

Machine Translation: Translating Automation into New Opportunities

By Ursula Schwalbach and Franco Zearo

Reprinted with permission from The ATA Chronicle (May 2006).

Who hasn't heard the following arguments? "Machine translation is getting better every day and will soon put human translators out of business." And, at the opposite end of the spectrum, "Machine translation is useless because a machine will never be able to satisfactorily translate a novel by William Faulkner." As translators, we choose to believe the latter because it feels safe, and we laugh at all the funny mistranslations circulating in our e-mail and on the Internet. We seem to assume that human translation is always perfect, which it is not. Both arguments miss the point.

This article will explore common prejudices and look at machine translation (MT) from a different perspective. When does it make sense to consider MT as a valid business option? What kind of investment is necessary to achieve acceptable language quality? What opportunities does such an investment present for language professionals?

In examining these questions, the article will not engage in a scientific analysis of the various types of MT engines. Rather, it will focus on the content being translated and how one can ensure that the end product is suitable for its intended purpose.

Audience and Circumstance

The intended purpose is the key to analyzing whether or not MT is a viable business option. Many types of computer-related content will always need the polished look only a language professional can provide, including product home pages, product packaging, marketing articles, and training materials. Other types of computer-related content, such as user instructions, are more repetitive. For these types of text, language automation is not only an option, but it provides the valuable benefit of stylistic and terminological consistency crucial to comprehension and knowledge acquisition on the part of the customer.

Case Study: Knowledge Base Articles

Ten years ago, computer software usually came with a huge help file integrated into the product itself. A manual weighing several pounds and describing every software feature in detail also accompanied the application. Because the shelf life of software and documentation tended to extend over several years, software companies invested large amounts of money in documentation production and accuracy because there was only one chance to get it right.

Now, fast-forward to the Internet age with its proliferation of content and much faster production cycles. All of a sudden, software companies are able to break information into more manageable chunks, update it more frequently, and benefit from the fact that computer experts anywhere can document their experiences in the form of supplementary content.

However, a difficult decision became apparent with the accumulation of these so-called "knowledge bases." How much of this content would be relevant for users in other countries; therefore, how much should be translated?

The amount of material officially translated is but the tip of the content iceberg. No company can afford to have it all translated. Were it not for MT, a great deal of content would remain a privileged experience savored only by those who speak the language in which the text was originally written.

Consider the following scenario. A college instructor in Madrid is working on course material to present the next day. She is using PowerPoint, and all her attempts at integrating complex text and pictures from a non-PowerPoint file keep failing. None of her Spanish product documentation covers the specific problem she has encountered. She finally turns to <http://support.microsoft.com> and finds a knowledge base article in English with an option to have it machine-translated into Spanish. The resulting text is coarse and linguistically awkward, but she can follow the instructions well enough to solve her problem. At the bottom of the article is a box that allows her to comment on her experience.

Applicability to Business

Knowledge base articles are also helpful to companies when evaluating their customers' experience. Did the article help solve a problem? Was the article too poorly translated to be helpful?

If a poorly translated article gets enough hits, a professional will eventually re-translate or post-edit it to make it more useful. This means that customers have access to much more content in more languages than ever before and have a better chance of getting their problems solved right away. At the very least, MT serves as a springboard for ranking information about customer needs and improving the language quality of those articles that are most relevant.

The Black Box Phenomenon

Those criticizing MT expect it to behave like a mysterious black box that consumes any type of source text and produces its beautifully translated target equivalent. The reality is quite different. Achieving satisfactory target results depends on carefully crafted source texts, which for the purpose of this article are assumed to be written in U.S.-English. However, this does not mean that creators of source text should be forced to resort to an artificial style of robotic writing. Even the mention of "controlled English" might produce negative sentiments on the part of writers. Instead, it is much wiser to call this effort "disciplined" writing.

