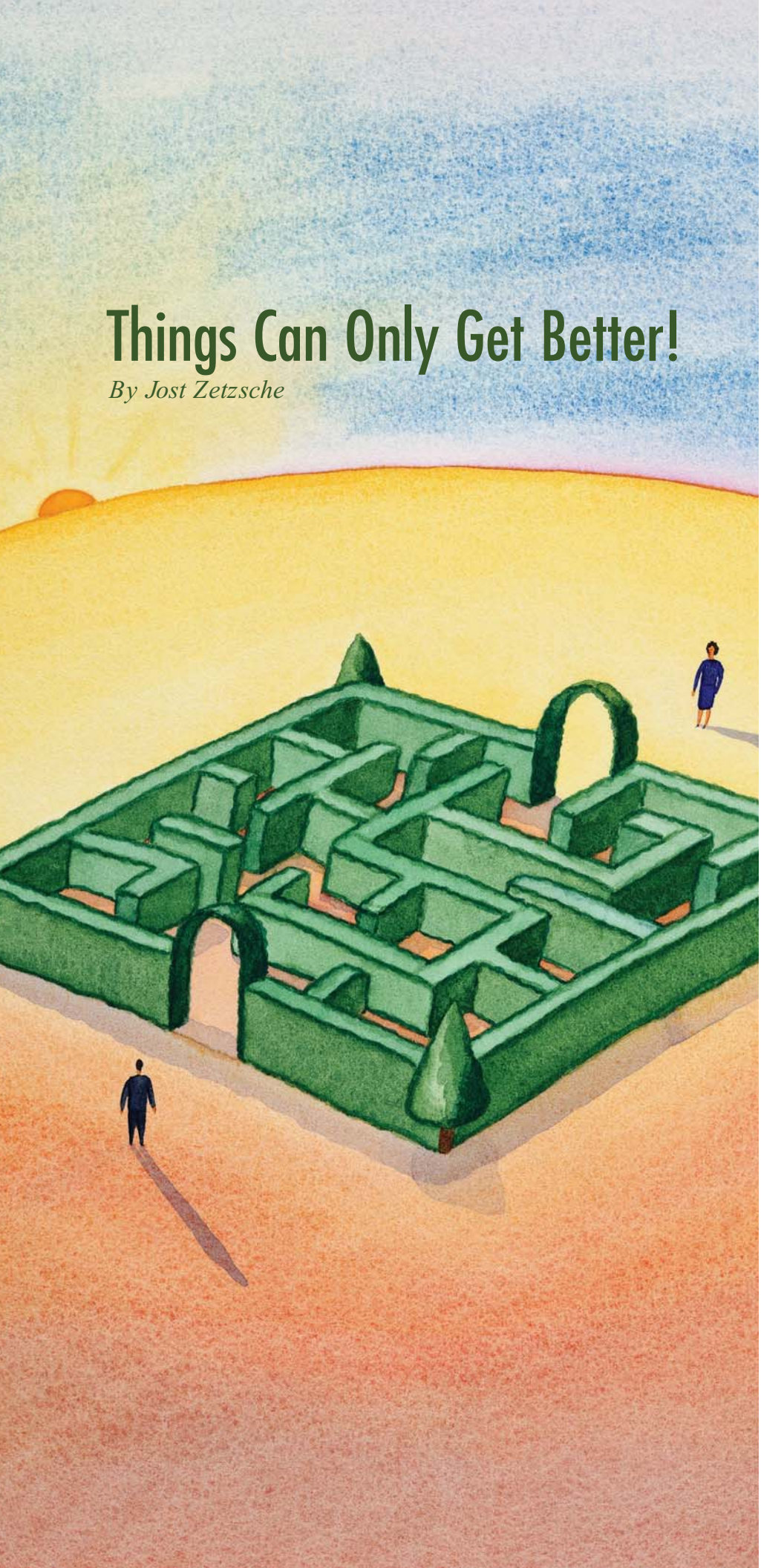
Things Can Only Get Better!

