



## **Commercial Machine Translation Today**

*Trends, Challenges, Opportunities, and a Case Study . . .*

Olga Pechenko, CEO  
© McElroy Translation 2010

## OVERVIEW

Machine translation (MT) is not the new technology on the block. However, in the last few years there has been incredible progress in the usage of MT within the commercial environment. Here we summarize the latest news on MT discussed at the Association for Machine Translation in the Americas (AMTA) conference held in early November 2010.

Insights include:

- Overall trends in machine translation
- Challenges
- Opportunities
- A case study demonstrating how MT can be leveraged today

## To sum it up...

As information continues to expand and translation budgets continue to shrink, more and more companies will be forced to turn to MT solutions. It is extremely important to remember, however, that training an MT engine to adequately translate your materials only makes sense for certain scenarios. You must first consider the following:

- What is the business purpose of the translation output?
- What is your current translation volume?
- How much translation memory do you currently own?
- Will your audience (be it internal or external) forgive “good enough” quality that comes with raw MT or post-editing? Do you or your customer understand the limitations and quality issues that come with raw MT?
- If you plan to use MT for external purposes, what consequences should be weighed? Is a low-quality translation better than no translation at all, or can a low-quality translation be more damaging to your company’s reputation than having no translation at all?

If done properly, training an MT engine will reduce translation costs and turnaround times, but it comes with an upfront investment of time and money to build an MT engine that will provide you with relevant raw MT and post-editable MT.

**MT** – machine translation, the production of text in one natural language from that in another by means of computer procedures

**AMTA** – Association for Machine Translation in the Americas

**RbMT** – rules-based machine translation built by using language rules and terminology

**SMT** – statistical machine translation built from a large corpus of previously translated texts

**LSP** – language service provider

**PE** – post-editing, a service provided that involves human editing of machine translation

**Light PEMT** – light post-edited machine translation to ensure correct translation of terminology

**Full PEMT** – full post-edited machine translation to ensure correct translation of terminology, as well as editing the overall document for legibility (Quality isn't as high as human translation, but document is functional.)

Machine translation (MT), at its basic level, performs a simple substitution of words from one language to another. At this level, an untrained engine will produce word-for-word translations that, when combined, often result in garbage. MT engines must be trained in order to provide any level of quality to the target document output.

There are multiple types of MT outputs to train an engine:  
 rules based (RbMT)  
 statistical based (SMT)  
 example based (EBMT)  
 hybrid MT (HMT)

For the purpose of our discussion, we will focus on RbMT and SMT. **RbMT is built by using language rules and terminology, while SMT is built from a large corpus of previously translated texts.** Each time new information is contributed the MT engine “learns” from said texts and delivers results based on a statistical assumption of what was previously translated.



The overall consensus at the AMTA was that initially researchers were focused on a scientific approach without gathering information from both customers and translators. With a shift in focus, MT quality has increased, and enterprises are now experiencing a true benefit for making more information available to their customers—information that would otherwise go untranslated.

**Why is machine translation necessary?**

The average company is seeing information double every 1 to 2 years. With the speed at which information is growing combined with decreasing budgets, companies are faced with the challenge of how to continually increase their revenue through new markets globally without spending more money. Fortune 500 companies have embraced this challenge by exploring MT engines and providers. Currently, it is estimated that an average Fortune 500 company translates roughly 5% of its content using MT and this is forecasted to increase to 30% to 50% over the next 5 years.

*But the questions linger: What exactly does this mean? Are corporations using free software (e.g., Google Translate), are they purchasing software and creating their own engine, or are they working with a language service provider (LSP)? Are they using raw MT or are they post-editing? If they post-edit, is it done lightly for technical accuracy or is it a full post-edit to provide near-human-quality translation?*

#### **Why not use a FREE engine?**

Companies like Adobe, Intel, Avaya, and Continental are forgoing Google's free engine for a couple of reasons:

- Content loaded into Google's engine becomes public domain, and is thus owned by Google.
- The engine is not subject-specific; i.e., Google Translate is not [and cannot be] trained according to subject matter and thus delivers mostly garbage when used for something technical or scientific.

Many of these companies, along with LSPs, are acquiring MT engines that enable them to keep their content private while controlling the training of the engine, be it client- or subject-specific. There are quite a few players right now who offer MT engines.

#### **What engine should I use?**

If you know you want to train an engine yourself, the first thing you will want to do is reach out to a few different suppliers to learn more about how their systems work and whether or not they are the right fit for you. Make sure if you plan on combining your translation memory (TM) with your MT engine that you take into account whether there is an open API and if the current TM system you have will connect. Here is a list of a few reputable MT engines:

- Privately owned MT engines: ProMT, AppTek, Asia Online, Systran, Language Weaver
- Free, open source MT engine: Moses

#### **This looks intense. Can I source the training to an agency?**

For companies not interested in devoting internal resources to developing an in-house solution, they can work with an LSP that offers MT engine development solutions. These LSPs will train an engine with your previously translated content and terminology and maintain it by continuously updating it as new content becomes available.

