Rethinking Interaction: The solution for high-quality MT?

Elliott Macklovitch
RALI Laboratory, Université de Montréal
C.P. 6128, succursale Centre-ville
Montreal, Quebec, Canada H3C 3J7
macklovi@IRO.UMontreal.CA
Antonio S. Valderrábanos
Software Engineering Division - Sema Group sae
Albarracín, 25, 28037 Madrid, Spain
antonio.valderrábanos@sema.es

Abstract

Our focus is on high-quality (HQ) translation, the worldwide demand for which continues to increase exponentially and now far exceeds the capacity of the translation profession to satisfy it. To what extent is MT currently being used to satisfy this growing demand for HQ translation? Quite obviously, very little. Although MT *is* being used today by more people than ever before, very few of these users are professional translators. This represents a major change, for a mere ten years ago, translators were still the principal target market for most MT vendors. What happened to bring about this change? For that matter, what happened to most of those MT vendors? The view we present is that the most promising strategy for HQ MT is to embed MT systems in translation environments where the translator retains full control over their output. In our opinion, this new type of interactive MT will achieve better acceptance levels among translators and significantly improve the prospects of MT's commercial success in the translation industry.

The Context

In this paper, we will point to what we think is the most promising strategy for bringing the production-enhancing potential of MT to bear on the problem of producing HQ translations of unrestricted texts. In order to place our vision in context, we will first outline in this section the trends that, in our view, will shape the translation market and industry during the next decade. Then, in the next section, we will define the position of MT in this general picture.

-High Quality. As a result of market globalization, translated materials will be a key component of every product and service sold worldwide. Quality standards will be set by end-users of those translations, not by other intermediaries.

-Industrialization and Automation. Translation is currently one of the major bottlenecks for shortening time-to-market cycles in different industrial sectors like the software industry. In an effort to overcome this impediment, the task of translation will increasingly be modeled as an industrial process. Standardized workflows will be set up to increase quality and reduce costs, and insofar as possible, every step will be automated.

-Leveraging. Every new translation task will build on previous translations (and other resources possibly extracted from these same translations). Leveraging will also facilitate quality controls and reduce costs.

The tool of the future

To the extent that they respond appropriately to these market trends, MT technologies will significantly increase their penetration in the commercial arena. However, acceptance by translators remains a key issue. Contrary to the use of MT as a gisting tool, collaboration between translators and MT technologies will continue to be indispensable for the production of HQ translation – and new models of collaboration will have to be developed.

We see the translation tool of the future as a new type of Interactive MT comprising the following elements:

-It should be driven by human translators, in order to ensure HQ translation levels. We do not see MT technology producing HQ translation without the help (control, assistance, training) of humans except in very restricted sublanguages, like weather bulletins (Chandioux 1989) or travel arrangements.

-It will incorporate machine learning techniques (like statistical or example-based MT) in order to better exploit previously translated materials and make the process more cost-effective and faster (as in any other industrialized task). The first forms of this type of technology are already being used to facilitate tasks like bilingual lexicon acquisition.

-It will include quality assurance and control tools (e.g. spell, grammar and style checking assistants; tools

for terminology consistency checking, term extraction, etc.)

Different research prototypes have been developed that target some of these points. For example, the TransType system (Langlais et al. 2001) integrates a statistical translation engine into a translator's editor. However, much work remains to be done before these prototypes find their way into the hands of working translators and can be integrated into standard production environments.

As for recent trends in technology, we see two significant developments. On the one hand, Computer-Assisted Translation (CAT) tools, like Translation Memory systems, have achieved significant commercial success in recent years and are becoming a standard for the translation industry (both for agencies and free-lance translators), despite the fact that these systems are relatively simple from a technical point of view, being based on string matching techniques. In our opinion, a significant part of their success lies in the fact that CAT tools give translators full control over the output text.

On the other hand, machine learning techniques offer the possibility to develop MT systems in a very cost-effective way, both in terms of time and money. We expect significant improvements in this area, not only in the development of statistical MT systems, as is currently happening (Al-Onaizan et al. 1999; Knight 1999), but also in the development of symbolic systems. We will see symbolic MT systems for which analysis or generation grammars have been acquired (semi-)automatically (Ramshaw and Marcus 1995; Veenstra and Daelemans 1999), and this trend could well extend to transfer modules. If so, this will significantly facilitate and speed up the production of linguistic components which are extracted from (or trained on) previous translations. As a result, it will be possible to customize these components for specific text types and styles, allowing for the production, for example, of corporation-specific resources. The resulting MT engines will have higher flexibility and will more easily adapt to the translator's needs. What are now expensive resources may become "cheap" components developed on demand.

Combining these two observations, we see that CAT tools provide a well-developed and widely used translation environment that enjoys a good reputation among translators and in the industry for their contribution to productivity. Hence, future versions of CAT tools should provide a suitable environment in which new MT engines can be integrated, working under the supervision of human translators. These MT engines will have to be able to learn from the translator's practice and adapt to her/his needs.

If this integration is achieved, we will see a new model of collaboration emerge for man-machine interaction. This model can be described as a new form of Interactive MT where the interaction is focussed on the target language, and not on the source side as has been the case in the past (Kay 1973). This new model will avoid the current problem of post-editing (where the translator has to correct the mistakes of an MT system) by allowing a more natural interaction between the MT system and the translator. Furthermore, it has the potential to significantly increase the output of human translators in the production of HQ translations of unrestricted texts.

References

[Al-Onaizan 1999] Al-Onaizan, Y., Curin, J., Jahr, M., Knight, K., Lafferty, J., Melamed, D., Och, F.-J., Purdy, D., Smith, N. A., and Yarowsky, D. 1999. Statistical Machine Translation, Final Report, JHU Workshop 1999. Technical Report, CLSP/JHU.

[Chandioux 1989] Chandioux, J. Météo: 100 million words later. In D.L. Hammond, editor, American Translators Association Conference 1989: Coming of Age, pages 449--453. Learned Information, Medford, NJ, 1989.

[Kay 1973] Kay, Martin. The MIND System, in R. Rustin (ed.), Natural Language Processing, Algorithmics Press, New York, 1973, pp.155-188.

[Knight 1999] Knight, K. 1999. A Statistical MT Tutorial Workbook. Ms, August 1999.

[Langlais et al. 2001] Philippe Langlais, George Foster and Guy Lapalme (2001), Unit Completion for a Computer-aided Translation Typing System, to appear in Machine Translation, Kluwer, 20 pages.

[Ramshaw and Marcus 1995] Lance A. Ramshaw and Mitchell P. Marcus, Text Chunking Using ransformation-Based Learning. Proceedings of the Third ACL Workshop on Very Large Corpora, Cambridge MA, USA, 1995.

[Veenstra and Daelemans 1999] Jorn Veenstra and Walter Daelemans, Cascaded Grammatical Relation Assignment. Proceedings of EMNLP/VLC-99, University of Maryland, USA, 1999.