The Value of Machine Translation for the Professional Translator

Elina Lagoudaki
Department of Humanities
Imperial College London
London, United Kingdom
elina.lagoudaki@gmail.com

Abstract

More and more Translation Memory (TM) systems nowadays are fortified with machine translation (MT) techniques to enable them to propose a translation to the translator when no match is found in his TM resources. The system attempts this by assembling a combination of terms from its terminology database, translations from its memory, and even portions of them. This paper reviews the most popular commercial TM systems with integrated MT techniques and explores their usefulness based on the perceived practical benefits brought to their users. Feedback from translators reveals a variety of attitudes towards machine translation, with some supporting and others contradicting several points of conventional wisdom regarding the relationship between machine translation and human translators.

1 Introduction

It was long thought that MT technology was something that should be kept separate from professional translation activity, as it aspires to be able to produce the same quality of translation as humans do, therefore posing a threat to their profession. However, things seem to be changing lately. As Champollion [2003] and O'Hagan and Ashworth [2002: 43] report, the use of MT is now considered a common practice among translators who prefer to have a rough draft of a translation before they produce a final translation, by editing the first draft. This also turns out to be the case within the translation departments of the European institutions as suggested by information from the EC Directorate-General for Translation [2007:10-11]. Evidence like this suggests that the user-base of modern machine translation applications is gradually expanding to include even their former rivals.

In the past two decades, several attempts have been made to combine TM and MT systems, to compensate for the limitations of the former and boost its productivity. A cursory look at the translation software market reveals that it is quite common to see TM systems with the ability to port to a MT system and vice versa. A successful example of such an integration of TM and MT is the EURAMIS system used by the European Commission since 1995 [EC DGT, 2007]. Translators who use this system are offered various match retrieval options, including the possibility for a match request from an integrated machine translation system.

The idea of bringing machine translation closer to TM systems has also existed for a long time in the realm of academic research as a possible way of expanding the capabilities of TM technology and of exploiting fully the available resources in the TM repository [see Hodász et al., 2004; Simard, 2003; Huang et al., 2003; Carl et al., 2002; Carl, 2000]. Gradually, TM developers started investigating how a TM system could employ MT techniques (instead of porting an MT system to a TM system), so that it would be able to suggest a match, in an autonomous way, when the TM repository did not contain an identical or similar target language segment. It did not take long before MT techniques found their way into commercial TM systems.

2   Commercial TM systems with MT capabilities

The first commercial TM system known to have implemented EBMT techniques in a dynamic sequential manner to construct a match is DéjàVu X[1]. When the system cannot find an identical or similar match in the segments of the TM repository using string-based fuzzy techniques, it then looks at the parts of each segment. If the system finds for a source segment two sub-segments existing in two different segments stored in the TM database, it puts them together to form a new segment that will be suggested as a match (the feature is called 'Assemble from portions'). The system can also carry out replacements of both invariable parts (numbers, proper nouns) and variable parts with declensions. However, this is a mere 'find and replace' process, which means that the new match has little chance of being a grammatically correct segment, especially if it has been generated from a highly inflected language.

The MT capability of the system draws on three essential prerequisites: a) that the system is able to identify matches at the sub-sentence level, b) that it is able to score multiple fragments found as a translation for a portion of the source segment, and pick the best and c) that a combination algorithm is employed for inserting the fragments in the correct place within the match that is going to be proposed. The second point relates to the issue of overlapping translation examples contained in a TM database, as identified by Somers [1999: 121]: "Some examples will mutually reinforce each other, either by being identical, or by exemplifying the same translation phenomenon. But other examples will be in conflict: the same or similar phrase in one language may have two different translations for no other reason than inconsistency." The same issue becomes even more complex when the system is confronted with word ambiguity (i.e. words that have many meanings, and which might therefore have many translations). Since no semantic analysis is performed, the system must rely on its statistical scoring system in order to decide which target fragment to use.

