



**A computer journal for translation professionals**

**Issue 13-11-229**

(the two hundred twenty-ninth edition)

## Contents

1. TaaS
  2. And This Is Where the Sharing Stops!
  3. Easy Handling of Web Translation (Premium Edition)
  4. Measuring Sticks
  5. New Password for the Tool Box Archive
- The Last Word on the Tool Box

## Celebrations


Hanukkah has already arrived around the world (Hanukkah Sameakh!), and here in the US we're celebrating Thanksgiving (Happy Thanksgiving!). Clearly it's time to make plans for what to give your clients and vendors for Christmas. I know, I know, many of you overachievers have already made up your minds or have already sent out presents or notes. But if you're like me, you still have a long way to go before making Christmas decisions, so I'm happy to share an idea with you!

I simply can't think of a better way to honor your colleagues and vendors and appreciate your clients than giving them each a copy of [\*Found in Translation: How Language Shapes Our Lives and Transforms the World\*](#).

Nataly Kelly and I wrote *Found in Translation*, published last fall by Perigee/Penguin, to shine a light on the compelling and essential work of translators and interpreters. The book demonstrates the importance of translation and its effects on every aspect of our lives. In your hands as a carefully chosen gift, *Found in Translation* speaks to both your colleagues and your customers: for your colleagues, it's an affirmation of their significance and expertise; for your clients, it's proof of the invaluable service you provide.

You can order it at Amazon or other booksellers, or you can also buy it in bulk directly from Penguin at a discount rate of 47% for purchases of 25 books or more. [Here](#) is the contact information.

Thanks for helping me lift up our profession and those who make it what it is. And happy holidays!



Before I forget: you may have noticed that I've renamed the "Tool Box Newsletter" to the "Tool Box Journal." The new moniker seemed a bit more appropriate for describing the format and content. Hope you like it.

## 1. TaaS

Here is the truth: At first I was a little skeptical when Peter Reynolds of [Kilgray](#) started to show me [TaaS](#) -- Terminology as a Service -- at this year's ATA. The lingo that surrounded the product was just a bit too bureaucratic and jargonic (and, just in case you wonder, yes, from now on that's an official word). Need some proof? "The motivation for the TaaS project is to address the need for instant access to the most up-to-date terms, user participation in the acquisition and sharing of multilingual terminological data, and efficient solutions for terminology resources reuse." All right, then.

But the more he -- and later Tatiana Gornostay from Latvian translation and technology provider [Tilde](#) -- showed me, the more I was won over by the depth and thoughtfulness with which this project was designed.

TaaS is a project that has received major funding from the European Union Seventh Framework Programme. It has five collaborators: Fachhochschule Köln, Kilgray, University of Sheffield, TAUS, and -- as the coordinator -- Tilde. The project is presently still in its beta phase (which was just launched on November 1), but it will eventually be a large, cloud-based terminology resource in all official working languages of the EU for translators, interpreters, terminologists, and technical writers ("language workers," according to their lingo -- a term I rather like).

Presently (November 27), English, French, German, Hungarian, Italian, Latvian, Lithuanian, and Spanish are supported; in just a few days, 16 more languages will be supported (Bulgarian, Croatian, Czech, Danish, Dutch, Estonian, Greek, Irish, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish, and Russian). Note that "support" for one language does not necessarily mean the same for every other, but before we go into this, we should probably see what this tool does in the first place.

Once you've registered, you can upload one or several files in various formats (PDF, DOC(X), XLS(X), PPTX, RTF, TXT, XLIFF, XML, or HTML), have terminology extracted from the file(s), apply content within existing terminology resources to those terms, select from the suggested translations and/or translate the terms, and then export it so that you can use it within your terminology database or glossary.

All this would still not be too overwhelming -- after all, there are plenty of tools that extract data -- were it not for a number of advanced tools that are (optionally) applied to this process. These include Tilde's wrapper system for CollTerm ([here](#) is some more detailed information about that), which performs a linguistic analysis (part of speech tagging, lemmatizers, morpho-syntactic patterns, etc.) as well as statistical analysis; Kilgray's terminology extractor, which also performs a language-independent statistical analysis; and a tool to normalize terms, which brings terms into their canonical forms (typically nominative singular or

infinitive). This latest, very cool feature is unfortunately only available for English and Latvian at this point (thus the aforementioned different levels of support).

