Translation Theories Explored

Sections to be added to Chapter 7 "Localization"

7.5.4 Data-based machine translation

The main recent advance in localization technologies has perhaps been the integration of machine translation into translation-memory systems. In some cases this is fairly simple. If the translation memory does not give you a full or fuzzy match, it can present a suggested translation drawn from an online machine translation system. The translator thus has to deal with the machine-translation suggestion only when the translation memory fails. However, there is a lot more happening in machine translation than this simple "Plan B" approach.

There are several different kinds of machine translation systems, and this is certainly not the place to go into all of them. The important difference these days is that some systems are "**data-based**" or "**statistical**." This means that, in addition to linguistic mapping rules, they are able to search through large databases of paired source-language and target-language phrases, propose the most statistically likely pairs, and determine which of them are well-formed in the target language. This might sound laborious, but its product is almost immediate, and the effect should be revolutionary. With these systems, the more the machine-translation tool is used, the more paired segments are fed back into the system, and the bigger the paired databases thus become. Why should that be important? Because the bigger the databases, the better the proposed translations. That is, the more these systems are used, the better they become, without the need for anyone to write complicated mapping rules.

Translators have spent several decades claiming that machines will never be able to translate. Now we have to start reconsider what that means. It is easy enough to feed text into an online machine translation system and make fun of the results. But for many language pairs we are now at the stage where it is quicker to revise (or "postedit") machine-translation output than to start translating from scratch, and the differences in quality may not be all that significant (Pym 2009, García 2010). In theory, all translators will sooner or later be postediting machine translation output.

To what extent will this fundamental change alter the way our societies view and use translations? That remains to be seen. The promise of high-quality machine translation would seems also to rely on **pre-editing** (rewriting source texts to make them syntactically simple), in addition to a lot of postediting, and it is not altogether clear how many people are going to be proficient at those very particular skills. For some localization projects, however, the consequences have been rather more immediate.

7.5.5 Volunteer translation and the most beautiful names for anonymity

If you have a system that improves the more it is used, you logically need a lot of people to use it. A system like Google Translator Toolkit, released in 2009, uses this logic by providing for free a online translation memory system that by default incorporates machine translation suggestions. That is, as you translate, you can build your translation memories at the same time as you postedit machine translation output (do not confuse this with the machine translation system Google Translate). In exchange

for this free tool, the translations you produce are by default fed into Google's databases, thus improving their system. The more people get involved in the system, the better it works, so the more people will be involved, and so on. This is how a private company can solve a lot of translation problems by giving us something for free.

This logic of public involvement can be seen in the settings of Google Translator Toolkit, which explicitly caters for the group translation of websites and Wikipedia articles. The system is set up for projects where translation is not only going to be done on a voluntary basis, but is likely to be done by a group of translators who communicate with each other online. The technology thus moves us toward new kinds of work arrangements, and a major challenge to the use of individual professional paid translators.

There are many names for the incorporation of volunteer translators in this way. Popular references are to "**user-generated translation**," "**crowdsourcing**" (as a poor rhyme on "outsourcing"), "**community translation**," "**collaborative translation**," or the abbreviation **CT3** (for "community, crowdsourced and collaborative translation"). All these terms have their relative advantages and drawbacks (cf. Pym 2011). None of them seems to focus on what might be the most innovative element: under all these rubrics, translative work is going to done by people who are not financially rewarded for their efforts. That is, the work is going to be voluntary, and for that simple reason we believe that "voluntary translation" is the most suitably provocative name for the novelty.

Volunteer translation is sometimes carried out by a community of users, as in the case of Facebook or Twitter. This makes some social sense. After all, the people who use these networking services are probably the ones best able to decide on the most appropriate translations, and who will most directly benefit from the results. In the case of the Facebook crowdsourcing system, users propose possible translations (mostly for less-than-transcendental sentences like "Who are you looking for?"), then the users themselves vote on the most appropriate suggestion. The translation process has thus been significantly socialized, in keeping with the social nature of the networking service itself. In more committed cases such as Greenpeace or Amnesty International, we might more readily say that the work of volunteer translators constitutes active intervention, an empowering democratization of translation technology.

In all such cases, various technologies are being combined to make translation far more than the individual professional activity that it is traditionally conceived as. Some professional translator associations have already begun to point out the dangers of placing undue trust in public technologies and unqualified translators. At the same time, however, there is little reason why interested users should not be well positioned to decide on the most suitable translations: Facebook fans know what works for their particular class and generation, and Greenpeace activists are likely to be quite good at finding the right ecological terms for their particular locales. Translation quality may ultimately not be the major problem. On the other hand, there might well be problems with respect to cross-product consistency, workflow deadlines, and corruption by intruders. Because of these problematic aspects, the way ahead would seem to involve various kinds of cooperation between volunteers and professionals, with both groups intervening at different stages in the workflow.

García, I. (2010) 'Is Machine Translation Ready Yet?' *Target* 22(1): 7-21.

Pym, A. (2011) 'Translation research terms: a tentative glossary for moments of perplexity and dispute,' in A. Pym (ed) *Translation Research Projects 3*, Tarragona: Intercultural Studies Group, pp. 75-110.