Best Practices for Enterprise Scale Post-Editing

Andrea Stevens & Rodrigo Fuentes Corradi

Language

Introduction

The volume of content that requires translation is growing faster than we are able to handle with conventional means. The data-driven world requires **speed and agility** and Post-Editing Machine Translation **(PEMT)** has proven the most effective way to meet the quality communication demands of our time.

But how do companies jump onto this fast track MT highway and successfully send their brand message to a global audience? The answer is to add a structured and controlled PEMT step to your translation workflow.

PEMT is the process of allowing machines to do the heavy lifting of translation with final editing and quality assurance being done by a trained translator. While this sounds easy, the reality can be challenging unless a holistic process is defined and followed at each step. The best results can only be achieved through an integrated team of data engineers, linguists and developers, **combining their expertise** to produce a relevant PEMT solution.

At every stage, the stakeholders must provide **insight and understanding** of all requirements to accelerate the MT process. These resources have to apply both a **short-term, quality-oriented** view, as well as a long-term view to create a **framework** able to accommodate future requirements. Without this framework, the ever-increasing volume of data will outstrip the current **capacity** of MT engines and editors.

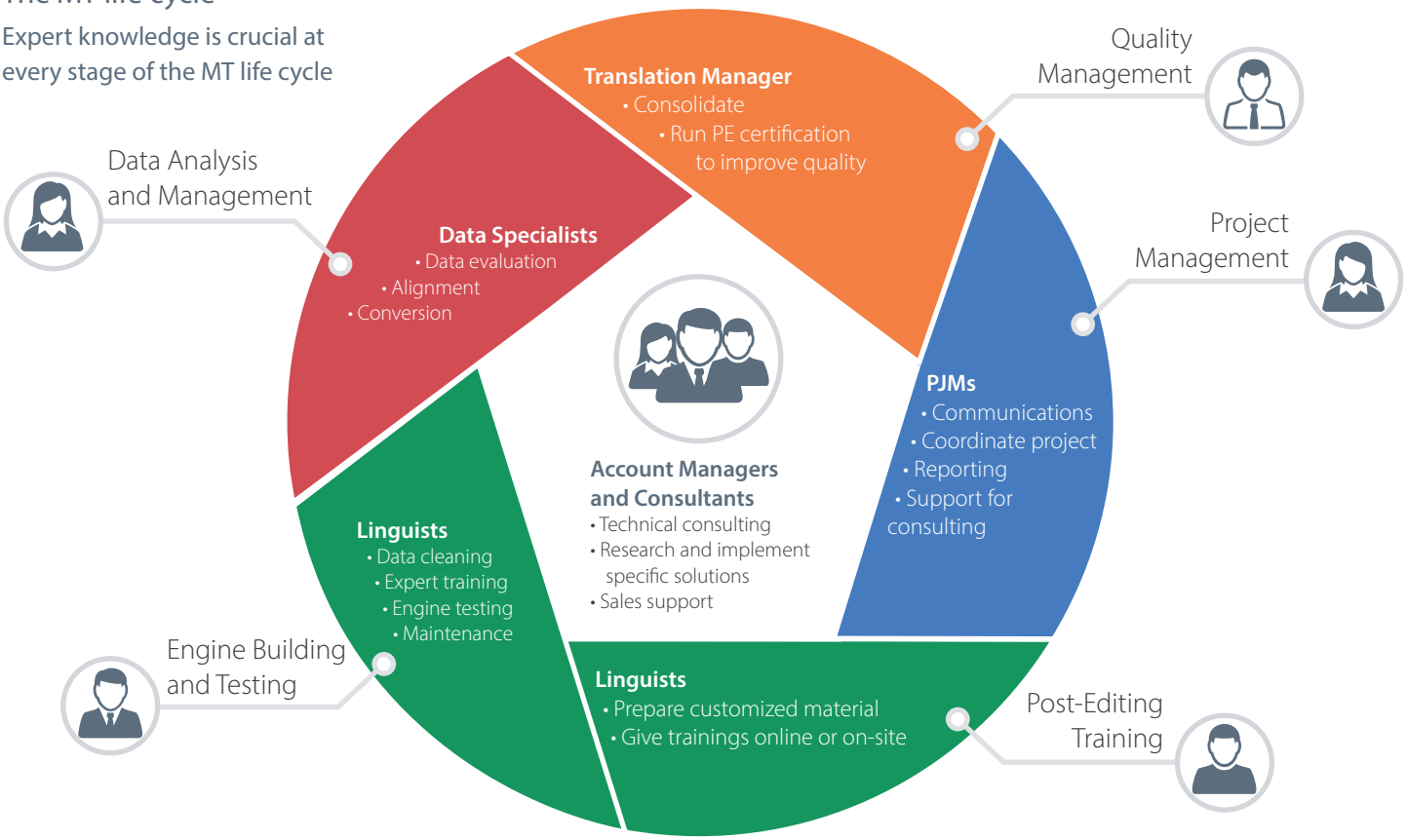
PEMT process

This paper looks to identify **key MT stages and stake holders** to provide clear insight into creating shared PEMT **objectives**. To illustrate PEMT best practices, the following model will be used:

1. Content evaluation
2. Engine creation
3. MT output evaluation
4. Post-editing machine translation
5. Quality assurance

The MT life cycle

Expert knowledge is crucial at every stage of the MT life cycle



1 Content evaluation process

Deciding what content can successfully be translated using PEMT is a commercial and financial decision. Different content types require **different translation approaches**. Some content is simply not suitable for a MT process as it requires a more nuanced approach with the skill that can only be provided by human translators.

A **content audit** allows a data review to establish MT suitability. A computational linguist can conduct the review of current and future needs by prioritizing on the basis of:

