MT Post Editing challenges: Training happy and successful posteditors.

Abstract

Today, we can state that machine translation post-editing has replaced traditional human translation assignments for many linguists, and great deals of translators are actually working, successfully or not, as posteditors. No doubt, this trend has been constantly increasing and it will continue doing so exponentially. If we focus on the posteditor's environment, even those who are successful and proficient posteditors have not received formal training on MTPE as part as their academic background. Sound and organized training on postediting and guided practice based on production goals and results are key components of the translator's toolkit to attain quality and quantity expectations, and also happiness while doing their work. This training should mingle diverse skills, such as learning about technologies for MT deployment and linguistic resources for MT engine training, tools and guidelines for estimating quality and attainable outputs, industry standards for MT, such as ISO, and structured feedback to increase productivity and improve MT raw output.

In this presentation, I will focus on the relevant skills and tools that linguist/translators need to become productive posteditors and reliable auditors for MT raw output and content suitability, based on its domains and linguistic characteristics. This learning environment should set clear goals and steps from the technical, linguistic and strategy perspectives, as well as raise broader awareness of automatic evaluation and quality estimation, corpus management and engine training. Posteditors should be trained so as to support all the players in the MTPE service chain, from content creators, translation salesforce, project managers and actual posteditors and reviewers of MT postedited content.

This presentation aims at setting the basis for formal academic training at translation colleges and certification institutions.

MT Post Editing challenges: Training happy and successful posteditors.

Postediting demands

Postediting requests have exponentially increased in the recent years, and most large language companies have massively changed their assignments from translation into postediting, as end clients and developers have learned about the advantages that Artificial Intelligence has brought to them in terms of time to foreign markets and translated content versus zero translation.

Empowerment and knowledge

Knowledge is power… Feeling comfortable with the task, understanding the underlying technology and making posteditors (PE) aware of the authority they have on certain areas or links of the production chain make them more involved and willing to work harder and in a more effective way when needed to attain an optimal output. Even though not all postediting tasks are performed by professional translators, translation courses should include this subject as part of their curriculum or at least as an extension course or postgraduate seminar.

Main aspects to be taken into account include the following:

Posteditor's profile

Translation background - Is a translation degree required?

Experience as a proofer / translator: The fact that many translators only accept translation as their working assignments should be taken into account, as proofing experience and skills may be relevant for optimal performance.

Technical background – Do posteditors have a good command of the different tools and mechanical processes that are involved in the postediting assignment? (Namely, terminology management, TM integration, basic programming skills, for macros, regular expressions, etc.). Are they familiar with chatbots and Artificial intelligence (AI)?

Basic Project Managenent – Can posteditors create and manage their own projects, with all the components needed to automatize terminology, checks, updates, unique segments, splitting or slicing projects taking internal repetitions/matches into account, etc.?

Subject matter specialization – Equally relevant as for Human Translation (HT)

Training posteditors has become a challenge for institutions which in most part of the world have only recently started covering CAT tools. Well-organized curricula should include

Companies salesforces and Project managers unaware of the posteditor skills or unable to empower posteditors: frustration

Introduction to Machine Translation (MT) and postediting

MT-related terms and definitions

MT history (main hallmarks regarding MT types): from statistic-based to neural and adaptive

Analysis of several types of engines – how they are trained, fed, and how raw output is assessed

Output error identification – Clear taxonomy of the types of expected errors, including linguistic, formatting and technical potential issues, and their proposed fixing strategies. Text linguistics

Content type analysis: MT raw output varies by content and text genre

ISO as a framework or TAUS guidelines as a roadmap – Standards and industry-based procedures and recommendations as a scaffold for PE training

MT technologies and related tool

Pre-PE and Post-PE steps - Controlled language and authoring, monolingual proofreading, cultural validation, formatting

Sampling guidelines for MT raw output and strategic decision – What is going to be light-postedited; what is going to be full-postedited and what is going to be outsourced based on the traditional human translation/editing /proofing (TEP) model

Actual postediting

Job requirements for the assignment should be clearly described and outcome expectations should be thoroughly set forth in advance, including the steps and strategies that correlate with them. Mostly, the following should be covered:

Level of postediting: light vs. full; user requirements, PE buyer's information, purpose and nature of the product/content object of the language transfer, and any other assignment-related particulars

Terminology management

Concepts of over-edits/under-edits – If the PE is used to self-edit his/her work as a translator (HT) or proofing, a clear limit to where edits should end should be clearly defined

MT workflow, including structured feedback

Expected quality – Industry-based, client-based, product type-based, and quality assessment of raw output

Checklists – To reinforce Quality metrics

Self-productivity assessment: time an quality metrics; mental concentration.