On the positive side, the match assembly capability of the system is further enhanced by the comparison of source segments not only to the TM and termbase entries, but also to a third database containing word lists called 'lexicon', thus exploiting more information. This helps, also, in building more complete target segments, since the 'lexicon' normally contains general purpose words like 'and', 'or', 'from', etc. Obviously, in order for the system to be able to suggest useful combinations of fragments from across databases, it is essential to have a sufficient number of entries stored in each database.

Déjà Vu X's example of integrating MT techniques into a TM system can be seen as an early attempt to combine the strengths of the two technologies in a seamless way for the benefit of translation professionals. While automation is attempted in a process reserved for the human intellect (i.e. text generation), the function is offered purely for the purpose of assistance and comes as an option, which may potentially add value without compromising the control of the user over the system. What is more important, Déjà Vu X does not make use of any external linguistic information, which means that the system offers the same matching function irrespective of the language combination.

It is interesting to note that Déjà Vu X chose to implement EBMT techniques instead of any other MT technique, namely rule-based, transfer-based or statistical. The obvious explanation is that an EBMT technique (just like a statistical technique) can work without the inclusion of language-dependent rules. It is also widely acknowledged that EBMT techniques could fit most appropriately in a TM model. Such an assumption derives from the similarities between TM and EBMT systems like the availability of a corpus of aligned translation examples in both systems and the common processes of segmentation, matching and alignment [Somers and Fernández Díaz, 2004]. Just like most TM systems, some EBMT systems also perform matching by comparing sequences of character strings [Somers, 2003:514], which clearly makes a possible convergence of the two systems sensible.

Another approach to match construction is the one followed by 'second generation' TM systems. Unlike the minimalist approach of Déjà Vu X, some TM systems have sought to exploit the TM databases based not only on the bare textual con-

---

[1] All information about Déjà Vu X has been drawn from the Déjà Vu X Professional Users' Guide (available from Atril's website: http://www.atril.com/)

tents but on the structural and syntactic contents of segments. To exploit a database in such a manner, however, requires the introduction of external linguistic knowledge that will help the system identify these structures and allow for a more sophisticated processing. Two commercial TM systems are known so far to belong to this generation: SIMILIS[2] and Masterin[3].

SIMILIS, in general, applies linguistic rules to a number of processes, including segmentation, alignment and automatic extraction of terms and phrases from translation memory content. After segmenting the source and target texts at sentence level, it runs a linguistic analysis and further splits each sentence into syntactic units ('chunks'), attaching grammatical annotations to them (this is performed with the help of monolingual lexicons and algorithms that can recognise grammatical categories) [Planas, 2005]. It then indexes those as translation units as well. So, every time the system searches for a match, it looks not only at the sentences, but also at the chunks (thus increasing the possibilities of finding a match), and especially those chunks that are in the same grammatical category as the source segment (thus increasing the possibility of finding the right one).

Masterin, on the other hand, segments the source and target texts in a flexible way according to the examples available in the TM database ('Knowledge Base') provided. Each segment is annotated with grammatical information (with the help of a POS tagger) and constitutes a grammatical 'translation pattern'. So, matches are sought by a deep-structure pattern recognition method that looks beyond the surface appearance of segments. If several matches are found, the system determines the best match by using semantics (with the help of a built-in lexicon) and/or examining their use frequency or domain information. In the case where no match is found, the system constructs and suggests a fuzzy match from the available resources in the database by applying translation heuristics [Grönroos and Becks, 2005].

Up to now, in the literature on second generation TM systems no evidence exists for any human evaluation of the systems. Consequently, it is difficult to determine if any of the systems outperforms the rest or, most importantly, if the systems of the

second generation surpass the systems of the first generation in usefulness. In contrast, what is very apparent from the relevant literature is the excessive insistence on the technical sophistication of the system (optimising the matching and reconstruction techniques in TM systems), without this being associated with recognized practical benefits for its users.

