

# The Savvy Newcomer

Tips and resources for translators and interpreters

[Home](#) [About Us](#) [Resources](#) [Write for Us](#) [Contact](#)

SEARCH

GO



November 16, 2021

## Savvy Diversification Series – Diversification into Machine Translation

The Savvy Newcomer team has been taking stock of the past year and finding that one key priority for many freelance translators and interpreters has been diversification. Offering multiple services in different sectors or to different clients can help steady us when storms come. Diversification can help us hedge against hard times. With this in mind, we've invited a series of guest authors to write about the diversified service offerings that have helped their businesses to thrive, in the hopes of inspiring you to branch out into the new service offerings that may be right for you!

Taking the pulse of the U.S. localization industry demonstrates what should be an economically prosperous period for qualified translators and editors. It's true that it doesn't sound great for the industry to be operating in what the Joint National Committee for Languages calls a period of "language crisis" in the United States. The materials distributed to U.S. lawmakers during the February 2021 Virtual Language Advocacy Days give alarming statistics: "9 out of 10 U.S. employers rely on employees with world language skills; and 1 in 3 foreign language-dependent employers reports a language skills gap; and 1 in 4... lost business due to a lack of foreign language skills" (JNCL-NCLIS, Legislative Priorities). That is to say, at the same time that the U.S. market is feeling repercussions for its lacking investments in multilingual education over the years, qualified language professionals are in high demand, while the roles being demanded by the market are becoming ever more technological in nature. In article "Future Tense: Thriving Amid the Growing Tensions between Language Professionals and Intelligent Systems," Jay Marciano points out, "The day-to-day work of the translator of today will be hardly recognizable to a language services professional in 2030."

Newcomers to the industry are at a particular advantage within these circumstances. During Slator's briefing for their *Pro Guide: Translation Pricing and Procurement*, Anna Wyndham noted that experienced buyers of localization services are less likely to adopt new pricing models, while new buyers from the tech industry and beyond are more open to and indeed may expect "human-in-the-loop" pricing models based on full integration with machine translation. Likewise, savvy newcomers to the translation profession are more likely to adopt machine translation as a reality of the role, while more veteran translators may feel less incentivized to go through the disruptive change of integrating Machine Translation (MT) technology into their everyday workflows. Newcomers and veterans alike who are looking to diversify now and have their services remain relevant for decades to come do well to incorporate machine translation before the learning curve has become so great as to effectively disqualify one from key markets.

This article outlines key MT-related services to include in your portfolio as 21<sup>st</sup>-century translators reinventing themselves as language technologists. As language technologists, your expertise in translation makes you an asset at MT-engine training, writing content for MT, and post-editing of machine translation (PEMT) stages. This article considers these services in reverse order, starting with the PEMT services that translators are most likely to perform, before shifting further and further upstream, first to writing for MT and then to training MT engines. The discussion of each service type addresses common misconceptions and key competencies so you can start developing the skills needed to add MT services to your field of expertise. Check out the additional resources section for further reading to continue your exploration of this dynamic service area.

### Service #1 – Post Editing of Machine Translation (PEMT)

In Episode 49 of The ATA Podcast, "A Look into the Future of Post-Editing and Machine Translation," Jay Marciano defines post-editing of machine translation as a "step that a professional translator takes to review and make corrections to machine translation output in the provisioning... high quality translation[s]" (Baird and Marciano). By rights, Marciano believes that this terminology "post editor" adds specialized meaning to what is already a post editing role. To summarize, traditional translation denotes not only the invention of completely new copy, understood to be the translation of "new words," but also the act of editing translation memory (TM) output at the segment level, the level of work involved depending on the quality of the contributors to shared, proprietary resources, and the level of match of the source segment for translation to existing segments within the TM, generally starting from 75% percent matches to above. Incorporating segments that have been pre-translated using MT adds another segment type for human post-editing, though the term "post-editing" itself is used exclusively to denote work reviewing machine translation output.

