# (More) Advanced Human-Computer Interaction for Translators

## A Conversation with Samuel Läubli and Nico Herbig

**Still on a high from the conversation I had a few months ago with Lynne Bowker, Vassilina Nikoulina, and Sharon O'Brien about "Women and Machine Translation"[1], I had another idea after I read about Nico Herbig's[2] work on a much more hands-on interface for post-editing (see links at the end of this piece) and Samuel Läubli's[3] ideas about more meaningful integrations of machine translation (MT) in the translation process. I reached out to them and—spoiler alert—it turned out to be another really great conversation.**

**Jost** Samuel Läubli and Nico Herbig, thanks for agreeing to have this conversation! I invited you because of your work on the usability aspects of translation environments. But before we start, would you like to quickly introduce yourselves?

**Samuel** Sure! I got really interested in MT when I studied computational linguistics in Zurich, so I went to the University of Edinburgh for my master's degree because I heard they had a fantastic research group. I graduated in 2014 and got a job offer from an American software company to train domain-specific MT systems. The goal was to increase translator productivity: the company was continually localizing more than 180 software products from English into 30+ languages at the time.

Initially, I was naive enough to think that I would be able to interact with the translators who used our pre-translations to get suggestions for improving the MT systems, but the company used multiple vendors who, often through additional sub-vendors, chopped up and distributed translation jobs to freelancers around the globe. I soon realized that MT quality wasn't the most pressing issue, even with the rather disfluent output that phrase-based systems produced at the time. At least that output could be deleted, compared to "exact matches" that were locked (i.e., uneditable) in the computer-assisted translation (CAT) tool.

I also saw how translators were tasked with translating strings from software user interfaces without any means of seeing what that interface looked like. MT wasn't (and isn't) necessarily good at translating "MENU," but neither were professional translators if they didn't know whether it was part of a navigation component or a description of an item in a virtual restaurant.

After two years, I was left with the impression that the way in which the translation industry builds and uses technology was just broken. So, I moved back into research—perhaps a bit naive again—with the hope

that I could gain a better understanding and then make things a bit better.

**Nico** I studied computer science and am currently working at the German Research Center for Artificial Intelligence (DFKI) in Saarbrücken, Germany. I spend most of my time on the Multi-Modal Post-Editing of Machine Translation (MMPE) project, which is funded by the German Research Foundation. Within the project, which is also the primary focus of my PhD, we investigate a broad range of explicit input modalities like handwriting or speech input to simplify the post-editing process. However, we also look at multi-modal implicit input, such as measuring pupil diameter or skin conductance to estimate cognitive load during post-editing. As this topic is at the intersection of human-computer interaction and language technologies, we work in tight collaboration with research departments lead by Antonio Krüger, chief executive officer of the DFKI and scientific director of the cognitive assistants department, and Josef van Genabith, scientific director of the multilinguality and language technology department. To retrieve input from domain experts, we ran our studies with professional translators in a user-centric approach.

**Jost** I think talking to both of you is turning out to be an even better match than I originally thought! Maybe I can drill a little deeper with each of you

about your projects first before moving on.

Samuel, let's start with you. In the *Routledge Handbook of Translation and Technology*[4], you co-authored a chapter with Spence Green, co-founder of Lilt, where you looked at the aforementioned human-computer interaction and evaluated a number of different ways to interact with the suggestions MT provides. This resulted in highlighting the need for adaptive MT systems. Another finding seemed to be that translators had a difficult time adopting a new working environment. Correct me if I'm wrong on that.

Just a few weeks ago you gave a talk that, according to the description below, covered the following:

*"Having faced tremendous resistance throughout the late 1990s and early 2000s, translation memories (TMs) are now considered indispensable productivity tools for professional translators. TMs are great at providing (partial) translation suggestions in the form of fuzzy or exact matches, but CAT tools are currently not too creative in utilizing these matches: they just display them to the user. In this talk, we take a look at how machine translation (MT) technology can ingest fuzzy matches to generate better and more domain-specific translation suggestions, or transform exact matches to comply with context-dependent linguistic requirements in the target language. We also discuss who's to blame about*

*the fact that these features are not yet available to professional translators."*

Naturally, we would all be interested in finding out who's to blame, but also whether you see widespread changes in a number of translation environments on the immediate or mid-term horizon. If so, what kind of changes, and will they be easily embraced by translators?

