Machine translation (MT) may not be an easy technology to love, but translators must learn to coexist with its development in the profession. At the very least, translators need to be able to discuss the issue when clients bring it up. How? As part of ATA’s webinar series, Laurie Gerber’s presentation—“Is Machine Translation Ever the Right Tool for the Job?”—explored the vocabulary and framework for talking to language services providers (LSPs) and end clients about how to get the job done with the optimum mix of computing power and human skill. Attendees learned the decision points to include in client discussions regarding MT, such as strengths and limitations, practical applications, cost savings, translation speed, and post-editing requirements. What follows are Laurie’s answers to some questions posed by webinar attendees. The transcript of this discussion is provided for the benefit of all members.

As a freelance translator, how can I prepare myself to use MT? I am usually dependent upon what the LSP hands me, so it is hard to be proactive.

Working with MT systems or output may happen in a variety of ways. For instance:

1. You license and customize an MT system and decide which jobs to run through it and whether/how to integrate it with your translation memory (TM) or other translation environment tools (TEnTs).

2. You use the “hooks” available in your TM system to send each sentence to an online/hosted MT system. The MT output appears in the current segment composition area or off to the side with other resources, such as fuzzy matches and terminology. (Note that you need your client’s permission to use this approach. See the discussion on page 22 concerning uploading entire documents from the client to free public MT systems.)

3. Clients send you translated drafts, probably in .tmx format, that are a combination of TM matches and MT output, possibly delivered together with other resources like terminology.

4. Clients send you translated drafts, probably in MS Word format, that

It is never appropriate to submit unreviewed, unrevised machine translation output to a client as a finished translation.
are either straight MT or some combination of TM and MT. You revise or retranslate the text they provide.

In the case of 1 and 2, you are in control and can build skills and prepare at your own pace. If you want to try 3 or 4—accepting post-editing work from clients—I expect that many clients will be happy to hear that you are open to doing post-editing and that you are “cultivating a new specialty.” If you can build some extra time into the first few projects in order to get comfortable with post-editing, you can discover ways to make it profitable and decide from there whether or not you want to continue.

If you decide that a job is suitable for MT, are you supposed to “announce” to your client that you are going to use MT? Do you openly offer the service as MT and post-editing?

First, it is never appropriate to submit unreviewed, unrevised MT output to a client as a finished translation. I do not need to tell ATA members that, but I hear complaints from my LSP friends that they occasionally receive poor quality translations from freelancers only to discover that they could have gotten identical output from Google Translate.

Except where a client specifies a particular tool (usually more for the convenience of project and file management than for expectations about quality), I believe there is no need to announce to your client what tools or resources you use to do the translation work. The tool does not alter the agreement with your client on the schedule and quality of the translations you have committed to deliver. However, if you want to try using the hooks in your TM system to upload the document to an online “public” MT system like Google, freetranslation.com, or Babelfish, you must ask the client whether it is okay to upload the text to a public MT system.

How do you make a client realize that MT is not adequate for every text?

As this question suggests, sometimes clients are too eager to adopt MT. Clients, who tend to be under pressure to shorten delivery times and cut costs in their own departments or services, are very susceptible to the appeal of “automatic” solutions, even if they suspect that it may be too good to be true. You can add a lot of value to your relationships with clients if you can help educate them about the appropriate use of MT.

Figure 1: When Is a Text Suitable for MT?

<table>
<thead>
<tr>
<th>Text Criteria</th>
<th>Consider MT</th>
<th>Do Not Consider MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anticipated Use of Translated Text</td>
<td>When users are only interested in searching or browsing for a few pieces of information (e.g., user manuals, legal discovery).</td>
<td>When the reader should be drawn into reading the full text in translation.</td>
</tr>
<tr>
<td>Communicative Goal of the Translated Text</td>
<td>When complying with information provision requirements, such as those found in time-sensitive e-mail or chat between cooperative parties.</td>
<td>When trying to persuade the reader to adopt an attitude, take action, etc. (marketing or editorial text).</td>
</tr>
<tr>
<td>Source Text Characteristics (where the translation must preserve all of the information and nuance in the source text)</td>
<td>When the source text is written clearly, with consistent use of terminology. (In these cases, consider dictionary customization.)</td>
<td>When the text is written carelessly, with spelling and grammar errors. (Information in this type of text can be of high value, for example, in questionnaire responses.)</td>
</tr>
<tr>
<td>Intended Audience for the Translation</td>
<td>When the audience is cooperative and has a positive relationship with the information provider (employees, business partners, and suppliers).</td>
<td>When the audience is likely to be critical and attentive to detail, such as market analysts and key customers in the buying and evaluation stage.</td>
</tr>
<tr>
<td>Time Sensitivity of Information</td>
<td>When the text is issued for immediate use and has highly perishable information.</td>
<td>When the information is enduring and rarely changes.</td>
</tr>
<tr>
<td>Text Volume</td>
<td>When the text contains over 10,000 words.</td>
<td>When the text contains up to 10,000 words.</td>
</tr>
</tbody>
</table>
Figure 1 on page 22 presents some criteria by which to judge whether a particular text is suitable for MT. No single criterion can serve as an absolute guide to tell you when to use MT, but in combination, you and the client can use the decision points in Figure 1 to help judge when it is time to consider MT.