Delta comparisons: PE should occasionally assess their performance comparing again their own HT productivity, when applicable.

Procedures and practice

The actual steps to be taken for the postediting assignment should be depicted, with a detailed description of the procedures and tools involves, as well as the "on-focus" skills for carrying out each of them.

2-Layer postediting (Bilingual postediting focused on accuracy + target only, focused on fluency and language (grammar/syntax, punctuation, spelling, etc.)

OR,

Full bilingual postediting, with or without a separate proofing step (like in HT, performed by another linguist)

Mechanical checks: black lists, advanced terminology management, regex, macros, consistency checks, etc.

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 58

Bilingual / Target only checks and fixing strategies

Hands-on Practice: from modular practical training (meaning, stepwise separation of tasks: raw output assessment, error taxonomy, text linguistics, etc.) to comprehensive postediting, covering the different content types, communicative purposes, participants and situations

Improve productivity – Is a better productivity attained in terms of speed, hourly rates, quality at large?

Conclusion

For optimal results, these curricula or programs should be designed as a multidisciplinary effort including linguists/language experts, subject matter experts (SME), MT trainers/developers, and terminologists. Training should be mostly practical, based on examples and the level of complexity of the content as well as the degree of requirements should be gradually increased, as for any other training environment. Academia and industry players should work together to set their goals, needs and expectations in order to reproduce a real-world scenario in the training workflow.

References

ISO 18587 Translation services — Post-editing of machine translation output — Requirements. First edition

2017-04

Guerberof Arenas, A, Moorkens, Joss. Machine translation and post-editing training as part of a master's programme. The Journal of Specialised Translation, Issue 31 – January 2019. Available at: https://www.jostrans.org/issue31/art_guerberof.pdf
TAUS (Translation Automation User Society). Evaluating Post-Editor Performance Guidelines. First published, February 2014. Available at: https://www.taus.net/academy/best-practices/postedit-best-practices/evaluating-post-editor-performance-guidelines

TAUS (Translation Automation User Society). MT Post-Editing Guidelines. First published, November 2010. Available at: https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines

TAUS (Translation Automation User Society). Post-Editing Productivity Guidelines. First published, December 2012. Available at: https://www.taus.net/academy/best-practices/postedit-best-practices/post-editing-productivity

Luciana Ramos's abridged bio

Luciana Ramos is a professional translator and interpreter with a long-standing presence in the Language and Technology industry, which, based on her sound linguistic background, broad experience in every link of the translation and localization supply chain (both as a linguist and a businesswoman), her specialist knowledge on language technologies and her research capabilities, offers a comprehensive array of professional services: scientific translation, software localization, telephone/on-site interpreting, professional training and coaching (for translators, vendor managers and project managers), as well as consulting on business organization and processes, Quality Assurance, and investments on technology.

Luciana Ramos holds a Master Degree on Biomedical Translation from the Universitat Jaume I, Castellon, Spain, and a Technical-scientific and Literary Translator Degree for English into Spanish, and Consecutive

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 59

and Simultaneous Interpreter Degree, both issued by Instituto de Enseñanza Superior "Olga Cossettini" Rosario, Argentina; 1992-1996. She has also been certified by the American Translators Association for English into Spanish Translations. As a member of Colegio de Traductores de la Pcia. de Santa Fe (2.º circunscrip.), Luciana Ramos does business under Professional License # 218/02. As a source of continuing education, market trends and business developments, Luciana is a proud member of Tremédica, GALA and ATA. Luciana has delivered, for international audiences, webinars, in-person workshops and courses, and conference presentations on several fields related to translation specialization and technology, with an important focus on Machine translation.

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, 1st Workshop on Post–Editing in Modern–Day Translation

Page 60