Once that is done, the extracted list of terms will be run against a number of (again, optional) resources in the following order: 1. your own personal resources that you might have collected on the site; 2. other users' terminology (I'll explain in a second); 3. the [EuroTermBank](#); 4. the EU's inter-institutional terminology database [IATE](#); 5. the [TAUS](#) corpus; and 6. the TaaS statistical database (SDB) that consists of aligned web data. Once these databases have been queried for translations, they will be shown as suggestions from which you can choose by just clicking on them and/or you can enter your own translation.

To test the system, I uploaded a rather technical 9,000-word English file out of which 430 terms were extracted. Of these 430 terms, approximately half were terms that were very good suggestions as terms for my termbase -- which I estimate to be a good average -- and of the remaining 200-some terms, about 50 had various translation suggestions into German, usually with one that I chose. The terms for which no translation was found included "stellar research challenge," "greater cost accountability," and "translatable content" (yes, it was a text about translation technology! ) -- so no surprise that these were not found in an existing termbase.

The suggested translation did typically come from the ETB and some (not particular helpful ones) from undefined web sources -- I suspect that the IATE was not properly queried because it is presently under maintenance (you might have noticed that also in your private searches), and the connection to the TAUS data is either still buggy or just takes so long that it really is not viable at this point.

Of course, one of the ideas behind this project is to make it possible to share terminology data. At the outset of each project you can enter a whole lot of optional data, but you will need to make a decision on the language combination, the domain of your text, and whether you want to share the data with other users. The shared data will not include the complete texts that you upload but only the term pairs that you will end up with in your termbases (and only on an individual term pair level rather than complete lists of term pairs). Presently, the suggested term pairs are provided anonymously, but at some point in the project the source of the respective data will be visible along with an option to contact that person or company. (I assume that it will also be an option whether I want to be contacted.)

The shared data will also be used for other purposes, including machine translation. Both Tilde and TAUS have a strong interest in machine translation (and so does the EU as the funder of this project), and high-quality termbases are naturally helpful for machine translation.

We will continue to see the tool as a standalone tool but also integrated into translation environments. Kilgray's *memoQ* will most likely be the first to offer a "plugin which will allow users to send a document from *memoQ* for term extraction there, and for users to use their TaaS termbases within *memoQ*" (quote from Peter Reynolds.) There is no doubt others will follow, though. Paul Filkin from SDL has already expressed some interest for *Trados Studio* (I imagine that in that case it will take the form of an app in the [SDL OpenExchange](#)) and others will follow, especially because there is a good API (application programming interface)

that is being made available.

I'm eager to see what kind of response this tool will get. Or let me correct myself: I'm pretty sure that it will get a good response from translators -- the extraction feature is just too clever (at least in English) not to be used widely. What I'm eager to find out is what kind of response the sharing feature will get. It seems tantalizing to be able to communicate with others about their terminology and to share or use others' experience in the form of their terminology. Will that also include our willingness to opt into sharing, knowing that the data will be used by machine translation engines?

## 2. And This Is Where the Sharing Stops!

I'm a very active member of a local church and, as in many churches, members are very, very generous. If there is a need somewhere, it just needs a mention (often not even that) and it is met by someone in the church. This is true for any kind of service-related activity and, in our case, even for financial matters. Still, I know that the pastor and the church leaders are very cautious about bringing up money too often -- even in a generous environment like ours, people just don't like to be overtly reminded to give.

This helps me understand why [Linguee](#) failed with its two paid models, *Linguee Premium* and *Linguee Professional*. The five- or ten-euro per month fee was just a little bit too much for us.

Don't really know what these services were? Here is what I wrote about them when they were first released:

My initial response when I heard about it was skeptical, but after having used it, I'm really impressed. You see, at first I thought this would just be another version of *IntelliWebSearch* in paid form -- that is, a way to quickly search for a term within any *Windows* application on *Linguee*. And that's true to some degree. You can search from almost any *Windows* application by simply holding the CTRL key, clicking on a term, and executing a search within *Linguee*. What makes it really powerful, though, is that it not only searches for the one word you just clicked on but for the context of that term as well, so the matches that are displayed are not random but specifically geared to the kind of text that you are presently translating. This is very, very cool!

