Bilingual concordancers and translation memories: A comparative evaluation

Lynne BOWKER
School of Translation and Interpretation,
University of Ottawa
70 Laurier Ave E., Rm 401
Ottawa, Ontario, Canada, K1N 6N5
lynne.bowker@uottawa.ca

Michael BARLOW
Department of Applied Language
Studies and Linguistics, University of
Auckland
Auckland 1001, New Zealand
mi.barlow@auckland.ac.nz

Abstract

Translators are increasingly turning to electronic language resources and tools to help them cope with the demand for fast, high-quality translation. While translation memory tools seem to be well known in the translation industry at large, bilingual concordancers appear to be familiar primarily in academic circles. The strengths and weaknesses of these two types of tool are analyzed in an effort to recommend those circumstances in which each could best be applied.

1 Introduction

Recent years have witnessed a number of significant changes in the translation market. Largely as a result of globalization, there has been a considerable increase in the volume of text to be translated. New types of text, such as Web pages, have also appeared and require translation.

The increased demand for translation has been accompanied by another trend: deadlines for translation jobs have grown shorter. This is in part because companies want to get their products onto the shelves in all corners of the world as quickly as possible. In addition, electronic documents such as Web pages may have content that needs to be updated frequently. Companies want to be sure that their sites reflect the latest information, so translators are under pressure to work very quickly to ensure that the up-to-date information is reflected in all language versions of the site.

Furthermore, it has been observed that in today's market, there is currently a shortage of human translators (e.g. Sprung 2000:ix; Shadbolt 2002:30-31; Allen 2003:300).

The increase in volume coupled with shorter turnaround times has resulted in an immense pressure on existing translators to work more quickly, while still maintaining high quality in their work. However, these two demands of high quality and fast turnaround are likely to be at odds with one another. Therefore, one way that some translators are trying to balance the need for high quality with the need for increased productivity is by turning to electronic resources and tools.

One type of language resource that has become popular is the bilingual parallel corpus, which is essentially a collection of texts in one language (e.g. English) alongside their translations into another language (e.g. French). The two sets of texts must be aligned, which means that links are made between corresponding sections (e.g. sentences, paragraphs) in the two languages.

Bilingual parallel corpora can contain a wealth of useful information for translators, but in order to be able to exploit these resources, some type of tool is needed. There are two main types of tool that can be used to search for and retrieve information from a bilingual parallel corpus[1]: a bilingual concordancer (BC) and a translation memory (TM). While these two types of tool have some common goals and features, they also have a number of differences.

As we will see in the upcoming sections, BCs can be considered to be "old technology" and they are not well known in the translation industry outside of academic circles. In contrast, TMs have garnered a significant amount of attention in the translation industry of late; they are very much in vogue and are considered to be leading-edge technology. Nevertheless, a number of translators have expressed frustration and disappointment when trying to apply TMs in certain contexts. It is possible that some of the frustration experienced by translators using TMs in certain situations could be alleviated by using BCs instead. The aim of this paper is to conduct a comparative analysis of the two types of technology in an effort to determine the strengths and weaknesses of each in order to determine those situations where translators would be best served by using a TM and those where they may be better off using a BC.

Following the introduction, the paper will be divided into four main parts. Part 2 provides some

---

[1] Note that while the same corpus data can be used with both types of tool, it is usually necessary to pre-process the corpus in a different way in order to render it readable by different tools.

background information, including a general description of how the two types of tool work, with reference to two specific tools – *ParaConc* and *Trados* – that are representative of the categories of BC and TM respectively. Part 3 contains a brief assessment of the place occupied by these tools within the translation industry today. Part 4 contains a more detailed comparative analysis of the features and associated advantages and disadvantages of each type of tool. Finally, Part 5 concludes with some general recommendations about which translation situations warrant the use of each type of tool.

2 General introduction to BCs and TMs

The general aim of both a BC and a TM is to allow a translator to consult, and if appropriate to "reuse", relevant sections of previously translated texts. In the following sections, BCs and TMs will be described with reference to *ParaConc* and *Trados*, which are representative examples of these respective categories of tool.

2.1 *ParaConc*: an example of a BC[2]

BCs, such as *ParaConc*, are fairly straightforward tools: they allow translators to search through bilingual parallel corpora to find information that might help them to complete a new translation. For example, if a translator encounters a word or expression that he does not know how to translate, he can look in the bilingual parallel corpus to see if this expression has been used before, and if so, how it was dealt with.

