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

Hostile Takeover? Welcome Addition?

Machine Translation Enters the World of the Translator

by Jost Zetzsche

*I*t's been an interesting phenomenon. Professional translators seem to loathe and fear machine translation as much as they ever have—or, because of MT's slow but steady encroaching into "our" territory, more than ever—but translation environment tools [TEnTs] left and right are including machine translation components into their workflow.

Whether it's *SDL Trados*, *Wordfast*, *Lingotek*, *Across*, *memoQ*, *Alchemy Publisher*, *MetaTexis*, *MultiTrans*, (naturally) *Google Translator Toolkit*, or even the made-by-translators-for-translators open-source tool *OmegaT*, they all use machine translation, all with connectors to *Google Translate*, some with additional connectors to other engines, and some even with customizable machine translation engines.

> The key to success seems to be whether the kind of data the machine translation engine was trained with matches the kind of data that is currently being translated.

As a translator and someone who readily embraces translation technology—if it's useful—I have no problem seeing the benefits (and, if the client asks for it, necessities) of using customized machine translation. But the generic *Google Translate* or *Bing Translator*? Though I have seen some benefits in my own language combination (EN>DE) in very specific texts (such as short software strings), I have not seen much benefit in other work.

So I asked the readers of my newsletter whether this integration of machine translation is just a cheap ploy by tool makers to score easy points or a real feature improvement.

Here is a sampling of the responses. It would be great if these could serve as a starting point for more discussion or—even better—a more comprehensive understanding of how machine translation at this point (summer of 2010) affects our professional work environments:

Marinus Vesseur's comments started some of the discussion. Here is what he had to say:

> I've been using [*Google Translate*] in [*Trados*] *Studio* and it helps me a lot, specifically for English into Dutch and specifically when the subject matter is not my specialization, which is unavoidable sometimes. It's like someone giving you his take on the matter, a different way of saying it, which I find very helpful at times. Plus it usually comes up with the proper legal vocabulary that I just don't master, but sometimes need. Must be careful not to lean on it too much, but that's why it can be switched on and off on demand. It keeps making grave errors as well, of course, and I'm kinda glad it does.

Here was a direct response to Marius' view by Robert Morin:

> My view is exactly the opposite of that of Marinus Vesser. I will accept some MT material as an aid to my translation project ONLY IF the subject matter of the text at hand is one that I am very experienced with, so I can stand back and look objectively at the preliminary translation provided by the MT tool. But the other way around, i.e., using MT as an aid for a subject matter which is new to me would be like walking blindfold in a marsh full of quicksand . . . very risky!

And following are other translators' views.

Mette Melchior:

> I have only recently started to experiment with MT and often find it more disturbing than helpful, but recently I have had the very dubious pleasure of revising translations that were obviously based on machine translation (which some random checks with *Google Translate* proved). In these cases some of the automatic translations had been slightly edited by the translator, but it was still clear that the translations were based on machine translation and they contained many errors both in terms of terminology errors, syntactic errors, and a generally poor and stilted style.
>
> While I think MT integration in TEnTs is a good idea, I fear we will see many more of such machine-based translations of poor quality. Personally, I am not against the use of MT, but I think much more importance should be given to the skills needed for post-editing (and good editing and writing skills in general). Without the proper editing skills and a well-developed awareness of the grammatical, syntactic, and stylistic features that characterize both the source and target language, I think MT can be a "dangerous friend."

Julio A. Juncal:

> I have been using machine translation in conjunction with *Wordfast* for some time. For a while, I have used the Pan American Health Organization's MTS (Machine Translation System), which, depending on the nature of the text, produces quite acceptable English-to-Spanish translation copy. This dongle-protected system works well and can be fine-tuned to the type of document you are going to work on (financial, reports, etc.). One feature of MTS is its ability to render English reported speech into the historical present tense in Spanish, a feature that comes in extremely handy when translating summary records.
>
> More recently, I have been using *Google Translate* via the *Google Translate Client*, also under *Wordfast*. Since I translate a lot of United Nations (or similar) documents (English and French into Spanish), *GT* works well because it uses a very large corpus of United Nations documents. Again, translation quality depends on the nature of the text. But in general, *GT* is very useful because it saves you a lot of keyboarding.

Paul Lebartz:

> I cannot speak for language pairs other than EN>FR, but I suspect from the various posts I see on various mailing lists that it is often similar for many pairs. While the results of the different MT systems (e.g., *Google*, *ProMT*, *Microsoft*) are far from perfect, they are good enough in many cases that they do not require much editing to be used. It speeds up the translation considerably.

> To be honest, I'm surprised that someone like you, much closer to various translation technologies than most, has apparently not been able so far to see the advantages provided by MT. In my opinion, a tool without MT capability is not "refreshing"—it's a tool with little future.

