Letting the CAT out of the bag - or was it MT?

Ian Gordon
Trados UK Ltd
Shepherd's Bank, Kettleshulme, SK23 7QU, England.

This paper is intended to alert all those who are involved directly or indirectly in the translation industry to the facilities offered by Computer Assisted Translation (CAT) and Translation Memory (TM). It considers some of the main functions offered by CAT in the form of Translation Memory and compares them with the basic characteristics and constraints of Machine Translation (MT). In particular it focuses on the facilities offered by Translation Memory software translation tools and assesses the benefits and implications for the translation industry and its customers.

Introduction

For nearly five decades any mention of automation in the professional translation process has generally focused on machine translation - and the man in the street has tended to assume, being unaware of the Alpac report of 1966, that the problem was already solved. In the sixties students of foreign languages were steered away from translation by careers advisors with the explanation that it was a profession about to be overtaken - and eliminated - by technology. Computational linguists, being naturally more modest, have assumed that the problem of translation was only nearly solved - or that it would be tomorrow or perhaps the day after.

In the meantime many translators have lived in fear of the carnivorous computer with its insatiable appetite for de-skilling and devouring professions. However, in the early eighties translators quickly came to love their Amstrads and PCs, in the form of wordprocessors with spellcheckers - and later add-on modems. But in other respects too the computer is now changing from being a rival to the human translator to being his reliable and cost-effective assistant.

CAT

These developments are commonly classified under the acronym CAT. Programs of this type are designed to assist the human translator in his task, typically through more or less sophisticated terminology assistance. Gradually they have acquired other functions such as supporting the translator by giving him immediate access to a perfect memory - Translation Memory. CAT in this form has developed into an intelligent

working environment for the professional translator. This integrated suite of programs may be referred to as a "workbench".

Translation Memory

The key component of a 'workbench' is the Translation Memory function.

Translation Memory is a surprisingly simple concept. As the translator translates the text, each sentence (better termed "translation unit" since it may also be a title or table cell text) is also saved automatically to a sophisticated translation unit database memory. As he translates, any similar sentence already in the memory will appear on screen for editing. Sometimes, in fact surprisingly often, there will be a 100% match. In the case of a second or a later version of a text the majority of sentences may well be 100% matches. Gradually, as the translator gains confidence in the system, he will usually allow the program to translate automatically until it encounters the next fuzzy match.

Translation memory is an astoundingly simple concept, yet in the best systems a highly sophisticated program is at work behind a simple, user-friendly interface.

A workbench of this type, especially if it is informed by a *proactive* terminology database such as MultiTerm®[1] enables the translator to achieve greater accuracy. MultiTerm® ensures he translates the term not only accurately and in context, but also with rigorous adherence to the preferred terminology of the specific client. Translation memory ensures that he does so the ***next*** time too, whether it is five minutes later, five hours, or five years.

Translators are often confronted with the problem of being unable to remember which of several synonyms they have used, but wish to retain consistency. One workbench already solves this problem with a *Concordance* feature[2]: the translator highlights the word in question, clicks on the Concordance icon and the screen fills with source and target sentence pairs, examples of the word translated in context. Suddenly the translator has power and precision at his finger-tips.
In the following it is the intention to examine the impact of Translation Memory, bearing in mind the perspectives of the freelance translator, the translation company, the staff translator, whether in a translation company or in industry, and of the end user company with *no* in-house translators.

The ISO 9000 Quality Assurance standard defines Quality as "The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs." (BS 4778 Pt 1) - and so does the accumulated experience of

---

[1] MultiTerm is integrated proactively in Trados Translator's Workbench
[2] Trados Translator's Workbench

commercially successful companies world-wide. Consequently our natural starting point is the customer and his specific needs.

Generally speaking the customer needs translations to be accurate, consistent, finished to perfection - & he needs them yesterday. He will take for granted that his preferred terminology will be used throughout and that the company house style will be replicated perfectly in the foreign language. Hitherto these expectations were often not fulfilled, though many end users were blissfully ignorant of it.

