



MT POST-EDITING GUIDELINES

Isabella Massardo, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri, Katrin Drescher

Published by TAUS Signature Editions, Keizersgracht 74, 1015 CT Amsterdam,
The Netherlands

Tel: +31 (0) 20 773 41 72

E-mail: memberservices@taus.net

www.taus.net

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the author, except for the inclusion of brief quotations in a review.

Copyright © 2016 by TAUS Signature Editions and contributors

The TAUS Guidelines included in this ebook are intended to be applied in combination with each other.

AUTHORS

Isabella Massardo, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri, Katrin Drescher

TABLE OF CONTENTS

Evaluating Post-editor Performance

Machine Translation Post-editing

Post-editing Productivity

Pricing Machine Translation Post-editing Guidelines

About the MT Guidelines

EVALUATING POST-EDITOR PERFORMANCE

An agreed set of best practices to help the industry fairly and efficiently select the most suitable talent for post-editing work.

WHY DO WE NEED GUIDELINES?

Machine translation (MT) with post-editing (PE) is fast becoming a standard practice in our industry. This means that organizations need to be able to easily identify, qualify, train and evaluate post-editors' performances.

Today, there are many methodologies in use, resulting in a lack of cohesive standards as organizations take various approaches for evaluating performance. Some use final output quality evaluation or post-editor productivity as a standalone metric. Others analyze quality data such as “over-edit” or “under-edit” of the post-editor’s effort or evaluate the percentage of MT suggestions used versus MT suggestions that are discarded in the final output.

An agreed set of best practices will help the industry fairly and efficiently select the most suitable talent for post-editing work and identify the training opportunities that will help translators and new players, such as crowdsourcing resources, become highly skilled and qualified post-editors.

SCOPE

These guidelines focus on the analysis and interpretation of productivity testing results combined with the evaluation of the final post-edited output and scores obtained via industry-standard automated scoring metrics in order to evaluate post-editors' performances. It does not cover design and execution of post-editing productivity tests, post-editing quality levels and pricing recommendations.

DEFINING GOALS

The suggested goals for the TRANSLATION SERVICE PROVIDER are:

Identify the best performers from the pool of post-editors, who deliver the desired level of quality output with the highest productivity gains; identify the “ideal” post-editor profile for the specific content type and quality requirements (linguist, domain specialist, “casual” translator).

Identify common over-edit and under-edit mistakes in order to refine post-editing guidelines and determine the workforce training needs to achieve higher productivity.



The suggested goals for the TRANSLATION SERVICES BUYER are:

Gather intelligence on the performance of the in-house technology used to enable the post-editing process, such as a translation management system, a recommended post-editing environment and MT engines.

Based on the post-editor productivity, **set the realistic TAT** (turnaround time) expectations, **determine the appropriate pricing structure** for specific content types and language pair and **reflect the above in an SLA** (Service Level Agreement).

STRUCTURING YOUR ANALYSIS

In order to select the top productivity performers and evaluate the quality of the output using reviewers we advise you to:

Select a subset of the content used for the productivity test for which the highest and the lowest productivity is seen (the “outliers”), and evaluate the quality of the output using reviewers and automated quality evaluation tools (spellcheckers, Checkmate, X-bench, style consistency evaluation tools). Make sure the final output meets your quality expectations for the selected content types.

Use multiple translators and multiple reviewers.

Make sure there is minimal subjective evaluation. Provide clear evaluation guidelines to the reviewers and make certain the reviewers’ expectations and the Post-Editors’ instructions are aligned. Refer to the known most common post-editing mistakes in your guidelines.

