



# First Date: Outreach from the Machine Translation Community to Translators

*By Laurie Gerber and Jay Marciano*

**In early 2008**, the leadership of the Association for Machine Translation in the Americas (AMTA), in preparation for the association's October conference, began to discuss how to reach out more effectively to other groups that have a stake in the future of translation technology. Machine translation (MT) has typically been viewed with skepticism, if not outright hostility, by translators, and for this reason, AMTA wanted to extend an olive branch.

Within the MT community, AMTA sees many ways to help translators. While many translators are not attracted to MT itself, many of the language technology components that go into this technology can improve and extend the existing tools that translators do like, including search capabilities. AMTA also wants to listen to translators' concerns about language technologies.

With this goal in mind, AMTA leaders proposed a session for ATA's 2008 Annual Conference entitled "First

Date: A Dialogue Between Translators and Machine Translation Developers." The relationship might not go anywhere, but it would at least be a chance to get acquainted! As president of the International Association for Machine

have not had the chance, or found a way, to influence the technologies that are supposed to help them. We found the notion of helping translators to become technology drivers a very timely and powerful way of capturing

---

**As ATA President Jiri Stejskal pointed out during the session, the use of MT in the provision of translation services is not a zero-sum game against translators.**

---

Translation (IAMT), I was asked to represent AMTA at the session, along with Jay Marciano, head of SDL's MT development group. As it happened, Donald Barabé, senior vice-president of technology at the Canadian Translation Bureau, gave an excellent and visionary presentation two days before our session, noting that translators have been technology-driven, but

the trend AMTA hoped to start. We borrowed his words in our own session, and in subsequent discussions within AMTA.

## **What Translators Want**

In the conference session, we asked translators to talk about what they would like a computer to do for them. We asked them not to limit

**Table 1: Topics Raised in the First Date Session and Their Frequency**

Number of Comments	Topic Category	Category Description
8	Communication	Better, clearer information and communication about MT technology, its uses, capabilities and limitations, best practices, and the evolving role of translators.
6	Resources and Search	Useful, high-quality reference glossaries and bitexts, in addition to intelligent tools for searching them.
5	Leverage	Translation tools that readily make use of existing terminology and translation memory resources, including format conversion and exchange.
3	New Capabilities	Some of the capabilities translators would like to have are not yet possible with the current technology.
2	Plug and Play	Tools that are easy to deploy and easy to combine into a workflow.
2	Better Standards	Advances in best practices or community standards.
2	New Tools	Suggestions were for software tools that simply are not available on the market.

themselves to what was available or possible. Surprisingly, the participants came up with a more or less continuous stream of ideas during the 90-minute session. We logged 28 different suggestions and later divided them into seven categories. These are presented in Table 1, together with the number of comments that addressed each topic.

**Communication:** This topic emerged toward the end of the session and then dominated the discussion. It was one of two topic areas that really focused on MT itself. Translators would like more definitive information about MT. They feel the need to understand the evolving translation market, to figure out whether they want to offer post editing as a service, and to arm themselves for conversations with clients who ask for information about MT or perhaps even justification for paying professional translator fees when MT is available as an option. Clients' ignorance and their hopes for an easy solution for translation problems have certainly fueled the tension between translators and MT.

**Resources and Search:** Translators have eagerly embraced terminology search technology as a way to access and expand their available reference materials. This topic brought out suggestions for advances that would enable online searches for bilingual, topic-appropriate text examples, and refinements of search tools and searchable resources that would fit into a translator's natural workflow. The group also pointed out the need for online access to the large corpora of material, as well as mechanisms for sharing translation memories (TMs) among translators.

**Leverage:** Translators have clearly made investments in accumulating high-value resources—TMs and glossaries—and want to be able to leverage those resources in any future work easily. The comments made hinted at past disappointments when a significant investment in creating such resources could not be transferred to a new tool or new working environment. Either the resources remained trapped inside a proprietary tool that would not export, or they could not be imported into a new tool because of formatting or some other issue.

**New Capabilities:** Participants said they would like MT systems that can learn from the translator's corrections. They would like such systems to correctly handle dates, currency, and numbers. They would also like to be able to give more feedback to the system. (For example, indicating the quality of a particular sentence output so that the system will learn to provide more translations that translators can really use.)

**Plug and Play:** Currently, many feel that an individual translator who wants to combine MT and TM—or any other tools that might operate in a workflow pipeline—needs to be a computer scientist to connect them. Translators want it to be easy and practical to combine tools in any order. Further, software options should be visible and easy to use, rather than hidden in places where only systems engineers can find them.

**Better Standards:** Participants expressed a desire for a standard document format that would allow them to bundle all language versions of a document into a single transferrable package. (Although this is not possible in Microsoft Office, this is what ➡

XLIFF [XML Localization Interchange File Format] accomplishes. It is importable/ exportable to/from many TEnTs [Translation Environmental Tools]). Participants also thought terminology lists and glossaries should preserve the identity of the creator.