By Jost Zetzsche



The market for translation environment tools is as crowded as ever. For every tool that falters, two new ones seem to show up. With so many choices, it seems logical that it would be easy for us to choose a tool that has all the features we have ever dreamed of, right? Wrong! When I sat down to enumerate my dream features, I came up with a whole list that are either missing or woefully underdeveloped. I will name just a few of those here.

Open Standards for Server-Based Systems

We have been talking about translation exchange standards forever: standards like TMX (Translation Memory Exchange) to exchange translation memories, TBX (TermBase eXchange) to exchange terminology databases, and XLIFF (XML Localisation Interchange File Format) to exchange the actual translation files. And there are others, such as SRX (Segmentation Rules eXchange), GMX (Global Mail Exchange), and xml:tm (XML-based Text Memory). These standards are really important for facilitating the exchange of translation projects between different versions of translation environment tools. Traditionally, the focus of these standards was the desktop, so that projects could be exchanged from one system to another. With the advent of server-based projects, however, assets (translation memories and terminology databases) are now stored online with real-time access to a translation professional. Suddenly, things are no longer quite so clear-cut. While in certain cases it is possible to download some or all of the data into one of these exchange formats, it defeats the purpose: the real-time collaboration between various translation professionals becomes impossible.

Ergo, the data is only accessible with the originating tool (which could be Trados, Déjà Vu, Across, memoQ, or any of the other tools that support this kind of workflow).

What is needed, therefore, is a standard that allows for any translation environment tool to hook into that workflow. In other words a tool that will query the underlying translation memories and terminology databases that remain on the server and write data to them as the project continues, thus enabling other folks who are working on the project to benefit from that data in real-time. Sound unlikely? Agreed. But it is no doubt possible if there is a real will on the side of tool makers. Is there? No, not really, and this is true for big and small vendors alike. Yes, they all support the above-mentioned translation data standards in some way or other and communicate this very publically and proudly. But with the future and present clearly going in the direction of server-based projects (at least for large projects), many vendors are only too happy to have found this new way of capturing their technology. You find that frustrating? I do, too. There is actually an e-mail address that ATA Director Alan Melby and I set up awhile back to allow you to voice your concerns so we can pass those on to the vendors (datastandards@atanet.org). Feel free to let tool makers know your thoughts by e-mailing this address, or talk to them directly. Are they going to listen? I think that depends on how many of us are talking.

Full Fruition of XLIFF

XLIFF is probably one of the best things that has happened to our industry in the past few years, and, ironically, it almost happened by accident. XLIFF really was only meant to support the translation process of software development formats, but it has long since morphed into an exchange format for all kinds of files. In fact, many tools use XLIFF internally for all their supported file formats.

It is a really cool standard. Yeah, it is great that we can exchange translation memories and termbases with standards, but how much greater is it to exchange the actual translation file between different tools? There is no more awkward converting of source files or the need for filters to support file formats of other translation environment tools if everything goes through XLIFF!

The problem with XLIFF is that it is extendable by definition. This means that you can create perfectly valid XLIFF documents that still cannot be read—or read in their entirety—by other tools. Now isn't that lovely! It tends to sound like an oxymoron when you have to create customized filters for an exchange format. The plea to tool vendors needs to be this: do not extend your XLIFF definitions. Use what XLIFF already offers. (And, yes, you might want to e-mail datastandards@atanet.org to voice that plea.)

> What is needed is a standard that allows for any translation environment tool to hook into an individual's workflow.

Wide and Customizable Access to Tool-External Resources

Some tools already offer the ability to link easily to online terminology resources and, in the case of Across, even hard-drive-based terminology resources. With tools like IntelliWebSearch, it is also possible to do searches on terminology resources for virtually anything and everything from within any tool in the Windows environment. Still, it would be a helpful and easy add-on for all translation environment tools to offer online dictionary and corpus searches as well as dictionary searches from within their environment. I would not be surprised to actually see this capability in upcoming versions of tools!