Examples of **full post-editing** known problem areas:

- Handling of measurements and locale-specific punctuation, date formats and alike
- Correcting inconsistencies in terminology, terminology disambiguation
- Handling of list elements, tables or headers versus body text
- Handling of proper names, product names and other DoNotTranslate elements
- Repetitions (consistent exact matches)
- Removing duplicates, fixing omissions (for SMT output post-editing)
- Morphology (agreement), negations, word order, plural vs. singular



Examples of **light post-editing** known problem areas:

- Correctly conveying the meaning of the source sentence
- Correcting inconsistencies in terminology
- Removing duplicates, fixing omissions (for SMT output post-editing)
- Morphology (agreement), negations, word order, plural vs. singular

ANALYZING YOUR RESULTS

In order to identify the common over-edit and under-edit patterns:

Obtain edit distance and MT quality data for the content post-edited during the productivity evaluation, using the industry-standard methods, e.g. General Text Matcher, Levenshtein, BLEU, TER.

If your productivity evaluation tool captures this data, obtain information on the changes made by post-editors, i.e. edit location, nature of edits; if your tool doesn't capture such data, a simple file difference tool can be used.

You will be able to determine realistic turnaround time and pricing expectations based on the productivity and quality data. Always have a clear human translation benchmark for reference (e.g. 2000 words per day for European languages) unless your tool allows you to capture the actual in-production data.

For smaller LSPs and freelance translators, tight turnaround projects or other limited bandwidth scenarios reliance on legacy industry information is recommended.



Make sure that the final output quality matches the desired quality level for the selected content type, differentiate between the full post-editing quality level and light post-editing quality level (refer to the TAUS Machine Translation Post-editing Guidelines for further clarification).

Assess individual post-editors' performances using side-by-side data for more than one post-editor; use the individual post-editor data to identify the most suitable post-editor profile for the specific content types and quality levels.

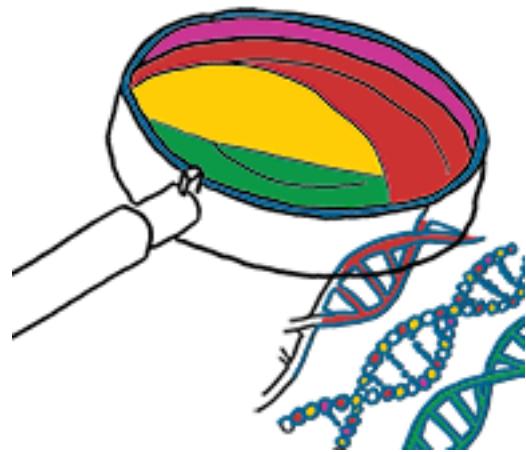
Calculate the mode (the number which appears most often in a set of numbers) based on scores of multiple post-editors to obtain data specific to certain content/quality/language combination; gather data for specific sentence length ranges and sentence types.

Do not use obtained productivity data in isolation in order to calculate expected daily throughputs and turnaround times, as the reported values reflect the “ideal” off-production scenario; add time necessary for terminology and concept research, administrative tasks, breaks and alike.

Identify best and worst performing sentence types (length, presence/absence of DoNot-Translate elements, tags, terminology, certain syntactic structures) and gather best practices for post-editors on most optimal handling of such sentences and elements.

Analyze edit distance data alongside with the MT quality evaluation data and assessment by reviewers to determine whether the edits made by post-editors were necessary for meeting the output quality requirements, gather best practices for specific edit types. Do not use edit distance data in isolation to evaluate the post-editor performance.

Analyze the nature of edits using the data obtained during the productivity tests, use top performer's data to create recommendations for lower performers; use the obtained information to provide feedback on the MT engine.



MATRIX

It is recommended that you create a clear matrix of post-editing productivity, quality, turnaround time and pricing discount expectations based on the results of your analysis. The proposed matrix structure is below:

- Content type and purpose of content
- Expected post-edited output quality level; risk assessment and error tolerance by error type for the specific content
- CAT-enabled environment or isolated post-editing environment; if the former – whether it will have an impact on the expected productivity gains
- Human translation daily throughput (use industry-standard daily throughout expectations as a benchmark)
- Post-editing daily throughput (based on productivity gain percentages)
- Productivity gain (per hour, per day)
- Expected pricing discount based on the factors above

MACHINE TRANSLATION POST-EDITING

Produced in partnership with
ADAPT Center, these guidelines
are aimed at helping customers and
service providers set clear expecta-
tions

OBJECTIVES AND SCOPE

These guidelines are aimed at helping customers and service providers set clear expectations and can be used as a basis on which to instruct post-editors.

