

Do You Have a Translation Tool Strategy?

ANALYSIS BY
TERENCE LEWIS

Machine Translation

When evaluating potential software tools in the translation process, it is safe to start with this guiding principle: don't use machine translation (MT) or other tools when they provide unusable output.

But usability is relative—it varies according to the purpose of the document, both for the author and the reader. And purposes may differ: a document that expresses a burning issue for the author may be considered irrelevant by the reader. Or a discarded draft dashed off by a specialist may be an essential piece in a particular reader's intellectual jigsaw puzzle. There are many cheap-and-cheerful translation packages on the market. The following translation was produced by one of them:

German original:

In jedem Fall ist eine Abstraktion und Reduzierung der Objektinformation vorzunehmen. Eine derartige Objektbeschreibung wird als Modellierung bezeichnet. Möglichkeiten zur Datenakquisition bieten Meßmaschinen, Methoden der Photogrammetrie und geodätische Verfahren.

English output:

An abstraction and reduction of object information is to be carried out always. A such object description is designated as a modeling. Possibilities for the data acquisition offer measuring machines, means of photogrammetry and geodesic procedures.

This is an example of very poor MT output that is still of some informational value to a reader with no knowledge of the source language. If that reader finds it gives him the knowledge he is seeking, he should use the MT software in question; if not, he should leave it on the shelf.

The above piece of information may fill a gap in somebody's knowledge. So one criterion might be that we should decide against using translation tools if they leave large gaps in our knowledge. A text of foreign specialist terms punctuated with translated determiners, verbs, and adjectives is hardly going to enlighten and inform. This degree of inadequacy is less likely in a controlled or restricted domain, where it is nearly always possible to generate useful machine-translation output. By definition there is near-total lexical, syntactic, and semantic coverage. Weather forecasts, avalanche bulletins, job descriptions, and highway traffic flow: in all these areas, few terms are involved and the output is likely to be reliable and valuable. But there would be little point in submitting an aircraft maintenance manual to the TAUM METEO system.

In practice, the use of any MT program does not make much sense if the lexical deficit is disproportionately large in relation to the length of the text. If we have a text of, say, less than five pages with more than 30 unknown words, it is usually more cost-effective to have the text translated by a human than to go through the business of inputting all those new entries. But that's a judgment call. If the subject is likely to crop up again, the dictionary update time might be considered a sound investment.

Most purchasers of MT programs are likely to use their software in a general translation environment. Experience has shown us that there are a number of disciplines where we are better off using approaches to translation other than MT. If we take a basic PC translation package straight out of the box, it will give a varyingly useful translation of a technical text. Legal texts, however, seem to provide problems for all systems that are not speci-

cally designed to deal with them. Consider the following example:

English original:

Agreements providing for both joint research and development and joint exploitations of the results may fall within Article 85 because the parties jointly determine how the products developed are manufactured or the processes developed are applied or how related intellectual property rights or know-how are exploited.

German output:

Übereinstimmungen, die für das beiden Gelenk sorgen, forschen, und die Entwicklung und exploitations von den Ergebnissen können innerhalb Artikel 85 fallen, weil gemeinsam die Parteien entscheiden, wie die Erzeugnisse, die entwickelt werden, hergestellt oder die Prozesse, die entwickelt werden, angewandt sind oder wie verwandte geistige Eigentumsrechte oder verwandtes geistiges Know-how ausgebeutet sind.

French output:

Les accords fournissant les deux développement et recherche conjointe et exploitations conjointes des résultats peuvent tomber dans L'article 85 parce que les fêtes déterminent conjointement comment les produits développées sont fabriquées ou les procédés développées sont appliquées ou comment liées des droits intellectuels de propriété ou connues-comment sont exploitées.

Sentences including phrases such as "how the products developed are manufactured or the processes developed are applied" would confound most off-the-shelf German-English packages, although the French-English package used here gets it right after completely misreading "parties" (perhaps the machine had one too many the previous night). In fact, we see that different sentences cause different problems for the two languages. Of course, the ridiculous German translation of "joint research and development" (the machine, perhaps still a bit tipsy, appears to be saying something about a knee joint, or something of the sort) could easily be avoided by coding the four words as a single entity.

Generally speaking, legal texts contain a multiplicity of clauses, often without any commas. This causes problems when translating into or from languages in which the verb is thrown to the end of the clause after a relative pronoun or a conjunction. Of course, the combination of MT and TM approaches would certainly reduce likelihood of errors. Strings such as "shall not deposit the proceeds of the sale" and "save in such circumstances" from a contract document could be entered into the translation archive and would then be delivered up as fixed entities. In our experience, this is certainly the most cost-effective way to "machine-translate" large volumes of legal text.

With legal translation, it is often a question of building a bridge between two quite different legal systems. False friends can very easily send the

reader down the wrong path. The distinction between the Spanish "Escritura de compraventa" and the "contrato privado de compraventa" is quite alien to English law. The idea that you can legally purchase a property under a private agreement while the title deed remains in someone else's name would strike most English lawyers as a preposterous proposition. Yet it is common practice in rural Spain. So even if an MT system got the translation of a term right we have to ask ourselves what use the translation would be to a reader who did not understand these differences between English and Spanish legal systems. Another aspect of legal translation is that sometimes much hangs upon the precise interpretation of a single word. In view of this, one would have to question the usefulness of a computer-translated legal text unless it were thoroughly revised by somebody with a knowledge of the two legal systems involved.

