

GeekSpeak

Jost Zetzsche

Hard Numbers

In a recent conversation,

David Canek, the founder of Memsource, shared some interesting observations with me that I would like to pass on. For those who do not know, Memsource is a cloud-based translation environment tool that has been around since 2010 and presently has about 30,000 registered users. That number might be a little misleading since it includes everyone who has ever registered, even if only to try the tool, but it is still a tool with a rapidly growing user base. And there are good reasons for that. While all of the resources are stored online (thus “cloud-based”), you can choose between a web-based and a desktop-based translation editor with virtually the same functionality. In fact, when I recently showed the web-based translation interface to European Parliament translators as one example of a user-friendly translation interface, you could almost see the unfortunate ghosts of clunky web-based interfaces of the past blowing out of their minds.

But I do not want to talk about Memsource, the tool, as much as about how users use Memsource. Since it is a cloud-based tool, it is easy for the Prague-based company to access aggregate statistical numbers across the entire cloud. David and his team recently looked at what percentage of users are using the translation memory (TM) and termbase features, what percentage are using machine translation (MT), and which MT engine (if any) they are using. Here are the results:

- **Translation Memory:** 28.8% do not use the TM feature (versus 71.2% who do).

- **Termbase:** 61.6% do not use the termbase feature (versus 38.4% who do).
- **Machine Translation:** 53.8% do not use MT (versus 46.2% who do).
- **Type of Machine Translation Engine Used:** Of all MT users, more than 98% use either Google Translate or the various options of Microsoft Bing Translator (versus less than 2% who use non-public engines, even though there are preconfigured connectors to the following MT engines: Apertium, Asia Online, KantanMT, LetsMT, MoraviaMT, NICT, PangeaMT, Systran, and Tauyou).

These are very valuable numbers. While I am sure they would differ to some degree in other tools, I would bet that the general direction is similar.

Let’s look at these numbers in more detail. The TM numbers are probably the most surprising. Almost a third of all users do not actually use TMs. In a certain sense that should make me happy as a long-time warrior against the term “translation memory tool,” but



it still seems odd that so many do not value TM. My guess is that most of these users are casual translators (there is a free Memsource edition for this type of user) who do not really understand the TM concept, or who are just interested in accessing MT content for the great variety of file formats that Memsource supports. If that is true, we need to keep that in the back of our minds for the remaining statistics.

The termbase numbers are somewhat unsurprising, although I might have expected an even lower usage of terminology databases. In this column (and elsewhere) I have often vented about the unfortunate underuse of terminology components in translation environment tools. In that sense, it is actually a positive development that more than a third of users are using the termbase—especially if our assumption about a segment of non-professional users is correct.

Now on to MT. There has been a chorus of voices in the past few years, especially in the MT community, claiming that even though so many translators are complaining about MT, a relatively large percentage of them are “secretly” using it. With almost half of all Memsource users proven to be MT users, I think we can say that there is some credence to that →

Information and Contacts

The GeekSpeak column has two goals: to inform the community about technological advances and at the same time encourage the use and appreciation of technology among translation professionals. Jost is the co-author of *Found in Translation: How Language Shapes Our Lives and Transforms the World*, a perfect source for replenishing your arsenal of information on how human translation and machine translation each play important parts in the broader world of translation. Contact: jzetzsche@internationalwriters.com.

Member News

- **Eriksen Translations**, of Brooklyn, New York, was awarded a Bronze Stevie Award in the 11th Annual International Business Awards. Eriksen received the award in the Best Overall Web Design category for the redesign of its website, www.eriksen.com.
- The following language services providers have been named among the 5,000 fastest growing private companies in the U.S. by *Inc.* magazine. The *Inc.* 5000 list represents companies that have had significant revenue growth over three consecutive years, are inde-

pendent and privately held, and are based in the United States.

Certified Languages International
Portland, OR

CETRA Language Solutions
Elkins Park, PA

CQ fluency, Inc.
Hackensack, NJ

Dynamic Language
Seattle, WA

Interpreters Unlimited
San Diego, CA

Language Training Center
Indianapolis, IN

Propio Language Services, LLC
Overland Park, KS

ProTranslating
Coral Gables, FL

TransPerfect, Inc.
New York, NY

U.S. Translation Company
Salt Lake City, UT

GeekSpeak Continued

notion. Granted, we might have to take some percentage points off to account for amateur translators, but it is safe to say that significantly more than a third of all professional users of Memsorce are using MT (more than use term-bases!). Of course, it would be interesting to analyze those numbers further to find out, for instance, which language combinations are being used and to what degree, as well as how the MT suggestions are actually being used. However, David and his team chose to focus only on a high-level view of usage data.

Perhaps more striking are the kinds of MT engines that are being used. Almost everyone is using generic public engines (Google, Microsoft) where the source segment is sent to the MT providers to be processed further. In fact, in Memsorce's particular case, a large percentage of users even select to send the target segments to Microsoft. This option is called

"Microsoft with Feedback,"¹ and goes back to an agreement between Memsorce and Microsoft whereby users do not pay any fees if they use that option. In exchange, users send Microsoft the target segment as well. Note that you can also use the regular Microsoft engine with Memsorce.

This makes me wonder about the answers to two questions. First, has our concept of data sharing evolved to the point where we feel comfortable with much of our data being shared with large corporations like Google and Microsoft? I am guessing that many of our clients would answer that question with a fairly resounding "No." I do not want to stir up a hornet's nest here, but I would likely respond to those clients like this: "So, do you not want me to use Gmail or other web services in which data is also being analyzed?" In a way, that should settle the first question.

But this brings us to the second question. If so many of us are using

MT, why are we not using customized engines? After all, the MT community has been telling us that using these engines is the only way in which we will see some real progress. My hunch is that while many MT vendors tell us that (theoretically) it is easy to build custom MT engines and access those services, in practice that is not really the case. These vendors still seem to have a lot of work to do. Whether that consists of making access easier or showing us that the results are actually superior remains to be seen, and chances are it will be a mixture of the two. ■

Note

1. More information on the "Microsoft with Feedback" can be found at http://wiki.memsorce.com/wiki/Machine_Translation.