1. General **suitability** and the chances of MT translation success (language/domain/content)
2. **Quality** requirements
3. Volume requirements that are too capital and time intensive for human translators to manage
4. The technical integration required to produce a production **workflow** (CAT tools/CMS & Workflow systems)

Linguistic and engineering resources need to apply their previous experience and data points to provide the **PEMT project team** with insight as to which content types to target. Companies can achieve the most significant gains through PEMT by targeting content that is both strategic and high volume with tight turnaround times.

It is critical for the evaluator to **identify and record** not just the content types where PEMT has been successful but also where PEMT has not met project goals. Collecting feedback from editors and users is critical for this to happen. A continual feedback loop will eventually **increase the speed** at which evaluations take place and help to more **correctly predict** translation outcomes.

Takeaways

- **ROI** goals need to be established early in a process
- **Difficult or technical content** may need attention from specialist linguists
- **Quality expectations** vary across content types and need agreement by all involved
- **Continuous project monitoring** enables future improvements

MT content evaluation

Guide to the right solution for your content

2 Engine creation

The MT engine and how it is trained is one of the most critical parts of the PEMT process. A mistake in engine training could result in repeated missed or inaccurate translations, forcing editors to make the same edits over and over again. This will have a negative impact on the financial ROI of the PEMT process.

Building an engine to handle large data sets across several file formats requires a structured process. Specialist tools at the **data intake** stage are required to optimize the data and ensure that it is understood, cleaned and prioritized. Engine design must also include a view of future translation requirements to ensure the ability to process data quickly and provide insightful and relevant analytics down the road.

Successful training is based on fully understanding the quality expectation for a particular project. This includes having access to **important project assets** such as terminology and style guides. These must be incorporated into the MT engine training as well as used to **measure the success** of the output.

