linguists go through a training programme, which includes theoretical and practical aspects of post-

editing. The training should prepare the linguists to perform post-editing efficiently while providing the quality expected by our clients. It includes practical tips on how to quickly decide whether a segment is eligible for post-editing, identify MT errors that are specifically difficult to see (these include missing negation, wrong numbers, mistranslations, among others), how to take advantage of their translation environment to be faster during post-editing, and how to make sure to deliver the desired quality of the final translation. This way we make sure that our linguists are comfortable with the task and have the right knowledge to use MTPE to their own benefit. Currently we have about 3500 linguists in our database who are certified for MTPE, which is more than a half of all our active linguists.

3. MT Evaluation

Evaluation of MT quality is performed on different stages of the workflow, it is necessary to decide where MT is suitable for a specific type of documents, and which MT engine is the best to use in each case. In MTPE projects, we use the post-editing distance (PED) as the main method of evaluation, as it measures the editing effort required. Tracking and storing the PED on a project level allows us know the amount of changes made in a specific job and confirm that we have paid the linguist accordingly. In addition, it allows us to track the performance of a specific MT engine over time, compare the PED in different languages, compare how much editing different post-editors do in the same project, and estimate how well a specific engine will perform on a specific content type by looking at the relevant historical data. Currently, the average PED of all projects is 22.21% and in one year it has decreased by almost 6%, which means that the overall quality of our MT systems is improving over time (Figure 4).
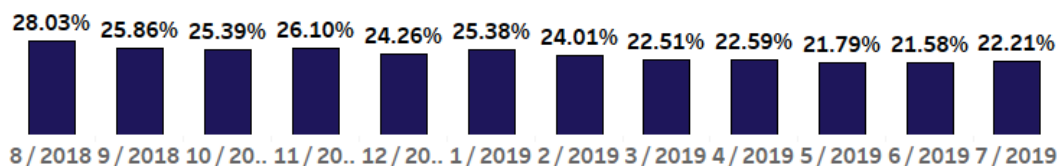

Figure 4. Average PED in the last 12 months.

4. Combining TM Matches with MT

The MT performance is crucial to the success of MT implementation but it is not the only requirement. It has to be used in combination with other tools and resources in the most efficient way. Typically, MT is applied on "low fuzzies", i.e. segments where the translation memory (TM) leverage score is lower than 75%. With the progressive improvement of the MT quality, however, the 75% threshold is being reconsidered, i.e. in many cases MT output is better than fuzzy matches. Our solution consists in comparing the average PED for MT suggestions with the average PED for TM matches and adjust the threshold as needed on an account level.

We have conducted multiple case studies on this topic, and one of the studies showed that for the specific account and language in question, MT requires less editing then almost any type of TM matches (including even 95%-99% matches). Table 1 shows the different TM match ranges and how the PED compares for MT and TM.

TM range	PED-MT	PED-TM
75-97%	15%	40%
80-84%	18%	35%
85-89%	17%	31%
90-94%	23%	21%
95-99%	*7%*	*16%*

Table 1. Comparison between the editing effort (PED) from machine translation (PED-MT) and TM matches (PED-TM) in a case study with English into Chinese translations.

5. What Have We Learnt?

One of the biggest lessons we have learnt is that MT is not a solution to all problems. On its own, it will make little difference in optimizing the translation workflow if it involves complex manual steps. It can help translators increase their productivity, but this is only one step in the workflow. Our goal to provide the best translation services to our clients in the most efficient way drives the need for automation of all the steps in the process that can be optimized. Automating certain processes can save as much time or more as an excellent machine translation system.

Another important component of success is training and support for all the people involved. Even at this advanced stage of implementing MT, it implies a change for many of our linguists and project managers, so we have to make sure that they are fully aware of all the processed and have all the knowledge needed to perform their task. We constantly work with all the roles involved (linguists, project managers, account managers, etc) by providing proactive training, answering requests, updating materials and making them easily accessible. Technology is only a tool to be used by people, and without the people its full potential will not be used.

Finally, properly using and collecting data is essential. By collecting information on how MT is used we can optimise many aspects of the process. Data on the PED and the time linguists take to post-edit can help us with MT quality evaluation, estimation of the translation budget, selection of the best MT engine, adjusting the TM match threshold any many more.

6. What next?

Our current work in progress includes automatic MT quality estimation (QE) on a document and on a segment level. On a document level, it will allow us to easily decide if the content in question is suitable for MT, and choose the best MT engine. On a segment level, it will allow us show the post-editor only the most useful segments, show the estimated quality score and know exactly where it is better to use the MT suggestion, the TM match, or start translating the segment from scratch.

Another development we are working on is Predictive MT, a predictive typing tool integrated with neural MT technology, which will allow the post-editors to see the MT suggestions that are being adjusted as they type.