Each company's post-editing guidelines are likely to vary depending on a range of parameters. It is not practical to present a set of guidelines that will cover all scenarios. We expect that organisations will use these basic guidelines and will tailor them as required for their own purposes.

Generally, these guidelines assume bilingual post-editing (not monolingual) ideally carried out by a paid translator, but that might in some scenarios be carried out by bilingual domain experts or volunteers. The guidelines are not system or language-specific.



RECOMMENDATIONS

To reduce the level of post-editing required (regardless of language pair, direction, system type or domain), we recommend the following:

- Tune your system appropriately, i.e. ensure high level dictionary and linguistic coding for rule-based machine translation systems, or training with **clean, high-quality, domain-specific data** for data-driven or hybrid systems.
- Ensure the **source text** is written well (i.e. correct spelling, punctuation, unambiguous) and, if possible, tuned for translation by MT (i.e. by using specific authoring rules that suit the MT system in question).
- Integrate **terminology management** across source text authoring, MT and translation memory (TM) systems.
- **Train** post-editors in advance.
- Examine the **raw MT output quality** before negotiating throughput and price and set reasonable expectations.
- Agree a definition for the **final quality** of the information to be post-edited, based on user type and levels of acceptance.
- Pay post-editors to give **structured feedback** on common MT errors (and, if necessary, provide guidance about how this should be done) so the system can be improved over time.

BASIC PE GUIDLINES

Assuming the recommendations above are implemented, we suggest some basic guidelines for post-editing. The effort involved in post-editing will be determined by two main criteria:

1. The quality of the MT raw output.
2. The expected end quality of the content

To reach quality similar to “high-quality human translation and revision” (a.k.a. “publishable quality”), full post-editing is usually recommended. For quality of a lower standard, often referred to as “good enough” or “fit for purpose”, light post-editing is usually recommended. However, light post-editing of really poor MT output may not bring the output up to publishable quality standards.

On the other hand, if the raw MT output is of good quality, then perhaps all that is needed is a light, not a full, post-edit to achieve publishable quality. So, instead of differentiating between guidelines for light and full-post-editing, we will differentiate here between two levels of expected quality. Other levels could be defined, but we will stick to two here to keep things simple. The set of guidelines proposed below are conceptualised as a group of guidelines where individual guidelines can be selected, depending on the needs of the customer and the raw MT quality.

“GOOD ENOUGH” QUALITY

“Good enough” is defined as comprehensible (i.e. you can understand the main content of the message), accurate (i.e. it communicates the same meaning as the source text), without being stylistically compelling. The text may sound like it was generated by a computer, syntax might be somewhat unusual, grammar may not be perfect, but the message is accurate.

Guidelines for achieving “good enough” quality:

- Aim for semantically correct translation.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling apply.
- No need to implement corrections that are of a stylistic nature only.
- No need to restructure sentences solely to improve the natural flow of the text.

HUMAN TRANSLATION QUALITY

This level of quality is generally defined as being comprehensible (i.e. an end user perfectly understands the content of the message), accurate (i.e. it communicates the same meaning as the source text), stylistically fine, though the style may not be as good as that achieved by a native-speaking human translator. Syntax is normal, grammar and punctuation are correct.

Guidelines for achieving quality similar or equal to human translation:

- Aim for grammatically, syntactically and semantically correct translation.
- Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling, punctuation and hyphenation apply.
- Ensure that formatting is correct.

POST-EDITING PRODUCTIVITY

These guidelines consist of two sets. The first is small-scale, short-term post-editing productivity tests. The second is large-scale, longer-term post-editing productivity tests.

