Machine Translation Post-Editing Levels: Breaking Away from the Tradition and Delivering a Tailored Service

Mara Nunziatini
Welocalize Italy S.r.l.
Via Alserio 22, Milan, Italy
mara.nunziatini@welocalize.com

Lena Marg
Welocalize Inc.
Frederick, MD, United States
lena.marg@welocalize.com

Abstract

While definitions of full and light post-editing have been around for a while, and error typologies like DQF and MQM gained in prominence since the beginning of last decade, for a long time customers tended to refuse to be flexible as for their final quality requirements, irrespective of the text type, purpose, target audience etc. We are now finally seeing some change in this space, with a renewed interest in different machine translation (MT) and post-editing (PE) service levels. While existing definitions of light and full post-editing are useful as general guidelines, they typically remain too abstract and inflexible both for translation buyers and linguists. Besides, they are inconsistent and overlap across the literature and different Language Service Providers (LSPs). In this paper, we would like to comment on existing industry standards and share our experience on several challenges, as well as ways to steer customer conversations and provide clear instructions to post-editors.

1 Introduction

As one of the largest multilingual LSPs, we have been offering machine translation post-editing services for many years, and our team supports more than 30 of our largest customers in the Enterprise or Regulated space with MT and post-editing programs in often 30+ language pairs. When implementing machine translation for a new customer, we always provide a post-editing training to the linguists working on the program.

During this training, among other relevant topics, we focus on the basics of post-editing and we explain what the client's requirements are regarding final translation quality.

Based on experience, we found that it can be very complicated to communicate what is expected of linguists in different post-editing levels. While it is easier to explain what is expected of light versus full post-editing, there are some grey areas that don't fall either into the full post-editing or the light post-editing service. Furthermore, our customers will often not be experts of translation quality assurance methodologies, and also not be familiar with the common definitions of the different levels of post-editing. As such, they are themselves often not entirely sure which approach would meet, exceed or fall short of their requirements. It is therefore crucial to guide them and define their requirements from the very outset, also in order to be able to clearly communicate them to the post-editors. This is extremely important since post-editors might feel confused if they do not receive clear instructions, and will probably end up delivering a quality that is either too high – in this case they will not be productive – or too low – and the clients' quality requirements will not be met.

Ultimately, the effort of the post-editor depends strictly on clients' quality requirements, therefore, it is not always advisable to rely exclusively on the current, most commonly used post-editing guidelines. In order to precisely define the quality requirements for each post-editing task, we reference the DQF-MQM error matrix and the TAUS DQF content types to align all parties on what types of errors are acceptable for a translation request given its purpose, target audience etc.

In this paper, we would like to share our experience on existing industry standards,

challenges, and ways to steer customer conversations and provide clear instructions to post-editors.

2 Existing Industry Definitions and Standards

As for quality, two error typologies were proposed in the last years: DQF by TAUS, 2011 (O'Brien et al., 2011) and MQM by QT21, 2014 (Lommel et al., 2014). These provide more flexible and dynamic ways to assess quality, and apply the same approach to machine translation and human translation. While they can also be used separately, the two typologies were brought together in 2015 into the MQM-DQF quality framework. These error typologies also aim to move away from the LISA QA model (LISA, 2006), used for a long time in the localization industry to rate translation quality. We like the DQF-MQM error typology because the error hierarchy, made up of well-structured main criteria and sub-criteria, allows for a granular categorisation of the quality issues in the translation.

Regarding post-editing, as mentioned above, it is common knowledge and generally accepted in the translation industry that there are different levels of post-editing, aimed at obtaining a final text that satisfies diverse predefined purposes and quality standards. However, there appears to be no recognised industry-wide standard and the definition and guidelines of each level of post-editing are inconsistent across the literature and different LSPs. Currently, the most commonly used and referenced definitions of light and full post-editing in the localization industry are probably those provided by the International Organization for Standardization, GALA, TAUS and Sharon O'Brien (O'Brien, 2010). While the last two were already analysed by Hu and Cadwell, we would like to summarize what ISO and GALA say on the different levels of post-editing, before we proceed with the comparative analysis.