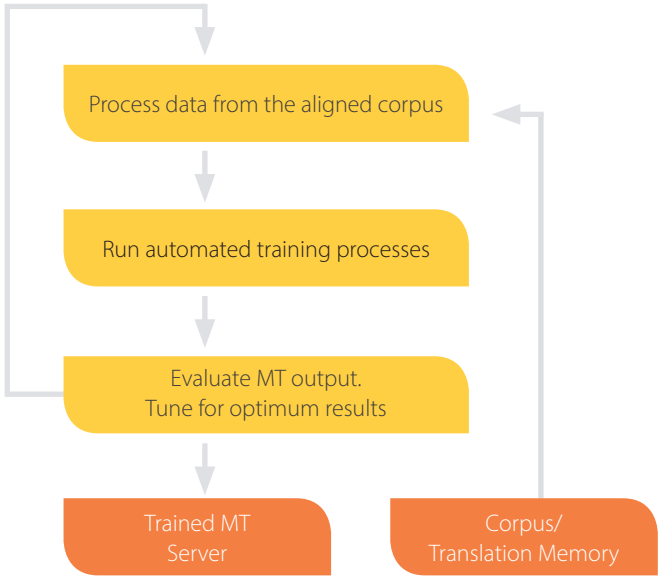
Once the training data corpus (a large collection of words) has been selected (typically Translation Memories), it is the role of the computational linguists to decide how to **optimize the data**, with the key objective of ensuring high-quality MT performance. Computational linguists, working with data specialists and engineers, need to view the data from a **high-level** vantage point but also **analyze the details** to understand the eventual translation output.

There is a close relationship between engine data and future client content. A combination of **automated** processes and **human** validation ensures that tests remain relevant for use with future data. Being able to **spot obvious issues** such as corruption in a million-word data set (impossible to the human eye) is critical to assessing content. **Recording the results** in a structured way will ensure that lessons are learned for future efforts.

It is likely that **several attempts** are needed (possibly simultaneously) in order to find the optimal engine. The choice of engine design has to be agile, based on experience, include previous data recipes and draw on an inherent understanding of Statistical Machine Translation behavior. Trainers must be able to incorporate industry standard automated evaluations to narrow down the best engine candidates for human evaluation.

Engine creation process

Components and workflow on Engine training



Takeaways

- For many terminology-rich commercial projects **superior MT** output is needed; the training phase must **select the correct data** to make it relevant to the PEMT user experience
- **A multi-discipline approach** requires collaboration between engineers, developers and linguists
- Agile and customizable tools allow computational linguists to categorize, observe and **optimize data** (generic/ domain/customer specific)
- The training environment needs to be capable of creating multi-scenario engines and **recording** and providing analytics for further understanding
- A clear approach on data with both a **wide and granular** view, with the ability to spot the trajectory of the final output, is essential

3 MT output evaluation

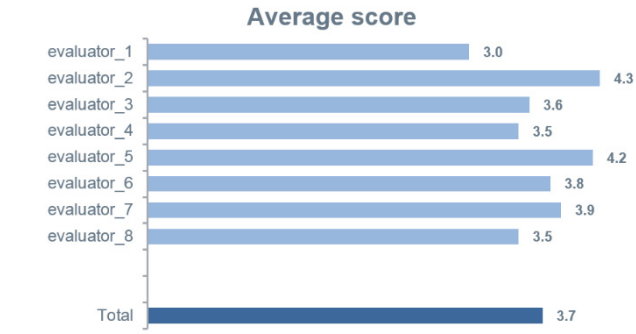
Evaluating the MT output to ensure it is suitable for post-editing can be done in many ways, using either standard industry **automated measures or several forms of human evaluation**.

Some of the factors that determine the type of testing required are effort, cost and logistics. Automated measures offer a quick, recognized and cost-effective way to analyze the potential quality of a MT engine. However, as the use case becomes **more complex** (such as PEMT), so does the requirement to test in a way that is **relevant** to the particular project.

For **PEMT**, meaningful measures need to have a link between test results and the final **production environment** of the user. This is particularly relevant given the exponential growth in post-editing requirements. Industry translators have already post-edited large volumes using MT and the test findings need to reflect their experience.

