ATA Position Paper on Machine Translation:
A Clear Approach to a Complex Topic

ATA's position on Machine Translation

Machine translation is one of several valuable tools that translators can use to produce a professional translation. For reliable and secure translation, machine translation should not be used without the ongoing involvement of professional translators. 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. Understanding is the role of professional translators.

This document is an attempt to clarify in easy-to-understand terms what machine translation is and how it realistically relates to professional translation. Here are the steps we'll take:

- Discuss expectations
- Define communication and translation.
- Describe the different kinds of machine translation by looking at how different approaches to machine translation parallel the evolution of computing power.
- Explore how machine translation is useful in various scenarios.
- Survey how professional translators use machine translation in their work.
- Demonstrate that in a scenario where reliable and secure translation is needed, the only way to use machine translation successfully is in combination with professional translators.

Expectations

In many ways, the story of machine translation (MT) is a narrative of fulfilled, failed, and adjusted expectations.

These expectations vary widely. A casual private user who needs to understand something in an unfamiliar language won't expect 100% accuracy—after all, they've read the hilarious media stories about machine translation failures. But they will expect to receive a general idea of what the otherwise inaccessible text says. Casual users also won't—and if they do, they shouldn't (!)—expect a high degree of privacy; the vast number of free web-based translations performed day-in and day-out proves that this kind of user is satisfied with a low privacy bar.

Business and other non-private users have different expectations, chiefly reliable, appropriate, and secure translations. Those that continue to use the free web-based approach to machine translation often end up frustrating their expectations and their end users. Most members of this group have learned to utilize machine translation only in combination with professional translators.

Professional translators' expectations of machine translation have shifted over time. What began with amusement and mockery gradually gave way to the perceived threat of a radically increasing quality of machine translation. Lately, however, the majority of translators have realized that machine translation is one of several valuable tools they can use to produce professional translation.

So what is machine translation?

Machine translation is the translation of a text from one language into another language by a computer without any human involvement during processing.

There is plenty of information about machine translation out there, and it often falls into one of three categories: it's highly technical and very challenging for the average reader; it assumes that machine translation has essentially solved the "problem" of translation—often referring to other literary fixes from science fiction movies and books (like "Babelfish" or "Universal Translator"); or it focusses solely on the failures of machine translation.

Understanding communication and translation

To understand the goal of machine translation, it helps to understand translation—the transfer of text from one language to another—as well as successful written communication.

Written communication is based on a number of assumptions. The most basic is that author and reader share a common language—including vocabulary, rules, and register—determined by the experience they both have with that common language. If one of these is missing, the text will be misunderstood or not understood at all. While these parameters might not seem particularly surprising, it's very helpful to remember that just because someone has written a text in a language you know, without these shared assumptions there is no guarantee that you will understand the text completely. Think of a legal document: it may be in English, but most people can't understand it.

Translation involves additional levels of communication. Translators not only need to understand the original text as it was originally intended, they also need to reproduce it in a way that elicits a similar response from the reader—or as similar as possible, given the differences in language and culture.

It's easy to see why this is a complex process. Elements that at first glance might seem easily transferrable, such as rules and vocabulary, are anything but straightforward. A complete match between one word in one language and its counterpart in another language rarely exists. (Compare the widely different images for an American, a German, and a French reader when they read the seemingly simple word "bread," even though these are all cultures where bread plays an important role. How much more extreme will this be in a culture where it does not play such a role?) And when it comes to different registers of language and the purpose of a text, things naturally get even more complex.

It's such a complex task that the idea of computerized translation becomes attractive.

The evolution of machine translation

From the very earliest days of computing, software developers have tried to use computers for translation. As a result, the evolution of machine translation has closely followed the development of computer technology and processing power.

