GeekSpeak | Jost Zetzsche

jzetzsche@internationalwriters.com

Training Machines

During a talk I gave last week about translation technology, I was criticized for painting a much too rosy picture of machine translation (MT). This frustrated me a little because I am not trying to convince anyone that MT is positive (or negative); instead, it is a reality with which many translators will have to deal or—more accurately—in which they will have to participate. I have been talking a lot about the interesting opportunities that may be connected with that reality. Here is the story of one.

I spent this past week training a rules-based MT engine, during which I finally came to understand a rather fundamental truth about MT. (I have never claimed to be particularly quick-witted!)

The translation results out of the box (language direction English-to-German) were really rather pitiful. And while they got better after changing some basic settings (such as the text genre and form of address for the user group), they still were not good and not much of a productivity boost for any translation effort. Plus, they were decidedly worse than what Google Translate would have given me.

But then I set out to work to improve the results. I imported large and high-quality translation memories the client had given me. I extracted terminology from the source texts that I translated based on the data in the translation memory and defined grammatically. I also imported and fine-tuned glossaries. Now the results were better, but still far from what they were supposed to be. The rest of the week was spent "translating" the source texts with the MT engine at its current state, hunting down terms that were not (or incorrectly) recognized,

and entering them into the dictionary. (Note that "entering them" does not just mean sending the source and target translation, but defining them grammatically as well.) I then spent time retranslating, hunting down more terms, and on and on.

At some point it was fun to see the results getting better each day, and at the end of the week the results were certainly better (though still maybe less idiomatic) than that of Google Translate. Assuming that the final product will be a translation of "high" quality, I would guess that the post editor will save somewhere between 20% to 40% of the time that it would take to translate the text within a traditional translation environment tool.

Clearly, this is not something that would make sense for a smallish kind of project, but it can be a large productivity boost for a very targeted project with a few hundred thousand words. (And if the goal is to deliver "only usable" translation, the savings would naturally be higher.)

But here is what I learned. I know that more and more translators are using tools like Google Translate or Bing Translator as productivity boosts. Aside from the confidentiality issues (your clients probably will not be too thrilled to see you uploading their data to Google or Microsoft), the results might look like decent productivity boosts at first glance (this obviously depends on your language combination and the kind of translation you are doing). But if you are inclined to use MT as a productivity tool and you generally work in the same domain, you will achieve greater boosts with MT tools that you can train. In turn, however, this means that you will have to make a heavy investment of time before you see any results.

ata