SMALL-SCALE SHORT-TERM POST-EDITING PRODUCTIVITY TESTS

WHEN TO USE SMALL-SCALE PRODUCTIVITY TESTS

Post-editing productivity measurement applies to scenarios where you might wish to use MT as a translator productivity tool. Generally, small-scale productivity tests should be used if you are thinking about getting started with MT in a particular language pair, and wondering whether you should invest further effort. The productivity tests will help you understand your potential ROI per language pair/content type.

You may also undertake such tests periodically, say annually, to get an indication of improvements (or not) in productivity.

Small-scale, short-term tests, by their nature, gather less data and are less rigorous than the larger-scale, long-term ones for which we provide a separate set of best practice guidelines.

DESIGN

Compare apples with apples not apples with oranges

Productivity measures are flawed when they don't compare like with like. When comparing post-editing against translation, it should be clear whether or not 'translation' means 'TEP' – Translation, Edit, Proof — and whether or not post-editing includes final proofing. If translators have access to terminology, so too should post-editors.

Employ at least three post-editors per target language

Productivity will vary by individual. By including a number of post-editors with varying translation and post-editing experience, you will get a more accurate average measurement.

Exercise control over your participant profile

Engage the people who will actually do the post-editing in live projects.

Include appropriate content type

MT engines will have varying degrees of success with different content types. Test each content type you wish to machine translate. For SMT engines the training data should not form part of the test set. Out-of-domain test data should not be tested on domain-specific engines.

Include a sufficient number of words

For each content type and target language, we recommend including at least 250 segments of MT output for short-term tests. The more words included, the more reliable the results will be.

Provide clear guidelines

If you want post-editors to adhere to your standard style-guide, then this should be clearly stated. In addition, guidelines for post-editing should be provided so that all participants understand what is/is not required. See the TAUS Machine Translation Post-Editing Guidelines. It has been noted that people sometimes do not adhere to guidelines, so the clearer and more succinct they are, the better.

MEASURES

Measure actual post-editing effort and avoid measures of ‘perceived’ effort

When people are asked to rate which segment might be easier to post-edit, they are only rating perceived, not actual effort. It has been frequently shown that agreement between participants in such exercises is only low to moderate.

Measure the delta between translation productivity and post-editing productivity

The delta between the two activities is the important measurement. Individuals will benefit to a greater or lesser extent from MT, so the delta should be calculated per individual and the average delta should then be calculated.

Extrapolate daily productivity with caution

For short-term measures, numbers of post-editors and words are often limited and so potential daily productivity should be extrapolated with some caution as you may have outliers who will skew results in a small group.

The TAUS productivity testing reports show on average and individual post-editor productivity, enabling users to determine the influence of outliers on results.

Measure final quality

Increased efficiencies are meaningless if the desired level of quality is not reached. Below are two examples of techniques for quality measurement.

- Human evaluation: For example, using a company's standard error typology or adequacy/fluency evaluation.
- Edit distance measures: Some research has shown the TER (Translation Edit Rate) and GTM (General Text Matcher) measures to correlate fairly well with human assessments of quality.

The TAUS quality evaluation tools enable you to undertake adequacy/fluency evaluation and error typology review. Again, reports show on average and individual post-editor productivity, enabling users to determine the influence of outliers on results.



LARGE-SCALE, LONGER-TERM POST-EDITING PRODUCTIVITY TESTS

The larger-scale, longer-term tests would be used if you are already reasonably committed to MT, or at least to testing it on a large-scale, and looking to create a virtuous cycle towards operational excellence, guided by such tests.

DESIGN

Compare apples with apples not apples with oranges

Productivity measures are flawed when they don't compare like with like. When comparing post-editing against translation, it should be clear whether or not 'translation' means 'TEP' – Translation, Edit, Proof - and whether or not post-editing includes final proofing. If translators have access to terminology, so too should post-editors.

Employ a sufficient number of post-editors per target language

Productivity will vary by individual. By including a broad group of post-editors with varying translation and post-editing experience, you will get a more accurate average measurement. For long-term measurements, we recommend employing more than three post-editors, preferably at least five or six.