3 Why is MT's usefulness for translators being debated?

There is a widely held assumption – used extensively by TM developers as a marketing hook – that combining translated segments from available TM resources increases the recyclability of existing content, improves productivity and cuts translation costs even further, compared to using a traditional TM system [Hunt, 2006]. The assumption is based on the idea that many translations are simple modifications of previous translations residing in one's archive.

However, this assumption does not generally hold true, unless certain special conditions are present: a) the TM repository must contain a large number of resources (in the form of bi-texts, translation units, glossaries or lexica) and b) the resources must be relevant to the translation one wants to produce; that is to say, they must fall into the same subject domain as the new text. The questions that arise from such a hypothesis are how often these conditions exist and who is more likely to benefit from MT capabilities. Answers to these questions are critical for TM developers who ponder the development and incorporation of MT technologies into a TM system and need assurances that their investment in TM research and development will eventually pay off.

4 TM Survey: Evaluation of the MT capabilities of TM systems by their users

With a view to investigating the usage and benefits of MT as an incorporated functionality in a TM system, a large-scale survey was conducted on TM use. An online questionnaire was posted on translators' fora, TM user groups and was distributed through translators' associations and academic institutions. Any translation professional using or having an interest in TM systems was invited to answer the questions. The survey, among other

---

[2] Developed by Lingua et Machina.
[3] Developed by Master's Innovations.

aspects, focused on the users' perspective on the existing MT capabilities of the TM systems they use and probed their personal evaluation via a series of open-ended questions (e.g. "which feature do you love the most in your TM tool?", "which is the most annoying thing that has happened to you while using this tool?", "is there any particular task that you would like to see a TM tool do in the future?").

4.1 Evaluation principles

Evaluating socio-technical systems such as MT is a complex endeavour necessitating the consideration of a number of system- and user-based quality aspects within the system's context of use. Broadly speaking, evaluation aims to find out whether the system serves its intended purpose of use and if so, to what degree. However, MT systems can be used in various different ways (e.g. for assimilation purposes, to produce a draft translation or to achieve bilingual text retrieval) by different groups of users. Therefore, evaluation is practised in accordance with each group's interests.

In fact, users evaluate MT systems very differently from MT developers. Developers can perform a selective evaluation (e.g. of match recall or match precision). They can elaborate metrics and produce accurate measurements. Users, on the other hand, are more concerned with usability issues; hence their evaluation usually includes so-called human factors (e.g. user interface, customisability) and is highly idiosyncratic.

Yet translators are even more different from the average user of MT. A holistic evaluation of MT by a translator consists of an assessment of the usefulness of the tool as one more means of assisting him to produce high-quality translations within tight spaces of time.

In fact, the general public and translators do not have the same requirements of MT. Generally, the latter are just looking for the gist of a text, or want a quick translation of a short-lived document. In contrast, translators are expected to produce an excellent quality translation on every occasion. They can use MT if they like, but the final result should be seamless. This means that the standard of editing is high, and if the MT output is poor, correcting it might take as long as starting from scratch. So, while the average user's MT tolerance threshold is normally average, the translator's

threshold is much greater, and what is useful for the one may not always be helpful to the other.

Another important issue that must be taken into account while evaluating MT capabilities in a TM system (in contrast to standalone MT systems) is that the performance of the system stems from a complex interaction between the processor and the input, i.e. the translator's own resources. The success of the MT functionality therefore depends considerably on the size and quality of lexical resources residing in the TM system. On these grounds, the surveyed TM users were not asked to rate the accuracy of match construction as performed by their TM system, but instead to offer comments and suggestions on the practical usefulness of MT functionality as part of a TM system.

4.2 Usefulness of MT functionality

Of the 874 translation professionals who completed the survey, 90% were translators and 73% were working as freelancers. The majority (89%) held a professional qualification relevant to their work and a large proportion (64%) rated their general computer usage competence as 'good' (30% rated their computer skills as 'excellent'). 61% reported specialising in the translation of technical texts with high levels of content repetition.