There may be times when the opposite question is relevant—“How do you make a client realize that MT is an adequate solution for some of the texts they have?” You may think that there will never be a day when you will want to make this argument, but what if you have a client with a large volume of text for which you are being pressured to provide “just a quick/rough translation” at a substantially reduced cost? Well, if all the client wants to do is find out if any of the information contained in the text is useful, why not suggest MT? For example, in a legal discovery/filtering process, much of the material will not be used, but portions of text may, for instance, become evidence in a trial. In this case, you can help clients make use of MT in the filtering stage to locate the pertinent information they will need. Then you can provide the polished translations of that material. Particularly if this becomes a frequent struggle with any of your clients, perhaps initially running the documents through MT is a solution that can make you both happy.2

Before purchasing any software package, it is very important to see whether the system has core capability to translate the language structure you need to handle. Here is a suggested two-stage test of the core translation capability:

1. Test out the MT system by translating a general text without specialized vocabulary (one containing short, declarative sentences).

2. Try translating a specialized text in the genre/register you will need (e.g., patents, which include long sentences and large compound technical terms).

In the general sample, ignore problems that arise purely from terminology that you can address by adjusting the dictionary. If the system cannot handle the structure of the language in the general text you submit, you will not be able to repair this with any dictionary building effort or statistical retraining. For the specialized text sample, the MT output will give you an idea of how much coverage the system already has and whether you like its translation choices. MT system dictionaries will let you override the translations used by the delivered system. They will also let you add new terms that are not yet covered and new translations for multi-word terms.

Could you name MT tools/software that you might recommend for purchase to freelance translators and mention their compatibility with Trados or other TM software?

The short answer is “yes.” (See the list on page 26.) But stick around for the long answer that follows below.

Before purchasing any software package, it is very important to see whether the system has core capability to translate the language structure you need to handle. Here is a suggested two-stage test of the core translation capability:

1. Test out the MT system by translating a general text without specialized vocabulary (one containing short, declarative sentences).

2. Try translating a specialized text in the genre/register you will need (e.g., patents, which include long sentences and large compound technical terms).

Before purchasing any software package, it is very important to see whether the system has core capability to translate the language structure you need to handle. Here is a suggested two-stage test of the core translation capability:

1. Test out the MT system by translating a general text without specialized vocabulary (one containing short, declarative sentences).

2. Try translating a specialized text in the genre/register you will need (e.g., patents, which include long sentences and large compound technical terms).

In the general sample, ignore problems that arise purely from terminology that you can address by adjusting the dictionary. If the system cannot handle the structure of the language in the general text you submit, you will not be able to repair this with any dictionary building effort or statistical retraining. For the specialized text sample, the MT output will give you an idea of how much coverage the system already has and whether you like its translation choices. MT system dictionaries will let you override the translations used by the delivered system. They will also let you add new terms that are not yet covered and new translations for multi-word terms.

You can conduct these basic tests using the free access to the systems that most developers offer online.

Which system is the best, and are systems compatible with each other?

Right now, there is no “best MT system,” in the same way that there is no “best car.” Perhaps someone will develop a straightforward feature comparison of these tools, which would be very useful.

What criteria are the best to use when trying to choose an MT package for various European languages?

MT software developers must develop each language pair direction separately. As a result, developers have likely made different levels of investment in each language pair direction. The high performance of MT software in one language pair direction is not always a strong predictor of performance in other language pair directions. However, almost all systems do pretty well with French↔English, and most developers have made a serious investment in Spanish↔English. Independent of the effort invested in development, some language pair directions are much harder than others. For instance, Japanese-, Chinese-, and Korean-to-English are hard to automate, regardless of the approach the MT developer uses, while French- and Arabic-to-English are relatively easy.
Questions and Answers about Machine Translation Continued

(See list on page 26.) Others may offer a time-limited, fully functional evaluation copy of the software. You can also have the MT vendors run a sample text for you.