2.1 ISO Standard No. 18587

The ISO standard defines the requirements for full and light post-editing, as well as post-editors' competences. According to the standard, the final output after full post-editing should be equivalent to human translation. Therefore, if we had to reference the DQF-MQM high-level error types, the post-editor should focus on Accuracy, Fluency, Terminology, Style and Design. Plus, post-editors should edit any inappropriate content (see Appendix B).

The standard is less precise regarding light post-editing, but it still calls out that the post-editor should focus on Accuracy and disregard Style. For both light and full PE there are some less clear instructions regarding inappropriate content, that should be edited, and restructuring of the sentence, which should happen only in case of unclear meaning.

2.2 GALA

GALA references an article from Juan Rowda (Rowda, 2016): *Better, Faster, and More Efficient Post-editing* to explain the differences between light and full post-editing. According to Rowda, full post-edited output should be close to human translation quality. During full post-editing, the linguist should focus on Language (grammar and spelling), Terminology, Style and Accuracy error types.

On the other hand, light post-editing should aim at fixing major/blatant errors only, while minor issues are acceptable. More precisely, during light post-editing, linguists should focus on accuracy. They should not focus on punctuation, style and spelling, and preferential changes should be avoided. While these guidelines are in line with other common definitions of light post-editing, they remain vague for a post-editor to implement. An interesting aspect of these guidelines is that the checklist for light post-editing also mentions that light post-editing should allow for a fast turn-around.

In addition to these, there are older and helpful guidelines found in translation studies publications, which we will leave aside here.

3 Challenges with Existing Definitions of Post-Editing

The main challenge with terms like "light", "medium" and "full" is that they remain very abstract. Hu and Cadwell showcased already in 2016 that the literature seems to offer inconsistent and/or overlapping nomenclature, definitions and guidelines for post-editing.

Having said this, it seems to be broadly accepted that light post-editing should focus on conveying the meaning of the source text in an accurate way. Therefore, if we had to use the DQF-MQM error types as a reference – instead of the categories from the LISA QA Model (Localization Industry Standards Association Quality Assurance Model) and SAE (Society of Automotive Engineers) J2450 translation quality

metric, as suggested by Hu and Cadwell – we could say that light post-editing should focus on fixing Accuracy error types, while it should not focus on Style, Design, Locale Convention and Verity error types, as long as the information is delivered accurately. Whether Terminology, Spelling and Grammar errors should be penalized in a light post-editing task seems to remain controversial and unclear – the requirements as for these error types are inconsistent (see Appendix A).

On the other hand, if we consider the findings from Hu and Cadwell as well as the ISO standard No. 18587, it seems to be broadly accepted that full post-editing should focus on readability. However, there appears to be no common agreement as to whether full-post editing should be of equal quality to conventional human translation from scratch. Considering the different guidelines we analysed, we could say that full post-editing should focus on Accuracy, Fluency and Terminology error types. Style is discussed controversially, as there is no agreement on its importance between all the different guidelines (see Appendix B). According to the TAUS guidelines, the style "may not be as good as that achieved by a native-speaker human translator",[1] while stylistic and textuality problems should be ignored according to O'Brien. On the other hand, we read that the ISO standard No. 18587 recommends that client's stylistic guidelines are followed, and highlights that the style should be appropriate for the text type. Lastly, GALA simply points out that the style should be consistent and appropriate.

Some LSPs also provide "medium post-editing" services, but the guidelines for this quality level are even more vague and inconsistent, and this level of post-editing is mentioned only sporadically in the literature. Generally speaking, when performing medium post-editing, we expect the post-editor to put more effort into editing Terminology, Fluency and Style compared to light post-editing, but not to the same extent as they would for full post-editing. There appear to be no medium post-editing definitions in the literature that we could reference here; the easiest way to derive a distinction between full and medium post-editing, for instance, might be via the text type and translation purpose, i.e. texts that

are more stylistically challenging and complex by definition would always require full post-editing, whereas text types with a simpler structure (often technical manuals) could fold into medium post-editing. However, this might ultimately be an unnecessary definition as such, that could also be covered by the full post-editing requirement for "appropriate" style.