Generally, machine translation appeared to be well received amongst translators who were familiar with it, with 45 of 145 TM users (using a TM system with MT capabilities) declaring it as their favourite feature in their TM. For these users, assembling sub-segment portions from their terminology database as well as their translation memory (thus rendering it possible to have a 'match' even if the exact source text is not in the TM) is an advantage that increases their productivity and saves them time. For some of these users, even when the results of the assembly are not 100% accurate, they are still satisfied, as a couple of translations – even if they do not make sense – can bring together a number of useful clues, helping them decide faster before they select everything, delete it and type in their new translation.

However, there were also quite a few respondents (10) who named the automatic assembly feature as the most annoying of all. These users were mainly dissatisfied with the repeated poor quality of match assembly. As one user observed, "…of the many previous translations for a given expres-

sion, some will fit in the current context. However, many examples that apply in specific contexts will also exist but they will not be fit because the terminology usage, style, and/or tone are inappropriate."

Another user, along similar lines of dissatisfaction, expressed his concern about the lack of linguistic, stylistic and tonal unity in a translation assembled from different pieces of text, something that does not only happen to literary translations, but also to non-literary ones. Indeed, most TM databases are made up of many different texts that could have been produced by many different authors and translators. Each text and each translator can have a different style, and when bits and pieces from a variety of texts are forcibly brought together, the resulting text can be a stylistic patchwork.

The survey also showed that not all translators are specialised in a particular technical field, spending all their professional life translating texts from this field. Perhaps there are hardly any translators of this sort, as people (the same being true for translation companies) tend to seek variety in their work, or they often grasp any opportunities for profitable jobs in other than their main domain of expertise. This means that TM tools with integrated MT functionality that rely on the principle of "you will never translate any previously translated text twice" (slogan of a well-known TM software brand) do not seem to produce benefits for a substantially large segment of translation professionals, not only those who translate non-technical texts but also all those who "translate in a variety of technical fields, sometimes all different throughout the year, hence repetition is not a major factor".

The attitudes of the respondents who commented on the MT features of a TM system varied, as expected. Some appeared to dismiss the whole idea of machine translation functionality in a TM system: "I don't think that a TM tool is supposed 'to bake a cake' while I am translating. There is a reason why it is called a translation memory tool and not machine translation. I believe that most of the things mentioned above [MT functions] are supposed to be handled by human translators. Sometimes less can be more."

On the other hand, other respondents expressed the desire to see more advanced MT capabilities: "A well designed translation program should not be based on the mere mechanical replacing of texts… but it must think, analyze and reconstruct the text in the destination language, using its database to enhance the translation."

The respondents of the second category were those who offered recommendations as to how MT capabilities could be improved to serve their needs in a better way. Some examples are listed below:

- A machine translation facility should be available when there is no match or the fuzzy match is below a user-predefined percentage – say 75%. The machine translation should get the information from a separate database which will be full of parallel corpora. In the TM options the user would be able to choose two databases: the client or project-specific one (as primary) and the master one (as secondary) which will contain the parallel corpus.
- The TM tool should have linguistic capabilites (e.g. recognizing singular-plural forms, inflections, etc.). When, for example, it finds fuzzy matches from the TM, it should have the ability to analyse and do some operations on the target segment as well (e.g. automatically correct the grammar).
- Incorporate MT at a user-defined level for sub-segment/new segment processing (e.g. search for terms in standard dictionaries).
- Have the option to turn any MT functionality on or off according to one's preference.
- The system should identify similar examples and make generalisations about them.
- Integrate automated and user-dependant feedback of new knowledge into the knowledge base in a dynamic and interactive manner.

By correlating the attitudes of translators (positive/negative) with various factors that may affect the way they view machine translation, some interesting findings emerged. There was no evidence to suggest that a positive or negative attitude is related to the age or the computer usage competence of the translator. However, attitudes appeared to be affected by the length of work experience of translators. Inexperienced translators seemed to favour machine translation and to tolerate incorrect assembly more habitually. Experienced translators,

on the other hand, with more than 5 years of translation experience, were those who voiced complaints and who seemed to be most annoyed with machine translation suggestions.