Integration of Common Word-Processing Features

This is an area that many tool developers have been working on, but it is also one that needs serious expansion. With MS Word all but out of the picture as the interface in which translation is being done, some features are missed by translation professionals in the tabular or otherwise structured interface of translation environment tools. (I know that many of you still use tools that allow for MS Word, but even you will have to admit that there is a general trend away from the program.) Many tools now have decent spelling checkers and AutoText features, some even have AutoCorrect, and an increasing number also have visible formatting (WYSIWYG, or What-You-See-Is-What-You-Get) for features such as bold and italic text, but that is not all that tools like MS Word or OpenOffice offer. What about grammar checks? What about track changes? What about more than just italics and bold as visible formatting? In many cases, tool developers

have to deal with limitations on the editing areas that their particular environment can handle, but maybe it is time to look for more highly enabled environments.

Morphological Tool Kits

I have had many conversations with tool vendors about this area. Why don't the tools know morphological rules when it comes to recognition of termbase and translation memory entries or the automatic adjustment once matches are found? Certain tools offer some functionality in this area in some languages, but the majority of tools do not. And why is that? It is too expensive. Plus, any kind of development in this area will be language-specific, so where do you stop? Which languages do you include?

My suggestion is for tool vendors to design tool kits for users to develop language-specific morphological rules that they can share with others or with their same language group, who also will have some contributions to make. Too simple a solution to make it work? Let's see whether we can find some takers among the tool vendors.

Rediscovery of Terminology and Integration into Translation Memory and Machine Translation Workflows

How long have we/you been preached to about the importance of proper terminology work? Forever? It sure seems like it, and yet I think it would not be presumptuous to say that most of us are still not particularly prudent about it. And being prudent can mean a lot of things. Do we take terminology work seriously and really invest into it? How do we use terminology in our workflows? Do we have manual lookup procedures or do our tools automate lookup and utilize their findings? Do the different components of our tools (translation memory, grammar checking, morphological "knowledge," machine translation, etc.) "talk" to each other and work in concert with each other?

> XML Localisation Interchange File Format is probably one of the best things that has happened to our industry in the past few years.

I think it is this last point where there can still be a lot done on the side of the tool vendors. Some tool vendors have offered features where termbases and translation memories "talk" to each other, so terms in a fuzzy translation memory match are switched on-the-fly if the old and new terms are "known" to the termbase (Déjà Vu does this). But there is clearly so much more that can be done. How about grammatical adjustments (singular versus plural or gender-specific changes in articles, etc.)? And how about communication between machine translation matches and terminology databases?

Some kind of machine translation feature is present in virtually all translation environment tools, but it is by no means advanced. The feature usually provides for a suggestion from a machine translation engine when no translation memory match is found, but there is no automatic lookup or even replacement of terms that are found in the termbase. And it would be an easy thing to do, especially if the same kind of communication already happens between the translation memory and the termbase. Provided we have good termbases(!), we might actually get decent hits even from machine translation engines like Google Translate or Bing Translator.

Now of course, there are ways to integrate terminology into machine translation processes, but only by training processes and not on-the-fly. I foresee this feature as a potential game changer on two fronts: 1) it could make machine translation a more reasonable additional feature for many of us, and 2) many of us might actually start to take terminology work seriously.

Honest Communication by Tool Vendors

This is pretty self-explanatory: we have been hurt enough by false promises. This is certainly true for machine translation technology, but also for the technology used in translation environment tools.

What about the much-heralded "support for PDF" in many tools? Give me a break. No tool truly supports PDF files, at least not in the way that other file types are supported. What about complete openness because of XLIFF support? Well, I already talked about that.

I would love to see communication between tool vendors and translators be direct and straightforward, just as one would hope for from a business-to-business relationship. And this goes both ways as we exert our influence in this business relationship by letting vendors know what we need.