

Translation technology comes full circle

Jost Zetzsche

I'm a historian by training who writes about the latest trends in translation technology. Some may see this combination as worldview schizophrenia, a perspective caught between the past and the future. I prefer to describe it like this: I study the past to gain a better understanding of the present and, hopefully, a better handle on the future.

With that in mind, allow me to give an overview of the short history of translation technology, especially the kind we find in computer-aided translation (CAT) or translation environment tools (TEnTs). We'll then look at what's happening presently and take a brave glance into the future.

In the 1950s and 1960s, translation technology was synonymous with machine translation (MT) or, more accurately, the idea of what MT would be able to do "in five years." As it became apparent that this five-year prediction was an ever-moving target, funding dried up and only a handful of academic and commercial attempts soldiered on.

Instead, attention turned to terminology in the form of dictionary applications and terminology tools. The first stand-alone terminology tool for the PC, called MTX, was launched in 1985 using a precursor to today's terminology exchange TBX format. Terminology management continued to develop (Trados' first commercial application was MultiTerm in 1990) as another technology received increasing attention from developers. Various developers were beginning to use a low-level form of MT called translation memory (TM), and they all released the first version of their products around 1992: STAR released STAR Transit, IBM launched its Translation Manager, TRADOS

introduced the Workbench product and Atril offered the first Windows-based commercial product, Déjà Vu, in 1993.

The stakeholders in the translation industry reacted to these releases in various ways that had a tremendous impact on the further development of the tools and their placement: translators largely rejected the new technology. Some language service providers (LSPs) used it as a competitive differentiator. The vast majority of translation buyers simply didn't even take notice – with the exception of the terminology components that were of interest to their terminologists.

The result? With the exception of Déjà Vu, the price of these early tools was so high that they were virtually unobtainable by translators. The tools' project concept was structured to match the needs of LSPs, and the terminology components were developed into high-powered applications with the needs of large corporations in mind. The following years produced next to no development of translation features, except the support of more languages with the advent and support of Unicode.

In the meantime, Déjà Vu and some newer tools, including Wordfast, had been targeting the freelance translator market relatively successfully, paving the way for other tool vendors to offer less expensive translator versions. In addition, the old business model of LSPs financing the expensive Trados or Transit translator licenses proved to be unsustainable. As a result, the use of CAT tools in some form or another became the rule rather than the exception, both in the freelance community and among LSPs. And more sophisticated customers were starting to expect differentiated pricing on the basis of TM leverage.

At the same time, a number of new players entered the market. Since translation buyers had become aware that there could be substantial savings by using technology, companies such as Uniscape, and later Idiom and GlobalSight, began offering large translation management systems (TMSs) that were first grandly called globalization management systems. Only later were they more aptly and humbly dubbed TMSs.

These large systems provided the workflow automation and transparency that translation buyers were looking for. Interestingly, the roles were suddenly reversed. The LSP was increasingly ceding control of the process – and to some degree the pricing – to the translation buyer. Naturally, at some point



Jost Zetzsche, MultiLingual editorial board member, is a translator and translation technology consultant. He is the coauthor of Found in Translation: How Language Shapes Our Lives and Transforms the World, to be released in October and already available for preorder on Amazon.

technology vendors also started to offer TMSs for LSPs, especially Trados/SDL (which had swallowed both Uniscap and Idiom), Across, memoQ and others.

And the actual translation technology? It stayed virtually the same throughout. Minor improvements were made with context-sensitive matching and some improved quality assurance processes, but the underpinnings of the foundational TM and termbase modules remained where they had been a decade earlier.

Then, soon after the turn of the century, something reawakened that many had written off as a productivity tool for the translation industry: MT. Three things prompted this resurrection. First, the events of 9/11 and its aftermath highlighted the desperate need for automated translation and opened subsequent government funding. Second, statistical machine translation (SMT) was “discovered” as a possibility to create MT engines relatively quickly for a large variety of languages. Thirdly and maybe most importantly, the concept of quality was replaced with usability – a more user-driven and much more variable concept of what the translated text needed to look like.