In addition, attitudes towards machine translation seemed to be related to the language combination a translator works with. For example, translators working with highly inflected languages, such as Greek or Polish, realise the fact that machine translation underperforms (and consequently is of little use), unless it can use language-specific assembly algorithms and rules. Their expectation is that when a TM system incorporates a knowledge base (grammar rules, lexicon) will produce more lexically and grammatically correct translations.

Finally and not surprisingly, the users' attitude towards machine translation appeared to be influenced by the area of specialisation of the translator. Those who regularly translate highly specialised texts or updates of previously translated texts seemed to be positively inclined towards the use of MT features and they would be interested in training their system, by supplying feedback, in order to improve its performance.

5 Discussion

As appears from the survey's findings, the majority of translators hardly ever need "useful" machine translations (that may allow them to get the gist of a text), as they prefer to resort to authoritative dictionaries and other sources of pragmatic information for that. Instead, they expect from a machine translation feature the ability to suggest correct translations: to construct not only a correct segment syntactically and grammatically, but also in terms of semantic equivalence to the source text. Then, the translator will be the one to decide whether the suggested translation is fit for context by looking at the full text in hand. Anything less than a correct translation can be misleading or, if it happens repeatedly, it can reduce the role of the translator to that of a post-editor of badly translated texts.

It is also evident from the users' recommendations that it is important for the system to be able to learn from the decisions/choices made by users (e.g. which potential translations are preferred, which were rejected and why), so that errors in future translation assemblies are reduced.

Finally, developers should never forget the intended use of a TM system – which is different from that of an MT system – which is to assist translators in their decision-making process (one way can be by offering valid pointers to potential translation solutions) and not try to automate this process (by mechanically constructing language to be offered as a translation solution), requiring translators to validate the system's decisions. At the other end of the spectrum, translators, for their part, should not over-rely on the MT capabilities of a TM system, as disappointing results can create a contemptuous attitude towards the technology which under certain circumstances might prove pain-saving. The key to a mutually beneficial relationship between a translator and an MT feature appears to be the cautious application of MT and the consultation of MT suggestions as if they were just another source of reference.

6 Challenges

The combination of translation units is considered a classic problem. It involves the application of rules (syntax, grammar) but also has to consider those aspects that are not covered by rules, that is, irregularities in grammar or complementation. Furthermore, semantic information as well as factual knowledge must be deployed to help disambiguation and correct match selection. The bilingual text corpus (available in a translators' TM archive) can provide all this data for solving the problem, determining any irregularities.

But what happens when the translator's archive is small or very diverse in subject (as is often the case for translators in the first few years of their career)? Obviously, the MT capabilities of his system are bound to be very poor to the point of being useless, unless the developer has incorporated language resources for his language combinations. For these TM systems, the problem is solvable, but adequate resources for this purpose have been developed for only a few language combinations so far. TM developers recognise the fact that language professionals do not need tools that function for all languages but tools that work well for their working languages. Subsequently, they develop (with substantial cost and human resources involved) language resources in an effort to address the needs of users in the major segments of markets they serve. However, when user needs are highly het-

erogeneous (there are countless language combinations for which language resources are needed), this approach leaves many seriously dissatisfied. In fact, by adding linguistic knowledge to a TM system, the product automatically loses its language-independence and its market share shrinks dramatically, especially since there are many translation companies which deal with multiple languages.

A possible solution in this case would be the availability of global TM systems and the development of language-specific add-ins (tools and language resources), that can be sold separately, based on a careful cost-benefit analysis which takes into account the high demand for a particular language combination. Obviously, such a solution would leave minority languages in a disadvantaged position, as the low demand for rare language combinations would not give sufficient incentive to develop costly language resources. In this case, TM developers offering MT capabilities should perhaps direct their research efforts to ways of acquiring quickly, easily and cheaply lexical resources from sources that contain readily available multilingual corpora. An enormous source for such material could certainly be the Web.