Language automation depends largely on consistent language patterns. Following some basic rules of style goes a long way in making source text more understandable, not only for the machine, but also for the human translator and for speakers of English as a second language. Among the known pitfalls are noun clusters, nested relative clauses, and those ubiquitous "-ing" constructions always in search of their proper agents. For a company, it means separating creative writing efforts (for example, marketing and advertising copy) from text intended for MT or other automation.

Another important factor in MT is domain-specific terminology. Documenting domain-specific terminology and feeding it into the automation process properly also requires the expertise of language professionals. Terminology is the glue that holds language automation together and will help the customer extract the knowledge, even if a text is stylistically or grammatically flawed. Reliable domain-specific terminology will

also make post-editing a much more pleasant task because the language professional is free to focus on polishing text instead of searching for proper terms and phrases.

Post-Editing Services

As the demand for MT grows, we will also see an increase in demand for post-editing services. Post-editing is fast becoming a highly specialized area of language expertise. Due to the increasing volume of text, most companies employing MT as part of their localization strategy have begun to rely on external business partners to provide this service. This section describes the various post-editing methods, which, as we stated before, need to be carefully chosen based on business circumstance and audience needs. Using MT only as a way to save money is the wrong approach. The successful application of each method depends on the superb language and project management skills of external business partners, and on the professionalism of post-editors.

Types of Post-Editing Services

A cursory survey of the post-editing topic shows that three levels of post-editing have been identified:

- Full (or complete) post-editing;
- Minimal (or partial) post-editing; and
- Rapid (or fast) post-editing.

Full post-editing is virtually the same as editing (reviewing) someone else's translation. The main difference from traditional editing is that the post-editor has to correct errors usually not encountered when editing a translation done by a fellow human being. Machine translation errors can be unpredictable, annoying, and even humorous at times. In addition, the post-editor can generally expect to have to correct a larger number of sentences than when editing a human translation.

Rapid post-editing is a "strictly minimal editing of texts in order to remove blatant and significant errors."¹ In rapid post-editing, the post-editors are limited as to the nature of interventions they can make.

Minimal post-editing falls somewhere between the full post-edit and the rapid post-edit. It consists of implementing a "minimum" amount of changes. The guidelines that dictate which post-editing interventions are obligatory versus those that are non-obligatory vary widely and are generally agreed upon between the client and the post-editor.

Rapid post-editing and minimal post-editing offer new and interesting opportunities in the translation field, both professionally and in translation formation curricula. The University of Exeter (U.K.) already focuses on rapid and minimal post-editing, "as the full post-edit was assumed to be indistinguishable from a full high-quality translation."²

Post-Editing Productivity

Post-editing is all about speed. Several experiments have been conducted, some touting savings of up to 70% over so-called "traditional" translation costs and times.³ We generally understand that a translator translates at a rate of approximately 250-300 words

per hour, depending on the translator's familiarity with the subject matter, the language combination, and other factors. An editor can work at speeds of about 1,000 words per hour when editing a translation done by a skilled human translator. Rapid post-editing and minimal post-editing should fall somewhere in between those productivity metrics. One study indicates that the productivity of a rapid post-editor is about 30 minutes per page.⁴ If we assume a standard page to contain 250 words, we are looking at a post-editing rate of about 500 words per hour.

There is one point that some people often overlook, which is that these claimed productivities for post-editing are not sustainable over a long period. In the future, we can foresee the use of teams of alternating post-editors similar to that of simultaneous interpreters working in pairs in the booth.

Post-Editing Skills

To use a sports analogy, post-editors are to translators what sprinters are to marathon runners. Both are athletes, but with very different abilities. Leaving natural talents aside, the distinction comes from different training regimens.