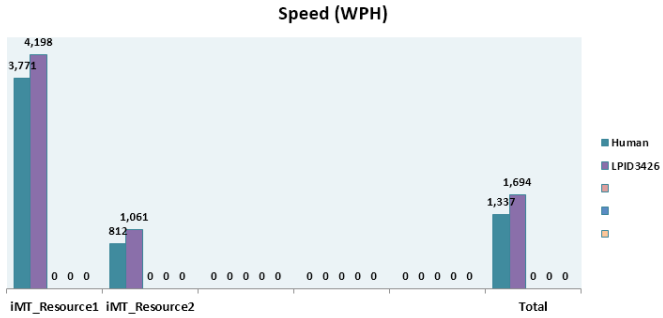
It is essential that testing matches real life while still focusing on the quality of the MT output. Some tests mimic all the features of a standard production environment whilst others focus exclusively on MT. Both methods have their benefits, but it is critical that they are transparent and correctly predict the increased productivity that PEMT can offer.

All **testing can be biased** by a number of factors such as inaccurate time keeping or the inability of testers to follow instructions. PEMT testing needs to demonstrate a **credible gain** over human translation for skilled translators to be motivated to take on PE assignments. It is therefore important to set up tests that are secure and produce easily understood, transparent and credible results. Again, we need to focus on tools, a structured testing environment and a **streamlined and executable process** as well as the required resources that can produce meaningful and authentic results.



MT quality evaluation

It has long been accepted that translators need an integrated **set of tools** to be productive; a workflow tool to receive and deliver files and a CAT (Computer Assisted Translation) tool in which to work with an integrated glossary. Likewise, a computational linguist needs an **integrated environment and process** flow, with tools that help create valid and **representative** test beds that can easily be used to test resources online – an environment that is capable of providing detailed analytics such as speed of manual translation versus post-editing.



Post-editing speed comparison

Most MT tests are carried out across a **number of testers**. This is to safeguard the results from individual user errors and inconsistent behavior. An **online testing environment** records a host of editing behaviors such as cut and paste (from other translation references), deletions and general key strokes which help to interpret the user experience. You need experienced linguists and translators with both **PE experience** and **project/domain knowledge**.

Maintaining a database of past translation **associates** can considerably reduce the risk of repeating costly tests due to invalid results. The design of a testing platform requires the **collaboration** of developers, engineers and linguists to ensure the environment is capable of producing **accurate results** that will lead to a successful project. Given that PEMT is aimed at large-scale projects encompassing hundreds of thousands of words and multiple **PE** resources, a state-of-the-art testing process is an essential ingredient for final success.

Test results need careful **archiving**, with the ability to quickly and easily **retrieve results** by project, domain or resource to **accelerate decision making** on future projects. A custom-designed testing environment can be expensive to set up, but potentially worthwhile when predicting the outcome on multi-million word projects and potentially **decreasing costs** by 30–50%. Archiving successful tests expedites future deployments without jeopardizing a **project start** or using the **wrong process**.

A streamlined testing methodology and the **efficient** capture of data patterns are more likely to emerge from an **integration of platforms**, rather than having disconnected stores of data across different departments. However, even if there is a single integrated platform, you still need **skilled computational linguists** to analyze data and deliver relevant and predictive observations.

Building MT engine solutions has become much more accessible with several **self-training** environments emerging across the MT industry. However, without the knowledge of how the **data influences the training,** it is not possible to exponentially improve such self-built trainings to aid the professional translator. **Linguists are at the heart of the decision-making process** and need to ensure that results are intelligible and actionable and match user expectations.

Takeaways

- Tests need to be able **to predict the user PE experience**
- Creating valid tests needs a **combination of tools and human validation** together with trained resources to obtain useful test results
- **Tests can be expensive**, especially with the inclusion of human evaluators; a purposeful **strategy** that combines automated tests with a variety of human evaluations will deliver **cost-effective and compelling** results
- Custom testing platforms, where test assignments are published to multiple groups of translators and key data is **consolidated**, ease the effort and long-term cost of testing

4 Post-editing Machine Translation

The processes described for evaluating engines are very closely linked to the task of post-editing machine-translated output **(PEMT)**. Language, domain and content suitability go hand in hand with creating high-quality output that allows **translators to focus** on the task of **refining** the pre-translated content.

