The Current State of MT

One Person's Perspective

Scott Bennett
scottb@logos-usa.com

Translation for Web sites with a single click! Free translation on the Internet! Instantaneous translation of e-mail and chat! Inexpensive translation software for your home or office computer! All powered by machine translation (MT).

The promises go on and on. Visit any of a dozen or more sites and see what wonders are available. It seems that the Net has brought us to the Golden Age of machine translation. For those of us who began working in the field under the shadow of the ALPAC Report, these seem to be heady times.

But before we get too excited, it is well to consider where we really are as we enter the new millennium. Machine translation has always suffered from false expectations both from its critics and from its proponents.

In this fairly short space I will give my views of machine translation as it stands today. These views are naturally general; there is insufficient space to cover each of the many products and MT-powered services now available .

Reality Check

To get a true picture of where machine translation is after over 50 years of research and production, it is necessary to do a reality check. I will present a series of realities which together yield a composite picture of the current state of the art in MT.

Reality 1: Machine translation systems are computer applications

This sounds obvious, but many people somehow miss this point. This simple fact means that MT programs run linearly and literally. Even parallel processing cannot get past this physical fact. Once the system has processed a given text segment (phrase, sentence, paragraph, whatever) it cannot return to reassess its initial analysis. Unlike human beings MT systems cannot revise a previous translation based on information found after the segment. All the clever algorithms – and they are clever – devised over the past decades cannot overcome this limitation.

As a result machine translation works best in those situations in which linear and literal translation is appropriate. Depending on the document this could be a substantial portion of what is to be translated. In such instances, MT serves to free the human translator of the onerous task of translating the unexciting and sometimes repetitive portions of the text.

Reality 2: Machine translation works best on unambiguous text

This has been a constant for a very long time in MT. The more ambiguous the text, the less likely the success with the software. Assuming a good system, machine translation is inherently well suited for technical documentation (paper or electronic) and other clear-cut writing. Conversely it is ill suited for writing in which subtlety or indistinctness is a virtue. Obviously literary pieces are going to fare badly with machine translation, as does much journalism and marketing prose.

And what about the Web pages for which we are offered translation with a single click? This is a gray area, depending mostly on the ambiguity of the page. Honestly, most Web sites are too cool for MT since they are filled with notoriously ambiguous marketing hyperbole, current slang and subtle cultural content. However, a site offering straightforward content may do well at the hands of an MT engine.

Reality 3: The dictionary is the key

At the heart of all machine translation systems lies the dictionary, the database containing the words the system must be able to process for any usable results. The issue is always coverage. However the dictionary is constructed, it must contain all the words which are in the text to be translated – and it must have the appropriate translation for those words in that text. Size is only an issue insofar as larger dictionaries tend to have better coverage. A 2000-word dictionary containing the requisite information for all the unique words in a text is sufficient for that text, although clearly not much use otherwise.

This has further implications in that it is necessary for the application to provide a means for the user to customize the dictionary for her or his needs by selecting the appropriate domain(s) and to update the dictionary fully to insure that the coverage is complete and suitable.

Reality 4: Not all languages are created equal (in MT)

Large-scale machine translation systems have been developed almost exclusively to meet the market demand. In addition to the obvious commercial markets, the various intelligence communities have created a market in their use of MT as part of their information-gathering endeavors. The result is that

only some language directions are available for robust machine translation .

It is easy to find the major commercial languages of Western Europe and Asia represented in multiple systems. The Foreignword.com list of available language directions indicates that around 50% of those available involve English, French, German, Italian, Japanese, Russian or Spanish. Beyond these, there is a sudden drop off in the availability of other European and Asian languages to say nothing of those of African and Middle Eastern countries. Some are being studied in academic institutions, but these are rarely available for general use. Addition of more languages to the mix is slow simply because the cost of development is high and the time needed long.

Reality 5: You get what you pay for

In spite of promises on the Web for inexpensive machine translation system for your home or office computer, the fact of the matter is that such systems rarely deliver what is needed for translation. No system, whatever the price, has reached the goal of FAHQT . The more robust system can achieve levels of output which make them potentially useful for translators.

The cost factor in machine translation often mirrors the robustness of the machine translation engine. Less expensive packages are generally based on what amounts to word-by-word translation followed by the application of algorithms which seek to get the right word order for the target language. More expensive packages tend to be based on parsing (analysis) of the source text followed by the synthesis (generation) of the target. These, then, seek to reflect in some way what a human translator does when she or he approaches the task.

One word of caution: price is relative. Most prices are fixed as much by the market as by the quality of the engine. Thus, inexpensive, with English <> Japanese systems tends to be higher than it is with English <> Spanish. The key is what the engine is doing; the price reflects this somewhat.

Reality 6: Translators need not fear the technology

For years MT has been viewed as a threat to human translators. This notion should be put to rest immediately. As is evident in the preceding sections, over fifty years of work have not brought this technology anywhere near the point that it can endanger the livelihood of human translators.

The fact of the matter is that in particular settings machine translation has proven to be a production tool for the human translator. Case studies, such as those presented in a number of language industry publications , highlight the use of machine translation as part of a corporate language translation environment. In such situations human translators combine with MT to produce more publication quality translation in shorter periods of time. The person remains in control; the system merely provides output which can be revised for publication.

In some instances machine translation may be useful for gisting or information level translation. Here the goal is not a publishable text, but simply to get an idea of what the content of the particular document is. The popularity of the AltaVista translation engine Babel Fish indicates that there is considerable demand for this sort of translation. However, this kind of work is not something professional translators do for a living.

Reality 7: Machine translation is not plug-and-play

In an era of easily installed applications, machine translation systems remain demanding. Most install easily enough, but then require considerable upfront investment before providing the return on investment expected. A key factor in this is the matter of the dictionary, discussed above. Someone will have to insure that the dictionary covers the necessary words – with the required translation for the domain. It may take hours to weeks to add the words depending on the system.

And there is the issue of the process. Machine translation systems may fit neatly into an existing translation workflow – or not. Adopting the technology to a work environment may mean significant adjustments to the environment.

Other factors may also come into play. The key is to have done one's homework in advance to insure that the technology will provide the sort of return-on-investment that warrants the investment of time and resources to get it going.

A New Millennium for MT?

My reality check provides a means to see where machine translation technology is in the year 2000. Years of research and development along with honest promises and less-than-honest hyperbole have brought us to a point where there are a lot of systems, some of which may provide some solutions to the challenges of the modern translation market.

I have attempted to provide an honest view of where we are now and what the technology can do for the language industry. We have come a long way in the technological aspects of the systems, but even more in the potential of the applications to assist human beings in getting translation done. Machine translation has proven to be useful both for publication level translation and in providing a gist of a particular document.

However, the bottom line is that machine translation is not for everyone in every translation setting. Intelligent and informed consideration of it is worthwhile for those who are facing tight deadlines and ever growing workloads.

There has been real progress in the first fifty+ years and more will come. While it is not for everyone, machine translation can rightly take its place as a tool for translation. ■