As we can see, these guidelines leave some grey areas when it comes to a hands-on post-editing task. For example, if I am performing full post-editing, should I check that bullets are consistent in the same list of items? That the headers are all title case? And what happens if I notice that the target language is using masculine form – for example, "amigo" in Spanish – when the source language might refer both to feminine and masculine gender – for example, "friend" in English? Should I edit all of these, or is it ok to leave those as they are?

There is also a potential problem in that existing definitions appear to assume that only one linguist should ever post-edit the machine translated output – irrespective of full or light. In other words, there appear to be no guidelines specifying how many linguists should be involved in the different post-editing levels, and the ISO standard No. 18587 does not set any requirement in this sense. It just mentions the requirement of a process to make sure that the final product meets the specifications. In the localization industry, however, it is still very common for translation buyers to enshrine a so-called "4-eye process" in the contract, i.e. irrespective of MT, that content needs to undergo post-editing plus a separate review or revision step, and potentially even a quality assurance step, which in some cases might be performed by a third party. In other words, customers still tend to buy a specific process (TEP, translation only, etc.), rather than an agreed service level or translation quality. This becomes even more stringent in the Regulated sectors, i.e. patent, life sciences, finance etc., where these additional steps can be mandatory to comply with other ISO standards and certification requirements.

Another challenge with the terms "light" and "full" post-editing is that often people misunderstand that these describe how much editing needs to be done, or in other words, how much effort the post-editor should put into the task, rather than what the final translation quality should be. More precisely, some people might erroneously

---

think that, if they were to translate the same content in multiple languages, depending on the quality of the raw MT output, some languages will require light post-editing while others will require full post-editing. For example, User-generated Content machine translated into Spanish will require light post-editing as the raw output's quality is good, while Finnish will require full post-editing, because the raw output isn't as good as for Spanish. This is a fairly common misunderstanding and yet another reason why we think it is better to focus on final translation quality requirements, than the vaguer definitions of light and full PE.

4   Challenges with Error Typologies

Since the lack of a clear, common approach highlighted by Hu and Cadwell (2016) is still a very present issue, translation service providers need to define their own methodology, in order to provide a flexible service offering, linked to transparent pricing for the client and fairer rates for the post-editors. What is ultimately needed is a highly flexible and granular approach, since the effort of the post-editor is essentially decided by the exact quality requirements of a given customer.

Rather than working with the somewhat vague definitions of "light", "medium" and "full" post-editing, we find that it is easier for all parties to define quality requirements by aligning on what types of errors are admissible for a translation request given its purpose, target audience etc. Considering the purpose of the text and the document type, and referencing the DQF-MQM error matrix, we help the clients choose what error categories are acceptable for them and what are not. Also, for each error category they decide how many (if any) major and/or minor errors they are admissible. We use the same framework for Quality Assurance (QA) steps to understand if the quality of the MTPE projects meets client's requirements – this way the linguists performing this task are fully aware of what they should focus on and we get full consistency as for quality requirements from the start until the end of the process.

In order to make this possible, first we created different groups of domains, considering the purpose of the document and the text type (based on TAUS DQF content types), and then we created different sets of standard checks for each one of these groups, aimed at getting a translation which is free of certain predefined unacceptable errors.

For example, for User-generated Content, we could propose a set of post-editing checks that focuses purely on the accurate transfer of meaning. User-generated Content would be an example of text type typically accepting a high error threshold – especially in light of the source input itself being known for being characterized by errors (O'Curran, 2014). At the other end of the spectrum we might find text types such as marketing materials with a focus on brand's style and tone of voice. We like to call the above-mentioned sets of standard checks "full", "medium" and "light" post-editing too, as our guidelines show some similarity with the most popular industry MTPE guidelines mentioned above. Then, building on these sets of pre-defined standard checks, we add or remove applicable error categories as per client's preferences, and we raise or lower the threshold of the acceptable number of minor and/or major errors.