> I would encourage you to start taking advantage of MT as fast as you can, because sooner or later the translation customers are going to want to take advantage of the increase in productivity we are getting from CAT tools integrating MT. The work of translating has changed, at least for many of us, and for better or worse it now includes the need for some post-editing skills. :)

Steven Marzuola:

> Much of my work is rather specialized and technical: oil and gas documents from Latin America. Online MT is rarely helpful due to the specialized vocabulary.

> However, a few weeks ago it was a different story. I was preparing for an interpreting assignment at a conference on corporate tax accounting and finances by studying advance copies of the speakers' presentations. When I came across an unknown term, I would look it up on *Google*. For most of them, I was able to find trustworthy translations very quickly (usually confirmed by searching for the Spanish term on its own).

> The reason, I think, was the subject matter. *Google* has evidently accumulated a large corpus of texts that are relevant to this subject.

> A few days later I was working on a finance and tax-related document, probably from Colombia. My TM program (good old *Déjà Vu 3.0*) does not have a built-in MT feature, so I used the *Google Translate Client*.

> It is definitely not my first option: I only called on it when there was no good fuzzy match and the results from *Assemble* were unsatisfactory. But it almost always gave me something useful. I also used it the next week on a similar job into Spanish. Not only for the terminology; I find it's very helpful in getting a better idea on how to organize a sentence that is closer to regular Spanish word order and farther from English.

And lastly, here is the perspective of a client that uses its own MT implementation:

> In a typical localization setup there are obvious data sensitivity issues with *Google Translate* or similar services. If we have NDAs in place with our vendors it's for a reason, and not compatible with sending our *source* content over the Internet, without encryption, to a company who can store, process, and distribute that data at will.

> My impression is that the quality of *Google Translate* in particular is generally very good—at least for the European languages I am familiar with—and I cannot see how it would not benefit translators when there is no alternative. Such an alternative can, of course, be MT engines trained specifically for the content being translated, as is our case here at [the client]. Such engines exceed the level of quality of *Google* for in-domain text (but are likely to be inferior for generic or out-of-domain text).

> In general, the statistical approach used by Google and others generally ensures that context is respected (...) But an issue with *Google Translate* is that only Google has control over it. So this may work well for one language but not another, and it may work well today but not tomorrow. Because of the generic nature of Google and similar engines, one company's product name may well end up being translated with another company's product (remember the noise around Google translating "Heath Ledger" into "Tom Cruise" a couple of years ago?) This could be pretty dangerous, and another type of problem you are unlikely to run into in traditional translation processes. Again, working with stable engines trained for a specific type of content helps rule this out.

So, to summarize, while there are clearly disagreements within this small sample of translators who in certain situations integrate machine translation into their workflow, the key to success seems to be whether the kind of data the machine translation engine was trained with matches the kind of data that is currently being translated.

In fact, in some cases it might work as a gigantic translation memory, as in this rather misleading example where the *New York Times* "tested" the MT prowess of *Google Translate* vs. others and *GT* essentially just used what it had in its memory from the many previous translations of *Le Petit Prince* it had captured:

Published: March 9, 2010

Putting Google to the Test in Translation

Google has poured resources into improving its automatic translation service. Some of these systems are better than others at handling particular pairs of languages, and it is impossible to evaluate them fully using small bits of text. But here are some quick comparisons of human translations and computerized versions from Google and two competitors.

Source	The Original Text	Human Translation	Google Translate	Yahoo Babel Fish via Systran	Microsoft Bing Translator
Le Petit Prince French "Le Petit Prince" ("The Little Prince") By Antoine de Saint-Exupéry	Le premier soir je me suis donc endormi sur le sable à mille milles de toute terre habitée. J'étais bien plus isolé qu'un naufragé sur un radeau au milieu de l'océan. Alors vous imaginez ma surprise, au lever du jour, quand une drôle de petite voix m'a réveillé. Elle disait: -Sil vous plaît... dessine-moi un mouton!	On the first night, I fell asleep on the sand, a thousand miles from any human habitation. I was far more isolated than a shipwrecked sailor on a raft in the middle of the ocean. So you can imagine my surprise at sunrise when an odd little voice woke me up. It said: "Please ... draw me a sheep." - Wordsworth Children's Classics, 1995	The first night I went to sleep on the sand, a thousand miles from any human habitation. I was more isolated than a shipwrecked sailor on a raft in the middle of the ocean. So imagine my surprise at daybreak, when a funny little voice woke me. She said: "If it pleases you ... draw me a sheep!"	The first evening I thus fell asleep on sand with thousand miles of any inhabited ground. J' stays much more insulated qu' a shipwrecked man on a raft in the middle of l' ocean. Then you imagine my surprise, at daybreak, when funny of small voice m' awoke. She said: - S' you like it... draws to me a sheep!	The first evening I thus fell asleep one thousand miles of any ground inhabited with sand. I stays much more insulated was shipwrecked man on a raft in the middle of the ocean. Then you can imagine my surprise, at daybreak, when black of small voice I awoke. She said:-If you like it... draws to a me sheep!

Good news? Bad news? I'm not sure. But it looks like news that we will have to continue to deal with.