**Samuel** When you think about it, TMs produce so many inadequate suggestions. Even exact matches often aren't too exact. Think about linguistic properties that are implicit (or undefined) in the source but explicit in the target language, such as choosing the appropriate pronoun forms for informal and polite address when translating from English into German. For example, "you" will be translated as either "*Sie*" (formal) or "*du*" (informal). But an exact match for "You can win fantastic prizes" that results in "*Gewinnen Sie fantastische Preise*" ("Win you fantastic prizes") on the target side really isn't that great in an informal context. Since the level of politeness can be controlled[5] within, and partial translations incorporated into, neural machine translation (NMT)[6], these "exact matches" could be adjusted automatically. However, the NMT system will need an indication of the desired level of politeness to adjust the match.

Since most CAT tools don't integrate but merely connect to MT systems, they only

## Meet the Interviewees

### Samuel Läubli

Studied computational linguistics in Zurich, graduated in 2014 with a master's from the University of Edinburgh in artificial intelligence

Received a job offer from an American software company to train domain-specific MT systems

Worked for two years before moving back into the research field

### Nico Herbig

Studied computer science

Currently working at the German Research Center for Artificial Intelligence in Saarbrücken, Germany

Investigates multi-modal input (speech, handwriting, touch, etc.) for post-editing machine translation

**RR**

> "Since most computer-assisted translation tools don't integrate but merely connect to machine translation systems, they only send very basic information."
>
> **Samuel Läubli**

send very basic information, typically the source segment to be MT-ed alongside two language codes. If CAT tools were that loosely coupled with translation memories and term bases (TBs), you would never see features like real-time subsegment matching or predictive typing. So, if you're asking yourself why MT doesn't update as you edit a target segment or doesn't respect the very terms displayed in the CAT tool's terminology pane, it's not because MT can't do that, but because the CAT tool doesn't send it to the MT system.

**Jost** Let me interrupt you briefly (we're still interested in who's to blame!). I think what you just said sounds a little too pessimistic regarding TMs and the current use of adaptive MT. On the latter, I agree with you that much more needs to be done, but between SDL's adaptive MT, Lilt's MT, and ModernMT's implementation in a number of CAT tools, there's some progress, right? And your examples on the weaknesses of TM matches make sense on a theoretical level, but I'm not sure how much of that is practically applicable when using project-specific TMs.

**Samuel** I absolutely agree. There certainly is progress being made. However, the idea of adaptive MT is currently centered on post-editing. For example, if the engine suggests an incorrect term, it will (hopefully) learn to avoid that mistake once the user has corrected it. But why does the system need to make a mistake in the first place? If the user works with a TB, the MT engine could (and can[7], technically) use correct terms right away. If the user works with a project-specific TM, the MT engine could (and again, can[8], technically) learn from the exact and even fuzzy matches in that TM before it provides a suggestion for the first segment. MT systems shouldn't just adapt only when the user corrects mistakes. They should adapt to project-specific resources upfront.

So, if you're asking who's to blame that modern MT features aren't available to translators yet, it's clearly the CAT tool manufacturers. I don't really see changes in widespread CAT tools on the horizon, which really puzzles me. Then again, this may be a chicken-and-egg problem: are these features unavailable because translators aren't asking for them, or are they not asking for them because they've never seen them implemented in a tool? Personally, I could well imagine that translators would embrace changes like neural fuzzy repair and NMT output toggles for things like honorifics to express politeness or other linguistic aspects as long as they're easy

to use and well visualized. But with fundamental design choices dating back to the 1990s, widely used CAT tools aren't exactly a prime example of effective, user-centered data visualization.

**Jost** That sounds like a perfect segue into what Nico is doing with his concept of the Multi-Modal Post-Editing of Machine Translation.