As mentioned above, the error categories are also based on the DQF-MQM error typology. The DQF-MQM framework involves the use of a list of error categories, and the content quality is judged based on the amount and severity of the errors found. The errors can have different severity levels: critical, major, minor and neutral. "Neutral" applies when an issue should be flagged to the translator but is not counted as an error and does not influence whether the translation is considered a PASS or a FAIL. During our QA step, a post-edited text (or a sample of it) is evaluated by a linguist who marks the errors; all errors are added up, based on severity, and output a PASS or FAIL score, depending on the defined threshold. The thresholds are flexible and depend on content type, text purpose and perishability of the text. In practice, this strategy is extremely helpful, as we can agree with customers, post-editors and reviewers at a very granular level what issues need to be addressed during post-editing, and which are of purely preferential nature.

However, while error typologies for quality assurance are fairly common among professional translators and reviewers, it can be trickier to agree on error categories and severities with translation buyers. This is primarily due to the fact that the owner of a given machine translation initiative on customer side may not be an expert in translation quality assurance methodologies. On the other hand, on post-editor and reviewer side, the main challenge is changing the mindset, and getting professional translators to accept that

for certain content types and translation purposes, it is acceptable to leave certain types of issues in the machine translation output unedited. However, by providing a granular breakdown of what constitutes an error in a given translation request, it is much easier to train and support post-editors, and to monitor their actual productivity for the task at hand.

5   Use Cases Examples and Strategies

As we have seen on a high level, clients often have specific requirements that cannot easily and universally be categorized with the typical definitions of full, medium or light post-editing. In the following section, we will showcase some examples to explain our approach: based on the purpose of the document, the content type and the error types that the client is willing to accept or not, we build custom requirements and instructions for post-editors.

5.1   Use Case 1

A good example for "light post-editing" presented itself with a client who needed to translate Knowledge Base content within a defined budget. For the content and purpose, light post-editing seemed the appropriate approach, as the main goal was to provide final translations that accurately transfer the meaning, while maximising translator throughput within a defined budget. However, for this particular client it was important that product names were handled correctly, in this case kept in English also in the target language. Light post-editing per se does not typically focus on terminology (Hu and Cadwell, 2016); this requirement therefore implied additional editing effort, especially in cases of product names that were unknown to the MT engine at a given point in time, or not handled consistently in the data used to train the MT system. In this case, we therefore added the specific terminology check requirement to the obligatory checks for post-editing, still classifying the task as light post-editing.

Below you can see an example in which the MT engine translated an unknown product name literally from English into Portuguese. Standard light post-editing instructions don't necessarily require post-editors to review such instances, and post-editors could be tempted to leave this unchanged.

Source	**Bugcheck** 7E
Raw MT	**Verificação de bugs** 7E
Final	**Bugcheck** 7E

Table 1: Example of correctly edited DNT ("Do Not Translate") term.

Post-editors working on this account received a list with all product names to be left untranslated, and before project kick-off they were also trained to perform light post-editing while still ensuring product names were in line with the client's requirements.

5.2   Use Case 2

Another client translating Online Help content wanted to have medium post-editing performed on the raw MT output: this was defined as providing usable and accurate translations, without a need for stylistic flourishes or lengthy terminology research. One requirement, however, was that the translations should all use the formal tone of voice, in line with the brand's style. This again goes slightly beyond what we would typically define as "medium" post-editing, so this instruction was added to the mandatory checks for post-editors; see an example below from English into Spanish:

Source	**Change** the size of the logo on the traveler ticket **if desired**.
Raw MT	**Si lo deseas, cambia** el tamaño del logotipo de la entrada del viajero.
Final	**Si lo desea, cambie** el tamaño del logotipo de la entrada del viajero.

Table 2: Example of correctly edited tone of voice.

In this case, raw machine translation output would have been accurate and correct according to the typical medium post-editing guidelines, however it would not have met client's requirements.

