Volume 13, No. 3
July 2009

Jost is an ATA-certified English-to-German translator and a localization and translation consultant. A native of Hamburg, Germany, Jost earned a Ph.D. in the field of Chinese translation history and linguistics in 1996. He began working in localization and technical translation in 1997. In 1999 he co-founded International Writers' Group (www.internationalwriters.com). Jost is also the publisher of the Tool Kit, a free technical newsletter for translation professionals (www.internationalwriters.com/toolkit). His computer guide for translators, A Translator's Tool Box for the 21st Century, was published in 2003. His latest endeavor is TranslatorsTraining.com, a site that offers in-depth comparisons of translation tools.

Jost can be reached at jzetzsche@internationalwriters.com.

Front Page

Select one of the previous 48 issues.
Select an issue:

- **Index 1997-2009**
- **TJ Interactive: Translation Journal Blog**

Translation Journal

Translators' Tools

The Google Translation Center That Was to Be

(Excerpt from the *Tool Kit—A computer newsletter for translation professionals*)

by Jost Zetzsche

I Remember about a year ago when news reached the public about a "*Google Translation Center*"? There was a lot of resulting hoopla from translators, tool vendors, and language service providers, and the responsible department at *Google* took a lot of heat. Back then I managed to talk to one of the folks in charge of the program, and it was one of the coolest conversations I ever had. In about ten minutes I asked all kinds of questions, and I always got the same answer: "Yeah, I can't really talk about that." It felt a little bit like talking to someone from the CIA: "Yeah, I could tell you, but then I'd have to kill you." (Though the *Google* guy was actually quite nice about it.)

Well, a lot of things have happened in the meantime, not least the radical change in the economy, and this may be one reason why Google's new *Google Translator Toolkit* (I'm going to call it *GTT* to avoid confusing it with an excellent newsletter I know of!) states this:

> *Google Translator Toolkit* is free, but in the future, we plan to charge users whose translations exceed high-volume thresholds.

This joins many other things that are quite different from the original scenario.

> Where GTT will have a good response is on the semi-professional or non-professional translation market.

But first things first: Here is what *GTT* is. It presents you with a rather well-designed front-end that allows you to do one of two things: You can either upload a file (in HTML, Word .doc [not .docx!], OpenOffice .odt, .txt, or RTF) or you can specify a URL and the corresponding HTML page will be uploaded.

When you select the file you need to select the language pair (the only available source language right now is English (!), but there is a choice of 48 available target languages,

Translator Profiles

- Success through Lifetime Learning
  by Gerardo Konig

The Profession

- The Bottom Line
  by Fire Ant & Worker Bee

In Memoriam

- In Memoriam—Ben Teague, 1945 - 2009
  by Gabe Bokor

Translation Nuts and Bolts

- What's Cooking: Translating Food
  by Brett Jocelyn Epstein

Medical Translation

- Physician Extenders—Who are they? Are they measuring up?
  by Rafael A. Rivera, M.D., FACP
- Translation of Medical Terms
  by Katrin Herget, Teresa Alegre

Cultural Aspects of Translation

- Cultural Untranslatability
  by Kanji Kitamura

Translation History

- The Issue of Direction of Translation in China: A Historical Overview
  by Wang Baorong

The Translator & the Computer

- Automatic Translation in Multilingual Electronic Meetings
  by Milam Aiken, Mina Park, Lakisha Simmons, and Tobin Lindblom

including double-byte and bi-directional languages). Then you need to choose whether this is a "shared" translation—i.e., whether you are using and contributing to a large anonymous translation memory or whether you would like to upload your own translation memory in TMX format. Here you can also upload or define a glossary. Should you choose to upload a glossary, it needs to be in CSV format with a strictly defined pattern, but it can have fields for part-of-speech or definition aside from source and target.

When you have made those choices the upload happens and your original file is displayed on the left pane of a split window (unless you choose **horizontal panels** in the **View** menu). On the right side you can see a pretranslated version of the file. The pretranslated material comes first from the translation memory(ies). If nothing is found there—and this is likely since at this point they don't seem to be particularly content-laden—a machine translation with *Google Translate* is performed. (If you choose to forego the machine translation step, you can select **Pre-fill with source text instead of machine translation** under **Settings**.)

Before you start the actual translation, you should click the **Show toolkit** button on the top of the window. This will open a new pane at the bottom of the window that contains four different tabs: **Translation Search Results** (TM hits), **Computer Translation** (results from *Google Translate*), **Glossary** (hits from your glossary—the number in parentheses indicates whether there are any and, if so, how many hits are being found), and **Dictionary** (this opens a search box in which you can enter single terms for lookup—no idea where that content comes from). According to which translation segment you highlight in the **Target** pane, the different tabs will show the corresponding content—if any is found.

The file is more or less displayed in WYSIWYG format, meaning that it is displayed the way you would see it in a browser or *MS Word*. Only when you select an individual translation segment (as in normal TEnTs, this typically corresponds to sentences) is the WYSIWYG view of that segment replaced with a text-only view in which inline codes are displayed as numeric codes with curly brackets ({1})—déjà vu, déjà vu! Unfortunately, it is not possible to enter the numeric codes with keyboard shortcuts. You will either have to use the ones that were automatically entered during the pretranslation and translate around them, or you can highlight the phrase that is surrounded by codes, select **Insert HTML tags**, select the ones you need, and they will be placed around the phrase (I'm not sure why it says "HTML tags" independent of the source format).

Since not all text is displayed in the normal WYSIWYG view of a document (think of keywords in HTML files or footnotes in *Word* files), there is a list of those at the bottom of the target pane under **Hidden Text** where they can be translated separately.