Exercise control over your participant profile

Engage the people who will actually do the post-editing in live projects. Employing students or the 'crowd' is not valid if they are not the actual post-editors you would employ in a live project.

Conduct the productivity measurement as you would a commercial project

You want the participants to perform the task in the way they would any commercial project so that the measures you take are reliable.

Include appropriate content type

MT engines will have varying degrees of success with different content types. Test each content type you wish to machine translate. For SMT engines the training data should not form part of the test set. Out-of-domain test data should not be tested on domain-specific engines.

Include a sufficient number of words

For each content type and target language, we recommend collating post-editing throughput over a number of weeks. The more words included, the more reliable the results will be.

Use realistic tools and environments

Commonly, MT is integrated into TM tools. It is recommended that if the post-editor is to eventually work in this standard environment, then productivity tests should be done in this environment, because more realistic measures can be obtained.

Provide clear guidelines

If you want post-editors to adhere to your standard style-guide, then this should be clearly stated. In addition, guidelines for post-editing should be provided so that all participants understand what is/is not required. See the TAUS Machine Translation Post-Editing Guidelines. It has been noted that people sometimes do not adhere to guidelines, so the clearer and more succinct they are, the better.

Involve representatives from your post-editing community in the design and analysis

As in the case of TAUS Machine Translation Post-Editing Guidelines, we recommend that representatives of the post-editing community be involved in the productivity measurement. Having a stake in such a process generally leads to a higher level of consensus.

MEASURES

Gauge the quality level of the raw MT output first

Productivity is directly related to the level of quality of the raw MT output. To understand PE productivity measurements, you need to understand the baseline quality of the MT output. Random sampling of the output is recommended.

Measure actual post-editing effort and avoid measures of perceived effort

When people are asked to rate which segment might be easier to post-edit, they are only rating perceived, not actual effort. It has been frequently shown that agreement between participants in such exercises is only low to moderate.

Measure more than words per hour

Words per hour or words per day give a simplistic view of productivity. The important question is: can production time be reduced across the life cycle of a project (without compromising quality)? Therefore, it may be more appropriate to measure the total gain in days for delivery or publication of the translated content. Spreading measurement out over time will also show whether post-editing productivity rates rise, plateau or fall over time.

Measure the delta between translation productivity and post-editing productivity

The delta between the two activities is the important measurement. Individuals will benefit to a greater or lesser extent from MT, so the delta should be calculated per individual and the average delta should then be calculated.

Measure final quality

Increased efficiencies are meaningless if the desired level of quality is not reached. Below are two examples of techniques for quality measurement. For longitudinal measures, the average final quality can be compared to see what improvements or degradations occurred:

- Human evaluation: For example, using a company's standard error typology. Note that human raters of quality often display low rates of agreement. The more raters, the better (at least three).
- Edit distance measures: Some research has shown the TER (Translation Edit Rate) and GTM (General Text Matcher) measures to correlate fairly well with human assessments of quality.

Measure opinions, get feedback

Some measurement of post-editor opinion/profile can be useful in helping to interpret the quantitative measures. Gathering feedback on the most common or problematic errors can help improve the MT system over time.

QUESTIONS ABOUT MEASUREMENT

Self-report or automate?

If a realistic environment is used, it is difficult to automate productivity measurements. Self-reporting is often used instead, where post-editors fill in a table reporting the number of words they post-edited, divided by the number of hours they worked. Self-reporting is error-prone, but with a large enough group of participants under- or over-reporting should be mitigated.

What about using confidence scores as indicators of productivity?

Confidence scores that are automatically generated by the MT system are potential indicators of both quality and productivity. However, development is still in the early stages and there is not yet enough research on the potential links between confidence scores and actual post-editing productivity.

What about re-training engines over time?

If SMT engines are re-trained with quality-approved post-edited content over time, then it can be expected that the MT engine will produce higher quality raw output as time progresses and that post-editing productivity may increase. This should be tested over the long-term.