Nico, I got really excited about your post on Kirti Vashee's blog[9] and the links to articles and videos you provided at the end of the post. A number of years ago, I wrote an article about a tactile approach to translation[10], and while I certainly didn't have all the tools in mind that you've made available, this was very similar to what I was thinking. The MMPE project includes computer interaction via the keyboard/mouse, touch, voice, and handwriting—what a great idea! A few questions come to mind. Is it correct to say that this seems to be quite language-specific since voice and handwriting only work for a select number of languages? And, will your prototype make it into existing translation environments? Or, to rephrase the last question, what would have to be done for that to happen?

**Nico** Indeed, post-editing requires very different interactions than traditional translation. We've seen a change from "production," where all text has to be entered manually, to "supervision," where the
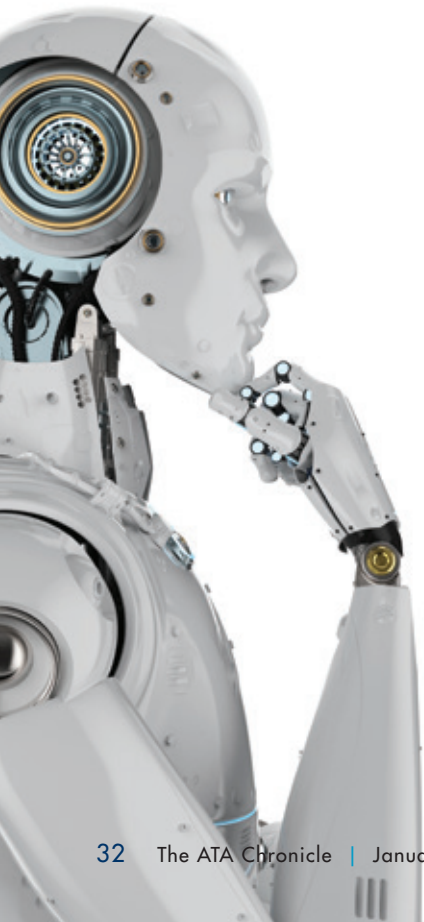
task changes to capturing and correcting mistakes, as well as manipulating and recombining useful suggestions. Naturally, this change already started with TMs, but the better MT gets, the more we move away from the production paradigm to supervision and collaboration with the machine. For example, we don't question that a mouse and keyboard are very good tools for production. However, we believe that other modalities could be very helpful for the changed interaction pattern in post-editing—not as a substitution for mouse and keyboard, but as a complement. This is what we've been exploring in the MMPE project. For example, we found that a digital pen and finger touch input are very well suited for deletion and reordering operations.

Regarding your question on language support, I would say that the transcription of handwritten text or speech input works well with many languages. You just need to exchange the underlying machine learning model with one that was trained on data in the target language. One would, of course, need to also define the speech commands in other languages, but we tried to keep our code rather flexible by having the commands in separate files outside the source code. Further studies would be needed to say for sure how well it works with other languages and to explore changed interface layouts for, say, right-to-left or logographic languages.

What will need to be done to integrate such modalities

into existing CAT tools? That's a good question. It depends a lot on the input modality. Most computers have an integrated microphone, and now with the pandemic, many people probably also own a headset. So, for speech input, the CAT developers can basically start integrating dictation and also speech commands. For example, memoQ is already offering an iOS app that transcribes your speech input and sends it to their CAT tool.

Pen and finger touch input could also be integrated rather soon. Many laptops now have touch screens, and tablets are becoming increasingly common. In general, I believe that with higher quality MT output, post-editing on tablets might become doable, especially with good handwriting, touch, and speech support. But I assume the market is currently too small for CAT companies to invest in this. Other modalities we're currently exploring, such as eye-based interaction (e.g., you look at a word and say "delete") or mid-air gestures (e.g., point at a word and do a hand gesture to delete), are interesting from the research perspective, but no one has these tracking devices in a standard office. So, I believe it will take a long time until we see something like this in commercial CAT tools, if at all.

