

Translation tools come full circle

Jost Zetzsche

I recently spent some time on the phone with a multilanguage vendor client who was desperate because he had to translate repetitive text without a translation environment solution in place. When I got off the phone, I realized how far we've come in our industry. I hadn't had a call like that for months! Although a handful of companies still steadfastly refuse to employ technology, these are in the minority, and calls like the one that morning are few and far between.

Translation environment solutions – a term that is much to be preferred to inaccurate and limiting terminology along the lines of translation memory (TM) or bitext tools – have reached a point of great maturity in terms of usage and product development. Today's translation buyers assume that language vendors use at least one solution. At the same time, tool developers are forging toward new horizons in the development of their tools.

While "traditional TM technology" dates back less than 15 years (TRADOS, STAR and IBM released their TM components in 1992 and Atril the first Windows-based product in 1993), its roots can be traced back long before that. As Jaap van der Meer points out in "Different Approaches to Machine Translation" (*MultiLingual Computing & Technology* #71, April/May 2005), TM is just a sub-category of machine translation (MT) and thus dates back to as early as the 1950s. Van der Meer rightly points to the root cause that makes us forget the close relationship between MT and TM: "The marketing message was tuned in to what the

professional translation industry wanted to hear: 'Forget about MT; it doesn't work. Instead, use our TM product because it leaves you in full control of the process.' The message worked well. Within a period of 10 to 15 years, TM products found their way to the workstations of more than 50,000 translators in the world. But the message has also caused a sort of 'cognitive disorder' in the translation industry, namely that TM is good and MT is evil, foregoing the fact that TM is just a new variant of MT. . . ."

More on MT later.

The last couple of years have been nothing short of fascinating in the translation tool industry. I won't bore you with yet another rendition of the "what the TRADOS acquisition by SDL means for our industry" litany. While this is an important development with a significant (and I think mostly positive) impact on the industry, there's a lot happening besides that. Let's start with the "failures."

IBM put its Translation Manager to rest in 2002. ALPNET/SDL gave up on its corporate TSS/Joust and Ampran tools as stand-alone tools. Lionbridge made LionLinguist/ForeignDesk open source and in effect gave up on it as well. And Cypresoft's Trans Suite 2000, the latest casualty, gave up in 2004 because, according to the developer, "knowing that almost 70% of our Trans Suites that are currently running around the globe are illegal, I can say that a lot of you helped us to reach the point where we had to close the company because we had lack of funding."

On the other side of the business, numerous tools have entered the market in the last couple of years (across, Heartsome and Fusion, among others) or are going to enter the market in the next two to six months. I know at least three tools in that category and wouldn't be surprised to see even more. And, by the way, this is where I see one of the most positive aspects of the TRADOS



Jost Zetzsche is a translator (English-to-German), a consultant in localization and translation, the author of The Translator's Tool Box for the 21st Century and one of the founders of TM Marketplace.

acquisition: it has invigorated a competition that previously was passive to the point of standstill.

There are numerous ways to classify the different commercial translation environment tools: TM/terminology database technology, translation interface, network support, supported platforms, APIs and so on. One helpful way to look at them is that some tools are almost exclusively aimed at the concept of "TM" (WordFisher, Wordfast, MetaTaxis), while others try to create a "translation environment" beyond that with features such as:

- advanced terminology database management
- processing of complex file formats
- analysis and quality assurance
- networkability and sharing of databases
- project management capabilities

It's exactly in this realm of a more complete translation environment that further development is happening and will continue to happen. Following are some of the primary areas where I see and hope for continuing developments.

Terminology management

Terminology management has to become stronger, and this is more a plea than a prophecy. More than a decade of steadfast refusal on the side of most translators to adequately use the terminology component of most translation environment tools (yes, I realize that there are exceptions to that, but they remain exceptions), coupled with often awkward methods of entering and retrieving data, has given way to some more encouraging developments. The Canadian line of bitext tools – MultiTrans, Beetext Find and LogiTerm – as well as a TRADOS tool and an SDLX tool allow for the harvesting of terminology from a bilingual environment, thus making it possible to (almost) skip the translator in the process. Other developments are an increasing number of terminology components that can be accessed online and from which terms can be entered seamlessly into the translation.

Déjà Vu is so far the only tool that allows for "assemble" processes, that is, the constructing of a "translation" from terminology and other segments if no direct match is found in the TM. It escapes me why no other tool has a comparable feature.

And last, tool vendors have been shockingly slow to implement TBX, the termbase exchange format. Only SDLX and Heartsome have been exemplary in supporting this format. When others follow their lead, terminology exchange beyond bilingual term lists may actually become possible.

Content management

For the service provider, content management could easily be the most daring development, and it is something that has been in the works for a long time.

Traditionally, there has been a clear separation between content creator and translation provider. Now this separation is in the process of crumbling. TRADOS/SDL already has a number of partnerships and integrations with content management providers; STAR Transit has an integrated authoring system for its corporate solution; and the newly released SDL AuthorAssistant and Iterotext's Authoring Coach TMX are simple tools that connect technical writers to a TM database so that the content will

be written with the greatest possible number of matches in the translation process.

The challenge to service providers is obvious. They either need to broaden their service portfolios to include authoring services at a much higher level or have the translation buyer much more directly involved in TM management.