5.3   Use Case 3

In this instance, we are using MT and post-editing for UI and UA content. Typically, for this content type we would recommend medium post-editing, as the focus is on accuracy and correct terminology, while style should not usually play a key role. However, this client wanted to also include stylistic requirements to reflect brand and voice. The impact of this was so significant, that this was ultimately classified as full post-editing. In the interest of maximising productivity, we typically train our post-editors to use as much of the raw MT output as possible, in line with the standard task definitions (TAUS MT Post-Editing Guidelines and ISO Standard No. 18587, just to mention two of those). In this case, however, they were instructed to make sure to always follow the client's preferred terminology and style – this implied editing the machine translation suggestions to reflect the client's style

guide, preferred terminology, punctuation, spelling (i.e. capitalization), tone and register.

Even when a client's requirements and post-editing guidelines are seemingly clear, we have experienced many challenges. For example, sometimes post-editors – also depending on how experienced they are (de Almeida et al., 2010) – find it difficult to understand what is expected of them and end up editing too much (over-editing) or not enough (under-editing).

5.4 How We Measure Adequate Editing Effort

If the post-editors are over-editing, they are not making an efficient use of the MT output because they are introducing unnecessary preferential changes. Generally speaking, we can recognise over-editing by comparing the raw MT output and the final post-edited files with our proprietary scoring tool and analysing common industry metrics like BLEU, GTM, Nist, Meteor, Precision, Recall, TER, and Levenshtein Edit Distance (Levenshtein, 1966). If we notice that the metrics are not in line with our expectations, and Edit Distance (ED) and/or TER are especially high compared to other target languages of the same project, or compared to what we usually see for a given language and domain, we might suspect that the post-editor is over editing. We would then check what was changed of the raw output, focusing on the segments with the higher edit distance, to find out where the post-editors are putting most of their effort and we investigate if the edits introduced are actually necessary to reach the agreed quality standards. In the example below, for instance, ED was particularly high:

Source	Milford, **MA**, USA
Raw MT	Milford, **MA**, USA
Light PE	Milford, **Massachusetts**, USA

Table 3: Over-editing in light post-editing.

The post-editor was instructed to perform light post-editing. Edits like the one in the example above are typically not in line with light post-editing expectations, as the raw output was perfectly understandable. If unnecessary edits like the one above are frequent in the final target text, it probably means that the post-editors were not clear about what was expected of them and were therefore unproductive.

On the other hand, if the post-editors are under-editing, they will deliver a final translation that does not meet the agreed quality standards and will fail Quality Assurance checks. If the Edit Distance for a given translation is suspiciously low, i.e. it is especially low compared to other target languages of the same project, or compared to what we usually see for a given language and domain, we would check closely the quality of the final translation to make sure the post-editor actually implemented all the necessary edits.

If we come across over-editing or under-editing issues, we follow up with the post-editors and provide feedback as well as extra training, to make sure they understand their task, mind their productivity and align with client's requirements.

6 How We Provide Guidance and Set Expectations

In order to help clients understand what service level best fits their needs and to make it clear to post-editors what is required of them, there are different strategies an LSP can put in place.

6.1 Supporting Post-Editors

To support post-editors and make sure they have a clear understanding of what is expected of them, we find it very useful to have meetings at the start of a new engagement.

On these calls we explain the project, the quality level agreed with the customer, we go through the post-editing guidelines (full, light or medium, depending on project requirements), the agreed quality assurance process and applicable error types, and we offer post-editors any extra guidance needed to reach the quality level, i.e. anything that would not be clear by simply reading the post-editing guidelines, or any exception: for example, the service required is medium post-editing but for the German target audience, the client insists on n-dashes being replaced by m-dashes. We also explain what MT engine we are using, how it was customized, its known strengths and weaknesses, and we discuss any areas the neural MT struggles with in general, and where the machine translated output might fall short of the client's particular requirements. This way, post-editors are aware of what is expected of them and know exactly what to look for in the raw output, we reduce the risk of misunderstandings and we also set expectations on the final quality of the output. These calls are also a good chance to clarify any doubt post-editors might have or answer their questions.

These calls are often followed by a quick questionnaire to make sure post-editors are clear on the topics presented during the call, as well as brief instructions summarizing the key take-aways. Once a program has started, we continue monitoring performance, typically via Levenshtein Edit Distance analysis, and check for unexpected behaviour. As mentioned above, if we notice anything unexpected, i.e. under-editing or over-editing, we get in touch with post-editors to explain what we observed and give them further support or correct any wrong behaviour.