To use *ParaConc*, the source and target texts must first be aligned, which means that corresponding text segments are linked together[3]. A semi-automatic alignment utility is included in the program to prepare texts that are not already pre-aligned. The initial part of the alignment process is carried out in three stages: first the texts are aligned based on headings, if any are present in the texts, then alignment is carried out at the paragraph level, and finally at the sentence level. The software uses the formatting information in files to carry out alignment of headings and

paragraphs. Alignment at the sentence level is achieved by applying the Gale-Church algorithm (Gale and Church 1993). To make adjustments to the alignment, the user can examine the aligned segments and either merge or split particular segments, as necessary. One important thing to note is that the aligned units remain situated within the larger surrounding text.

Once the texts are aligned, the translator can consult the corpus. By choosing the basic search command, the translator can retrieve all examples of a word or phrase (or part of a word) from the corpus. As shown in Figure 1, the search term "head" has been entered and all instances of "head" from the English corpus are displayed in the upper pane (here in a KWIC format). The corresponding text segments from the French corpus are shown in the lower pane.
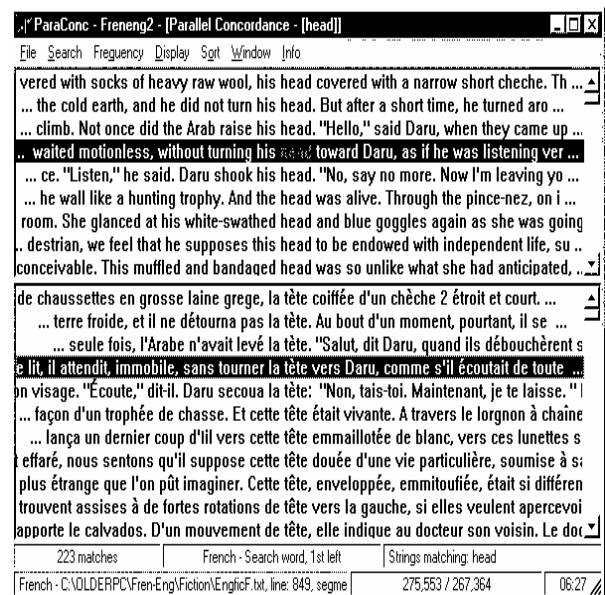

Figure 1. A *ParaConc* results window.

The concordance lines can be sorted in various ways (e.g., primarily 1st left and secondarily 1st right) in order to group similar phrases together and therefore make it easier for a translator to spot linguistic patterns. Clicking on a concordance line in the upper pane will highlight that line and also the corresponding text segment in the lower pane. Double-clicking on a line will bring up a window containing the segment within a larger context.

Suggested translations for the English "head" can be highlighted by positioning the cursor in the lower French results pane and clicking on the right mouse button. A possible translation of "head" such as "tête" can be entered. The program then simply highlights all instances of "tête" in the French results window, which can then be displayed (and sorted).

It is also possible to use a utility that presents a list of "hot" words in the French results pane,

---

[2] In fact, *ParaConc* could more properly be termed a *multilingual* concordancer, since it is possible to consult texts in up to four languages at once. However, in the context of this paper, we will refer to it as a BC and discuss its use for comparing texts in two languages.

[3] A detailed description of alignment techniques is beyond the scope of this paper; however, alignment is a non-trivial matter. Problems can arise, for example, if a single source text sentence has been translated by multiple target language sentences, or vice versa, or if information has been omitted from or added to the target text (e.g. to handle cultural references).

including possible translations. Some or all the words listed can be selected and they will then be highlighted in the results.

Finally, more complex search commands can also be used if desired. Some of the possible advanced search options are: Text search, Regular expression search, Tag (part-of-speech) search, Batch search, and various heading-sensitive and context-sensitive searches. Of particular interest to translators is a Parallel search, which allows the user to enter both an English and a French search word and to retrieve only those occurrences that match both (e.g. only instances where "head" is translated by "tête" and not by "chef").

2.1.1 Potential limitations of BCs

There are a number of potential limitations that are often associated with BCs: 1) the limited degree of automation; 2) the nature of the search item; and 3) the nature of the matching process.

With regard to degree of automation, when using a BC, it is up to the translator to decide what word or expression to look up, and he then has to manually type this into the search engine.