There are many academic **definitions of post-editing** that all share one central commonality – the human element. At its core, PEMT simply implies that a human translator is **editing or revising** text that has been translated automatically by **machine translation** software.

When providing a **light post-editing** service, quality is defined by acceptable and **actionable** translations, **not perfect** grammar and style. It is the responsibility of the post-editor to ensure that the customers' instructions are followed to produce a final result that achieves the right balance of information transfer and language standards.

The requirements for post-editing to **publishable quality** (or full PE) on the other hand are very similar to those of conventional human translation. Common translation principles such as correct grammar and spelling as well as consistent use of terminology and style are required.

Post-editing in all its variations has quickly become an important part of mainstream translation. However, with digital **content volumes ballooning**, demand for PEMT solutions is growing exponentially. According to Common Sense Advisory, businesses need their website to cover 14 languages to reach 90% of the most economically active international markets.

The continuing strong demand for PEMT triggers two key questions:

1. How do we ensure the **availability** of enough trained and experienced post-editors?
2. How do we **engage** new and existing **post-editors** and keep them motivated?

Given the need for both experienced post-editors to accept more PEMT tasks and to recruit new post-editors, the answer to our first question must include **training and knowledge transfer.** Our goal is to **retain** the knowledge and experience of mature and often specialist **translation professionals** whilst harnessing the **curiosity** and eagerness of **newly qualified translators**.

Reaching out to translation professionals through PEMT-focused trainings, workshops or webcasts is one avenue to **promote PEMT** within the freelance community. This provides a platform to explain the upfront investments for productive post-editing in terms of time and practice. Post-editors need to have the tools to work productively with MT to enable an **easy transition** from purely human translation-related tasks.

For a wider appeal, SDL's **PEMT Certification** offers an introductory course to the techniques and skills required in post-editing MT output. It gives post-editing best-practice recommendations and practical examples. The course is aimed at anyone interested in PEMT and enables translators to meet the demands of a changing market and future-proof their offerings. SDL's **training courses** equip translators with core **skills** and strategies for effective post-editing as well as explaining known MT behaviors and **common MT errors.** Armed with this wealth of knowledge, they are ready to experience post-editing first-hand.

This is where our second question comes in. If we want translators to have a great post-editing experience what can we do to make **PEMT both rewarding and sustainable?** The simple answer is to provide a **solid, streamlined and proven MT process,** including the five key stages of the MT lifecycle described in this document:

- **Evaluate content** and translation assets for suitability
- **Train MT engines** for specific purposes
- **Test MT quality** and productivity
- **Train translators** to become MT Champions
- **Monitor quality and retrain** MT engines when appropriate

Takeaways

- **PEMT** is closely linked with **key process stages** and should not be considered in isolation
- Different post-editing solutions cater to **different client requirements**
- **Availability** of experienced and trained post-editors – training and recruitment
- **Rewarding** and sustainable post-editing through solid MT **process**

5 Quality assurance

An efficient and well thought-out MT process not only facilitates sustainable post-editing but includes quality assurance based on number of key factors.

From a post-editing viewpoint, the overriding goal is to create and **maintain high-quality engines** which enable post-editors to focus on processing the output in the **most productive** way possible.

MT engines are created for specific use cases based on customer data. Quality assurance is a key tool to maintain or improve engine quality through:

- **Technology** improvements
- Additional or **updated user content**
- **User feedback**

MT quality and feedback are closely related. Issues with the MT output need to be reported as soon as possible to allow for a prompt resolution.