6.2 Supporting Clients

It can be very difficult for clients to understand the distinction between the different definitions of post-editing service levels. The differences between light and full post-editing are easily enough understood where content types very clearly require different approaches, e.g. user-generated content versus patents or branded website content. However, it is harder to explain the different requirements for technical content and stylistically demanding content, especially if the person overseeing the MT effort at client end is not familiar with different quality assurance methodologies. It still remains crucial to clearly define the client's requirements, so that they will know what they are buying, and what contributes to the productivity gains and compensation models. For this purpose, it can be useful to provide samples of the text to be translated with different post-editing approaches, and applicable error categories. This way they will see how the target text changes and choose what service they prefer:

7 Conclusions

There is no gold standard for post-editing guidelines nor universally applicable definitions of different post-editing services. While still being useful for initially steering conversations, we saw that the generic guidelines overlap in key aspects. At this point in time, we find that in order to effectively communicate with different stakeholders in the localization industry, it is necessary to refer both to definitions of light, medium, full post-editing, but to also supplement these with very hands-on, practical definitions of what constitutes an error in a given scenario, and how quality assurance is provided. Instructions, error categories and penalty thresholds need to be defined on a case-by-case basis with customers and need to be communicated very clearly to post-editors. Metrics such as TER or Edit Distance can help analyse and monitor the actual post-editing effort, and can be used to fine-tune and revisit requirements, productivity expectations and fair compensation.

References

De Almeida G. and S. O'Brien 2010. Analysing Post-Editing Performance: Correlations with Years of Translation Experience. Accessed January 2020. Available at http://www.mt-archive.info/EAMT-2010-Almeida.pdf.

Hu, Ke and Patrick Cadwell. 2016. A Comparative Study of Post-editing Guidelines. *Baltic J. Modern Computing*, Vol. 4 (2016), No. 2, 346-353.

International Organization for Standardization. 2017. Translation services — Post-editing of machine translation output — Requirements (ISO Standard

Source	Raw MT	Light PE	Medium PE	Full PE
Dopo la cottura, la "verace pizza napoletana" (vera pizza napoletana) presenta un diametro variabile che non deve superare 35 cm, con il bordo rialzato (cornicione) e con la partecentrale coperta dai condimenti.	After baking the "real Neapolitan pizza" has a variable diameter that must exceed 35cm, the edge (cornion) and the participant covered by the seasonings.	After baking the "real Neapolitan pizza" (original Neapolitan pizza) has a variable diameter that must not exceed 35cm, the raised edge (cornicione) and the central part covered by the seasonings.	After baking, the "verace pizza napoletana" (original Neapolitan pizza) has a variable diameter that must not exceed 35cm, the raised edge (cornicione) and the central part covered by the seasonings.	After baking, the "verace pizza napoletana" (original Neapolitan pizza) has a variable diameter that must not exceed 35 cm, a raised edge (cornicione) and the central part covered with toppings.

Table 4: Different levels of PE.

It is important to guide the client and provide recommendations in order for them to get the appropriate post-editing level for the content type and translation purpose they are looking to address, and to help them achieve the cost and time savings they were hoping to see.

No. 18587).

Levenshtein, Vladimir Iosifovich. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*.

LISA. 2006. LISA QA Model 3.1: Assisting the localization development, production and quality control processes for global product distribution (press release). Romainmôtier: LISA.

Lommel, Arle and Uszkoreit, Hans and Burchardt, Aljoscha. 2014. Multidimensional quality metrics

(MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica: tecnologies de la traducció*, (12):455–463.

O'Brien, S. 2010. Introduction to Post-Editing: Who, What, How and Where to Next? Accessed January 2020. Available at http://amta2010.amtaweb.org/AMTA/papers/6-01-ObrienPostEdit.pdf

O'Brien, S., Choudhury, R., Van der Meer, J., Aranberri Monasterio, N. 2011. TAUS Dynamic Quality Evaluation Framework: TAUS Labs report, Accessed January 2020. Available at https://www.taus.net/thinktank/reports/evaluate-reports/translation-quality-evaluation-is-catching-up-with-thetimes

O'Curran, Elaine. 2014. Machine Translation and Post-Editing for User Generated Content: An LSP Perspective. Proceedings of AMTA 2014, vol. 2: MT Users Vancouver, BC.