In terms of the nature of the search item, BCs are generally designed to search only for words or very short phrases. It is true that, in principle, a BC could be used to search for an entire sentence or paragraph; however, the fact that the search pattern must be manually entered tends to discourage this type of use because it would be extremely time-consuming and error prone (e.g. typos).

Finally, BCs are sometimes criticized because of the nature of the matching process that they use. By default, these tools basically search through the corpus for occurrences that match the entered search pattern *precisely*. For example, if the translator enters the search pattern "flatbed colour scanner" into the concordancer, it will retrieve only those occurrences that match that pattern exactly. It will not retrieve an example that contains differences in punctuation, spelling or morphology (e.g. "flat-bed color scanners"). However, as noted in section 2, some BCs, such as *ParaConc*, have added more advanced search features to improve the flexibility of searching.

2.2 *Trados*: an example of a TM

Like a BC, a TM is a tool designed to help translators identify and retrieve information from a bilingual parallel corpus. However, one of the motivating factors in developing TMs was to overcome some of the seeming limitations of BCs as described in section 2.1.1. Consequently, TMs are more automated, can search for longer segments, and employ fuzzy matching techniques.

The data contained in a conventional TM, such as *Trados*[4], are organized in a very precise way, which differs somewhat from the way in which data are stored for use with a BC. *Trados* divides each text into small units known as segments, which usually correspond to sentences or sentence-like units (e.g., titles, headings, list items, table cells). The source text segments are linked to their corresponding target text segments and the resulting aligned pair of segments is known as a translation unit (TU). Each TU is extracted from the larger text and stored individually in a database. It is this database of TUs, not the original complete text, that is later searched for matches.

When a TM, such as *Trados*, is first acquired, its database is empty. It is up to the translator to stock the database. This can be done interactively by having the translator add each newly translated segment to the database as he works his way through the text, or it can be done by taking previously translated texts and aligning them using the accompanying automatic alignment program. It is important to note, however, that in order to ensure that the automatic alignment has been done correctly, manual verification may be required.

When a translator receives a new text to translate he begins by opening this new text in the *Trados* environment. *Trados* proceeds to divide this new text into segments. Once this has been accomplished, the tool starts at the beginning of the new source text and automatically compares each segment to the contents of the TM database. If it finds a segment that it "remembers" (i.e., a segment that matches one that has been previously translated and stored in the TM database), it retrieves the corresponding TU from the database and shows it to the translator, who can refer to this previous translation and adopt or modify it for use in the new translation.

Of course, language is flexible, which means that the same idea can be expressed in a number of different ways (e.g., 'The filename is invalid' / 'This file does not have a valid name'). Consequently, a translator cannot reasonably expect to find many exact matches for complete segments in the TM. However, it is highly likely that there will be segments in a new source text that are similar to, but not exactly the same as, segments that are stored in the TM. For this reason, *Trados* also employs a feature known as fuzzy matching. As shown in Figure 2, a fuzzy match is able to locate segments in the TM that are an

---

[4] Note that *Trados* is actually a suite of tools that includes, among other things, an automatic aligner, a terminology manager and a TM.

approximate or partial match for the segment in the new source text.

Segment from new source text	The specified operation was interrupted by the system.
Fuzzy match retrieved from translation memory	EN: The operation was interrupted by the application. FR: L'opération a été interrompue par l'application.

Figure 2. Fuzzy match retrieved from the TM.

If more than one potential match is found for any given segement, these are ranked by the system according to the degree of similarity between the new segment to be translated and the previously translated segment found in the database. Note that the similarity in question is a superficial similarity (e.g., the number/length of character strings that the two segments have in common) and not a semantic similarity (thus "gone" and "went" will not count as similar despite the similarity in meaning of the two words). The match that the system perceives as being most similar to the new source segment is automatically pasted into the new target text. The translator can accept this proposal as is, edit it as necessary, or reject it and ask to see other candidates (if any were found).

Trados also works in conjunction with termbases; however, it is important to note that these need to be manually pre-stocked by translators with specialized terms and their equivalents. By searching in the termbase – if one exists – *Trados* can locate matches at the term level and present them to the translator. Nevertheless, there is still a level of linguistic repetition that falls between full sentences and specialized terms – repetition at the level of expression or phrase. This is in fact the level where linguistic repetition will occur most often.