7 Conclusion

Although when MT systems first appeared they provoked negative feelings on the part of threatened translators, the latter have now started to realise the benefits from using MT features, as long as *they are in control* of a) the repository content (thus ensuring reliability of results) and b) the operation (by defining the rules of match construction and even opting out from MT matching when it is not necessary).

Translators also seem to be coming to terms with machine translation as an alternative means of translation production and appear to feel comfortable in their powers, knowing that an MT system is unlikely to ever produce high-quality translation of a text, if unsupervised. This derives from their belief that although computer programming can address even the most complex equivalences with appropriate descriptions of all possible discourse environments/situations, it can never predict accurately the aims, purposes, intentions, complex strategies, changing tactics and fine choices required by the creative character of a high-quality

translation. Furthermore, translators know better than anyone else that the linguistic choices in a translation are a function of language rules and the translator's creativity, a factor which knows no rules and therefore cannot be replicated.

However, since not all texts entail elements of creative writing (many are rather repetitive, tedious and written in controlled language), translators have no reason not to appreciate the assistance of machine translation in dealing quickly with such texts, which offer little joy of intellectual challenge in any case.

References

Carl, Michael, Andy Way, Reinhard Schäler. 2002. Toward a Hybrid Integrated Translation Environment". In proceedings of the *5th Conference of the Association for Machine Translation in the Americas, AMTA 2002*, 8-12 October 2002, Tiburon, CA, USA.

Carl, Michael. 2000. Combining Invertible Example-Based Machine Translation with Translation Memory Technology. In proceedings of the *4th Conference of the Association for Machine Translation in the Americas, AMTA 2000: Envisioning Machine Translation in the Information Future*, 10-14 October 2000, Cuernavaca, Mexico.

Champollion, Yves. 2003. Convergence in CAT: blending MT, TM, OCR & SR to boost productivity. In proceedings of the International Conference *Translating and the Computer 25*, 20-21 November 2003, London. London: Aslib.

European Commission, Translation DG. 2007. *Translation Tools and Workflow*. Directorate-General for Translation of the European Commission, February 2007.

Grönroos, Mickel and Ari Becks. 2005. Bringing Intelligence to Translation Memory Technology. In proceedings of the International Conference *Translating and the Computer 27*, 24-25 November 2005, London (London: Aslib, 2005).

Hodász, Gábor, Tamás Gröbler and Balázs Kis. 2004. Translation memory as a robust example-based translation system. In proceedings of the *9th EAMT Workshop, "Broadening horizons of machine translation and its applications"*, 26-27 April 2004, Malta, pp.82-89.

Huang, Jin-Xia, Wei Wang and Ming Zhou. 2003. A Unified Statistical Model for Generalized Translation Memory System. In the proceedings of *MT Summit IX*, 23-27 September 2003, New Orleans, Louisiana, USA.

Hunt, Timothy. 2006. Will Language Search Engines replace Translation Memory Systems? Article avail-

able on Lingotek's website at: http://www.lingotek.com/lsereplacetm.pdf.

O'Hagan, Minako and David Ashworth. 2002. *Translation-mediated Communication in a Digital World: Facing the Challenges of Globalization and Localization (Topics in Translation S.)*. London: Multilingual Matters.

Planas, Emmanuel. 2005. SIMILIS: Second-generation translation memory software. In proceedings of the International Conference *Translating and the Computer 27*, 24-25 November 2005, London (London: Aslib, 2005).

Simard, Michel. 2003. Translation Spotting for Translation Memories. In proceedings of *NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*. Edmonton, Canada. May 31, 2003.

Somers, Harold and Gabriela Fernández Díaz. 2004. Translation Memory vs. Example-based MT: What is the difference? *International Journal of Translation* 16.2, pp. 5-33.

Somers, Harold. 2003. Machine Translation: Latest Developments. In Ruslan Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press; pp. 512-528.

Somers, Harold. 1999. Review Article: Example-based Machine Translation. *Machine Translation* 14: pp. 113-157.