Linguistic feedback includes language-related issues such as wrong words, bad grammar or syntax/word order issues. When reporting linguistic feedback, the user is asked to **bear in mind expected MT behavior** and focus on systematic and recurring errors.

Technical feedback describes process-related issues such as missing tags or error messages. SDL has put in place a streamlined process to **channel feedback** and support our clients' MT roadmap beyond the initial creation of engines.

SDL's internal point of contact will provide a comprehensive **feedback template** and request any supporting materials as required. Linguists and data engineers are on hand to carry out **investigation or data analysis** as needed.

SDL's POC will be in charge of client communication and remain in close contact throughout to report findings and **manage timelines**. This includes the timeframes for implementing potential fixes as these can vary depending on analysis results or technical complexity. Whilst some issues can be fixed without delay, others must be implemented as part of a mid- or long-term product improvement strategy.

Feedback integrates MT customers firmly into the development process. Feedback concerning the technical capabilities of a MT solution or the linguistic quality of machine-translated output helps shape the process of redefining MT roadmaps and generate momentum and speed for inclusive and far-reaching MT solutions.

Takeaways

- Quality Assurance is a key tool to maintain or **improve engine quality**
- Feedback is divided into **technical** and **linguistic** categories and processed accordingly
- SDL's **process** channels feedback efficiently
- Feedback is an **opportunity to redefine** and accelerate MT roadmap

Conclusion

Whilst Machine Translation and PEMT are constantly evolving, the concept of creating an all-encompassing **MT life cycle** can be locked down within the framework of the key stages described in this White Paper. **Data catalogs and experience** are the building blocks of this holistic approach, mistakes made along the way have helped to shape and **refine future processes** and methodologies.

With the MT market no longer in its infancy, **mature MT** players and users alike share technology, goals and communities but utilize different approaches. **PEMT** requires clear and fluent **control mechanisms** to future-proof MT solutions in line with the level expected by big corporations.

Next-generation MT no longer implements a one-size-fits-all approach, but acknowledges previous attempts – and failures – to redress the balance between **corporate MT** goals and end **user experience**. MT is moving out of the research lab into a bigger, more critical playing field, powered by **multidisciplinary teams**, end-user satisfaction and a rewarding PEMT experience. The combination of structured user **feedback** loops and the latest, best-of-breed MT **technology** put the post-editor firmly at the heart of the MT life and improvement cycle.

A look into the future

Thirty years **after CAT tools** marked a revolution, **real-time MT learning** mechanisms are the basis for a radical recalibration of the relationship between MT and translation experts. These technology advances allow translators and post-editors to play a pivotal role in the acceleration of MT roadmaps across the industry.

The latest best-of-breed technology now goes beyond a focus on quality through a structured feedback loop: it aims to directly learn in real time from the choices made by legions of skilled and dedicated translators.

To see how SDL can meet your Machine Translation needs, contact **imt@sdl.com** to discuss an analysis of your projects and data.

Andrea Stevens is Translation and Quality Manager for SDL's intelligent Machine Translation group. She trained as a professional translator in her native Germany and has worked for SDL for over 15 years in a variety of senior linguistic roles.

Andrea has been involved in Machine Translation at SDL from the very beginning, when SDL pioneered a high-quality translation process incorporating Machine Translation technology.

With a wealth of experience in localization, language services, translation resource and quality management as well as Machine Translation, she works with SDL's internal and external clients to promote the benefits of MT and help translators to integrate MT in their daily work.

Rodrigo Fuentes Corradi has 15 years of experience in the localization industry including 10 years in strategic and operational roles. He joined SDL in 2003 and in 2007, moved to the MT division of SDL. Here, he helped commercialize MT Post-Edit with many of SDL's major clients.

During his time at SDL iMT, he has helped drive PEMT usage in SDL from 3% to 27% while maintaining the high quality standards required for publishable content. Currently, SDL post-edits around 200 million words per year for their clients.

SDL_wp_iMT_best practices_EN_A4_050216