**Jost** That makes sense, but what I had in mind was whether it would somehow be possible to use MMPE or aspects of it and essentially connect it to existing translation

> "In general, I believe that with higher quality machine translation output, post-editing on tablets might become doable, especially with good handwriting, touch, and speech support."
>
> **Nico Herbig**

environments via an application programming interface or some other mechanism. Otherwise, I think that translators would have to wait an awfully long time for existing tools to implement it. Also, I'm not sure I completely agree that the interface you're proposing isn't necessarily suited for TM-based work.

**Nico** Integrating aspects of the MMPE project into existing CAT tools is probably not that easy, which is why we also chose to start from scratch. As Samuel already said, many CAT tools still follow outdated design patterns, making them look more like spreadsheets and not like modern websites and applications. For example, consider handwriting. We rely heavily on the MyScript application programming interface here, which is working great and could also be integrated into existing CAT tools. But if you try to handwrite into the small space that most translation environments provide for

> "MT systems shouldn't just adapt only when the user corrects mistakes. They should adapt to project-specific resources upfront."
>
> **Samuel Läubli**

editing, there's no way that it's going to work well. The same holds true for touch reordering, eye tracking, or mid-air gestures: stronger interface changes are required. If you just try to squeeze it into existing tools, I believe the user experience will suffer so strongly that you'll stick to your mouse and keyboard. However, we just open-source released MMPE on Github in the hope that people will try it out and give us additional feedback. (Visit: https://github.com/NicoHerbig/MMPE) Who knows? Maybe even some CAT developers might decide to build certain aspects into their tools, which I would love to see.

I don't disagree regarding your comments about TM. The newly explored modalities might also help with TM-based work, especially when the match scores are high. We just haven't tested that, so I can only guess here. For our study, we also chose to pre-fill the editing box with the MT suggestion. If you would do the same with highly matching segments from TMs, it should basically be the same. I just believe that the new interaction possibilities mostly make sense to quickly fix a variety of smaller changes.

In TMs, one part of the sentence might be perfectly matched while another part isn't matched at all. If you then need to insert 10 words, typing or maybe dictation are great, but handwriting and finger touch input might be less helpful in this setting. However, for low-quality MT, you might want to re-translate larger portions of the segment as well, where again typing and dictation are probably better than other modalities. So, I would rather say that the new modalities show their benefits for highly matching segments from TM or high-quality MT because they allow you to very quickly change the few remaining mistakes, like quickly grabbing a few words and moving them somewhere else. Here, you produce less and supervise the machine more.

**Jost** Why are we stuck with the concept of post-editing one MT suggestion? Why are we not, for instance, looking at how we could harvest several MT suggestions simultaneously by using mechanisms like auto-suggest (which would mean that we don't even have to look at the various MT suggestions—we just see what matches our keystrokes)? Also—and I think that Samuel already alluded to this—why don't we look at a closer integration of our three most important resources (TM, MT, and TBs) with each other and achieve better results that way?

**Nico** Indeed, we've been asking ourselves the same thing. Therefore, we're currently adding multiple MT proposals in MMPE, where we penalize similar MT outputs. No one wants to see almost the same suggestion three times, since a normal post-edit of a single suggestion would be quicker than that. But we believe offering multiple high-quality and diverse outputs might help. Especially for shorter sentences, a translation very similar to what you aim for is most likely among the suggestions. For long sentences, however, mentally processing multiple suggestions might just take longer and be more cognitively demanding than directly post-editing a single suggestion. At least this is what I would expect now. We'll know more when we run a study on this.

In parallel, we're also looking at more interactive ways to post-edit, where you click on parts of the MT suggestion you don't agree with and get alternatives. I believe that this, in combination with touch input and handwriting, could really be a nice approach to post-editing and could also work well on tablets.

**Samuel** Nico brings up an important point. Showing too many alternatives could lead to cognitive friction. The prototype of Lilt[11] offered both what you refer to as auto-suggest—a single suggestion for word, phrase, or sentence completion that adapts to the

user's input, rendered as ghost text—and multiple word or phrase translation alternatives presented in a dropdown menu. The latter were used so rarely that they didn't make it into the final product.

