

Managing change when implementing MT systems

Michel Lopez

e2f translations

2901 Tasman Drive #222

Santa Clara, CA 95054

mlopez@en2fr.com

Abstract

Managing large scale MT post-editing projects is a challenging endeavor. From securing linguists buy-in to ensuring consistency of the output, it is important to develop a set of specific processes and tools that facilitate this task. Drawing from years of experience in such projects, we will attempt here to describe the challenges associated to the management of such projects and to define best practices.

1 Introduction

Implementation of MT systems usually relies on two very separate teams.

- A linguistic engineering team, usually composed of a project manager, several engineers and one or several computational linguists. Its mission is to configure the MT engine and produce for each source file an “MT-engine output” (i.e. an engine-translated target file or a corresponding translation memory that can be reapplied to the source file).
- A post-editing team, usually composed of a project manager and several linguists. Its mission is to edit the MT-engine output and to produce final target files.

In most cases:

- The first team is located within a large MLV (multi-language vendor) or end-client.
- The second team is an ad-hoc team created either within the MLV or end-client, or most frequently within a contracted SLV (single-language vendor).

- Communication between the teams is very limited and infrequent, usually through project managers and not at team members level.

Whereas numerous papers and studies have focused on the tasks associated to MT engine configuration, we will concentrate here on the management of the post-editing team, whose success is equally crucial to the quality and timeliness of the overall project.

2 Background

e2f translations is a translation company specialized in English->French projects. It is composed of 4 offices worldwide (located in California, France, Mauritius and Madagascar), and has a 60%/40% mix of in-house linguists and freelancers.

The largest worldwide subcontractor in its main language pair, e2f has managed for several years a large number of post-MT editing projects, mostly on behalf of large MLVs.

Many of these projects were large and quick (several hundred thousand words in a few weeks to several million words in a few months).

This has enabled us to identify the multiple challenges linked to the management of such projects and to find appropriate solutions.

3 Challenges

In the course of managing MT post-edition projects, we have been faced with a number of challenges.

3.1 Finding linguists

Whereas it is relatively easy to find linguists for a standard translation project, it is much harder to build a team around a large post-edition project.

Some translators refuse outright (“MT will never replace human translation”) and it’s very difficult to convince to give it a try.

Many others initially show interest in machine translation, want to understand how it works and how it can help them in their daily work. But a large number of them have had early on a very bad experience on a MT post-editing project, where they have ended up retranslating the whole document, while being paid their review rate.

They have then sworn that they would never do that again, and it is difficult to convince them that the quality has improved a lot over the years, and keeps doing so.

3.2 Keeping linguists involved

It can be equally difficult to keep post-editors working on a project.

The nature of the job can be daunting, particularly when the linguists are faced with the same grammatical or terminology issues day after day, which they need to correct, and have very little influence on the quality of the MT output.

3.3 Managing quality

Post-MT projects are generally very large, and it can be difficult to maintain consistency throughout the project.

The fact that the initial output is automatically generated by the engine seems to facilitate the consistency, but in real life, as the quality of the output is always imperfect and needs to be edited, it’s quite possible that different post-editors edit in different ways, and it’s impossible to skip some kind of harmonization step.

3.4 Managing productivity

End-clients and MLVs consider Post MT projects as review projects rather than translation projects. As such, they are not willing to pay for multiple steps, regardless of the actual process needed.

In order to remain within the budget, this forces the post-MT team to implement strict productivity control measures.

4 Overcoming challenges

4.1 Securing linguists buy-in

It is important to keep explaining skeptical linguists that while it is true that MT will probably never completely replace human translation, it is becoming a tool that linguists have to learn to work with.

We found that it is useful to remind linguists that, whereas machine translation does not replace human translators, it takes away the boring and repetitive part of translation and helps them focus on the most interesting aspects.

It is also possible to involve senior linguists in the overall project, but mostly by giving them linguistic coordination tasks, rather than post-editing tasks per se. They bring experience, credibility and cohesion to the team.

4.2 Keeping linguists involved

In order to keep linguists involved, it is crucial to understand what are the obstacles they face during post-editing, and to make sure that the MT engine configuration team takes into account their remarks about terminology, etc. in such a way that they don’t see the same errors over and over again, which leads to frustration and ultimately defection.

It is always important to insist that the client reviews the source files before sending them to the MT engine, as typos and other issues in the source files get magnified by the MT process and contribute to the alienation of linguists, who feel comforted in their initial opinion that “MT is not ready” and tend to bail out.

Of course, if the source text has been written in “Simplified English” (no complex sentences, passive voice, use of synonyms, compound ambiguous phrases or culturally-specific sentences) or has been pre-processed through a software such as Acrolink, the satisfaction of linguists increases and it becomes easier to maintain morale within the team.

4.3 Managing quality

Quality is a direct function of individual post-editors level as well as overall consistency. We found that the only way to increase both is to foster communication within the team and designate senior linguistic leads in charge of the linguistic aspects of the project.