The belief that it takes less skill to post-edit machine translation than it does to produce traditional human translation is a misconception that has circulated in the translation field since the advent of MT. This misconception is tied to several factors. Among those is the outdated perception that MT produces poor quality output that is too repetitive to be interesting for humans to review. Older rules-based or statistical models indicated performance better for content that corresponds to lower levels of the Interagency Language Roundtable (ILR) scale for translation performance. The ILR scale is comprised of 5 levels, with level 2 and below indicating limited or minimal performance, and level 3 and above indicating levels of professional performance. Traditionally, rules-based and statistical models have been best geared for texts that correspond to level 2 of the ILR scale, or straightforward texts like sets of instructions produced using controlled language that leaves little room for creative interpretation. ATA certification is a mid-career certification that demonstrates that a translator performs at (at least) a Level 3 of the ILR scale, and older MT models could not at all compete with professional humans for content characterized by the abstract language, implication, and nuance that requires a human mind to be parsed. However, machine translation technology has evolved at light speed, and even if MT cannot surpass the quality produced by human translators, the levels of fluency and correspondence it is possible to achieve using artificial intelligence and neural machine translation is remarkable. The linguistic challenges encountered in this work are interesting for those who enjoy studying the intersection of human and machine-produced languages too.

No matter the complexity of the content that a machine translation engine is designed to pre-translate, MT engines are far from replacing humans. According to the *ATA Position Paper on Machine Translation*, this is because "Computers can be very sophisticated in calculating the likelihood of a certain translation, but they understand neither the source nor the target text, and language has not yet been captured by a set of calculations." While the results of MT are getting better all the time, when confirmation of any degree of accuracy or polishing is needed, a professional post editor is the one to do that job. According to *ISO 17100:2015(E)*, the professional competences of translators are: translation, linguistic and textual competence in the source and target language, research, information acquisition and processing, and cultural, technical, and domain competences (3.1.3.). Professionalism is a competence added to the translator competences indicated in ISO 17100 for MT post editors according to *ISO 18587 – Translation services – Post-editing of machine translation output – Requirements*. That professionalism entails a knowledge of MT technology, common linguistic errors produced by MT, and Computer-Assisted Translation (CAT) tools, and the ability to post-edit linguistic analysis, provide structured feedback to improve MT output over time, and interact with terminology management systems ("5 Competences and qualifications of post-editors" ISO 18587).

To undertake the linguistic challenges that post editing of machine translation presents requires a thorough understanding of key post-editing concepts and how those concepts relate to post-editing specifications. To review, specifications outline the requirements of buyers and expectations of target users that define how localization services are produced. With regards to machine translation, the value proposition of the content being produced will determine whether light post-editing or full-post editing is needed, that is, whether what the *TAUS MT Post-Editing Guidelines* refers to as "good enough" or "human translation" quality is needed. If light-post editing is called for, such as in circumstances in which speed of delivery takes priority over fluency and style, the post editor will intervene minimally in the raw MT output to make corrections to inaccurately rendered meaning, grammar and spelling errors, and culturally offensive content. If full-post editing is called for, greater checks for consistency in terminology, product names, and mechanical aspects of the text are also employed.

Within either light or full post-editing models, discipline is key, and in post-editing, discipline is demonstrated by using the least number of keystrokes to make only the necessary corrections. Experienced post-editors can quickly distinguish among segments that are good enough, segments that require minor edits, and segments that need to be started from scratch.

Localization managers use post-editing distance – or the measure of the change between raw MT output and post-edited content – to gauge the overall quality of the MT engine and the post editor's work and to identify instances of over-editing and under-editing. According to Silvio Picinini of eBay, low edit distances can be an indicator of both quality and productivity, since if both the MT engine and the post editor have been well trained, that should result in lower edit distances. For those who are interested in working as post editors or in training post editors, Sharon O'Brien recommends the following curriculum in the 2002 paper "Teaching Post-editing: a proposal for course content"; "Introduction to Post-editing, Introduction to Machine Translation Technology, Introduction to Controlled Language Authoring, Advanced Terminology Management, Advanced Text Linguistics, [and] Basic Programming Skills" (103).