What's the difference between the two products? Essentially, one is created for the non-professional translator whereas the other -- yes, it's the more expensive one -- is for the pro. The difference between the versions is that only the *Professional* version is enabled to work within translation environment programs. It allows you to create a profile, and your services will be advertised a certain number of months on the site. There will also be a search history (kind of like *Google* does to target search results more specifically), and you'll be able to search within certain preferred sources.

So far this offer is available only in the EN<>DE language directions, but the other supported languages will follow later this year after the German version is successfully launched.

Well, not only will there be no other language version, but (unless you have already paid for some service that is still owed to you) these versions of *Linguee* have been completely stopped. It makes me really sad (and maybe even a little angry.) Not at *Linguee* mind you, no -- at ourselves.

When I wrote those paragraphs above, I myself had not entirely comprehended the power of the tool. *Linguee*'s corpus is large, and it's nice to be able to search through it on the basis of a single term -- that's what you do in the web interface. But to be able to search on the basis of a term and have the complete sentence taken into consideration while doing so gives you immediate access to the matches you need, or it lets you know that there are no good matches, each of which is helpful information.

I'm just not quite sure why this wasn't worth the price of a couple of fancy coffee drinks a month for us. . .

My analogy above was based on giving to benevolent causes, and it falls apart here, of course, because this was not for a benevolent cause. Quite the opposite, in fact: The speed of the lookup and the quality of the results should have provided a return on investment of the monthly fee within just a couple of days. It would have been a good business decision for many of us -- and it's too bad we didn't make it, because now no one can.

I talked to Gereon Frahling, *Linguee*'s CEO, about this, and he was pretty pragmatic -- his offer was not accepted, he couldn't afford to bleed money, so he had to pull the service. He will continue to offer the "normal" web-based service (which will add a large number of other language combinations in a very short time.) Still, while for him this is only a missed attempt at an opportunity, for us it's a truly missed opportunity!

### 3. Easy Handling of Web Translation (Premium Edition)

A few weeks ago I mentioned in the Premium edition of this ~~newsletter~~ journal that the new community model of the *Microsoft Translator* widget allows for a very low-cost way to translate smallish kinds of websites. No experience is needed, and if it's done right, you'll end up with a decent product (see issue 227 in the archives). The solution that I want to introduce this time presents a real possibility for translators or translation agencies without much background in web technologies to offer professional-grade and profitable web translation services.

In the last few years, a new kind of website translation has slowly made its way into the mainstream, and I assume that this happened without many of us actually noticing it.

Companies like Lionbridge and TransPerfect are offering it as a service, *Systran* offers it in the form of [SYSTRANLinks](#) as a machine translation service, and larger technology companies like Smartling and MotionPoint are offering it as a product. What is it? It's proxy-based website translation. This means that translations for a website in its original language are being produced without actually getting into the source of any of the translatable materials. Instead, the user browses in a cloud-based version of the site without actually



realizing it. The cloud-based site continuously sends queries to the original, untranslated website, which in response serves pages that go through the cloud-based layer where they are translated on the fly and appear in a different language.

In a way, it's not so unlike what is being done to a webpage when it's translated by *Google Translate*, only what we are talking about here is not machine translation, and the results are controllable beyond the mere translation. This means that the layout as well as the text being used in the localized website is customizable, and you are free to choose what kind of URL you want to use (provided they're available, of course).

Since the process is truly dynamic, it's also easy to understand that a notification system for newly introduced text becomes a no-brainer: If there is no translation for something in the cloud, the translator (or client) is notified.

Hmm.

I talked with Balázs Benedek from the Hungarian [Easyling](#). *Easyling* offers a solution much like its big competitors, but it really focuses on website translation without intermingling it with services, and it's priced to be affordable for smaller companies. According to Balázs, he started working on his product several years ago after realizing how cumbersome the traditional translation process for some of his [own websites](#) was.

You see, translating file-based, static websites in the 1990s and early 2000s was relatively straight-forward, but today's dynamic websites that are fed from content in content management systems (CMS) are much more complex to translate. You have to find ways to get into the CMS to get the translatable content and place the translated version back into the system, and then it's hard to find what changes are occurring on a regular level (they're not called dynamic for no reason). Also, since you don't necessarily just want to have a carbon copy of the site in the translated language (you might want different graphics, maybe a different layout, or maybe there's text that should not be in the new site), you will also have to build that from scratch.