Until recently, *Trados* permitted phrase or expression searching only though a feature that resembled a BC. In other words, a translator could manually select an expression, and *Trados* would search through the database of TUs to find examples. In the most recent version of *Trados* (v6.5), however, an auto-concordance function has been added, which, when activated will automatically go on to search for text fragments when no segment-level match is found.

Once the translator is satisfied with the translation for a given segment – which can be taken directly from *Trados*, adapted from a *Trados* match, or created by the translator from scratch – the newly created TU can be added to the TM database and the translator can move on to the next segmentôt. In this way, the database grows as the translator works. *Trados* can also be networked so that multiple translators can search and contribute to the same TM.

3 BCs and TMs in the translation industry

A literature survey indicates that BCs and TMs are both widely used in academic settings for translator training. A long list of researchers (e.g. Bernardini 2002; Hansen and Teich 2002; Palumbo 2002; Pearson 2000; Tagnin 2002; Zanettin 1998) have shown that using BCs in conjunction with parallel bilingual corpora can help students with a range of translation-related tasks, such as identifying more appropriate target language equivalents and collocations; coming to grips with difficult grammatical points (e.g. prepositions, verb tenses, negative prefixes); identifying the norms, stylistic preferences and discourse structures associated with different text types; and uncovering important conceptual information.

With regard to TMs, meanwhile, many translator trainers (e.g. Austermühl 2001; Bowker 2002; DeCesaris 1996; Kenny 1999; L'Homme 1999) are now using TMs for tasks such as getting students to analyze and evaluate different translation solutions; helping students to learn more about inter- and intra-textual features by examining source texts and evaluating their characteristics in an effort to determine whether or not they can be usefully translated with the help of a TM; and conducting longitudinal studies of students' progress over the course of their training program.

In contrast to the academic setting, where both BCs and TMs are well known and widely used, the situation in the professional setting is somewhat different: TMs are very popular, but the existence of BCs does not seem to be widely known.

For example, TMs are discussed frequently in the professional association literature. According to newsletters/programmes circulated to members, translators' associations such as the *American Translator's Association* or the *Association of Translators and Interpreters of Ontario* have provided their members with opportunities (e.g. demonstrations, workshops, professional development seminars) to learn about TMs.

In addition, some professional translators' associations, such as the *Ordre des traducteurs, terminologues et interprètes agréés du Québec*, also publish magazines aimed at language professionals, and in recent years, these have included a number of discussions on TMs (e.g. Bédard 1995, 1998; Arrouart and Bédard 2001; Lanctôt 2001).

In those same publications, however, considerably less attention has been paid to BCs: only one event focusing on these tools was reported (Evans 2002).

This raises the question as to why BCs appear to have received a less enthusiastic welcome in the professional world than have TMs. One factor that may have led to a difference in uptake of these two tools is the ease of access to such tools.

Firstly, it should be noted that BCs have long been known in fields such as language teaching or second-language learning (e.g. Johns 1986, Mindt 1986, Barlow 2000), but it is only more recently that their potential as translation aids has been recognized. Academics working in the field of translation are often involved in, or have colleagues who are involved in, language teaching, and as such they may have gained exposure to BCs in this way. Many of the existing BCs were initially developed by academics who work in language training[5] often as a means of helping their own students. This means that while such tools are generally very reasonably priced and may be easily accessible within the academic community, they are sometimes not widely advertised or distributed to the professional translation community because the people who have created these tools have full-time teaching jobs. In contrast, tools such as TMs, which have typically been developed in the private sector by companies that have professional full-time programmers, technical support staff and generous advertising budgets, are more actively marketed to working translation professionals. The fact that BCs do not seem to be well advertised in the professional setting may explain, in part, why translators and translators' associations seem to be more aware of the existence of TMs than they are of BCs. This situation may change in the future, however. As noted above, the use of BCs in translator training institutes has become firmly established since the late 1990s. This means that, at present, most of the translators in the workforce will have received their education during a time when BCs were not part of the translator training curriculum. However, over the coming years, the number of BC-saavy graduates will increase and they will bring to the workforce their knowledge of BCs. They will be able to share their experience with their colleagues and employers and gradually, more and more companies will have translators on staff who have an understanding of such tools.

---

[5] For example, *ParaConc* was developed by Dr. Michael Barlow, who works in the Department of Applied Language Studies and Linguistics at the University of Auckland; *MultiConcord* was developed by a consortium based in the Centre for English Language Studies at the University of Birmingham.