### Service #2 – Writing for Machine Translation

In a world in which more-and-more data is being authored on a daily basis that could ever possibly be translated by humans, the authors of a great percentage of that data may not be good writers at all, much less good writers of content intended for translation. Within workflows that incorporate MT, professional linguists have an opportunity to get involved before any content is imported into the engines that produce the raw output for PEMT. Just like workflows built around human translation benefit if the source content is written for translation, workflows that incorporate machine translation benefit from increased efficiency and quality if the source content is written expressly for that purpose. Localization workflows for human translation already incorporate copy-editing of source content to promote smooth processing during translation, especially where multiple target languages are involved. This copy-editing stage decreases the need for clarification mid-workflow and prevents the extensive rework that results from misunderstandings and poor comprehensibility by identifying and correcting ambiguities and inconsistencies in source content prior to sending that content for translation.

Once post editors have a good sense for the errors that are common to a language pair, subject field, and text type, they will be more equipped to customize recommendations for how to best write for machine translation, and for certain text types and subject fields, the professional recommendation may just be that MT will not suffice. Ambiguities and inconsistencies that should be flagged prior to both human and machine translation include unclear referents, the use of synonyms, long compound nouns, and the misinterpretation of homonyms, among many other textual features. Examples of some common sources of translation errors are provided below.

- Unclear referent: Group A and group B compared their results, and **they** (Group A, Group B, or Group A & B?) decided to make changes based on finding C.
- Potential synonym use: The **drying process** should take so many days. Once the **dehydration process** is complete, do this next. (Are drying and dehydration separate processes, or do both refer to the same process?)
- Misinterpretation of homonyms: Our earnings for this quarter are as follows. (Depending on the context, the best equivalent for "earnings" may be an equivalent that conveys one of these senses: pay, profits, returns, income, etc.)

When getting started with writing for MT, the principles from controlled language and plain language have good general rules that can be applied too. Uwe Muegge's *Controlled Language Optimized for Uniform Translation*, for instance, includes such guidelines as expressing only one idea per sentence, using simple yet complete grammatical structures, limiting the use of pronouns by restating nouns instead, and using articles so that nouns can be easily identified; and Plain Language Association International recommends that jargon be avoided and that simple words be employed ("What is plain language?"). The rules for controlled language and plain language may imply that these forms of communication are easy to use, but even identifying the myriad of textual features encompassed by these principles takes a great deal of study, practice, and experience. The Simplified Technical English, a controlled language of the AeroSpace and Defense Industries Association of Europe, for instance, consists of sixty-five writing rules in nine different categories and a dictionary of nearly 1000 approved words.

### Service #3 – Training Customized MT Engines

The invention of machine translation has largely remained in the realm of programmers and engineers. Despite the noticeable lack of linguists involved in MT development, so much high-quality data is needed to train customized MT engines that getting corpus linguists involved before undertaking what can be expensive, manual data collection processes makes perfect sense. A corpus is a collection of texts that have been selected for a specific purpose. A general language corpus will include many millions of words, while a corpus of specialized texts written by experts from a specific subject field may include only hundreds of thousands of words to start. Parallel corpora of translated and aligned segments are most frequently sought when training MT engines, whether rules-based, statistical, or neural models. However, high-quality parallel corpora take a long time to build and are exceedingly hard to find in any off-the-shelf format. Because high quality parallel corpora are so hard to find, those training MT engines may turn to comparable corpora, or collections of similar texts in multiple languages, for languages with less resources.

When building monolingual corpora, linguists will be able to identify the characteristics of the most representative data to collect for each corpus upon which the MT engine will be trained. Corpora might include one technical corpus of general content written by subject matter experts in a specific subject field per language and one client-specific corpus of proprietary product documentation per language. Since MT is trained using human produced language, it therefore replicates human biases. Linguists can help identify and mitigate the race and gender biases that manifest in large data sets by identifying specific populations, geographical regions, or language dialects not adequately represented in a corpus. They can help by eliminating any content from the corpus that is not fit for use too. Thus, MT users will not be made to feel insulted by offensive language produced by an MT engine and MT developers can avoid alienating MT users. Salvador Ordóñez gives several examples of high-profile manifestations of racial and gender bias in MT and how to overcome it in the article "Avoiding Bias and Discrimination in Machine Translation" published via Forbes.