The translation supplier is now in a position to offer the customer enhanced quality which is a definite benefit for both parties - especially in terms of the reduced risk of litigation. The fact that the translation is of a higher overall quality through enhanced accuracy and consistency is a direct benefit to the translation supplier in that he is less vulnerable to litigation, as is of course the end user. Professional translators pride themselves on the quality of their work and can test it by using the bilingual Concordance feature in Translator's Workbench. Indeed as a translator becomes accustomed to the security of using Translation Memory he is likely to feel very vulnerable when working "freehand".

In addition the translation is not only better - but produced quicker. There is a significant acceleration in the translation process which the customer appreciates, especially when he realises it is accompanied by enhanced quality

And Quality...?

For translation suppliers or users who are operating ISO 9000 Quality Systems there are distinct benefits to be obtained from using Translation Memory software. It can, for example, provide accountability and traceability to a high level. Translation units can be stamped *automatically*[3] with details of translator, translation date, reviser, revision date, date last used, domain, customer and total number of times used. This is particularly useful for example in a translation company where translation and revision will naturally be carried out by two separate translators whose separate contributions remain identifiable. The final revision of the translation may even be carried out in Word or WordPerfect and this final draft used for updating the Translation Memory at the last moment before cleaning up the hidden text.[4]

Is it possible to compare MT with TM?

First we must ask what MT has achieved. The proceedings of the ASLIB *Translating and the Computer* conference over a period of 18 years will provide the evidence. Perhaps in this context a crude generalisation may be forgiven: suffice it to say that

---

[3] This feature is available for example in Trados Translator's Workbench

[4] Translator's Workbench allows translation and editing of the document in Word or WordPerfect with automatic update of the translation memory before finalisation.

MT can be successful in an institutional or large company limited-domain context and, depending on the sophistication of the system, can produce quick results for post editing, or just a raw draft for an intelligent and imaginative reader - who has no thoughts of litigation.

Some of the basic differences between MT and TM could be summarised as follows:

MT	TM
no 'real world' knowledge	'real world' knowledge supplied by human translator
'learning' does not occur except through human intervention in a complicated and skilled process	'learning' is automatic with every Translation Unit (TU)
post-editing often regarded as tedious (repair work - uncertainty of discovering all the landmines)	whole process self-esteem enhancing (constructive)
similar to Quality Control - product of unknown quality checked, then rejected/improved after creation	compatible with Quality Assurance - product right first time
can be used by non-translator to generate low grade 'alien' translation	can be used by non-translator to eliminate the cost of repeated text
translation of text version 2 case: MT mistakes recur exactly as in raw draft of version 1	translation of text version 2 case: TM builds on previous achievements
MT tends to be a mindlessly fast blunt instrument requiring a slave post-editor	TM is a fast and precise tool in the hands of the professional translator
inflexible system	highly flexible system
all output is machine created	all output is human output
post-edited MT final version 'mortal'	TM final version 'immortal'

The major advantages of Translation Memory over machine translation are firstly that all output is human output and secondly that the quality of previous final translations is retained or improved with every translation.

This can easily be put to the test: imagine the results if the same version 2 of a manual is translated in parallel using MT and TM. MT will make a fresh attempt at translating the text, unaware of the post-editing changes which were necessary to make the first version acceptable. In stark contrast TM will remain 100% faithful to version 1 and identifying the changes requiring translation. These will be translated in context by the human translator with the support of fuzzy matches. There is no post editing stage. The customer, like the translation supplier, has the advantage of knowing that the translation is reliable and can be trusted.

Admittedly this is an extreme case - but it is useful in that it highlights an essential difference in the output of MT and TM.

Some might argue that MT is inherently suspect, littering a text with land mines for discovering if possible during post editing, whereas TM is inherently safer with the result that very few changes are required at a later stage.

Some systems give the translator the opportunity to combine the advantages of both Translation Memory and Machine Translation[5]. In Translator's Workbench the translation units are colour coded: 100% matches are green (for go), fuzzy matches are yellow (caution) and the optional Transcend MT suggestions, which are triggered below a selected fuzzy match similarity percentage, are grey - so perhaps programmers do have a sense of humour. A major advantage is that the MT program is seamlessly integrated so there is no software learning curve when Transcend is used in this context. Significantly, Transcend is fed automatically with all the subject and client-specific terminology resident in the MultiTerm database. Once the translation suggested by Transcend is edited it becomes a future 100% match, so avoiding the 'mortality' of machine translation final versions.