And while you are working on your translation (or before, or after), you can also invite others to participate in your translation/editing efforts by selecting **Share> Invite people**.

Arts & Entertainment

- On the Dubbing of Humor: Tidying Up the Room
  Juan José Martínez-Sierra, Ph.D.
- Doblaje audiovisual y publicidad—Reflexiones en torno al concepto de manipulación
  Isabel Cómitre Narváez

Literary Translation

- Chosen Aspects of the Polish Translation of J.K. Rowling's *Harry Potter and the Philosopher's Stone* by Andrzej Polkowski: Translating Proper Names
  by Anna Standowicz
- A Key Word in Gabriel García Márquez's *One Hundred Years of Solitude*
  by Dr. James McCutcheon

Translator Education

- Communication Strategies Do Work! A study on the usage of communication strategies in translation by Iranian students of translation
  by Sahar Farrahi Avval
- The Applications of Keywords and Collocations to Translation-Studies and Teaching—A Tentative Research on the Parallel Corpus of the 17th NCCPC Report
  by Dai Guangrong

Translators' Tools

- The Google Translation Center That Was to Be
  by Jost Zetzsche
- Thirteen Days in June—Adventures with SDL/Trados
  by Danilo Nogueira and Kelli Semolini

All they need is some kind of *Google* account.

When you are finished with the work on your file, you can download the translated file in the original format (and formatting). And be happy ever after. Or not?

Let's look at the tool in more detail aside from its functional aspects.

The first thing I noticed when I started to look at it—with the original *Google Translation Center* still burned in my memory—is that it is much less project management- and process-oriented than what was originally planned. This new tool is aimed at the translator rather than the translation buyer. (Remember that glimpse of *Google Translation Center* that we all saw, where translation buyers could upload a document and then choose translators to work in it? None of that anymore.)

The official Google blog says this:

> At *Google*, we consider translation a key part of making information universally accessible to everyone around the world. While we think *Google Translate*, our automatic translation system, is pretty neat, sometimes machine translation could use a human touch. Yesterday, we launched *Google Translator Toolkit*, a powerful but easy-to-use editor that enables translators to bring that human touch to machine translation.

The only argument that most of us would have with the statement that "sometimes machine translation could use a human touch" would be the "sometimes" (and that it's often more than just the "human touch" that's required). But the statement still tells us quite a bit about the immediate intent: to add the human expertise to *Google's* machine translation efforts. Remember, *Google Translate*, unlike engines like *FreeTranslation* or *Babel Fish*, uses a statistical machine translation engine that relies on good bilingual data—lots of good bilingual data.

And I don't think that there is anything wrong with it if it's just that I am adding data from the translation that I am currently working on. What I did not like was this: As mentioned above, it is possible to upload TMX translation memories, and as you upload you can select whether you want to use it yourself or with others—and that only seems fair. What I did not read before I uploaded a fairly large TM for testing purposes was this:

> By submitting your content through the Service, you grant Google the permission to use your content permanently to promote, improve or offer the Services. If Google publicly displays any of the content you submitted through the Service, Google will display only portion(s) and not the entirety of the content at one time.

This means that even though I can now go ahead and delete my TM (so that I and possibly other users won't have access anymore), *Google* will continue to use it. That strikes me as

- Translators’ Emporium

Caught in the Web

- Web Surfing for Fun and Profit
  by Cathy Flick, Ph.D.
- Translators’ On-Line Resources
  by Gabe Bokor
- Translators’ Best Websites
  by Gabe Bokor

- **Call for Papers and Editorial Policies**

odd—to say the least.

Also, as you know, TMX stands for translation memory exchange, but it is not possible to get your TM (or your glossary) out of *GTT* once it's in there. The only thing you can get out is your translated file, though you can, of course, continue to use the TM and glossary content within *GTT*—but only there.

To be fair, this was a problem in the first incarnation of *Lingotek* as well and they fixed it rather quickly, but we're talking about *Google* here, and I'm sure they don't do things without thinking them through. (Though other large companies do, and must also humbly concede defeat—see below.)

Aside from the thorny intellectual properties issue, there are a number of other things that make me think that *GTT* in its current state is not for the professional translator (unlike what we saw—or imagined seeing—in the *Google Translation Center*):

- The only source language (and UI) is English (though I imagine that this will change in no time).
- There is no possibility to alter the segmentation.
- There is no possibility for concordance searches.
- There is no way to manage fuzziness and it's not very clear how fuzziness is decided on.
- There are no QA tools and the only spell-checking is that of your browser.
- The number of file formats is much too limited (no XML, *Office 2007*, *InDesign*, etc.).
- You can only work on single files at a time.
- There are no project management facilities.
- There is no link to professional translation formats aside from TMX (such as TBX, TTX, or bilingual *Word* docs).

Where I think *GTT* will have a good response is on the semi-professional or non-professional translation market. The *Wikipedia/Knol* feature that allows you to take an English page, translate it, and publish it right to the server will certainly play a role in this (though I admit that I could not make this feature work!).

In its present form, though, this will not be a threat to small translation agencies or sites like *ProZ* or *TranslatorsCafe* as was feared with the first version. However, it is a "beta version," and my feeling is that if you were to ask *Google* where this is headed, they could tell you—but they'd (regretfully and politely) have to kill you.

Google
YouTube
0:00 / 0:00

© Copyright *Translation Journal* and the Author 2009
URL: http://translationjournal.net/journal/48google.htm
Last updated on: 06/19/2009 02:38:29