

Translation Memories, Linguistic Assets and the Commoditisation of Translation

João Carlos Antunes Brogueira

Faculdade de Letras da Universidade de Lisboa, Portugal

PhD program in Translation and Intercultural Studies, Universitat Rovira i Virgili, Tarragona, Spain

icbrogueira@octante.net

Abstract

Translation technology has come a long way over the last three decades. Along with the development of technologies such as translation memories (TM), terminology managers, machine translation (MT) systems, content management systems (CMS), and globalisation management systems (GMS), huge repositories of translation knowledge have been built. Since the information stored in these systems is intended for reuse, it is crucial that it be error-free. Therefore, it was only a matter of time until some companies started offering specialized TM services that include management, maintenance and quality control. More recently, linguistic asset brokerage was introduced in the marketplace¹. Are we now on track to see Arabic or Romanian listed alongside Arabica or Robusta on euronext.liffe? This paper will look at how translation technology is enabling a commoditised perception of translation in software localisation. We will analyse to what extent so-called linguistic assets might be en route to becoming exchangeable commodities and how this may impact «developing languages» and the status of translation.

Key words: technology, translation memories, commoditisation.

What is a commodity?

A commodity is «any tangible item that can be bought and sold»². Commodities are usually basic resources or raw materials traded in bulk, «primarily on the basis of price, and not on differences in quality or features. (...) Manufactured goods are said to be commodity goods if purchasing decisions are made almost solely on the price of the product»³. Commodities can be divided into hard (e.g., mining products) and soft commodities (e.g., agricultural products).

Commodities are traded in specialised exchanges such as the NYMEX, LIFFE or TOCOM, usually via futures contracts.

¹ <http://www.tmmarketplace.com/>

² <http://www.lasallebank.com/investments/glossary.html>

³ <http://www.mastercardbusiness.com/mcbizdocs/smallbiz/finguide/glossary.html>

A market for linguistic commodities

Fuelled by an ever increasing pressure to deliver fast-growing volumes of multilingual content across several regions in less time, while striving to keep costs under control and ensuring high quality standards, several companies have ventured into what has come to be called the «localisation industry». In little over twenty years, these new market players put on an impressive display of mergers and acquisitions, which have completely changed the face of the translation business. The traditional small and medium-sized enterprise business environment developed in the direction of more mature business areas and recent buyouts have created the first corporations in this segment.

Equally impressive was the development of technologies to assist in the processing and delivery of multilingual content. Localisation was «born» in the USA when the IT industry became aware that products needed to be made «linguistically and culturally appropriate to the target locale (country/region and language)» (Esselink 2000: 3) in order to increase sales overseas. This association with IT enabled the localisation industry to benefit from the hype surrounding the technology sector in the late 1990's, getting funding and support from major players and also from investors anxious to put their money into anything related to the Web and globalisation. When the Internet bubble ⁴ burst in late 2000 and through 2001, cost-effectiveness became a very serious issue for the clients of the localisation industry.

Conversely, localisation companies had to find new ways of expanding their businesses in order to ensure growth and profitability. They thus enlarged their portfolio of services to include the full cycle of content globalisation and localisation, developed new technologies and refined workflows. Software developers, on the other hand, were working hard to develop and refine systems that maximize the reuse of previously translated chunks of information, including computer assisted translation (CAT) tools, CMS and GMS and, to a lesser extent, MT systems. With projects in localisation nowadays reaching in excess of tens of millions of words per language pair, it is not hard to imagine the breadth of translation knowledge accumulated in the repositories fed by the above mentioned systems. It was just a matter of time until someone came up with the idea of trading this data.

Technologies and standards enabling the commoditisation of linguistic assets

In their often two-faced advertising strategies, CAT software publishers make abundant use of taglines such as «Never translate the same sentence twice» (Trados) and «You only need to translate the segments and words that are different» (Star). Although most translators know that such statements are not entirely true, this mindset about translation appears to be gaining ground among translation and localisation clients. Most CMS and GMS suites actually rely on the principle that once a particular segment (a user-defined set of bi- or multilingual data) is committed to the TM and validated, it will never have to be checked again and it can be reused regardless of context and target text production conditions. Some TM systems even feature functions that will only export untranslated or fuzzy matched segments and send these loose bits of source text (ST) to the translator. The conscientious translator, however, is not discharged from «reading many paragraphs to understand the context each time» (Translator X 2004), often in a third party application. The gain is thus partly achieved at the expense of the translator's productivity.