4   Comparative analysis of BCs and TMs

On the surface, it may seem to be an obvious choice for a translator to select a TM over a BC since a TM includes the basic functions of a BC, as well as a number of additional features (e.g. automated searching, segment-level matching, fuzzy matching). However, if one looks beneath the surface, it seems that while TMs may be favourable in some circumstances, there are other situations where a BC may be the preferred tool. In the following sections, we will examine the strengths and weaknesses of BCs and TMs, using *ParaConc* and *Trados* as representative examples of these respective categories of tools.

4.1   Automation

Automation is an oft-touted advantage of TMs. In principle, automating the search feature should speed up the process; however, this may not always be the case. As pointed out by Bédard (1995:28), it is possible to approach automation in one of two ways: 1) an ambitious or high-tech approach, using very sophisticated and highly automated tools, such as TMs, or 2) a more modest or low-tech approach, where the tools (e.g. BCs) are simpler and require more user input.

In the case of the highly-automated approach, there can be hidden costs. Because the tools are more sophisticated, they may require a greater investment of time and effort in learning how to use them, which may prompt users to ask "What have I got myself into?". The pre-processing steps (e.g. alignment) may also be more demanding because an automated system depends more heavily on correct alignment. As noted in section 2.2, in the case of *Trados*, if a translator wishes to ensure that the alignment is absolutely correct in order to prevent misaligned TUs being presented, he must manually verify, and if necessary correct, the alignment – a process that can be extremely labour-intensive if the database is large. In contrast, since the data generated by BCs is designed for consultation by a human user, not a computer, the alignment requirements are somewhat less stringent. A certain number of alignment errors can be tolerated in a BC because the danger of "automatically" retrieving misaligned segments does not exist, and if an error does occur, the translator can simply look to the preceding or following text to find the corresponding segment because a BC does not extract the segment from its surrounding text. Because BCs can tolerate a certain margin of error, the translator need not bother to manually verify every alignment segment prior to beginning to use the tool, which can represent a significant time saving.

Another potential drawback of automation is that the system searches for all matches, even in cases where the translator may not need help with a particular passage. For example, if the auto-concordance feature in *Trados* is activated, it may retrieve and display matches for phrases such as "because of the" or "in order to", for which an experienced translator is unlikely to need assistance. This can be distracting because the fact that information has been retrieved means that the translator will probably at least have a brief look at what the system has proposed, which takes time and is disruptive to the translation process. And the return on investment is bound to be low for time spent looking at matches for segments for which no translation assistance was required in the first place. In contrast, when working with a BC, the translator initiates the searches and therefore only looks for passages for which he requires help.

In addition, the fact that many TMs, including *Trados*, automatically copy and paste fuzzy matches or term matches directly into the target text can sometimes be a hindrance. Depending on the amount of editing required to produce a desirable target segment, it may actually be faster for the translator to type the translation from scratch rather than editing the proposed segment. In contrast, a BC does not automatically paste any text directly into the target document, which can be a good thing or a bad thing depending on the quality of the match retrieved.

A small point, but one that is worth mentioning nonetheless is that TMs often require a great deal of user-initiated clicking in order to view or use the "automatically" retrieved information. For example, in *Trados*, when working in interactive mode, the user must click in order to instruct the system to conduct a search for each new segment. Once the search has been conducted, only the highest-ranked match is automatically presented to the user, but depending on the translator's needs, this is not necessarily the match that will be the most helpful. There are extra clicks involved in pulling up and viewing additional matches. Lastly, when the auto-concordance feature is activated, if the system does not find any sentence-level matches for the current segment, it automatically opens the concordance window and displays the results; however, in so doing, it makes the concordance window the active window, so the translator has to make a point of clicking back in the target field before starting to type, otherwise the text will be inadvertently written to the search field of the concordance window. It is true that there is also typing and clicking to be done when using a BC, but the point we want to make here is that BCs such as *ParaConc* do not profess to use

automation as a time-saver. Moreover, the lack of automation may actually save time in some cases. For example, in *ParaConc*, all the matches are displayed at once and the user can peruse them at a glance instead of having to click through them.

Finally, it should be noted that not all features of TMs are in fact automated. In *Trados*, for example, the termbase that is used to identify term matches must be manually pre-stocked with term records by the translator prior to beginning a translation job. However, as pointed out by Arrouart and Bédard (2001:30), when a translator consults a parallel bilingual corpus using a BC, he has at his disposal a sort of "full-text glossary" which, by its very nature, contains countless "term records" that the translator has not yet had the time to formalize. Arrouart and Bédard go on to observe that one day, such resources may well supplant carefully managed collections of term records.

