success with MT

MT on a Shoestring · -Part 2

By
Mike Dillinger, PhD,
Translation Optimization Partners

MT is hot. Google has done a great job both of drawing attention to MT and of making better quality automatic translations available for an amazingly broad array of languages. Microsoft made the connection to its in-house MT servers open to the public, too. Translation providers SDL and Lionbridge both are developing MT systems in house. Organizations like Cisco, Symantec, Intel, Adobe, the European Union, and the Pan American Health Organization all have installed MT servers. The Association for Machine Translation in the America's conference last year in Hawaii had twice as many presentations and participants as usual. Everyone seems to want to know more about MT.

But most small and medium organizations don't have the resources to invest US $100,000 in an enterprise server for one or two language pairs, even when they do enough translation work to get a return on that investment. What are the other options available for trying MT, when you don't have much of a budget?

Freeware

Free translations have been available on the web for more than 10 years. So, what's new? For one thing, there's the possibility of connecting your programs to the translation services to feed them documents automatically. For another, the quality of the translations and the range of languages are much better than 10 years ago.

I've heard several people think aloud saying, "Let's just get draft translations of our clients' documents for free with Google or Microsoft. That'll save us lots of money." There are a couple of very important issues with freeware, though. Your clients might not be too happy that you're circulating their documents out on the wild, wild Web to be stored on someone else's servers. You also don't want to build a business that depends crucially on a service that might be cut off at any time without warning. Murphy's law suggests that they will discontinue the service, as soon as you get addicted to it. Finally, think about the freeware approach in practice: you'll end up fixing the same translation mistakes over and over and over again, paying for the same work many, many times.

The same mistakes will keep coming back to bother you because you can't do anything to improve the translation service or adapt it to a particular client's project.

I call this approach black-box MT: the translation service is like a sealed black box and you don't know and can't influence anything of what's happening on the inside. Black-box MT will work for the casual translator or for small projects, but it's not a good process to rely on and you can't transfer what you learn to an in-house MT system.

Desktop software

MT software vendors like Systran and PROMT offer high-end desktop software that is a much better option to evaluate how MT will work in your organization. The software, however, won't magically teach you how to set up a reliable translation process to get the most out of the software! For that, you need help from an experienced MT consultant.

Which translation engine? For whichever vendor you prefer, pick the top-of-the line desktop product, for three reasons. Reason 1 is that it'll be more similar to the enterprise version when you decide to upgrade later. Reason 2 is that you absolutely need the dictionary tools that come only with the high-end versions. Without these tools, you'll be stuck back in black-box MT, with no control at all over the system and with no ways to improve it. Reason 3 is that you're getting functionality that's very similar to an enterprise server for less than 1% of the price. A $50 translation system just won't do the job.

For about US $1,000 (list price per license), you can get the Systran Premium Translator with the English World pack (English <> 12 other languages), the PROMT Expert 8.0 Translator with the English Giant pack (English <> 6 other languages), or Word Magic's Premium Translator for English <> Spanish. There are also additional dictionaries that you can buy that will improve translation right away if you need to focus on a specific topic. You can choose fewer languages for a lower price, according to your needs, but you definitely need the Premium or Expert editions that have the dictionary tools.

How to use it. Set up one machine with an MT license. One person will get familiar with the software while others will build the processes around MT. The most important trick is NOT to look at your document translations right away! (If you do, then everyone moans and groans, uses expletives in multiple languages, and comments on what a waste of time this is! So, be patient.)

First, scan your documents (using the dictionary tools) for terms that are not in the MT dictionaries yet. Develop the discipline to do the prep work necessary to bring the dictionaries up to speed before looking at any translations. Remember that you will reuse this dictionary building effort every single time that you translate with MT. Dictionary building is a very good investment: the better your dictionaries are, the less you'll spend later on post-editing. Then, connect the translation memories that you already have to the MT system - usually you can **import them as TMX files.** Yes, these high-end desktop systems can re-use your translation memories, too. You get the best of translation memory together with the best of machine translation.

Now that the system has a better chance for success, run your documents through again and look at the draft translations. Yes, there will be errors; some will be stupid, others will be funny. But the point is that you get a full draft translation in seconds and for free. No matter how long the document is, the MT engine zips through it at a few hundred words per second. That will save you time and money, even if you have to re-translate some of the sentences.

NEXT STEPS

Just because you got a pretty good translation doesn't mean that you're finished. Now you need to organize your work to get better and better translations each time. You need to adapt your translation quality control procedures to deal with MT. You have to be able to measure how good or bad the draft translations are. You need to develop procedures and train people to build better **dictionaries. You need to find coaching to develop skills in error analysis so that you can fix the translation engine** once rather than edit the drafts forever. You need to train people how to post-edit the MT output.

The good news is that the dictionaries and processes that you develop around the MT engine save you more time and more money as they improve. It will take you quite a while before your other processes can catch up with the speed of a desktop MT engine. These processes are assets and intellectual property that can become a competitive advantage - today, very few organizations can manage machine translation well. You can use your processes if you change MT vendors, buy more licenses, or add new topics. And they are the same processes that you will use when you expand your operations to include more source-language documents, more target languages, and, eventually, a shiny, new enterprise translation server.