On top of this rather mechanic perception of translation being fostered by some technologically-oriented visionaries in the localisation industry, recent developments in character set standardisation (Unicode) and the latest Translation Memory eXchange (TMX) specifications have enhanced the portability of TM contents among TMX-compliant translation memory systems. Based on XML, the TMX standard offers great flexibility both in

⁴ http://en.wikipedia.org/wiki/Internet_bubble.

terms of character set compatibility (Unicode) and also in terms of extra information that can be inserted in the file as <prop> tags (Musale 2004: 9). It is now technically possible to totally or partially import or export translation memories using any TMX compliant tool, so why not multiply the profit by reusing its content as well as by selling it? Or save time and money by acquiring ready-made memories?

Trading knowledge or information?

Terminology lists and databases, dictionaries and lists of words (for spellcheckers, e.g.) have long been available on the market for whoever is willing to pay for them. Some are also available for free, such as EuroDicAutom or Microsoft glossaries. Translation memories, on the contrary, are usually for internal company use only and translators are often required to sign non-disclosure agreements to be able to access the memories. Not least of all because of the delicate situation in terms of Intellectual Property, both pertaining to source and target text (Zetzsche 2005).

This difference in treatment of seemingly similar linguistic assets coincides with the distinction between information and knowledge as laid out by Budin (2002). For terminology or dictionary entries are information that is converted into knowledge during the cognitive appropriation process of translation and also during text production, which is then stored in the TM.

«The focus and real goal of knowledge management is actually on content [...] the concepts and the messages. When knowledge is then packaged as a product for a certain audience, presented in certain media presentation forms, then we can speak about content» (Budin 2002)

Moreover, according to Cadieux (2004), «there is a parallel between content creation and translation since translating is really a form of authoring». In this sense, then, translation is knowledge. Another striking difference between terms (information) and translations (knowledge) lies in their very nature: terms are self-contained discontinuous entities whereas translated segments require careful attention to context and co-text to ensure consistency and continuity. In other words, it is much simpler to successfully use terms as raw materials for translation than previously translated segments, as these require a much higher degree of correspondence to be cost-effective.

My style, your style?

Zetzsche claims that «no reasonable translator tries to introduce a new style or terminology just because it is different from a competitor unless the source text makes him or her do that» (2005). While we do not wish to challenge Zetzsche's positive view of the reasonability of translators, his claim is oversimplistic. In the Portuguese market, for example, Oracle and Microsoft use different sets of equally valid terminology, and the differences go as far as using different translations for 'Save'. When laying the foundation of their localisation programmes into Portuguese (European), most companies developed an internal style guide that tends to follow Microsoft's. However, Microsoft's style guide for Portuguese (European) has proved to be quite volatile and over the last two years two major changes have wrecked havoc in TMs: a change in the translation of 'double click, to' and the adoption of ST capitalisation rules. As a result, most style guides are now out of sync. TMs have followed the same path, and while Microsoft may be willing to repair the translations in their own TMs, other vendors have been reluctant to change their existing translation memories and pay for a change they never asked for. Buying a TM from a competitor under these conditions seems to be a big step in the direction of failing Quality Assurance (QA) requirements.

It is thus clear that for translation memories to become highly valuable assets and worthy of being traded in a specialized market, they must first comply with one of the basic

definitions of commodity, namely that they do not display differences in quality or features. This would entail the highest degree of standardisation in the production of content across any given industry, as well as a similar degree of standardisation in the localisation of such content. This standardisation might, in turn, be perceived by the end client as an uniformisation of content (and features), thus resulting in reduced visibility for R&D efforts and advances in technology. And if there is one thing that IT companies want to convey in their products, it is their technological supremacy.

On top of that, a considerable part of the so-called «functional texts» have moved in the direction of greater user-friendliness, especially in the mass-consumer market, hence less standardized. With companies in the same industry competing for different segments, standardisation might be more of a pitfall than an aid. And to what extent are companies willing to relinquish a substantial part of their brand image in exchange for cost reductions in the development cycle of a product? According to Brooks (2000: 45), translation represents 35 % of the total cost of globalisation spending at Microsoft, which, in turn, is a fraction of the total cost of developing a product. By releasing or selling their TMs, large corporations would be inviting freeloaders to capitalize at the expense of their own translation knowledge, which combines language, marketing and technical skills.

Widening the digital divide

If all standardisation requirements are met, purchasing a translation memory might actually prove to be a cost-effective alternative for Small and Medium Enterprises (SME) to use in purely functional texts, instead of jump-starting and maintaining their own memories. However, such resources are only widely available for so-called major languages (European languages, Arabic, Chinese and Japanese), as well as the tools that use them. In fact, «there isn't much evidence that the major providers of LE software will turn their attention towards NIMLS [non-indigenous minority languages]» (Somers 2003: 88) and, e.g., only recently has support for Bengali – a «minority» language of 100,000,000! – been implemented in the SDLTrados product line (formerly Trados)!