Regarding specialized glossaries, several MT developers have explored specialized glossaries as a higher-value component to their offerings. Some developers may offer “topical glossaries” for subject areas as an included feature of their higher-end products. Others may also offer glossaries as add-on products for a separate fee. If this is an important part of the offering for you, it is important to test these glossaries for your language pair and subject area. The extent of investment in each subject area and language pair direction may vary considerably among vendors, and even within a single vendor between languages and subject areas.

People planning to choose an MT system always hope that there will be an objective way to make comparisons across the board, but the only way to do this is to compare uncustomized general-purpose systems. The most important factor to consider is how difficult customization is going to be and whether you can get the system to the level of quality you need.

Finally, if you plan to do dictionary customization, it is important that the MT system be able to import your terminology lists—regardless of whether you have them in a terminology management system or Excel—in order to save manual dictionary building effort. You may still need to do some work to get the full power of an MT system’s dictionary functionality, which should be able to recognize and generate any form of any word (unlike TM systems, which typically do not handle morphological processing). It is equally important to be sure that the MT system will let you export the dictionaries you have built up in it at a later time. If you want to change MT systems later, you need to be able to take your investment with you.

I find it hard to learn about MT tools all by myself since there are no workshops or training sessions out there. Could you advise on educational programs in post-editing?

Neither my colleagues nor I are aware of any regularly offered training. Unfortunately, the tutorials/seminars offered at conferences tend to stay at the introductory level. However, there is clearly a lot of interest right now, and I expect that conferences and schools will keep trying to build content in this area.

When teaching translation, I do not let students look at any MT until we have finished about two-thirds of the course. I want them to learn human translation first, and then look at MT to be able to evaluate. Do you agree with this approach?

I agree that it is important for translators to master the core discipline and critical skills, which can only come from doing it themselves, before they discover the temptation of automation. The future of translation as an art depends on students making this investment in learning the skills necessary. On the other hand, I would be extremely grateful if translation instructors did expose their students to the tools that are available. In another direction, teaching organizations also have a very important role in teaching students for this reality and discuss the range of legitimate needs that clients of language services have and how to serve them (gist, summary, MT+post-editing, whatever). Our clients’ requirements are consistent with the needs of their businesses, and the community of language professionals is best positioned to find a way to satisfy these needs.

MT for Trip Advisor appeared successful, but not for an auction site. Both are for persuasive purposes, so why do you think MT for the auction site did not work well?

Businesses like Trip Advisor succeed when people spend time on their websites and look at ads, occasionally clicking through for more information. Site visitors are happy to browse information that may help them make decisions about where to stay or go without feeling any particular risk in accumulating such data points. Site visitors also do not hesitate to click through on hotel or other ads, and each click generates ad revenue for Trip Advisor. However, an online auction site succeeds only when site visitors buy things. If translations do not engage the site visitor’s attention or give them the confidence to follow through with buying things, MT cannot work.

Having started with Systran, I find it is a lot of work before I get to see real results from my investment in MT customization.
How many months do you think are reasonable to vet the system (customization) before it can be used quickly and effectively?

Mike Dillinger, vice-president of the Association for Machine Translation in the Americas (AMTA) and principal of TOPs Globalization Consulting, suggests that it should take less than a month to see benefits from customizing MT. Adding 20% of the unique terms in your source language text to the system’s dictionaries is enough to yield a significant improvement. For example, a 10,000-word corpus may have 1,000 unique terms. Of these, around 200 terms may have a frequency of three or more occurrences. Adding these 200 terms to the dictionary up front will make a big difference. LSPs might consider outsourcing dictionary customization either to the MT developer or to a consulting company that can advise on technology choice and do the customization.

Isn’t it easier to do a good translation from scratch rather than edit a bad translation?

This, plus MT’s inability to handle language nuances and culture, are condemnations of MT that most translators find quite satisfying. I offer three ideas for your consideration below.

1. Post-editing MT can be unpleasant, which makes it seem like it takes much longer. At the Translation Automation Users Society (TAUS) conference in October 2010, Mirko Plitt, of Autodesk, described Autodesk’s experience. He explained that while translators often feel that post-editing MT takes longer than translating from scratch, timing their work showed that post-editing MT actually yielded a 30%-40% productivity increase. This will not be true for all post-editing jobs, but I know three freelance translators (all ATA members) who have taken matters into their own hands and done MT post-editing voluntarily to increase their earnings and speed their work. Two accept MT output from clients and produce translations much faster. Even at a discounted per-word rate, they find this type of work quite profitable. The third translator realized that he did not like typing out long terms repeatedly, so he invested in his own desktop MT system. By building up the term lists gradually, he is able to save a tremendous amount of time both in terms of typing and dictionary work.