4.4 Managing productivity

On many projects, we noticed that productivity improves if linguists are aware of MT processes and follow the basic rules of MT post-editing. If they are not ready, they will look at every segment for several minutes and finally decide to retranslate it from scratch. In this case, productivity can be worse than with traditional translation.

4.5 Managing resources

For the success of these large projects, it is crucial to include in the team a few very senior linguists who understand the benefits and inevitability of machine translation.

It is also beneficial to include junior linguists, who are more technologically-savvy and are more likely to accept to work in a post-MT environment.

5 Case study

5.1 Background

In this section, we'll take the example of a very large project that we have been working on in the past 6 months, and consisting of batches of 700k to 1M words to be processed each month.

The end-client is a large pharmaceutical company, and we are translating a very large array of e-learning material describing their internal SAP implementation.

After initial missteps, we have been able to put in place a structure that is delivering final files of the expected quality while managing to stay within deadlines and budget.

5.2 Post-MT editing team

The team is composed of the following members:

5.2.1 Project Manager

A senior Project Manager with experience in the management of large projects, has the following responsibilities:

- Liaise with client's project manager
- Manage project resources, in collaboration with Linguistic Lead and Resources department
- Manage overall project steps and deadlines
- Ensure quality of deliverables

5.2.2 Linguistic Lead

A senior Linguistic Lead with experience in post-edition and large projects, has the overall linguistic responsibility of the project and performs the following tasks:

- Liaise with client's linguistic resources
- Review all glossaries
- Make all global linguistic choices
- File bugs in client's system
- Maintains team cohesion and team spirit
- Provide feedback to linguists
- Include/select/eliminate linguists in sub-teams

5.2.3 Senior Linguists

Several Senior Linguists are each in charge of a domain-specific sub-team, and perform the following tasks:

- Ensure consistency within domain
- Answer linguistic queries from linguists
- Refer choices to Linguistic Lead in order to guarantee overall consistency

5.2.4 Post-Editors

Post-Editors are mostly junior linguists specialized in a domain, they perform the following tasks:

- Post-edit individual files
- File linguistic queries

5.3 Process

The overall process of each batch is the following:

5.3.1 Post-Editing

This is the step normally associated with post-editing projects. Linguists edit segment by segment the MT output, while following glossary, translation memory, instructions, style guide, and perform a spell check before delivery.

5.3.2 Review

In the same fashion as is done on a TEP project, Senior Linguists review segment by segment Post-MT editing output, checking for accuracy, grammar, typos, fluency of output, and adherence to all references.

As is explained in later sections, this step is not performed for all files, as it would otherwise be impossible to stay within project budget.

5.3.3 Quality Assurance

In order to guarantee accuracy and consistency, we then run a software that automatically detects remaining typos, grammar errors, inconsistencies, terminology issues, etc.

Implementation of corrections is done by Junior Linguists.

5.4 Tools

Apart from standard CAT tools, we are using a number of Project Management tools.

5.4.1 Facebook group

In order to foster team spirit and increase exchanges within the team, a Facebook group has been created.

When we started using this group, the lead linguist was making recommendations/suggestions, but after a while, linguists have started asking more questions, which other linguists answered with suggestions, without needing senior linguists involvement for all points.

We found that using Facebook rather than email for this type of communication fosters team spirit and encourages everybody's participation.

5.4.2 Google documents

Because whereas a Facebook group is a very good way to quickly share news and ask questions, it can become cumbersome as a repository of structured data.

This is why we are using several shared Google spreadsheets to share structured information about the project, for example:

- All project files, color-coded to denote stage of each file within the project (Post-Editing, Review, QA)
- Consistency, where linguistic leads indicates all linguistic choices in a formal way
- Q&A, for questions that none of the linguists could answer within one day
- Source code issues, to report to client

5.5 Team Management

Junior linguists have been selected on the basis of tests. The first post-edited projects of each linguist are thoroughly reviewed by a Senior Linguist.

It is well known in the industry that average reviewer speed should be about 1,000 words per

hour. If after a few projects and proper feedback/comments from reviewer, the reviewer speed for a particular linguist is still below 700 words per hour, then the reviewer and project manager meet to decide whether to keep the linguist on the project.

We found that reviewer speed is a better measure of post-editor quality than other methods, and it reflects an economic reality as it directly measures the cost of review.

Conversely, if reviewer's speed increases well beyond 1,000 words per hours and the reviewer indicates that a full review is not necessary anymore, the linguists files are not thoroughly reviewed anymore, but rather directly sent to the Quality Assurance step.

We have found that this method gives us a good compromise between overall quality and productivity.

5.6 Lessons learned

Finally, each batch is reviewed in a post-mortem session.

We review processes, tools configuration and linguists. A brainstorming session helps us decide on improvements to include in next batch.

We also suggest improvements to the client, particularly about the source files issues and MT engine configuration.

6 Conclusion

Machine translation is here to stay, and the whole industry needs to learn how to manage post-MT editing projects in a fast and efficient way.

Most of the challenges linked to the management of post-MT editing teams are similar to those that arise in traditional TEP processes, but some are new, and translation companies will need to adapt their tools and processes accordingly.