For a post-editor to be efficient, the following skills are necessary:

- **Professional translation competence:** Defined as mother tongue proficiency in the target language and an excellent command of the source language.
- **Keyboarding skills:** Efficient post-editors are very skilled at using keyboard shortcuts. Post-editors should also possess some familiarity with programming macros to expedite the most tedious tasks.
- **Subject area expertise:** Post-editors who are familiar with the subject matter can work faster and are, therefore, more proficient.
- **Lexicography expertise:** In some cases, the post-editor will collaborate in the creation of the glossary that the MT system will use. Lexicography and terminology management skills will thus play an important role.

Who Makes the Best Post-Editor?

Some experts believe that junior translators might have better potential in becoming post-editors because they do not have to “untrain people on what they have done for years.”⁵ Junior translators, it is argued, come with “less baggage.” In our experiments, however, senior translators generally fared better than the junior translators. As in many other crafts, experience does play a role and has its advantages. Nevertheless, more experienced translators might have some preconceptions toward MT. A low tolerance for the types of errors encountered and a negative attitude toward MT are definite obstacles that will prevent a translator from becoming a successful post-editor. We will discuss these issues later in this article.

How to Become a Post-Editor

One factor that hindered the development of post-editing in the past was that mostly in-house staff performed it. Because employed translators receive a fixed salary,

they have little or no incentive to increase their production output. Freelance translators, on the other hand, are generally paid by the word. Even when the job is paid by the hour, it is not unreasonable to demand a higher hourly rate, given the specialization required. This category of language professionals, therefore, has a great incentive to benefit from the increased productivity offered by post-editing and might be more receptive in accepting this role.

If freelancers are the category that stands to benefit most from becoming post-editors, the obvious question is, “Where can they learn about post-editing and pick up these new skills?” Unfortunately, here is where we encounter an obstacle. Besides encouraging professional associations like the ATA and its regional affiliates to begin offering workshops and seminars on post-editing, here are a few ideas:

- **Read about it:** Technology evolves and keeps changing our lives every day. For many of us, commercial MT systems were not available when we completed our translation studies or when we first started working as translators. The references cited at the end of this article constitute a good starting point to become acquainted with post-editing. A rich electronic repository of articles, books, and papers on topics in MT and computer-based translation tools can be found at www.mt-archive.info. Information specific to post-editing can be found by visiting Jeffrey Allen’s pages at www.geocities.com/mtpostediting.

- **Increase your typing speed:** If you type using two fingers and find yourself looking at the keyboard, learn how to touch-type with all 10 fingers. Some adventurous translators might even want to try their hand at learning the Dvorak keyboard or stenocaptioning. According to the National Captioning Institute, “[h]ighly skilled captioners [...] phonetically stroke words on a stenograph machine (like court reporters use) at speeds of over 225 words per minute.”⁶

- **Edit online:** If you like to print translations for review purposes, try to become more comfortable doing online edits. There are several advantages to performing some translation tasks on paper, but now post-editors are expected to work exclusively on screen.

- **Learn sight translation:** Interpreters possess translation skills that can be useful in the MT post-editing process. Sight translation is one type of exercise that might help translators think faster and become more efficient at post-editing.

Even if you decide not to become a post-editor, you stand to benefit from incorporating these speed skills in your everyday practice.

Finally, the most important component in becoming a post-editor is attitude. If one is prejudiced against MT, he or she will not go very far in becoming a successful post-editor. One can always learn new skills, but a positive attitude is something that has to come from within. A good post-editor must understand the quality standards expected from the process, and one’s role in the process. Most translators already work with Translation Memory (TM) software. Fixing a fuzzy match segment in a TM environment is not very different from repairing a MT segment. Jaap van der Meer acutely observed

that, “TM is just a new variant of MT.”⁷ His point is that TM enjoys a widespread acceptability on the part of the translation community largely thanks to the marketing message adopted by the developers of TM software, helping them to avoid the stigma attached to MT in the translation industry. Fortunately, in the recent decade this attitude has been changing, both on part of the MT developers and of the translation community.