So Balázs (and a number of other companies around the same time -- see above) started to build products where you never actually have to touch any content management system or source files (yes! -- no asking for the dreaded "source files" anymore). All you need to worry about is the content that the original website presents to the user, make decisions on which of it is applicable, translate it, and store it in the cloud. The next time the Brazilian user pulls up your website in Portuguese, that's exactly what he sees: a website in Portuguese.

The "cloud" that *Easyling* is using is the [Google App Engine](#) (or it's also possible to select an on-premise solution), and the actual translation that you perform takes place in a WYSIWYG (what-you-see-is-what-you-get) environment in the HTML of the translated website. The benefit of this is that you see the context and can automatically adjust your text so that it fits into predesigned spaces, or change the spaces so it will fit whatever text you deem necessary. The text that is not directly visible (like dropdown lists or keywords) is presented in a separate tabular interface. Naturally, since this is a translation environment, it comes with TM support and a termbase.

If you read the description above carefully, you'll have stumbled on one of the limitations of the system: Only HTML-based content is translatable. This means that content like graphics, PDF files, Flash files, or any other multimedia files can't be touched with this process. While this makes sense if you think about it, it's an important consideration when you promise your client wonders.... (Image files, by the way, are displayed in a separate interface, so you can decide which needs "traditional" DTP/translation and which not.)

I've talked to three different users of *Easyling*, two freelance translators and one larger LSP. None of them had used it for a very long time or for many sites, but all three mentioned that the tool had given them a foot in the door with clients that they would not have had access to otherwise. One of them said that her clients ended up being hesitant to actually put the proxy-based system into practice (that client asked advice of his web designer, who preferred to copy and paste everything over into a more traditional model -- note to self: never give your web designer too much power), but even in that case it ended up being a great sales tool to start with.

While the idea of working directly in the actual view of the website is nice, it's also possible -- at least with some of the higher-priced plans -- to export the translatables into an XLIFF file, translate it within the translation environment tool of your choice, and then reimport it into *Easyling*. One of the folks who I interviewed reported that she found it much easier to work that way. (If you work with *XTM*, there is a more direct integration than just through the exchange of XLIFF files, and there is another direct integration in the works with *memoQ*.)

And [talking about prices](#), there are a number of plans out there, starting from \$12 a month and going all the way up to \$240 a month depending on the number of pages and features such as change notifications and XLIFF handling. But what you need to keep in mind with this pricing is that since it includes the hosting of your client's websites, it's something that you can easily pass on.

#### 4. Measuring Sticks

While there are many websites and tools that perform all kinds of conversions between measurements, it's still helpful to have a little freeware utility like [Convert](#). *Convert* not only allows you to convert between an incredible multitude of measurements, it even lets you define your own parameters. (For example, if you happen to stumble on the formula for converting light-years into gallons -- which seems to be about the only conversion that is missing -- you can enter it right into the Custom tab.)

It's no wonder that this very comprehensive tool was created by Josh Madison. After all, he's also the guy who keeps very careful track of all the fortune cookie fortunes he receives and [shares them with all of us](#), as well as letting us be part of his very careful and personal [toilet paper usage analysis](#).

All hail the über-nerd!

#### 5. New Password for the Tool Box Archive

As a subscriber to the Premium version of this journal you have access to an archive of Premium journals going back to 2007.

You can access the archive [right here](#). This month the user name is toolbox and the password is uebernerds.

New user names and passwords will be announced in future journals.

### The Last Word on the Tool Box Journal

If you would like to promote this journal by placing a link on your website, I will in turn mention your website in a future edition of the Tool Box Journal. Just paste the code you find [here](#) into the HTML code of your webpage, and the little icon that is displayed on that page with a link to my website will be displayed.

Here is a webpage that proudly display the Tool Box logo:

[www.intexts.com](http://www.intexts.com)

If you are subscribed to this journal with more than one email address, it would be great if you could unsubscribe redundant addresses through the links Constant Contact offers below.

Should you be interested in reprinting one of the articles in this journal for promotional purposes, please contact me for information about pricing.

© 2013 International Writers' Group