If post-editing is seen as a problem in MT, perhaps this is the solution: MT tamed, edited immediately and in context and reusable reliably in the future.

A translator using Translation Memory can translate desktop published material without having to re-format the text afterwards, so saving a complete stage[6]. For in-house translators this may all be grist to the mill but for freelance translators and translation companies it represents considerable added value and the end user benefits from accurate formatting and perhaps also reduced costs. A further advantage is that text can be edited right up to the last moment before finally being used to update the Translation Memory for the next job.

Document production benefits in a number of ways from translation memory. There is no need for a controlled source language since inconsistencies, intentional or otherwise, will be dealt with as fuzzy matches. Here there is a new development: it is likely that for the first time the target text will be more consistent than the source text. There have been cases where bemused customers have been informed of inconsistencies and errors in source texts which they have been using for years without being aware of the blemish. Needless to say, provided that such information is offered in all humility, the translation supplier can gain considerably in credibility through his critical awareness of the source text.

---

[5] Translator's Workbench v. 1.08 features a Transcend interface as standard. A Logos interface is also available.

[6] The range of formats covered by Translator's Workbench includes for example Interleaf, Ventura, SGML, HTML, DCF, TROFF and RC Help files.

Customers are always confident that they can identify all the changes in a later version of a document for translation - until it is retranslated using Translation Memory. Then many of the more subtle but perhaps significant changes come to light and once again reputation of the translator is enhanced: he is seen as having unusual powers of perception - even beyond those of the author.

In some cases confidentiality may be a matter even of national security, in which case the customer running his own Translator's Workbench can pre-translate the 100% matches, leaving the translator only with the remainder. In this case there is a trade-off between the need for confidentiality and absolute consistency.

Like the freelance translator, the staff translator will enjoy enhanced self-esteem and professionalism. Through being employed he is free from the need to invest in the technology, but he may wonder what he stands to gain from the new technology. Initially he may fear that increased productivity means a higher workload. More significantly he may fear that as he gradually creates a large and effective Translation Memory he may become expendable and find himself redundant. He might even fear that he could be replaced by a less experienced and cheaper translator who could lean on the expertise of his predecessor. One suspects that the way ahead for the staff translator will not be to retreat into a defensive position, but to embrace the new technology, adding it to his repertoire of skills.

This fear of being drained dry of all that you know and can translate is a fear shared by freelance translators. In practice TM allows the translator to concentrate his creative efforts on new material for translation, rather than drowning in inconsistent repetition, rather like the errant schoolboy copying out lines incessantly. Essentially the computer has a better memory than the human, and the human vastly superior powers of perception and creative linguistic skill.

It is interesting that for a translator or translation company the most valuable asset could soon become the Translation Memory itself. In some companies it has even been known to replace an uninterruptible power supply unit: after the power cut the translator simply retranslates the text on 100% matches at high speed or automatically, until he returns to the point he had reached.

Change means winners and losers

As a critical link in the translation supply chain the translation company is at one time both supplier and customer, thus sharing in common with the freelance translator and the end user many of the benefits - and perhaps concerns. If translation companies justify their existence in the supply chain by offering project management, continuity and a quality filter, the question arises as to what extent a freelance translator using the customer's Translation Memory and terminology database can challenge the translation company in these functions. With the increasing use of Translation

Memory there may be a reassessment of the role of the translation company in the supply chain, although it seems likely that for large scale projects only a well organised translation company using Translation Memory will be able to offer the required quality at the right price.

Since Translation Memory enhances productivity it naturally increases turnover of the translator and translation company - indeed it is not unknown for a translation company to more than double it's turnover in a 12 month period on adopting Translation Memory software.