PRICING MACHINE TRANSLATION POST-EDITING

These guidelines aim to help you understand how to arrive at a suitable pricing model for machine translation post-editing (MTPE).

AIMS

These guidelines aim to help you understand how to arrive at a suitable pricing model for machine translation post-editing (MTPE). Unfortunately, there is currently no single method to determine MTPE pricing. Instead a combination of approaches is needed to calculate post-editing effort and set pricing accordingly. Pricing may be set on a per-hour or per-unit basis.

CAVEAT EMPTOR

This Guideline focuses on the analysis and interpretation of productivity testing results combined with the evaluation of the final post-edited output and scores obtained via industry-standard automated scoring metrics in order to evaluate post-editors' performances. It does not cover design and execution of post-editing productivity tests, post-editing quality levels and pricing recommendations.

GUIDING PRINCIPLES

Whatever combination of approaches you decide to use, your model should be:

Predictive. A model for pricing MTPE (Machine Translation Post-Editing) helps to predict the cost.

- Your model should help establish pricing up-front. Therefore, a model should either allow for extrapolation or be able to calculate the cost of a particular volume of text instantly. Remember, pricing may change each time you evaluate and deploy a new version of an engine.

Fair. A model for pricing MTPE provides buyers, language service providers and translators with a reliable measurement of the required work.

- All parties involved in the translation process, for example, translators, language service providers and buyers should be involved in establishing your approach.
- All parties should agree that the pricing model reflects the effort involved.
- Translators should be provided with training on post-editing and realistic expectations must be set. See [TAUS Machine Translation Post-editing Guidelines](#) for more detailed information on quality expectations.
- It can be difficult to demonstrate you are always being fair, because circumstances will serve to undermine the assumptions in certain cases. We ask that you share those experiences with us (academy@taus.net) so that we can create a [public knowledge base](#) over time.

Appropriate. A model for pricing MTPE considers content characteristics.

- Content type. MT output quality might greatly vary depending on content type. Similarly, different content types might require differing post-editing levels.
- The language pair involved in the translation process will affect the quality of MT output.
- Hence pricing may differ by language pair and content type.
- When you undertake evaluations to help establish pricing make sure
 - You test the model on representative test-data, i.e. the quality of the test-data has the same characteristics of that used in the real setting;
 - You use a representative volume of test-data to allow for a comprehensive study; and
 - The content-type in the test-data matches that of the real setting.

MT quality should be established in one comprehensive evaluation early in the MT adoption/implementation process for EACH engine and content type, and not continuously, unless you are able to establish a cost effective operating model to capture the right data points on an ongoing basis to set pricing.

In a post-editing scenario, spot checks to monitor quality is advised, and feedback from post-editors should be collected — keeping the dialogue open, acknowledging and acting on feedback where possible.

Your method for assessing quality and establishing pricing should be transparent.

APPROACHES

This section introduces the approaches that should be used. The links below in this document will guide you to more detailed information.

You will need to combine a number of approaches to achieve a predictive, fair and appropriate model. This may involve combining automated and human evaluation, and undertaking a productivity assessment. A productivity assessment should always be used

A combination of these three approaches is recommended:

- Automated quality score (GTM, TER, BLEU, MT Reversed Analysis)
- Human quality review
- Productivity assessment (post-editing speed)

AUTOMATED METRICS

We are only outlining two automated metrics. There are many others. GTM and MT Reversed Analysis and can be used in combination with productivity assessments and human review to help set pricing. Both require human reference translations.

GTM

GTM (General Text Matching) measures the similarities between the MT output and the human reference translations by calculating editing distance.

MT Reversed Analysis

This approach aims to correlate MT output quality with fuzzy-match bands. It calculates the fuzzy-match level of raw MT segments with respect to their post-edited segments. The approach relies on a well-established pricing model for TM-aided translation. The process runs as follows:

- Post-edit the raw MT output. Apply a fuzzy-match model to the raw MT and post-edited pairs as it is done in TMs. Assuming that a particular engine will behave the same way in similar scenarios (content type, language pair, external resources), establish expected fuzzy-match band proportions and rates for each band.
- To calculate cost savings, you can compare: (1) the hypothetical price for the source and the final translation (post-edited version of the source) obtained through a fuzzy-match pricing model, and (2) the cost of post-editing the raw MT output through a productivity assessment to test the results and refine assumptions.