#### **How can you trust a machine to perform at a level that will provide a readable output?**

You shouldn't. Even a well-trained engine will have quality issues with flow and grammar. If you plan on using the machine-translated documents for external use, you will need to have the documents edited. **Post-editing (PE) is a service that involves human editing of the raw MT output by a bilingual or sometimes even just a monolingual editor.** There are two types of PE: light and full. A light edit provides a review to ensure the accuracy of technical terminology within a document; a full edit provides a medium-quality output that includes correct terminology in addition to correctly arranging the words to improve the flow of the text. Both services cost less than human high-quality and draft translations and increase the rate of output 4 to 5 times. PE is incredibly important for increasing the value and usability of your MT engine, as well as providing a legible output.

## So, what type of documentation is a good candidate for MT?

You can see a very general categorization in a table below.

LEVEL	TRANSLATION	CONTENT	QUALITY
1	Human Translation	Marketing materials, translations requiring certification, highly technical documentation for external use, home pages, landing pages	High
2	Post-edited MT (PEMT)	Internal use documentation (scientific, technical), web pages with high click-through rates (e.g., services, contact us)	Medium
3	MT + terminology, TM*, and/or dictionaries	Discovery documentation, for your information web pages	Below Average
4	Raw MT	Social media, digital media, consumer information	Poor

\* Translation Memory

There is a saying in the MT world: the quality of the output depends on the quality of the input. Terminology consistency is easily one of the greatest challenges in engine development. Large enterprises struggle with varying departments disagreeing on terminology, compromising the quality of the output. Another big challenge is information redundancy. In order to get a quality MT, your technical writing departments should keep sentences short and simple (which is desirable for your customer as well). The more complicated the sentence becomes, the more room for error MT faces.





## Intel Case Study presented at AMTA

Marketing Solutions: Translation Innovation & Services

Will Burgett, product manager for Intel Corporation, presented on Intel's MT solution for their customer support department. Eighty percent of Intel's revenue is from international sales. In four years, Intel's customer support department has seen its budget cut in half, while its translation volume grew 350%. This translation volume includes 1200 to 1500 files, into 9 languages, requiring revision every 10 days.

Prior to implementing machine translation, it took 10 days for a human to translate these files. By the time translation was complete, revised files were on deck for another round of translation. Through implementing and training an MT engine, translation time (for raw MT) was reduced to 24 hours, and translation costs were reduced by 95%. Intel is currently processing 60,000 words/hour.

Intel's MT engine is maintained and updated every 6 months by retraining the engine with all human translation that has been acquired within that time. Engine training is a complicated task that requires Intel to employ an internal team; however, this expense has been offset by the incredible ROI they have experienced.

But what about the actual experience of Intel's customers? Intel has a customer feedback field on each web page containing raw MT translation to monitor customer satisfaction. Less than 1% of the comments they've received have been complaints pertaining to the quality of the translation.

It should be reiterated that in order for Intel to experience the success that they've had in creating machine translations that are "good enough," they had to begin with an incredibly large corpus of quality human translations and have put an immense amount of time and energy into training and re-training that engine.

## About McElroy...

McElroy Translation provides quality translations that drive the global economy. **Since 1968**, McElroy has provided technical, medical, legal, business, and marketing translations for **more than 3,000 clients** in almost every market imaginable for **more than 200 languages**. We continue to earn the business of many long-term clients and currently work with a roster of more than 850 clients. Our workflow system delivers translation projects efficiently and is easily customized to integrate with client systems. We use native-speaking translators and editors, incorporating strict quality control measures, and our processes allow our clients to view the current status of their projects online at any time during translation via their client portal. McElroy invests strategically in talent and technology to ensure ongoing language industry leadership. **McElroy is a WBENC (women-owned) and HUB certified business.**

To learn more about machine translation,  
email us at [MT@mcelroytranslation.com](mailto:MT@mcelroytranslation.com).

### GET STARTED!

Submit an Order: [orders@mcelroytranslation.com](mailto:orders@mcelroytranslation.com)

Submit a Quote: [quotes@mcelroytranslation.com](mailto:quotes@mcelroytranslation.com)

### GENERAL INQUIRIES, INFORMATION, & TECHNICAL ISSUES

Customer Service: [customerservice@mcelroytranslation.com](mailto:customerservice@mcelroytranslation.com)

### INDUSTRIES

Legal Division: [legal@mcelroytranslation.com](mailto:legal@mcelroytranslation.com)

Life Sciences Division: [lifesciences@mcelroytranslation.com](mailto:lifesciences@mcelroytranslation.com)

Technology Division: [technology@mcelroytranslation.com](mailto:technology@mcelroytranslation.com)

Energy Division: [energy@mcelroytranslation.com](mailto:energy@mcelroytranslation.com)

Consumer Products Division: [consumerproducts@mcelroytranslation.com](mailto:consumerproducts@mcelroytranslation.com)

### CORPORATE HEADQUARTERS

McElroy Translation

910 West Avenue

Austin, Texas 78701

+1 512. 472.6753, 800.531.9977

Facsimile +1 512.472.4591