If the translation supplier is producing a better product quicker, what is it actually costing him to do so? Firstly he needs a well specified computer, he needs the Translation Memory software and he needs the expertise. Provided that the texts he wishes to work on are available electronically or as good quality hard copy for scanning, his investment will provide a relatively short payback time, in some cases very short (in one case under a week). Freelance translators are traditionally reluctant to invest in themselves yet like to regard themselves as professionals. In comparison a taxi driver would willingly spend several times the cost of the most expensive translation memory system on his taxi and regard it as a good investment. For a translation company the investment may be costed and payback times calculated bearing in mind that translation company will be able to use the software intensively both with its in-house translators and freelance professionals. In the latter case the company may work with freelance professionals who have the software, loaning the software for a contract or even sending out a text for over-typing[7] which can later be dumped into the Translation Memory.

Before committing himself to an investment in Translation Memory software a prospective purchaser can analyse texts considered for translation, perhaps a set of manuals, to establish the element of 100% matches and fuzzy matches[8]. He can do this both within an individual document, across a family of documents, and relative to an existing Translation Memory. In this way the payback time can be estimated with some degree of certainty.

This feature is also of tremendous value in enabling a translator to price jobs for quotation. If he wishes to depart from the value added concept he is actually in a position to trim his price with considerable precision, for a better chance of obtaining the contract - and also operating profitably. He will also be able to schedule his milestones with greater confidence.

In addition to increased output of a higher quality, perhaps at a premium price, the freelance translator enjoys an improved sense of professionalism and self esteem: his

---

[7] Translator's Workbench allows a text to be translated or edited remotely using Word or WordPerfect and then re-imported for updating the translation memory.
[8] The Analyse function in Translator's Workbench.

product is appreciated and his skills valued. For the translation company there is a tight-rope to walk: whilst avoiding the impression that translation is now automatic at the touch of a button, it will want to maximise the marketing benefits of its investment. There is a major benefit for freelance translators, who traditionally are workaholics, in that they have the opportunity to enhance their earnings whilst working less anti-social hours.

It goes without saying that in a market situation the benefits of Translation Memory will ultimately be reflected in pricing structures. There is the possibility of the value added premium but customers are also likely to analyse their texts for repetition and be interested in negotiating an appropriate price. Translation companies will naturally resist the idea that 100% matches should not be chargeable since they are handling them and by definition they have a responsibility for them in the new context. They represent part of their turnover on which they pay Liability Insurance premiums. To this dual pricing system (100% matches and the rest) may be added a third layer in the form of a separate price for fuzzy matches. This type of graduated pricing is not unknown in the software localisation industry and may spread.

Before the advent of Translation Memory any translations from previous years were of limited informal value only but now it is possible by using a text alignment program[9] to recycle past *approved* translations; naturally there is no point in recycling low grade material. As in all recycling the aim is to reuse as much as possible as cost effectively as possible. Typically 70-80% of previous translation units will be accepted as reliable at the first pass. The user can increase this percentage by fine tuning the system but he would not normally expect to achieve 100%. Any recycling at all is of benefit both financially and for consistency of terminology and house style. In companies with an ISO 9002 Quality System it is likely that texts from several previous years will be available for recycling. Aligned translation units are added to the Translation Memory, with details of their origin being added automatically to every translation unit. Despite the computer origin of the alignment process, quality standards are maintained by the system awarding the aligned text units only a 99% rating, this avoiding the risk of automatic substitution occurring without translator approval. When an aligned translation unit appears on screen during the translation process as a suggestion for editing, the translator will use his judgement and either simply accept it or edit it if necessary. The next time this unit reappears it will be a 100% match - because it already has the blessing of the human translator.

Conclusion

It is easy to identify winners - but losers are elusive. Translation Memory offers so many advantages in so many areas that there is scope for all those in the supplier/purchaser chain to benefit. How this works out in detail will depend on

---

[9] e.g. Trados TAlign

market conditions but it seems likely that whoever takes the initiative will be empowered by the new technology - empowered in terms of both the quality and speed of translation - and hence customer satisfaction and commercial advantage. For the cognoscenti it is a state of the art secret weapon, closely guarded to ensure they retain the competitive edge.