This means that in the foreseeable future high-street, TMX-compliant translation memory packages are likely to remain too expensive or non-existent for «minority» languages.

Conclusion

Heralded as an economic opportunity with «astronomic» potential (Zetzsche 2005), sharing translation memories still faces many hurdles until it becomes a viable alternative to the traditional process of creating and maintaining a TM. On the one hand there are still a number of unresolved Intellectual Property issues; on the other hand there is little empirical evidence to support such bold statements as Zetzsche's. Moreover, the standardisation that has been mentioned in the previous sections also implies accepting the dominant positions of the key market players. The success of Windows terminology suggests that companies feel comfortable with incorporating third-party standardized information in their products. Whether they are willing to abide by someone else's authoring and localisation guidelines and relinquish all decision power regarding source text production as well as its localized versions (for source and target must match the TM supplier's style and terminology for optimal results) is a different matter.

From the perspective of translation, this attempt at commoditising linguistic assets represents the death of the holistic dimension of the text. Translating becomes the mere replacement of text strings and the translator fills in the gaps and, eventually, polishes the final product. Since some localisation and globalisation tools enable the leveraged content to be locked, thus preventing anyone from making any changes, the translator is further disempowered in favour of standardisation and reduced indeterminacy.

One final word of scepticism regarding the interest of major companies in trading their TMs. Large corporations would probably be best placed to share their translation memories and make a profit out of this venture. However, in recent years the trend has been to outsource all localisation work and translation is not seen as a core skill in these companies (Brooks 2000: 50). Could the potential profit from this business really be that significant for these companies to be interested in this non-core operation?

References

- [1] Austerlühl, Frank, *Electronic Tools for Translators*, St. Jerome, Manchester & Northampton, 2001.
- [2] Brooks, David, «What Price Globalisation? Managing Costs at Microsoft», in *Translating into Success. Cutting-edge strategies for going multilingual in a global age*, Sprung, Robert C. (ed.), John Benjamins, Amsterdam and Philadelphia, 2000, p. 43-57.
- [3] Budin, Gerhard, «Global Content Management – Challenges and Opportunities for Creating and Using Digital Translation Resources», <http://www.ifi.unizh.ch/cl/yuste/postworkshop/repository/gbudin.pdf>, 2002, visited September 2005.
- [4] Cadieux, Pierre, «Globalisation Is Here!», *Globalisation Insider XIII* (2.1), <http://www.lisa.org/archive/newsletters/2004/2.1/CattindesBois.html>, April 2004, visited July 2006.
- [5] Esselink, Bert, *A Practical Guide to Localisation*, John Benjamins, Amsterdam and Philadelphia, 2000.
- [6] Gordon, Raymond G., Jr. (ed.), *Ethnologue: Languages of the World, Fifteenth edition*. SIL International, Dallas, 2005. Online version: <http://www.ethnologue.com/>.
- [7] Musale, Shailendra, «Getting More From Translation Memory», *Localisation Focus* 3 (1), Limerick, March 2004, p. 9-10.
- [8] Pym, Anthony, *Contemporary Translation Theories. A Coursebook*, unpublished manuscript, forthcoming.
- [9] Somers, Harold, «Translation technologies and minority languages» in *Computers and Translation: A translator's guide*, Harold Somers (ed), John Benjamins, Amsterdam and Philadelphia, 2003, p. 87-103.
- [10] Sprung, Robert C. (ed.), *Translating into Success. Cutting-edge strategies for going multilingual in a global age*, John Benjamins, Amsterdam and Philadelphia, 2000.
- [11] Translator X, «Towards a more equitable pricing structure: A translator's perspective», *Globalisation Insider XII* (1.4), <http://www.translationdirectory.com/article439.htm>, February 2003, visited July 2006.
- [12] Zetzsche, Jost, «Translation Memories: The Discovery of Assets», *Multilingual Computing & Technology*, 16 (4).

BIOGRAPHICAL NOTE

João Brogueira holds a degree in Modern Languages and Literatures from the University of Lisbon, where he also pursued a post-graduation in translation. He is currently enrolled in the PhD program in Translation and Intercultural Studies at the Universitat Rovira i Virgili, Tarragona Spain. His research interests include localisation, translation technology and norms. He teaches translation technology at the Faculty of Letters of the University of Lisbon and he also works as a freelance translator in the fields of localisation, finance and architecture.