Machine Translation: Friend or *Foe*?

By Jiri Stejskal, CETRA, Inc.

Marketing research companies conducting research in languages other than English, face several options when preparing their surveys in foreign languages. Often, cost is a major consideration, and the possibility of having the translation done for free, or for a nominal fee, is very alluring.

In order to cut costs, some companies turn to their bilingual employees, and others try machine translation (MT). Chris Durban, an American-born translator now based in Paris, succinctly describes the risks of using bilingual employees who are not trained translators. She observes that "Bilinguals speak two languages fluently, but are not necessarily good at moving information between the two, especially in writing. And experience shows that many people described as bilingual overestimate their communication skills altogether."

Risks Of Using MT

Machine translation has been around for longer than most people think. The first proposal for using MT was submitted by a researcher at the Rockefeller Foundation in 1949, and in 1954 the first public demonstration of an MT system was held in New York at the IBM headquarters. Much hope had been put into MT during the Cold War years, with disappointing results. Today, it is clear that while MT is here to stay, it has limited use and will not replace human translation anytime soon. Let us look at "Free Online Language Translator" by WorldLingo to demonstrate the limitations of MT.

Consider a fairly simple question that could be used in any questionnaire:

> *Thinking about this feature, please indicate your level of agreement with the following statements. Use a 5-point scale where a "1" represents "Strongly Disagree" and a "5" represents "Strongly Agree."*

After we pasted the text into the Free Online Language Translator, and select Spanish as the target text, we receive the following result:

> *Pensando de esta característica, indique por favor su nivel del acuerdo con las declaraciones siguientes. Utilice una escala de punto 5 donde discrepa un "1" representa "fuertemente" y "5" representa "convienen fuertemente."*

Using the same tool, but now translating the Spanish rendition back to English, we get this remarkable statement:

> *Thinking of this characteristic, it please indicates its level in the agreement with the following declarations. Use a scale of point 5 where a "1" differs represents "strongly" and "5" represent "agree strongly."*

Besides the obvious grammatical and stylistic errors both in the Spanish and the back-translated versions, we see that the **meaning** of the original statement has changed significantly. "Strongly Disagree" became just "strongly," thus rendering the scale meaningless. The request ("please indicate…") became a statement ("it indicates…") with a nonsensical "please" in the middle of it. In addition, during the translation process, the user had no way of indicating who the target audience would be. In the case of Spanish, it makes a big difference whether the translation is geared toward a panel in the US, Mexico, South America, or Spain.

GLOBALLY SPEAKING

Is Free Translation Too Good To Be True?

So Is MT A Foe?

There are situations for which MT is appropriate and can bring positive results. There are MT tools available that are far more sophisticated than the WorldLingo application we used in the example above, and many other "free" programs such as Google Language Tools and AltaVista Babel Fish. SYSTRAN provides off-the shelf products, and companies such as LanguageWeaver are offering licenses for the MT engines they develop for a substantial fee.

MT has been used successfully in a controlled language environment by Caterpillar and other companies. With vocabulary restricted to some 800 words and rigid syntax rules, Caterpillar's technical writers were able to produce English documents suitable for MT, yielding remarkable accuracy.

MT is also used by the US government to scan large amounts of data in order to identify certain topics or keywords; such documents are then typically submitted for human translation.

The European Commission uses its Machine Translation Service for a similar purpose, and informs the users that "[MT] cannot compete with the accuracy or quality offered by professional translators. However, it can still be helpful, particularly when time is at a premium. Machine output may be adequate, for example, when you need a quick overview of a document written in a language you do not understand."

"the 'Lead Story' becomes a 'Story of Lead,' as in lead paint."

It is not wise though, to try to turn this idea around to give a "quick overview" of your text to the world at large. Take for example the official site of the Philadelphia city government, which uses machine translation to translate its Web pages into multiple languages. The city's mayor, Mr. Street, becomes Herr Straße, Monsieur Rue, and Señor Calle, among other things, and the "Lead Story" becomes a "Story of Lead," as in lead paint.

To quote again from Chris Durban's booklet, "As a general rule of thumb, do not use raw computer output for anything outbound without the express agreement of your clients. It is simply not suitable: you run the risk of looking inarticulate. Even stupid."

In the marketing research industry, which typically involves creative writing, the use of MT is very limited. However, surveys do follow a particular structure and include repetitive language that is suitable for Computer-Assisted Translation (CAT). The basic difference between CAT and

MT is that in the former, the computer program supports the translator, whereas in the latter the translator supports the machine (or not, if the resulting translation is not edited). CAT is a broad term covering a range of applications used by translators to improve efficiency and quality of the translation process. The most popular applications are translation memory tools such as SDL Trados, Déjà Vu, or WordFast.

CAT tools are widely used by language services providers (LSPs) as productivity tools that enable the LSP to use consistent language within a project and across multiple projects. Because of the significant up-front investment and a steep learning curve, these are not viewed as cost-saving tools, but rather a means for providing consistent linguistic output.

Additionally, it is important to note that the translation memory is only as good as the translations entered into it by the translators, and that strict quality control is essential. As with MT, CAT tools also have limited use. For example, they are not intended for translation of creative text seen in marketing pieces or for translation of open-ended verbatims that are often already riddled with errors and lack any kind of structure.

Suggestions For Using MT and CAT Tools:

• Resist the urge to do it yourself, unless you have a translation department with a staff that understands the intricacies of MT and CAT tools.

• MT and CAT are productivity tools, not cost-saving tools. Free MT applications do not produce usable results; sophisticated MT and CAT applications that produce good results have a high price tag and a steep learning curve.
• MT and CAT tools are suitable for certain types of text only. Creative writing is least suitable for this type of automation.
• Un-edited MT output is a recipe for disaster. Even the best MT applications require editing by human translators.
• Both MT and CAT can be used successfully if utilized by language professionals who understand the limitations and inherent risks.

Evidently, machine translation is both a friend and a foe. Just like any complex technology solution, it can enhance productivity and quality, but can also do the exact opposite if not used properly. Much research and development is needed to design a good MT program capable of recognizing context and following syntactical rules, and therefore it comes with a price tag. Free MT programs which operate on a word level disregarding the context are a good source of office ridicule, but certainly not a tool for a professional researcher or writer.

Jiri Stejskal, Ph.D. is the President of CETRA, Inc. He can be reached at 215-635-7090 or jiri@cetra.com.