### Pricing MT Services According to Skill

In summary, to diversify into the MT services that are already a nearly ubiquitous part of the provisioning of human translation services, translators should develop advanced skills in CAT tools, technology in general, and linguistic post editing, the ability to match services rendered with the quality expectations conveyed in specifications, and knowledge of controlled languages, corpus building and analysis, TM management at scale, terminology management, and data security. Regardless of the wide range of competencies necessary to work in MT, translators should be aware that traditional buyers accustomed to per-word pricing models tend to see the incorporation of MT as an opportunity to purchase translation services at further discounts to MT-pricing models. As Slator emphasizes in the *Pro Guide: Translation Pricing and Procurement* new buyers mean that new pricing models are possible. When working with new buyers, shift to value-based pricing models that more adequately compensate you for your rich expertise where you can. Above all, remember that in the design, implementation, and review of MT, teaching the parrot to talk is among the goals, but it is much more valuable if you can teach the parrot to say the correct thing.

### Works Consulted & Recommended Resources for Further Reading

Aslan, Sölen. "9 Types of Data Bias in Machine Learning." TAUS, 2021 Mar 22, <https://blog.taus.net/9-types-of-data-bias-in-machine-learning/>. Accessed 2021 Apr 12.

"ATA Position Paper on Machine Translation: A Clear Approach to a Complex Topic." American Translators Association, 2018 Aug 13, <https://www.atanet.org/client-assistance/machine-translation/>. Accessed 2021 Apr 1. <https://www.atanet.org/position-papers/ata-position-paper-on-machine-translation-a-clear-approach-to-a-complex-topic>

Baird, Matt and Jay Marciano. "E49: A Look into the Future of Post-Editing and Machine Translation." The ATA Podcast, Episode 49, 2020 Sept 24, <https://www.atanet.org/podcast/e49-a-look-into-the-future-of-post-editing-and-machine-translation/>.

Berger, Carola F. "An Introduction to Neural Machine Translation." American Translators Association, ATA 59<sup>th</sup> Annual Conference, October 2018, [https://ata-divisions.org/5\\_TD/wp-content/uploads/2018/11/ATA59\\_An\\_Introduction\\_to\\_Neural\\_Machine\\_Translation.pdf](https://ata-divisions.org/5_TD/wp-content/uploads/2018/11/ATA59_An_Introduction_to_Neural_Machine_Translation.pdf). Accessed 2021 Apr 10.

"ILR Skill Level Descriptions for Translation Performance." Interagency Language Roundtable, <https://www.govlink.org/Skills/AdoptedILRTranslationguidelines.htm> (Links to an external site.). Accessed 2021 Mar 30.

ISO 18587:2017, Translation Services – Post-editing of machine translation output – Requirements. International Organization for Standardization, Geneva, Switzerland, 2017, [http://www.iso.org/iso/iso\\_18587.htm](http://www.iso.org/iso/iso_18587.htm)

Legislative Priorities of the Language Enterprise-17<sup>th</sup> Congress. Joint National Committee for Languages and the National Council for Languages and International Studies (JNCL-NCLIS), 2021 Feb, handout.

Marciano, Jay. "Future Tense: Thriving Amid the Growing Tension between Language Professionals and Intelligent Systems." *The Chronicle*, American Translators Association, July/August 2020, 29-32, <https://www.nxtbooks.com/nxtbooks/chronicle/20200708/index.php>. Accessed 2021 Apr 12.

Massardo, Isabella, et al. MT Post-Editing Guidelines. TAUS, 2016, <https://info.taus.net/mt-post-editing-guidelines/>

O'Brien, Sharon. "Teaching Post-editing: A Proposal for Course Content." European Association for Machine Translation, 2002

Ordonó, Salvador. "Avoiding Bias and Discrimination in Machine Translation." Forbes, 2021 Mar 1, <https://www.forbes.com/sites/forbesbusinesscouncil/2021/03/01/avoiding-bias-and-discrimination-in-machine-translation/>. Accessed 2021 Apr 12.

Picinini, Silvio. "Going the Distance – Edit Distance 1." eBay blog, eBay Inc., 2019 Aug 8, <https://tech.ebayinc.com/research/going-the-distance-edit-distance-1/>. Accessed 2021 Mar 31.

Pro Guide Briefing: Pricing and Procurement. Slator, 2021 Apr 7, Webinar.