Rowda, Juan. 2016. Better, Faster, and More Efficient Post-editing. Accessed January 2020. Available at https://www.gala-global.org/publications/better-faster-and-more-efficient-post-editing

TAUS. 2010. MT Post-editing Guidelines. Accessed January 2020. Available at https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines

Appendix A. Comparative Analysis of Light PE Guidelines based on DQF-MQM framework

Error Type		O'BRIEN 2010	ROWDA 2016	TAUS 2016	ISO 2017
Accuracy	Addition	The message transferred should be accurate	Accuracy is key	Ensure that no information has been accidentally added or omitted	Ensure that no information has been added or omitted
	Omission	The message transferred should be accurate	Accuracy is key	Ensure that no information has been accidentally added or omitted	Ensure that no information has been added or omitted
	Mistranslation	The message transferred should be accurate	Accuracy is key	Aim for semantically correct translation	Restructure sentences in the case of incorrect or unclear meaning
	Over-translation	The message transferred should be accurate	Accuracy is key	Ensure that no information has been accidentally added or omitted	Ensure that no information has been added or omitted
	Under-translation	The message transferred should be accurate	Accuracy is key	Ensure that no information has been accidentally added or omitted	Ensure that no information has been added or omitted
	Untranslated text	The message transferred should be accurate	Accuracy is key		
	Improper exact TM match				
Fluency	Punctuation		Variations in style, punctuation, and spelling are OK		
	Spelling	All basic rules regarding spelling still apply	Variations in style, punctuation, and spelling are OK	Basic rules apply	
	Grammar	Not a big concern, unless grammatical problems interfere with accuracy		May not be perfect	
	Grammatical register				
	Inconsistency				
	Link/cross-reference				
	Character encoding				
Terminology	Inconsistent with termbase	Do not spend time researching terms			
	Inconsistent use of terminology	Do not spend time researching terms			
Style	Awkward	Ignore stylistic problems	Variations in style, punctuation, and spelling are OK	No need to implement corrections that are of a stylistic nature only	Need not be stylistically adequate
	Company style	Ignore stylistic problems	Variations in style, punctuation, and spelling are OK	No need to implement corrections that are of a stylistic nature only	Need not be stylistically adequate
	Inconsistent style	Ignore stylistic problems	Variations in style, punctuation, and spelling are OK	No need to implement corrections that are of a stylistic nature only	Need not be stylistically adequate
	Third-party style	Ignore stylistic problems	Variations in style, punctuation, and spelling are OK	No need to implement corrections that are of a stylistic nature only	Need not be stylistically adequate
	Unidiomatic	Ignore stylistic problems	Variations in style, punctuation, and spelling are OK	No need to implement corrections that are of a stylistic nature only	Need not be stylistically adequate
Design	Length				
	Local formatting				
	Markup				
	Missing text				
	Truncation/text expansion				
Locale convention	Address format				
	Date format				
	Currency format				
	Measurement format				
	Shortcut key				
	Telephone format				
Verity	Culture-specific reference	Edit any offensive, inappropriate or culturally unacceptable information		Edit any offensive, inappropriate or culturally unacceptable content	Edit any inappropriate content
Other			Fix major/blatant errors only. Minor issues are acceptable. Avoid stylistic and preferential changes.		
		Throughput expectations: very high Quality expectations: low	Fast turn-around	Use as much of the raw MT output as possible	Use as much of the raw MT output as possible

Appendix B. Comparative Analysis of Full PE Guidelines based on DQF-MQM framework