Starting in the 1950s, machine translation consisted of rules- and dictionary-based machine translation systems (rules-based machine translation or RbMT), followed by systems based on massive amounts of mono- and bilingual data that was fragmented and reassembled for translation (statistical machine translation or SMT). Neural machine translation (NMT), an "artificial intelligence" (AI) technology, processes the same data fragments using neural networks that analyze the text to be translated as part of the larger text and suggest more context-based translations. Each of these approaches matched the amount of computing power available to larger organizations at the time of its development.

Interestingly, each of these approaches is still being used, and each has its own weaknesses and strengths. Rules-based systems tend to deliver better results if languages are similar; these systems are more easily customizable, but they tend to be very labor-intensive to set up. A statistical approach is quicker to set up—if there is enough high-quality, professionally translated data as training material— but it tends to struggle with languages that are structured differently (such as English and Japanese). Neural systems handle these different language combinations better, but their more fluent output can make it difficult to spot regularly occurring textual omissions and mistranslations—mistakes don't stand out as well.

Where does machine translation play a role?

Different machine translation programs can be used in a wide variety of ways. Publicly available programs from search engines Google, Microsoft Bing, Yandex, and Baidu are designed to allow for generic, ad-hoc communication between users of different languages on the internet where neither strict accuracy nor confidentiality is of central importance. However, all of these tools are unsuitable for confidential and/or professional communication because they capture all data transmitted through their services for further training purposes (unless you pay for specialized tools to keep your data private).

Custom-programmed machine translation engines—regardless of their underlying technology—are trained by a specific organization for a specific purpose with specific language material, using either proprietary rules and dictionaries and/or professionally translated texts or even non-translated texts. Results from these programs are much more reliable when it comes to terminology, and confidentiality is addressed by limiting access.

While unedited output from these programs is sometimes used in public-facing knowledge bases, low-level support material, or internal company communications, it is understood that the output of these customized engines doesn't meet the same level of reliable quality as competent work done by a qualified professional translator. The reason for this failure is the same across the different technologies: 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.

Understanding is the role of professional translators.

Human or machine?

Professional translators interact with machine translation output in a variety of ways. First and foremost, they can help with the fundamental decision of whether to use machine translation at all, and if so, for what kinds of material. Even companies that have invested in a machine translation system do not use the system for all content. For instance, a program that generates acceptable technical content

may not work well for legal content. One that is acceptable for legal content will be sub-par in medical translations. And neither of those will be suitable for very creative content such as marketing material. If the translation buyer and translation professional determine together that machine translation generates enough helpful output to improve translation productivity, that data can be used in a number of ways.

In "post-editing," the professional translator edits machine translation output. This might make sense when the training data is of very high quality and the domain of both training data and translatable data is very well defined. If any of those parameters is not in place, any productivity gains made in using machine translation would be negated by increased professional human editing time. Increasingly, post-editing of machine translation is becoming a sub-category of translation. Developments like adaptive machine translation, which dynamically learns from the corrections that the post-editor makes, particularly help this process.

Professional translators also increasingly use machine translation data in concert with the many other resources at their disposal. This includes terminology databases, translation memories (databases of previously translated and verified material), and a wide range of quality assurance tools. Many machine translation programs base their translations on stored phrases or fragments. Translators can access these phrases or fragments automatically to increase both translation speed and consistency. A professional translator can also use machine translation as reference material or as a source of alternative translation suggestions (and output from more than one machine translation program can be used in parallel).

These professional translators use either customized machine translation engines that are inherently secure or generic systems that, while also open to the public, provide secure and paid access for professionals.

Human <u>and</u> machine!

In any of these constellations, professional translators and machine translation engines work together very well. And that brings us to the ultimate takeaway of this complex topic: If reliable and secure translation is desired, machine translation should not be used without the ongoing involvement of professional translators.

For translators, this means that their tool sets have been expanded by yet another potential resource that will prove valuable and increase productivity if used appropriately. And for purchasers of translation who are interested in using machine translation as part of their translation processes, it's one more qualification they need to look for in a suitable translator: someone who can evaluate the potential benefit of using machine translation, can provide guidance in a choice of a system, and knows how to use it to our mutual benefit.