



Do-it-yourself machine translation at Autodesk

Autodesk, based in San Rafael, California, publishes 3D software for design, engineering and entertainment. Autodesk products are localized from English into as many as 20 languages. Mirko Plitt is senior manager of language technologies at Autodesk as part of the localization services team in Neuchâtel, Switzerland.

Thicke: Autodesk was one of the first enterprises to deploy machine translation (MT) internally. Why do you suppose actual MT deployments are so thin on the ground?

Plitt: My impression is that MT has gained a lot of traction in recent years, but you still have to choose between different enterprise systems and open source, and that is pretty involved no matter what. It's difficult to choose the MT approach that best meets your needs, and it's not easy to put together a team that has the required skill set to drive that decision.

Also, there is still a lot of resistance to MT. For instance, discussion is being hindered by subjective statements about MT quality and productivity. We have found that translators often have a wrong perception of their own productivity. Some participants in our productivity tests claimed they were slowed down by MT. But when we actually measured it, we found that they had increased their speed. Such subjective opinions contribute to a resistance to MT.

Thicke: Do you find translators by and large resistant to post-editing MT?

Plitt: No. Some translators even adopted MT before the enterprise. There are translators who are using Google Translate and finding it useful. When translation memory started, there also was a lot of resistance against it, not in all cases because translators didn't believe that it could work, but also because some had their own ways of dealing with repetitions. Even so, only a minority of our translators tell us they prefer post-editing, whereas most still would opt for translating from scratch – even



Mirko Plitt, Autodesk.

some who admit that they're faster post-editing.

Thicke: You are known in the localization industry for being ahead of the curve. Tell us about your MT innovations at Autodesk.

Plitt: Innovation may not be the word for it, but one thing that's unique about what we are doing is that we probably have the biggest Moses-only production deployment in the industry. So you could say we were one of the first ones to see the potential of the Moses MT toolkit to work in a business context. We have figured out ways of making it work, integrating it and building workflows around it. We have been running Moses in production at a pretty large scale for the last couple of years.

Thicke: How have you integrated MT into your workflow?

Plitt: We've created a Moses infrastructure by adopting scripts that were already out there to develop our own servers. We have around 30 servers running Moses in parallel with load balancing among them. Requests come in from either WorldServer or via Passolo clients and the servers deliver the translations. We basically leveraged our existing processes.

Thicke: What types of content do you translate using MT?

Plitt: We use post-edited MT for all our documentation and user interface text. We initially hesitated to use it for the user interface (UI), but as UI tends to have shorter sentences, the quality turns out relatively high.

Thicke: Why did you choose Moses?

Plitt: The question really is, why did we choose open source? The main driver for our decision to go to MT was to reduce our localization costs. It was also in the context of the financial crisis and pressure to reduce our translation spend – though of course translation cost is not the only driver of localization costs. So return on investment was important to us. We looked around for open source so we could invest without engaging significant amounts of money. And we moved relatively quickly from exploration to

Lori Thicke is cofounder and general manager of Lexcelera, cofounder of Translators without Borders and a member of the MultiLingual editorial board.

productivity experiments to deployment.

Thicke: You said you have 30 servers running Moses. What other resources did you need internally to deploy MT?

Plitt: We were doing it with limited resources, but we were able to leverage our existing team. At Autodesk we have a pretty sophisticated ecosystem with authoring systems and translation management systems that require a team to maintain. We were able to leverage that team and skills for our Moses deployment. If that had not been the case, we would not have gone down that path. When we started this initiative, not everyone agreed that this was the best way to go.

Thicke: What kind of companies would be as successful as you have been in a "do-it-yourself" approach to MT?

Plitt: Probably companies with a relatively strong software or IT background, as long as they identify language-related services as an important part of what they do. As a global software company, we have a lot of people who are excited when they can play around with new software tools, and we have people who are passionate about language. You need people who are passionate about being at the forefront of localization.

An open source system such as Moses requires additional energy for getting it to work. In terms of

resources and investment, we were lucky in the sense that we have a fairly large in-house localization team and part of that team is dedicated to maintaining complex systems. We found we could leverage that infrastructure to get this to work.

Thicke: So Moses was a good fit for Autodesk, with the resources you have available?

Plitt: For us, it was the right choice at the time. I'm not saying that open source is what everybody else should be doing. And we're not going to develop our own MT technology; we're not big enough to do what Google or Microsoft do. At the same time, we have the means and data to get more complex things to work that smaller companies or companies with less in-house expertise do not have the resources for.

Measuring productivity gains

Thicke: You and your colleague François Masselot published a paper about your approach to measuring post-editing productivity.

Plitt: In terms of innovation, something that distinguishes our approach to MT is that we have been quite thorough in measuring post-editing productivity. That really has helped us to take our MT as far as we have taken it. We know what's working without spending a lot of

effort discussing with vendors whether the quality is good, whether it slows translators down or not.

Thicke: Why did you set up this study?

Plitt: There is not much publicly available data on post-editing productivity. And as most of the data has not been acquired under controlled conditions, it was impossible to apply it to our specific situation. We felt that we needed that type of data, so we had to gather it ourselves.

Thicke: How did you measure productivity gains?