HUMAN QUALITY REVIEW

Human review can be used to assess the quality of MT output, assess the validity of mapping MT output quality to translation memory match rates and also assess whether the final quality of post-edited content is up to the desired level.

POST-EDITING PRODUCTIVITY ASSESSMENT

This approach calculates the difference in speed between translating from scratch and post-editing MT output. The results may vary each time you create a new engine. Therefore in order to be predictive, fair and appropriate, you would need to rerun productivity evaluations each time you create a new production ready engine. Depending on how you combine methods to establish pricing, you may undertake small-scale short-term productivity tests or larger-scale longer-term assessments. A link is provided to Best Practice Guidelines for Productivity Evaluations later in this document.



HOW TO COMBINE APPROACHES: EXAMPLES

Determine a threshold for automated scores above which a minimum acceptable level of/improvement in quality for post-editing has been achieved. Or undertake human review to determine that a minimum level of improvement in quality has been achieved. Undertake productivity assessment to determine the (added) speed from post-editing and determine a pricing model. You will need to be familiar with the nuances of automated metrics. You will need to undertake productivity assessments over a period of weeks to establish a predictive, fair and appropriate pricing model.

Determine a threshold for automated scores, above which a minimum acceptable level of improvement in quality for post-editing has been achieved. Or undertake human review to determine that a minimum acceptable level of improvement in quality has been achieved. Post-edit a sample of representative content. Undertake reversed analysis of the post-edited content to map to fuzzy match price band rates. Undertake a small-scale productivity assessment and human review to validate and refine the conclusions. The errors produced by MT are different from that found in fuzzy matches, hence productivity tests and human review are necessary. Combining these approaches each time you have a new engine should ensure your pricing model is predictive, fair and appropriate.

ABOUT THE MT GUIDELINES

A thank you to everyone who helped put the Machine Translation Post-Editing Guidelines together.

ACKNOWLEDGEMENTS

Thanks to everyone who helped to put these guidelines together. We were very fortunate to have the help of TAUS members, governmental institutions and translator organizations. Details about the project team and process for arriving at these guidelines [can be found here](#).

MACHINE TRANSLATION POST-EDITING

A special thanks to Sharon O'Brien, Dublin City University and CNGL, and Fred Hollwood, Symantec and TAUS Advisory Board for their dedication and support in putting these guidelines together.

PRICING MACHINE TRANSLATION POST-EDITING

A special thanks to Nora Aranberri (TAUS Labs) and Katrin Drescher (Symantec) for drafting these guidelines.

Also, the following organizations reviewed and refined the guidelines at the TAUS Quality Evaluation Summit 15th March, 2013, Dublin:

ABBYY Language Services, Amesto, Capita Translation and Interpreting, Concorde, Crestec, EMC, Google, Intel, Jensen Localization, Lingo24, McAfee, Microsoft, Moravia, Pactera, R.R. Donnelley, Sajan, STP Nordic, Vistatec, Welocalize and Yamagata Europe

ABOUT TAUS

TAUS is a resource center for the global language and translation industries. Founded in 2004, TAUS provides insights, tools, metrics, benchmarking, data and knowledge for the translation industry through its Academy, Data Cloud and Quality Dashboard.

Working with partners and representatives globally, TAUS supports all translation operators – translation buyers, language service providers, individual translators and government agencies – with a comprehensive suite of online services, software and knowledge that help them to grow and innovate their business.

Through sharing translation data and quality evaluation metrics, promoting innovation and encouraging positive change, TAUS has extended the reach and growth of the translation industry.

FEEDBACK

To give feedback on improving the guidelines, please write to
academy@taus.net