In summary, while less-automated tools such as BCs appear to achieve less, they may be quicker to provide translators with results they can actually use, and they are likely to be more tolerant of unexpected situations. Of course, using such tools may call for a higher level of inventiveness or creativity on the part of the user, but thankfully, these are qualities that translators typically possess.

4.2 Search flexibility

It was noted in section 2.1.1 that one of the perceived limitations of BCs is the nature of the searches that can be conducted. Typically, BCs search for occurrences in the corpus that precisely match the search pattern entered by the user. In contrast TMs can make use of a fuzzy matching technique that can identify patterns that are similar to, but do not precisely match, the source segment.

However, a fuzzy match is not a panacea. When using fuzzy matching techniques, the translator can set the sensitivity threshold of the match; in other words, the translator can decide how similar the two segments must be in order for a TU to be retrieved and displayed. Setting the appropriate sensitivity threshold can actually be quite tricky: if the threshold is set too high (e.g., 95% similarity), then potentially useful matches may be overlooked and the translator will be forced to do unnecessary independent research. But if it is set too low (e.g., 30% similarity), then irrelevant segments may be erroneously retrieved and the translator will waste time weeding through the non-pertinent data. In addition, as noted in section 2.2, even if a fuzzy match has a high percentage of similarity, it may not be that useful to the translator since the matching is based on surface structure similarities rather than semantic similarities. For instance, the following would be retrieved as a good match in a

TM since the two segments strongly resemble each other on the surface, differing by only two characters: **File the form. / Fi<u>ll</u> the <u>d</u>orm.**

In contrast, the following pair would not be retrieved because they are not superficially similar, though they are closely linked semantically: **File the form. / <u>He</u> <u>is</u> <u>re-fil</u>ing th<u>ose</u> form<u>s</u>.**

A translator who is looking for an equivalent of a given segment would find the translation of a semantically-related segment to be more useful than that of a segment which bears only a superficial resemblance to the source text segment. With a BC, a translator could use his own knowledge of semantics to try to formulate more relevant queries, but with a TM, the translator has no input into the search patterns used.

Moreover, as mentioned in section 2.1, many BCs have developed a number of additional flexible searching techniques which, though still manually initiated, can approximate to some extent the results of a fuzzy match. For example, *ParaConc* offers the possibility of using operators such as wildcards as part of a search. If used properly, these operators can increase the flexibility of a search (e.g. by finding inflected forms). However, as was the case with fuzzy matching, they can also lead to problems if they are not used rigorously. For instance, in an effort to retrieve examples of all forms of the verb "to enter", a translator may input a pattern such as "enter*" where the * can be used to represent any string of characters. However, this pattern will also retrieve occurrences of all other words beginning with the string "enter" (e.g., "enterprise", "entertain"). As a result, the translator may inadvertently be presented with irrelevant data.

The nice thing about working with a BC, however, is that the translator does have control over the search pattern that is entered, so by learning the proper search syntax and by gaining some experience, translators can learn which types of patterns are likely to produce valuable information and which are likely to waste time. When working with a TM, however, the translator has no control over the search pattern that is used. For example, as mentioned in section 2.1, the parallel search offered by *ParaConc* allows a translator to limit a search to a given word sense, whereas this cannot be achieved using a TM.

4.3 Consistency

Another highly advertised feature of TMs is that they promote consistency in translation. The question that has been raised by some translators, however, is whether this is always desirable.

Merkel (1998:143) conducted a survey of 13 translators using TMs to carry out the translation of software manuals. One of the questions asked was whether they preferred consistent translations of a given source segment in two different contexts. The choice of answer was either "yes" or "no", with space for the respondent to elaborate on the motivations for his/her choice. Upon examining the completed questionnaires, Merkel noted that "it became apparent that there was a need for a third response, in between 'yes' and 'no', namely a response which we can call 'doesn't matter'. This applies when the translator in the justification for the choice has indicated that the translation could be consistent, but that it would not matter whether the source segment was also translated differently." This raises an interesting point: in contrast to what many TM vendors would have us believe, while consistency may sometimes be desirable, it may not always be strictly necessary.