Ethical Considerations

Finally, post-editing of MT raises some ethical considerations. When the customer does not have high expectations and quality standards are relaxed, one could be tempted to take some “liberties” and then blame the MT for the poor quality. The American Association for Machine Translation in the Americas (www.amtaweb.org) has reportedly set up a special interest group to determine post-editing guidelines with a view to establishing a post-editing qualification program⁸, although we can find no mention of this initiative on the association’s website at the present date. Certification from professional associations such as the ATA might be one guarantee against malpractice by unscrupulous individuals.

Conclusion

In sum, customers are driving the demand for more and more translated content. Language automation has the potential to satisfy this demand. When coupled with terminology management, and post-editing services, MT can provide an attractive cost/benefit solution. Corporations and translation services companies have a mandate to provide these solutions in order to help their clients. Everybody in the language profession stands to benefit from this increased demand for new language services, and our role is to help the language professionals to develop the skills necessary to meet these new requirements.

Notes

1. Allen, Jeffrey. “Post-editing,” in *Computers and Translation: A Translator’s Guide*. Edited by Harold Somers (Amsterdam: John Benjamins Publishing Company, 2003).
2. Belam, Judith. “Bringing Up to Falling Down: A Deductive Approach to Teaching Post-editing,” in *Teaching Translation Technologies and Tools, Proceedings of the Machine Translation Summit IX* (New Orleans, 2003).
3. Van der Meer, Jaap. “Machine Translation Pilot.” *Translation Automation, Cross Language*, Issue 4 (February 2005).
4. Krings, Hans P. *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. Edited by Geoffrey S. Koby (Ohio: Kent State University Press, 2001).
5. Allen, Jeffrey. “Who Should be Doing MT Post-editing?” (2003). www.mail-archive.com/mt-list@eamt.org/msg00578.html.
6. National Captioning Institute. www.ncicap.org/FAQ1.asp (retrieved August 29, 2005).

7. Van der Meer, Jaap. "Different Approaches to Machine Translation." *Multilingual Computing & Technology*, Vol. 16, Issue 3 (Sandpoint, Idaho: Multilingual Computing Inc., 2005).

8. Allen, Jeffrey. "Post-editing," in *Computers and Translation: A Translator's Guide*. Edited by Harold Somers (Amsterdam: John Benjamins Publishing Company, 2003).

Additional References

Arnold, Doug, Lorna Balkan, et al. *Machine Translation: An Introductory Guide* (1995). www.essex.ac.uk/linguistics/clmt/MTbook/HTML/book.html.

"Machine Translation." http://en.wikipedia.org/wiki/Machine_translation (retrieved August 30, 2005).

Melby, Alan K., and Terry C. Warner. *The Possibility of Language: A Discussion of the Nature of Language with Implications for Human and Machine Translation* (Amsterdam: John Benjamins Publishing Company, 1995).

Ursula M. Schwalbach holds a B.A.-equivalent in translation from the Übersetzer- und Dolmetscherinstitut Würzburg, Germany, and is in the process of leaving Microsoft to complete her graduate studies. After joining Microsoft in 1992 as an in-house translator, she worked for several years as a localizer and localization lead, and became one of the first terminologists at Microsoft. Her final position was managing a newly formed team of terminologists focused solely on source-language integrity. She is ATA-certified (English-to-German, German-to-English). Contact: uschwalbach@comcast.net.

Franco P. Zearo is the worldwide director of Language Services with Lionbridge Technologies. He holds a degree in translation from the Advanced School of Modern Languages for Translators and Interpreters at the University of Trieste, Italy, as well as an MBA from the University of Phoenix. Before joining Lionbridge in 1996, he worked as a freelance technical and medical translator in Italian, English, and Russian. At Lionbridge, he is responsible for quality assurance and process improvement for language services for all corporate offices worldwide, and interacts with translation professionals in Europe, Asia, and the Americas. He is an ATA-certified (English-to-Italian) translator. Contact: franco.zearo@lionbridge.com.