A further problem is that of cross-disciplinary documents—for example, a legal document concerning highly specialized photogrammetric equipment to be used in geological surveys of rock faces. If the MT system does not allow the user to select at least three customizable subject dictionaries, the computer version of such a document would be of questionable value and the translation would be more usefully entrusted completely to a human translator.

Challenges and Constraints

We have looked at domains and disciplines. A further factor that plays a part in the decision on whether to use MT or any other translation tool is the purpose of the document and the purpose of the translation. A translation may face numerous challenges and constraints:

The translation may be needed to assess the creativity of the original.

A creative translation may be required.

The translation should surpass the original.

The translation will become the new original.

Cultural bridging is required.

Let us look at each of these points. If we need to assess the creativity of the original—say a sales brochure—it's unlikely that the computer translation is going to capture this. Only a human translator would be able to recognize creativity and interpret it for the target language. Then, the translation itself may need to be creative. It may need to relate to the target audience, live and breathe the culture of the target audience. No matter how good or bad a computer translation is, it will shape the final draft of the document. In fact, the better it is, the less likely it is that the reviser will make any drastic changes. In our experience, the good machine translation seduces the reviser into complacency. If the translation is grammatically and semantically correct, the MT reviser is loath to scrub it out and recast it sentence by sentence.

(continued on page 18) ▶



Terence Lewis

The Bottom Line:

.....
*If a computer
can beat a chess
grand master,
why not teach
it to translate?
One reason: there
seem to be faster
ways to make
a buck.*



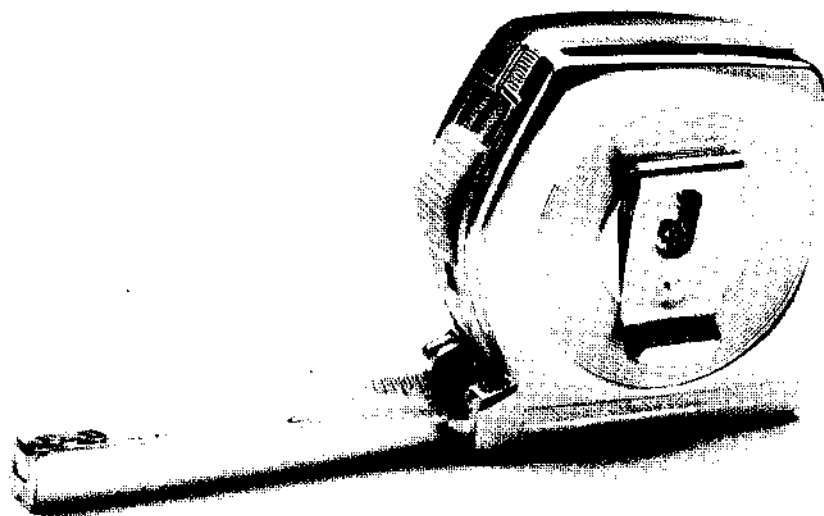
Another thing the MT program cannot do is to assess how much of what is assumed in the source text needs to be brought out or clarified for the target audience.

In some situations, the translation will become the new original. This might happen in a multinational company where the author does not write the corporate language well enough for his report to serve as a master document that will be translated into many languages. If the requirement is for a vibrant, confident, authoritative text, the translation would better be done by a human translator from the outset.

As for cultural bridging, it involves identifying concepts for which there is no precise equivalent in the target language, or references which the intended readership might not fully understand. If we are translating the term "Notaire" from French into German, no cultural bridging is needed. But "Notaire" does not fully correspond to the English "notary" and it would be misleading if not outright wrong to translate it as such. If a document contains a large number of terms that would have to be explained in detail to the target audience, it is probably more useful to have it translated by a human translator. On the other hand, if we were designing a system to be used exclusively for legal translations, such explanations and distinctions would ideally be included in our sub-language dictionary module. Situational decision-making is extremely critical when assigning tasks to a set of translation tools.

Where the Buck Starts

Basically, however, it all comes down to money. If Deep Blue can search through the ramifications of hundreds of thousands of chess moves in seconds, it is possible to search corpora containing all possible combinations of a limited number of words and select the statistically most appropriate solutions to a particular lexical or semantic problem.



Why hasn't anyone built a commercial program to do this? There seem to be faster ways to make a buck.

In work environments, translation managers have to make the most cost-effective use of the translator's time. If a translator can produce 1,000 words of fully checked text an hour using MT or TM to generate a draft document and is costing me, his employer, around £20 an hour, and I can sell that text at the rate of £150 per thousand words, that translation tool is giving me a gross margin of £130 per 1,000 words.

When considering whether to use or not to use translation tools in a professional translation environment, we have to decide whether:

- a translator with a workbench tool,
- a translator revising MT output,
- or a translator who knows the subject and uses dictation software

is going to produce the greatest quantity of acceptable translation at the lowest cost.

We come back to the question of what kind of translation is needed. Even poor MT output can be rendered usable with enough post-editing. If the customer will pay £50 per thousand words for lightly revised computer output and the translator can post-edit at a rate of 2,000 words an hour, I am grossing less than when I can sell "ready-to-use translation" at £150 per thousand words. But if the customer will pay £15/1,000 for pretty raw computer output and the system can produce 20,000 words an hour, that is far more profitable. At the end of the day, the question of when and when not to use MT comes down to what the customer wants and what he's willing to pay for.

Terence Lewis is a director of Hook & Hatton Ltd, a UK company specializing in technical translation and MT services via email. He gave an earlier version of this article as a talk at the European Association for Machine Translation meeting held earlier this year in Copenhagen.

"At the end of the day, the question of when and when not to use MT comes down to what the customer wants and what he's willing to pay for."