Workflow components

The need for workflow/project management solutions has finally become apparent to the language industry. While most of the large service vendors responded to this long ago by purchasing or developing their own tools, many of the smaller and mid-sized companies are still looking for ways to automate project management and translation processes. The two lines of tools that are offered in this area are either tools such as LTC Organiser and]project-open[that support and automate project management on the financial and personnel level (bidding, assigning personnel, tracking progress, invoicing) or tools that are closely integrated in the translation environment (such as TRADOS TeamWorks or Idiom WorldServer). It remains to be seen whether there will ever be a tool that truly combines these aspects or whether the API of tools will be so readily accessible that an integration can be easily done. But there is no doubt that the market is ready for it.

Another aspect of managing the workflow is online access to TMs and terminology databases, which, of course, is already a reality in many workgroups. More needs to be done, however, especially on a noncorporate level as shown in some of the newer tools such as Fusion or across (or Logoport, which was bought by Lionbridge). These tools have recognized the need for easily accessible online access and have built their tools around this core feature.

MT components

Some of the attempts at a renewed fusion of MT and TM on a commercial tool level have failed (think of SDLX's MT component a few versions ago). However, on a service level they have become reality, especially among some of the larger service vendors and translation buyers.

Tools such as Déjà Vu have long used algorithms that on a sub-segment level attempt MT-like processes. While these could and should be enhanced, they will become more prominent in other tools with much larger pre-configured language-specific MT engines behind them.

Open source

Though open-source applications are still somewhat at the fringes in the world of translation environment applications, there's enough out there to make one sit up and take notice. Besides Lionbridge's first attempt at releasing ForeignDesk as an open-source application, OmegaT has also been around for a number of years and translates HTML, text and OpenOffice .org files. Sun's newly released Open Language Tools converts SGML, XML, HTML, OpenOffice.org and a number of software development formats to XLIFF and translates those within its own TM environment. And just weeks ago ENLASO also released its tools (Rainbow for the conversion of a large number of files into

Terminology management has to become stronger, and this is more a plea than a prophecy.

TM-applicable formats and Olifant for the maintenance of TMX databases) to the open-source community.

Though it is a project management application rather than a translation environment tool, [project-open] is commercial open-source software for the translation industry, with revenue primarily achieved through implementation, consulting and support. This is certainly an interesting approach, and it wouldn't be too surprising if translation environment tools followed this pattern as well.

A different attempt at pricing for translation environment tools is offered for Lingua et Machina's Similis. Though it's now possible to directly buy the tool, users can also buy and update "cartridges" for a certain number of translated words instead.

XLIFF

After endless (though important) discussions, the TM eXchange format (TMX) has not yet had the impact in the translation tool market that many had hoped

for. However, XLIFF may actually now be poised to finally break open the market. A number of tools (Heartsome, Open Language Tools) already rely exclusively on XLIFF as their translation interface format, SDLX supports it, and ENLASO's now open-source product, Rainbow, supports it as one of its main conversion formats. The power of XLIFF lies in the fact that, if implemented adequately, translatable content will be completely exchangeable between supporting tools, giving both vendor and buyer a much higher-level exchangeability of data.

TM exchange

As one of the founders of TM Marketplace, I admittedly have a personal stake in this area, but I can't but believe that data exchange on a TM level may have a greater impact on our industry than any of the other areas that I've mentioned.

Think of how the situation today has radically changed from that of five or ten years ago when the discussion

first began. TM data in TMX or other exchange formats long ago went beyond the traditional perfect and fuzzy match schemes in TM tools. Tools now are specifically designed to harvest terminology from these databases; TMX databases are used for the authoring process; and MT engines require bilingual or multilingual data to enhance their algorithms. And the amount of data that is held by translation buyers and the cost of the investment into creating that data has reached a magnitude that requires new ways to benefit from that process. Whether it will be the licensing approach to TM data that my company proposes or some other paradigm, TM data will be exchanged in an increasingly regulated manner within the next couple of years.

Conclusion

While few translation tools focus only on the TM aspect, there is an even stronger move toward tools or tool suites that create an environment which covers many more aspects of the translation process. These include more sophisticated terminology management features, project management capacities, MT plug-ins and content creation/management. This creates a higher level of sophistication for the service vendor, but it also presents rather unique challenges that may include a shift in expertise and service portfolios.

In the area of more openness and exchangeability, it isn't too surprising that the open-source movement is leading the way. TBX and XLIFF represent the opportunity to forge ahead and realize what was started with TMX years ago: a meaningful and realistic independence of tools and an increasing focus on features rather than marketing prowess. TM data exchange approaches exchangeability from a different angle: the actual content. The impact that this may have on the language industry and the translation tool industry could be enormous.

This article was originally titled with a line from Bob Dylan's song "The Times They Are A-Changin'." And while that is true in a certain sense, it's just as true that we are simply coming full circle, back to the visions of early computational linguists half a century ago who trusted the computer and its ability to translate perhaps a little too much for their time. Be that as it may, these are interesting times, and I can't think of any industry that I'd rather be part of. **M**

We speak
your
language.

Everywhere.

ADAPT
LOCALIZATION

www.adapt-localization.com
adapt@adapt-localization.com