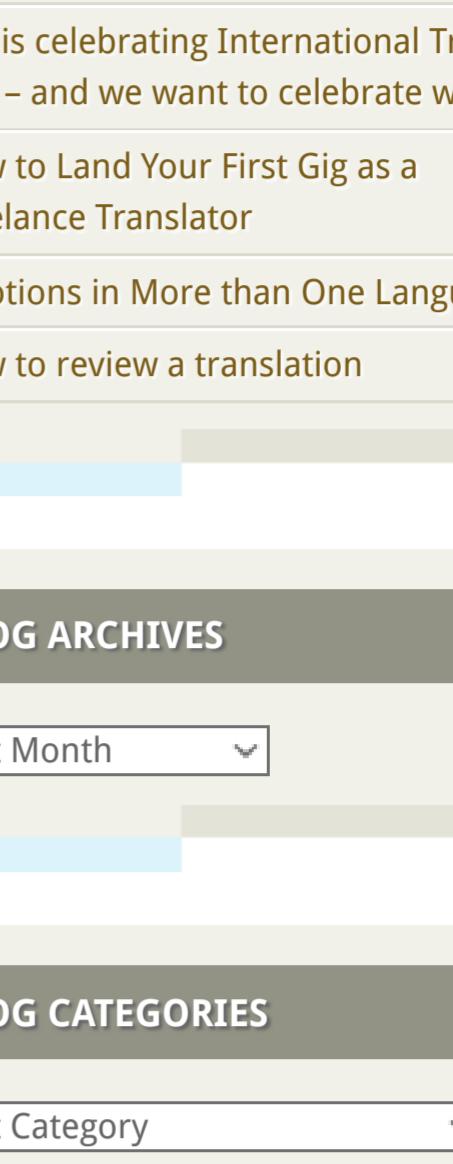
Pro Guide: Translation Pricing and Procurement. Slator, 2021 Mar 19, <https://slator.com/data-research/pro-guide-translation-pricing-and-procurement/>. Accessed 2021 Apr 12.

Simplified Technical English Specification ASD-STE100. AeroSpace and Defence Industries Association of Europe, Issue 7, 2017. PDF.

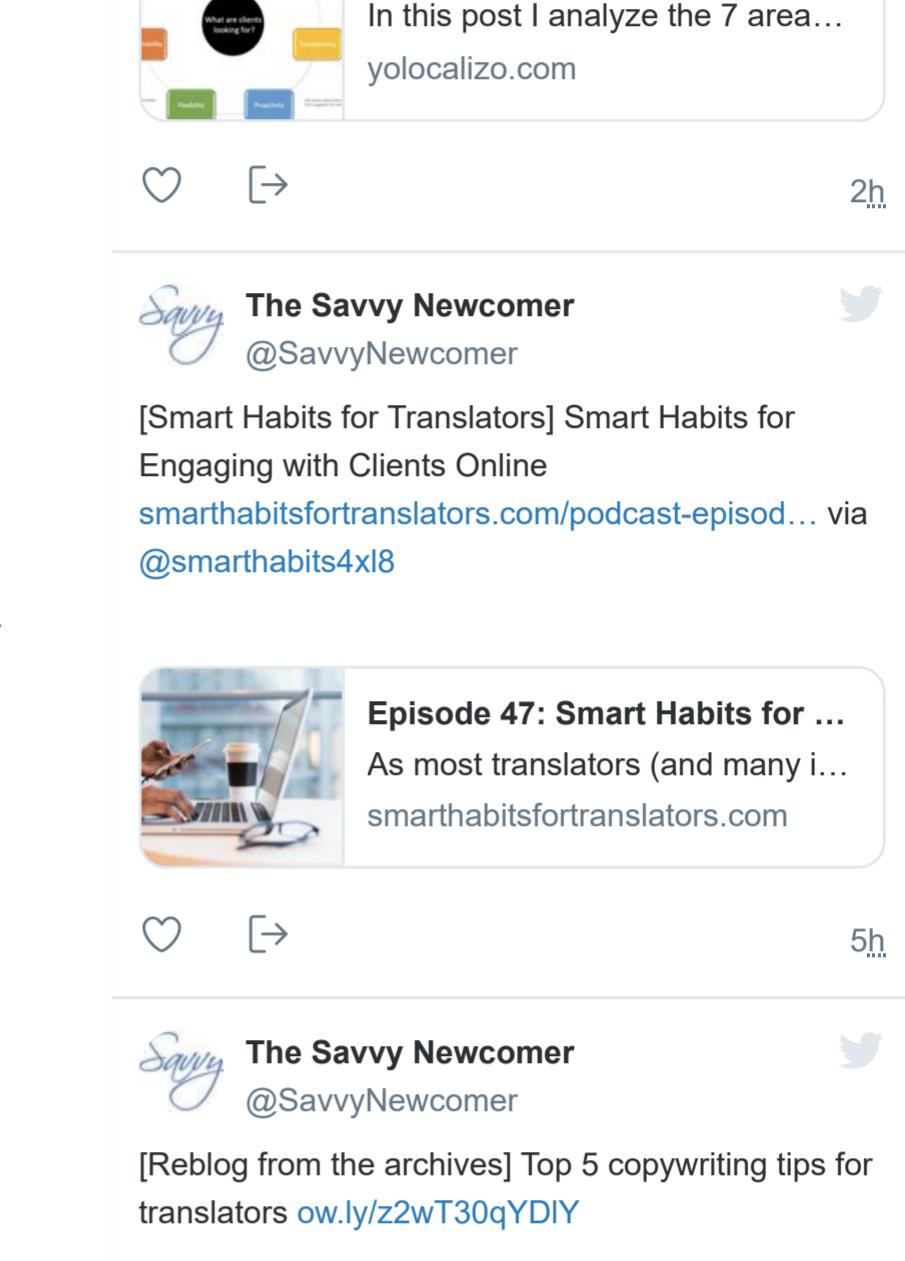
"What is plain language?" Plain Language Association International (PLAIN), 2021, <https://plainlanguageinternational.org/what-is-plain-language/>. Accessed 2021 Apr 12.