Error Type		O'BRIEN 2010	ROWDA 2016	TAUS 2016	ISO 2017
Accuracy	Addition	The message transferred should be accurate		No information has been accidentally added or omitted	No information has been added or omitted
	Omission	The message transferred should be accurate		No information has been accidentally added or omitted	No information has been added or omitted
	Mistranslation	The message transferred should be accurate	All mistranslations fixed	Aim for grammatically, syntactically and semantically correct translation	Restructure sentences in the case of incorrect or unclear meaning
	Over-translation	The message transferred should be accurate	All mistranslations fixed	No information has been accidentally added or omitted	No information has been added or omitted
	Under-translation	The message transferred should be accurate	All mistranslations fixed	No information has been accidentally added or omitted	No information has been added or omitted
	Untranslated text	The message transferred should be accurate		Untranslated terms belong to the client's list of "Do Not Translate"	
	Improper exact TM match				
Fluency	Punctuation	All basic rules regarding spelling, punctuation and hyphenation still apply		Basic rules regarding spelling, punctuation and hyphenation apply	Apply spelling, punctuation and hyphenation rules
	Spelling	All basic rules regarding spelling, punctuation and hyphenation still apply	Detailed corrections, no grammar or spelling errors should be ignored	Basic rules regarding spelling, punctuation and hyphenation apply	Apply spelling, punctuation and hyphenation rules
	Grammar	Grammar should be accurate	Detailed corrections, no grammar or spelling errors should be ignored	Aim for grammatically, syntactically and semantically correct translation	Produce grammatically, syntactically and semantically correct target language content
	Grammatical register			Aim for grammatically, syntactically and semantically correct translation	Ensure that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed
	Inconsistency				
	Link/cross-reference				
	Character encoding				
Terminology	Inconsistent with termbase	Ensure that key terminology is correctly translated	Accurate terminology	Key terminology is correctly translated	Adhere to client and/or domain terminology
	Inconsistent use of terminology	Ensure that key terminology is correctly translated	Accurate terminology	Correcting inconsistencies in terminology, terminology disambiguation	Adhere to client and/or domain terminology
Style	Awkward	Ignore stylistic and textuality problems	Style should be consistent and appropriate	May not be as good as that achieved by a native-speaking human translator	Ensure that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed
	Company style	Ignore stylistic and textuality problems	Style should be consistent and appropriate	May not be as good as that achieved by a native-speaking human translator	Ensure that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed
	Inconsistent style	Ignore stylistic and textuality problems	Style should be consistent and appropriate	May not be as good as that achieved by a native-speaking human translator	Ensure that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed
	Third-party style	Ignore stylistic and textuality problems	Style should be consistent and appropriate	May not be as good as that achieved by a native-speaking human translator	Ensure that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed
	Unidiomatic	Ignore stylistic and textuality problems	Style should be consistent and appropriate	May not be as good as that achieved by a native-speaking human translator	Ensure that the style appropriate for the text type is used and that stylistic guidelines provided by the client are observed
Design	Length			Ensure that formatting is correct	Apply formatting rules
	Local formatting			Ensure that formatting is correct	Apply formatting rules
	Markup	For tagged formats, ensure all tags are present and in the correct positions		Ensure that formatting is correct	Apply formatting rules
	Missing text			Ensure that formatting is correct	Apply formatting rules
	Truncation/text expansion			Ensure that formatting is correct	Apply formatting rules
Locale convention	Address format			Handling of measurements and locale-specific punctuation, date formats and alike	
	Date format			Handling of measurements and locale-specific punctuation, date formats and alike	
	Currency format			Handling of measurements and locale-specific punctuation, date formats and alike	
	Measurement format			Handling of measurements and locale-specific punctuation, date formats and alike	
	Shortcut key			Handling of measurements and locale-specific punctuation, date formats and alike	
	Telephone format			Handling of measurements and locale-specific punctuation, date formats and alike	
Verity	Culture-specific reference	Edit any offensive, inappropriate or culturally unacceptable information		Edit any offensive, inappropriate or culturally unacceptable content	Edit any inappropriate content
Other		Retain as much raw translation as possible		Use as much of the raw MT output as possible	Use as much of the MT output as possible
		Throughput expectations: high Quality expectations: medium	Close to human translation quality		Produce an output which is indistinguishable from human translation output