However, regardless of how suggestions will be visualized, it's vital that they combine all the resources available to translators: TMs, TBs, and MT. Jost, I like your idea of using multiple MT engines for auto-suggest. At TextShuttle, we're using a technique called Diverse Beam Search[12] to produce diverse translation variants with a single engine. The rationale is that even MT engines from different providers typically produce very similar translations for many sentences. And because NMT systems always generate multiple translation variants behind the scenes as they generate a target sentence, enforcing variability comes with almost no computational overhead. It's easy to generate multiple translation variants for a given source sentence with a single NMT system, but as long as CAT tools don't query and visualize them, there's no way for professional translators to take advantage of them.

**Jost** Thank you so much for this, Nico and Samuel! It feels like we could continue talking about this for a long time. But it seems even more important at this point that translators start considering some of the things we've discussed, and that tool developers start a dialogue among themselves to see whether they can implement some of the changes mentioned here. Or maybe there's even a team of developers that will read this and say, "Yeah, there are so many good ideas in this that I think I can build something new and interesting and become super-rich selling it!" (Well, the latter is not going to happen, but the former might).

My biggest take away from our discussion is that just because we think we've found a widely accepted way of working, doesn't mean that we couldn't and shouldn't be questioning it on a continuous basis and, well, making it better. All this excites me greatly, partly because it goes to show that professional technical translation (and by "technical" I mean to include virtually everything non-literary) is alive and well, even as research tries to find better ways to facilitate it. ATA

**Jost Zetzsche, CT** is chair of ATA's Translation and Interpreting Resources Committee. He is the author of *Translation Matters*, a collection of 81 essays about translators and translation technology. *jzetzsche@internationalwriters.com*

Remember, if you have any ideas and/or suggestions regarding helpful resources or tools you would like to see featured, please email Jost Zetzsche at **jzetzsche@internationalwriters.com**.

**NOTES**

1 Zetzsche, Jost. "Women and Machine Translation," *The ATA Chronicle* (November/December 2020), 25, http://bit.ly/women-MT.

2 You can find Nico Herbig on Twitter at https://twitter.com/nico_herbig.

3 Samuel Läubli is on Twitter as well: https://twitter.com/samlaeubli.

4 O'Hagan, Minako (Editor). *Routledge Handbook of Translation and Technology* (Routledge, 2020), https://bit.ly/Routledge-translation.

5 Sennrich, Rico, Barry Haddow, and Alexandra Birch. "Controlling Politeness in Neural Machine Translation via Side Constraints." In the *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Association for Computational Linguistics, 2016), http://bit.ly/politeness-NMT.

6 Bulte, Bram, and Arda Tezcan. "Neural Fuzzy Repair: Integrating Fuzzy Matches into Neural Machine Translation." In the *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, 2019), http://bit.ly/fuzzy-matches.

7 Sennrich, Rico, Barry Haddow, and Alexandra Birch. "Controlling Politeness in Neural Machine Translation via Side Constraints." In the *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Association for Computational Linguistics, 2016), http://bit.ly/politeness-NMT.

8 Bulte, Bram, and Arda Tezcan. "Neural Fuzzy Repair: Integrating Fuzzy Matches into Neural Machine Translation." In the *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, 2019), http://bit.ly/fuzzy-matches.

9 Herbig, Nico. "The Evolving Translator-Computer Interface," *eMpTy Pages* (October 21, 2020), http://bit.ly/Herbig-eMpTy-Pages.

10 Zetzsche, Jost. "Getting Physical," *The ATA Chronicle* (August 2013), 29, http://bit.ly/getting-physical.

11 Albarino, Seyma. "New Research Flips the Script on CAT Tools—Literally," *Slator* (November 25, 2020), http://bit.ly/Slator-CAT.

12 Ashwin K., Michael Cogswell, Ramprasath R. Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. "Diverse Beam Search: Decoding Diverse Solutions from Neural Sequence Models," http://bit.ly/diverse-beam-search.