Plitt: We developed a simple web-based post-editing interface that allows us to measure the speed of a translator. We have done two tests so far, one when we deployed our model, and the second one to evaluate additional languages. Using this interface we were able to measure the time spent on a sentence-by-sentence basis. We conducted our tests over three days each time. The first time we had 12 participants working into four languages. Last year it was 32 participants and eight languages.

Thicke: What were your findings?

Plitt: In our first test, MT allowed all translators to work faster, though in varying proportions. On average, MT led to a productivity gain of more than 70%. The results of our second test were more mixed and on average lower because of the bigger distance between the English source and most of the new target languages.

Thicke: How do productivity gains like this translate into cost savings?

Plitt: Productivity gains and cost saving gains are not the same numbers. It is common for people to miscalculate the savings potential that you get from post-editing, thinking that a productivity increase of 50% would justify paying half price. Also, there will always be a margin of error. You should allow for some flexibility there. But most importantly, you should start from the principle that gains in efficiency must benefit everyone involved.

Thicke: Did you find the same productivity gains across all translators?

Plitt: There is a big difference between individuals, so one has to be careful when working with averages. We found that some translators improved their throughput by more than 130%, meaning they more than doubled their productivity, whereas others had much

wordbee
Manage. Translate. Save.

Awarded the Seal of e-Excellence at CEBIT

Want to see the future of Translation Technology?

- Simplify your translation process
- Collaborate in real time
- Implement within hours, not weeks

The future of translation technology **today**. Wordbee brings you all the functionality you need in an **approachable**, **affordable**, yet **powerful** enterprise translation management system.
For more information, and to download a free trial www.wordbee.com

Europe & ROW: sales@wordbee.com
USA : usa@wordbee.com

See us at Localization World, Santa Clara

more modest gains. The benefits from MT were generally greater for slower than for faster translators.

Thicke: So you measured translation speed before you measured post-editing speed?

Plitt: If you don't measure translating productivity, how can you compare it? It's too easy to assume 2,500 words a day. We thought it important to establish a benchmark for translation speed, then post-editing speed. Not all translators have the same pace, so if we are looking for productivity increases we thought we should not assume the standard productivity.

Thicke: Did translators diverge much from standard translation rates?

Plitt: Yes, a lot, from 360 words an hour to more than 1,000 – in translation, not post-editing. But you cannot extrapolate these numbers across typical work days.

Thicke: Did you measure anything else?

Plitt: We measured many different aspects of translation – for instance, the time spent on sentences and the number of words they contained. An optimum throughput appears to be reached for sentences around 25 words for translation and 22 for post-editing.

Thicke: What were the most surprising findings from your second productivity test?

Plitt: It was interesting to see that Chinese worked particularly well. Also Eastern European languages – we expected worse results in Russian, Czech and Polish, but they were not that far off from French, Italian, German or Spanish. It was, however, disappointing to find that Japanese and Korean are not good enough to go into production. We use Moses for gisting in those languages, but they're not yet part of our localization process.

Another interesting result is that we found for Portuguese, the productivity gain was actually quite low, and that was in contradiction to what you hear. We noticed that the translation throughput without MT is already high, even higher than French. Post-editing productivity is still reasonably good, but since the translation speed is already quite high, there is maybe only so much more you can gain. The feedback we got from the Portuguese translators was also positive.

Thicke: What other differences did you find?

Plitt: We don't rely on automated met-

rics to assess MT quality, but find them useful to track what changes people are making. We found that Chinese post-editors were moving words rather than changing them. They were focused on word ordering. In Russian, they spent their time changing word endings. Japanese post-editors did not make many more changes than Chinese ones, but it took them more time.

Also, as we were looking at the results we saw that people have different ideas of translation quality. This is true for both MT and translated text. We used five different scoring categories: unacceptable, poor, average, good and excellent. We also found that reviewers score differently. Korean texts were never judged as excellent, while Czech texts were excellent 70% of the time. Evaluations are subjective, so you should be careful when relying on human evaluations.

Thicke: Have you benchmarked Moses statistical machine translation (SMT) against any other systems?

Plitt: We compared Moses output to certain other products, two commercial hybrid systems – rule-based machine translation (RBMT) and SMT – which were trained on our data, and also Google Translate in the case of Japanese. We wanted to know where we stand in terms of commercial offerings. We did better than Google in post-editing productivity in Japanese, but Japanese is still not good enough to go into production.

When we tested the first hybrid systems for one language, we found that Moses was exactly on par with it, with exactly the same post-editing throughput. For the other language, Moses was clearly inferior to the second hybrid system and way superior to the first. In that language Moses was about 20% better than one RBMT hybrid but 15% slower than the other.

Thicke: Is a 15% improvement significant?

Plitt: If the MT is used raw in a customer-facing context, then the additional 15% may correspond to a significant improvement. But in terms of savings from post-editing productivity, it doesn't represent that much additional gain. We couldn't really get our vendors to drop their prices more.

Thicke: You have produced one of the industry's most extensive measurements of MT productivity gains. What about quality?

Plitt: To measure quality, we did sample checks of the translation at the end. In general, we found that the number of errors is slightly lower after post-editing than after translation – that is, we identified fewer errors on post-edited MT than on a fully human translation. And our reviewers couldn't tell the difference between the two. **M**

over 60 languages
on budget
beyond expectations

www.net-translators.com

sales@net-translators.com

Net-Translators
Your Vision. Worldwide.