Zetsche, Jost, Lynne Bowker, Sharon O'Brien, and Vassilina Nikoulina. "Women and Machine Translation." *The ATA Chronicle*, American Translators Association, Nov/Dec 2020, Volume XLIX, Number 6. Print. Also available via: <https://www.atanet.org/tools-and-technology/women-and-machine-translation/>.

JOIN THE ATA



UPCOMING ATA WEBINAR



RECENT POSTS

[How and Why You Should Diversify Your Freelance Translation Business \(COVID-19 Series\)](#)

[Found in translation. The invisible art of translation deserves wider recognition.](#)

[Savvy Diversification Series – Diversification into Machine Translation](#)

[How to Make Prospect Conversations Easier \(and Land More Clients\) with the "Selling Staircase"](#)

[Inclusive Language Resources](#)

[Buddies Welcome Newbies: A conference event you don't want to miss](#)

[ATA is celebrating International Translation Day – and we want to celebrate with you!](#)

[How to Land Your First Gig as a Freelance Translator](#)

[Emotions in More than One Language](#)

[How to review a translation](#)

BLOG ARCHIVES

Select Month

BLOG CATEGORIES

Select Category

FOLLOW US ON TWITTER @SAVYNEWCOMER

Tweets by @SavvyNewcomer

The Savvy Newcomer @SavvyNewcomer 7 areas that a Localization Service Buyer considers important when working with a Language Provider [yolicalzo.com](https://yolicalzo.com/localization2/) #10n

7 areas that a Localization Service Buyer considers important when working with a Language Provider [yolicalzo.com](https://yolicalzo.com/localization2/) #10n

The Savvy Newcomer @SavvyNewcomer Smart Habits for Translators Smart Habits for Engaging with Clients Online [smarthabitsfortranslators.com/podcast-episode-1/](https://smarthabitsfortranslators.com/podcast-episode-1/) via @smarthabits4x18

Episode 47: Smart Habits for Translators As most translators (and many...) smarthabitsfortranslators.com

The Savvy Newcomer @SavvyNewcomer [Reblog from the archives] Top 5 copywriting tips for translators [ow.ly/z2wY30qYD1Y](https://ow.ly/z2wY30qYD1Y)

Top 5 copywriting tips for translators This post originally appeared on Anja Jones ... [alasavvynewcomer.org/](https://alasavvynewcomer.org/)