2. MT post-editing is a cultivated skill (see Point 3 below). Embarking on any new activity where we lack “fitness” is frustrating, embarrassing, and feels very uncomfortable. Together with translators’ natural ambivalence about MT, the very suggestion that you should develop yet another set of skills to cope with the “abomination” of MT probably makes you angry! But the full-time post-editors at the Pan American Health Organization—which only accepts applications from translators who will post-edit the output of its homegrown MT system—actually build post-editing skills and enjoy their work. The PAHO MT development team solicits and acts on translator suggestions for how to improve the MT system because it reflects their suggestions.

3. Using MT output need not be limited to revising everything the MT system generates, even though this is what the term “post-editing” implies. The following are two post-editing skills that are worth considering.

- **Quick decision making:** If you can make a quick decision about whether a sentence from the MT system can be revised efficiently, you can get to work on it or discard it and translate from scratch. What is time consuming and unpleasant is trying to untangle misinterpretations or infelicities that the MT system has introduced. The larger challenge is to look at an entire post-editing job as it comes to you from a client and decide whether you want to accept it. Regarding sentence length, longer sentences are much more likely to include confused, disordered ideas. Do not get bogged down in untangling them. If there are many clauses, the earlier clauses are more likely to be usable.

- **Look for anything you can use:** TAUS has put out short and long guidelines/best practices for post-editing. The short
guidelines include suggestions for coming to a shared understanding with the client up front. The longer best practices guide (which may be available only to members) makes a suggestion that I have seen used effectively in several TM interfaces: do not worry about revising the output, just grab anything from the MT output you can use. For example, the Lingotek, MemoQ, Fluency, and other TM interfaces present MT output off to the right of the screen, together with additional fuzzy matches and terms that match against the current segment. In the special “TM showcase” workshop at the AMTA conference in Denver in 2010, five translators/organizations demonstrated the MT systems they used and the post-editing process. The one that struck me as most readily applicable to freelancers was Dorine Ooze Vermeulen’s Lingotek+Google Translate demo. Dorine is an ATA member who translates between English, Dutch, and Hebrew. In her demonstration, she explained that for each sentence, she read the source, skimmed the target, and made a decision whether or not to post-edit the MT output, or simply copy and use some portion of the output or work from scratch. (See Figure 2 above.)

More Information

In closing, here is a list of vendors, along with other websites, where you can become acquainted with the capabilities of various systems and follow more discussions on this topic. Please note I have no ties (other than professional friendships with former colleagues) to any MT developer/seller listed below.

**MT Products with a Single User/Small Office Version**

( Customizable version for production translation will run $300-$1,000.)

**ASIAN Language MT**

[www.aamt.info/english/nts.htm](http://www.aamt.info/english/nts.htm)

( For MT into/out of/between Asian languages, see the directory on the
website of the Asia-Pacific Association for Machine Translation.)

LEC
www.lec.com
(Many languages. English-to-Japanese developed by LEC, and others are licensed from other developers.)

Linguatext
www.linguatext.net
(Many languages, with special expertise in German.)

Morphologic
www.morphologic.hu
(Primarily Hungarian<>English)

ProMT
www.promt.com
(Many languages)

Systran Software
www.systransoft.com
(Many languages)

Word Magic
www.wordmagicsoft.com
(Spanish<>English)

More Information
See the Compendium of Translation Software. The 2009 and prior versions are available for free.* The current version is available to members of:

Asia-Pacific Association for Machine Translation
www.aamt.info

Association for Machine Translation in the Americas
www.amtaweb.org

European Association for Machine Translation
www.eamt.org

Notes

2. This example is based on a thread from the ATA Business Practices listserv in March 2011, http://finance.groups.yahoo.com/group/ata_business_practices.

3. This past June, Mike Dillinger presented a webinar, “MT On a Shoestring,” for LSPs that are interested in getting started with MT. For more information, visit the Globalization and Localization Association’s website: www.gala-global.org/gala-webinar-mt-shoestring.


5. Short videos of all the “Collaborate to Innovate” presentations at the 2010 Translation Automation Users Society conference can be found on YouTube at: http://bit.ly/taus2010program.


8. www.hutchinsweb.me.uk/Compendium.htm

American Foundation for Translation and Interpretation: Help Yourself by Supporting the Professions

The American Foundation for Translation and Interpretation (AFTI), a 501(c)(3) charitable organization, was established to help translation and interpreting professionals preserve their history, assist in their present work, and stimulate research and scholarship.

Give something back to the professions. Please make a tax-deductible contribution to AFTI today.

Send your check, payable to AFTI, to:
American Foundation for Translation and Interpretation
C/o American Translators Association
225 Reinekers Lane, Suite 590
Alexandria, VA 22314