



What's Wrong with Translation Technology?

When Lucy Brooks from eCPD Webinars contacted me some-time last year to offer me a slot (or two) for a webinar, I tried to come up with something a little different. And I think I did.

Rather than talk about what I know about translation technology, I wanted to talk to other translators about where they felt we are missing out in translation technology. Then I would take those suggestions to the translation technology developers and get their responses—whether they could see that an implementation could happen, whether they might already have something like that in place, or whether they might have other suggestions themselves. Finally, two weeks after the first webinar, I would do a follow-up webinar to report on the developers' responses and come up with an action plan of sorts.

I am very happy to report that all of this happened just as planned and with very tangible results.

For the first webinar, entitled “Translation Technology—What’s Missing and What Has Gone Wrong?” (you can view it for free at <http://vzaar.com/videos/1464723>), I had already polled the subscribers of my *Tool Box Journal* and my Twitter followers. I was able to present the webinar attendees with some categories of missing or underdeveloped features—voice recognition, access to external resources, termbases, translation memories/corpora, machine translation, general user-friendliness, exchange standards, and “other”—that we then used to flesh out during the webinar. (You can find the specific criteria in a recent copy of the *Tool Box Journal* at <http://tinyurl.com/missingfeatures>.)

This long list of proposals and queries went out to more than 20 translation technology providers (essentially everyone I could think of who makes software relevant to freelance translators). The following providers responded, some with very comprehensive answers: Atril (Déjà Vu), KantanMT, Kilgray, Lingua et Machina (Similis), MemSource, Multicorpora (MultiTrans), SDL (Trados Studio), Star (Star Transit), Tauyou, Terminotix (LogiTerm), Wordbee, Wordfast, and XML-INTL (XTM Cloud). Among those that did not respond were some longshots like Microsoft and companies like the Ukrainian AIT, whose owners probably have their minds on more elemental issues right now with recent events in Ukraine. On the other hand, the non-response of other disappointing MIAs may reveal a certain dismissive attitude toward their users.

As I said, some of the responses were rather detailed, so it would go beyond the allotted space of this column to give you all of them, but I am happy to share the compiled results as a large Excel spreadsheet. (Just e-mail me at jzetzsche@internationalwriters.com.) In the meantime, though, here are some highlights.

Overall, our suggestions were not only welcomed but deeply appreciated by most vendors, who typically were very honest in their assessments. Take, for instance, Wordbee’s introductory remark: “Our development plan is quite full. I would be lying if I said it’s possible to turn it around now. Therefore, I can only say ‘we will do it’ for a very small part of your list. But all points seem ‘logical’ pertaining to stuff to be done in the mid- or long-term.” Great: if that is the result, we will take it!

A number of tool vendors also used this opportunity to show off some of the features of their tools. Rather than this being annoying, however, it was helpful to see that in some categories we might already have made more progress than we (or I) thought. For instance, consider the automatic fixing of fuzzy translation memory matches and/or machine translation (MT) matches with termbase data or other material. I was aware that Déjà Vu had been doing this for some time (this was easy to see in the wording of the proposal), but it turns out that Wordfast Pro, MultiTrans, and Star Transit are already using this important feature as well. And just as importantly,

Continued on page 24

Information and Contacts

The GeekSpeak column has two goals: to inform the community about technological advances and at the same time encourage the use and appreciation of technology among translation professionals. Jost is the co-author of *Found in Translation: How Language Shapes Our Lives and Transforms the World*, a perfect source for replenishing your arsenal of information on how human translation and machine translation each play important parts in the broader world of translation. Contact: jzetzsche@internationalwriters.com.

memoQ, XTM, and Wordbee are working on implementing it.

Interestingly, MemSource responded to that particular item with this: “None of our clients have asked us to implement fuzzy match repairs yet.” If I had any doubts about the importance of this whole exercise, that response convinced me of its value. Far too often, technology (and other) vendors focus on their existing customers rather than assessing what other potential users would like to see in a technology they might want to later adopt. At least now they all know what freelancers would like to see.

This is exactly why the developers of relatively new tools were much more open to suggestions. If you look through the Excel spreadsheet, you will find that Wordbee, XTM, and MemSource provided answers to the tune of “Great idea, we’ll work on it” much more frequently than more traditional companies. This is not to say these companies ignored our suggestions. Star, for instance, pleaded for more suggestions of regular expression uses and usage examples. (We had suggested that the use of regular expressions may be powerful, but also counter-productive because it essentially creates two different classes of users: the ones who are not afraid of using computer language to control their tools, and the rest of us.) In addition, Kilgray, Atril, and SDL showed themselves to be very open to creating user interfaces based on user profiles (something else we suggested). MemSource indicated that it wants to take that a step further by analyzing user actions over a period of time to create a user profile.

Overall, SDL had a slightly different strategy when responding to our requests. Rather than promising to do this or that, it pointed consistently to its OpenExchange app store, where third-party developers offer all kinds of apps, many of which offer, or could offer, the very features for which we asked. SDL’s Daniel Brockmann even coined a new term: TEnP (Translation Environment Platform). This is juxtaposed with TEnT (Translation Environment Tool)—the term we use to refer to what were formerly called computer-assisted translation tools. Daniel has a valid and interesting point. In many ways, SDL Trados Studio is (potentially) more able to respond to the needs of users by relying on third-party developers who recognize that very need and develop solutions for it. The OpenExchange has been around for long enough to have gained some traction, and it will be interesting to see whether other developers—I am thinking especially of Kilgray—come up with something comparable.

There were a couple of other interesting tidbits. When it came to the topic of terminology management, the responses of the tool vendors made it very clear that there really are two very different approaches to terminology: the glossary-like approach assumes that a termbase consists essentially of a source term/target term terminology list, whereas the much more complex terminology approach satisfies not only the immediate need of the translator but also that of the terminologist. Clearly, tools like SDL MultiTerm, Star TermStar, Kilgray qTerm, and LogiTerm fall into the

second category, while the various Wordfast products (Wordfast Pro, Classic, and Anywhere) fall into the glossary category.

The two (statistical) MT providers that responded (KantanMT and Tauyou) are now at least aware of the requests that freelance translators have for their tools, in particular that there is an immediate need for the MT engine to learn when the translator adds corrections in the post-editing process. (KantanMT already has an Instant Segment Retraining technology, whereas Tauyou is more cautious: “In some cases, [instant training] is possible, while not in others (or too dangerous).”

Beyond these and many other results (which you can find in the Excel spreadsheet), there were two more immediate action items.

Since there was a strong feeling among participants that it would be beneficial to have an exchange standard for keyboard shortcuts between different tools, I provided the tool vendors with a list of the 20 most important processes within a TEnT that would be helpful to be exchanged. I have also passed on a plea to revive the SRX standard, which is the standard that is concerned with exchanging segmentation rules between different tools. While the first of these requests is self-explanatory, the second was brought up in connection with the desire to have different sets of segmentation rules for different types of texts (and, of course, languages). It would be great if these kinds of rules did not have to be developed for each tool but could be shared between the different technologies. SDL Trados Studio does not currently support SRX, and this would be a very helpful change.

Oh, and there is one more outcome from my first webinar. We will have another comparable webinar in January 2015 to see what progress we have made with our requests and whether there are new ones that have risen to the top of the list. Stay tuned for the announcements. ■

ATA Member-to-Member Discount Program

To see what discounts your fellow members can offer through ATA’s Member-to-Member Discount Program, visit www.atanet.org/member_provider.