Many different MT applications have emerged in the last few years, from raw output of a trained MT engine for knowledge bases, to post-editing MT output in various degrees, to the increasingly specialized training of MT engines. But MT’s most surprising effect may have been the transformation of CAT tools’ stale translation features.

The most obvious change was the addition of tool-internal connectors to online translation tools such as Google Translate or Microsoft’s Bing Translator, or other commercial and open-source machine translation systems. Virtually all tool vendors quickly implemented these. The logic behind the reunification of these long-parted siblings of MT and TM goes something like this: if no match in the TM is found, propose a match from an MT engine that then will have to be edited like a fuzzy match. There’s nothing too exciting in that, but in combination with the next development, something truly new was created – hold that thought for a second.

The value of TMs was also re-examined. With the increasing necessity to feed data to SMT engines, the need to subsegment existing TMs became a primary concern. This was especially

voiced and championed by TAUS. With the exception of a small number of tools, most importantly MultiTrans, existing technology only gave manual access to data below the level of a complete segment, typically a sentence, even though it had long been obvious that below the sentence is where the true linguistic treasure of TMs was buried.

Responding to the increased pressure of their user groups, most tool vendors have now started to dig deeper and give translators materials at their fingertips that had always been there, just not in an accessible way. It’s fascinating to watch this evolution. While many of the earlier paradigms of finding whole-segment matches and using a separate terminology database as a reference were virtually uniform across the different technology solutions, the subsegmenting approaches are almost as varied as the number of tools supporting them. Because we are still in the infancy of these developments, even more creative approaches will likely be put forward.

One sign for how new and disruptive this concept of subsegmenting is can be seen in the fact that most tools have not yet completely grasped that this new approach to data brings forth two major paradigm shifts. First, the newly required quality control of TMs needs to become much more sophisticated. The old model of garbage-in/

garbage-out has been replaced with garbage-in/every-little-piece-of-litter-in-the-garbage-on-the-carpet out, which asks for much more in-depth pruning and control of TMs. Second, the concept of terminology has shifted, with terminology now automatically being extracted from TMs. While the specialized termbase applications of most TEnTs will not just go away, their usage and design will have to adapt to the new reality.

Let’s return now to the introduction of MT output into the TEnT workflow. In combination with subsegmenting, MT will now start to play a significantly much greater role in the normal, non-MT-centric project. MT will provide those subsegments that cannot be unearthed from the TM. Depending on the quality of the underlying MT engine, this has the potential to give an immediate boost to translation productivity, with MT as one tool of many in the translator’s TEnT.

Translation technology is poised to come full circle. MT is about to return as a productivity tool. And those tools that started out as translation tools but lost their true calling are reembracing their identity. In the process, they’ve rediscovered their formerly evil sibling: MT.

Will they live happily ever after? Only time will tell. But as a historian and a futurist, I’m watching the story unfold with rapt attention. **M**

+ Software localisation.
+ Web site localisation.
+ Technical and general translation.
+ Interpreting.
+ Third-party translation review.
+ Style guide creation.
+ Desktop publishing.
+ Linguistic advisory.
+ Terminology and document management.
+ Technical writing.
+ Multimedia translation.
+ Web site design, development and internationalisation.
+ Linguistic, typographic and style revising and review.
+ Video and audio tape transcription, including studio dubbing and voice-over.
+ Training on translation and localisation.

HERMES, THE SPANISH EXPERTS IN SPANISH TRANSLATION AND LOCALISATION:

PLEASED TO MEET YOU.

HERMES
 TRADUCCIONES Y SERVICIOS LINGÜÍSTICOS

Founded in 1991
 Córcega, 6, portal 2 - 3.º I. Edificio Prisma,
 28230 Las Rozas, Madrid - SPAIN. Phone: (+34) 91 640 7640

Parque Tecnológico de Andalucía Juan López Peñalver, 17, 3.º, ofc. 6
 Edificio Centro de Empresas 29500 Campanillas, Málaga - SPAIN
 Phone: (+34) 952 020 525

TÜV Rheinland® CERT
 Certified service of translation
 EN-15936

TÜV Rheinland® CERT
 ISO 9001

Fundéu
 CERTIFICADO DE CALIDAD
 TRANSLADORA

Email: hermestr@hermestrans.com www.hermestrans.com