**New Tools:** Many of the suggestions above would be new offerings in the marketplace and can be built with existing technology. The additional suggestions in this part of the discussion were for tools that could be used outside of a workflow. (What if you want a professional caliber terminology management system but not TM?) Jost Zetzsche chimed in for translators who work in less common languages, suggesting that high-quality tools are needed that can handle more languages well (e.g., optical character recognition, terminology management, TM).

In general, translators would like convenient task-specific widgets.<sup>1</sup> For those who do not want to buy into a big, expensive suite of functionality, they would welcome simple applications that do very limited things, such as an add-in that does a customizable online search of highlighted material. Simplicity is the key.

### What Is Available Now?

Above, we summarized translators' comments from the session. Here we offer some suggestions on what might be available regarding the first two topics in Table 1. For the rest, it remains for technology developers to respond.

**Communication:** Writings and presentations on MT have generally not been aimed at translators. AMTA is very interested in continuing to participate at ATA conferences to provide information about MT and to gain a

better understanding of how translators work and what they need. In addition, AMTA is incorporating more content aimed at translators to be used at its own conferences. The organizers of the upcoming MT Summit, to be held August 26-30, 2009 in Ottawa,

there is an opportunity to commercialize them.

In the area of large bitext corpus resources and mechanisms for sharing TMs, there is more commercial activity. For example, the natural language research group (*Recherche*

---

## Clients' ignorance and their hopes for an easy solution for translation problems have certainly fueled the tension between translators and machine translation.

---

Canada, have taken this to heart, and are planning conference sessions and tutorials directly aimed at translators, as well as sessions that educate technology developers about how translators work. For more information, check out <http://summitxii.amtaweb.org>.

**Resources and Search:** Some of the capabilities that translators need are already available on the market, though not necessarily in the form of tools aimed at translators. Naomi Sutcliffe de Moraes wrote two extremely helpful articles on terminology search tools in the July<sup>2</sup> and September<sup>3</sup> 2008 issues of *The ATA Chronicle*. She covers tools and suggestions on using search tools to help understand a term in the source language and to find the appropriate target-language term. Identifying bilingual text sources has become a specialty in the statistical MT research and developer communities, but tools tend to be aimed at researchers working in Unix/Linux and are not generally available for Windows, nor are they very precise. It seems that

*appliquée en linguistique informatique*) at the University of Montreal has a number of online tools that provide access to many Canadian monolingual and bilingual text resources. Monolingual concordance search is available for free. Bilingual concordance search is available for \$129.95 per year for an individual subscription, and provides access to 452 million words in French and English on government and legal topics. TM Marketplace and the Translation Automation User Society Data Association offer access to bilingual corpora aimed primarily at language service providers, corporations, or statistical MT development efforts. TM Marketplace buys and sells individual TMs; the Translation Automation User Society Data Association offers members access to TM resources contributed by its membership. There are also places where translators can contribute and share TMs, such as the Wordfast Very Large TM Project, which is free and anonymous.

### Further Thoughts

As representatives of AMTA, we were honored by the open-minded



## Links of Interest

**Association for Machine Translation  
in the Americas**  
[www.amtaweb.org](http://www.amtaweb.org)

**International Association for  
Machine Translation**  
[www.eamt.org/iamt.php](http://www.eamt.org/iamt.php)

**MT Summit**  
<http://summitxii.amtaweb.org>

**Recherche appliquée en  
linguistique informatique**  
<http://rali.iro.umontreal.ca>

**TM Marketplace**  
[www.tmmarketplace.com](http://www.tmmarketplace.com)

**Translation Automation User Society  
Data Association**  
[www.translationautomation.com/tda](http://www.translationautomation.com/tda)

**Wordfast Very Large TM Project**  
[www.wordfast.net/?whichpage=jobs](http://www.wordfast.net/?whichpage=jobs)

reception we received during the First Date session. Concerning the historic tension between translators and MT, we realize that translators are not so much against MT as against: 1) having their skills and services compared to MT, and 2) the assumption that post editing MT output is the same thing as translation. We do not think that translators' jobs are threatened by MT. As ATA President Jiri Stejskal pointed out during the session, the use of MT in the provision of translation services is not a zero-sum game against translators. Done correctly, the use of MT will expand the translation market, but not necessarily eat into the professional translation market.

Technology developers need and value exactly the kind of input that translators gave during the First Date session! We will share the comments with the MT community in writing and at the August MT Summit. We look forward to continuing the dialogue and hearing more from language professionals at future ATA conferences, and perhaps at the MT Summit, to keep the ideas flowing!

### Notes

1. A widget is a program that performs some simple function, such as providing a weather report or stock quote, that can be accessed from a computer desktop or webpage, usually by clicking on a

button or scroll bar. For more information: [http://en.wikipedia.org/wiki/Widget\\_engine](http://en.wikipedia.org/wiki/Widget_engine).

2. Sutcliffe de Moraes, Naomi J. "IntelliWebSearch: A Configurable Search Tool for Translators." *The ATA Chronicle* (American Translators Association, July 2008), 26.
3. Sutcliffe de Moraes, Naomi J. "The Translator's Binoculars, Part II: Desktop Search Tools and How They Can be Used to Search Reference Texts." *The ATA Chronicle* (American Translators Association, September 2008), 32.

*ata*