Furthermore, there may even be cases where consistency is not at all appropriate. For instance, the translators consulted as part of Merkel's survey warn that there is a need to evaluate a proposed match within the new context, and that it may not always be automatically acceptable. This is particularly true in the case of different structural contexts (e.g. sentence vs heading vs table cell), where caution should be used in applying consistent translations (Merkel 1998:145).

4.4 Other quality-related issues

In addition to the question of consistency, other quality-related issues have been raised by translators working with TMs. One of the most significant, which was briefly introduced in section 2.2, is the fact that TM databases store isolated segment pairs, rather than complete texts. In the words of Arrouart and Bédard (2001:30), a TM is actually a memory of sentences out of context.

This can be problematic because the sentences in a text generally depend on each other in various ways. For example, when we read/write the third sentence in a text, we can refer back to information already presented in the first two sentences, which means that it is possible to use pronouns, deictic and cataphoric references, etc. However, if we take that third sentence in isolation, it may not be clear what the antecedents of such references are.

In addition, because languages do not have a one-for-one correspondence or the same stylistic requirements, translators who are trying to convey the overall message of a text may map the information to the sentences in the target text in a way that differs from how that information was originally dispersed among the source text sentences. The result is that even if the two texts are considered to be equivalent when taken as a whole, the sentences in a translation may not

depend on each other in precisely the same way in which the source text sentences do (Bédard 2000).

In order to maximize the "recyclability" of a text, a translator working with a TM may choose to structure the sentences in the target text to match those in the source text, and he may choose to avoid using pronouns or other references. According to Heyn (1998:135), the result may be a text that is inherently less coherent or readable, and of a lesser overall quality. Bédard (2000) describes this as a "sentence salad" rather than a text.

The sentence salad effect is exacerbated when the sentences in a TM come from a variety of different texts that have been translated by different translators. Each text and translator will have a different style, and when sentences from each are brought together, the resulting text will be a stylistic hodgepodge. It is highly unlikely that the source text has been created in such a fashion (i.e., by asking a variety of authors to contribute individual sentences), so it is questionable whether this approach should be used to produce a translation, which is also a text in and of itself.

Another quality-related problem is that errors contained in TMs may come back to haunt a translator if the database is not scrupulously maintained in order to correct such errors. Lanctôt (2001:30) provides the following account of a translator who carefully stores all his translations in a TM, but who does not update the contents to reflect corrections made by the client to the final document. When the client sends a document that closely resembles a version of a document previously translated the year before, the translator uses the TM and blithely reproduces the same errors in the new translation. The client is irritated because the same passages that were corrected last year need to be corrected again. This is not the kind of added value the client was looking for.

It is worth pointing out that a BC will also produce less-than-satisfactory results if the contents of the corpus are not of high quality. The main advantage offered by a BC in this regard is that it is much more straightforward to update the corpus with a corrected text than it is to fix erroneous TUs in a TM.

4.5 Translators' attitudes and satisfaction

An important point to consider with regard to any tool is whether or not the intended users enjoy working with it. In the case of TMs, Merkel (1998:140) observes that some translators "fear that translation work will become more tedious and boring, and that some of the creative aspects of the job will disappear with the increasing use of translation memory tools." Merkel (1998:141) goes on to note that there is concern that a translator

who works with a TM may be reduced to somebody who simply has to press the OK button.

In a similar vein, Bédard (2000) expresses concern that translators may lose motivation when working with a TM because they risk becoming "translators of sentences" rather than "translators of texts". In order to maximize recyclability when working with a TM, translators are encouraged to translate one source text sentence by one target text sentence. However, as noted in section 4.4, the aim of most translators is not to translate sentences, but rather to translate a message. To do this effectively, translators often need to work outside the artificial boundaries of end-of-sentence markers, and they may therefore feel constrained by the sentence-by-sentence approach imposed by TMs. In contrast, Arrouart and Bédard (2001:30) have observed that when working with a BC, few constraints are imposed by the tool and translators are therefore more free to work as they wish.

Another difficulty that may be faced by translators working with TMs is that they may be biased by what the system presents. In other words, after a translator has seen a suggestion from the database, it may be difficult to think of another way of expressing that thought, so he may use the suggested translation even if it does not fit very coherently into the text as a whole. When using a BC, however, a translator is more likely to be seeking inspiration for handling a shorter term or expression, rather than a complete segment match, so he is less likely to feel unduly influenced by the overall structure of the sentence contained in the corpus. He is also more likely to find examples of that term used in a variety of ways, so he can pick the usage that is most suitable for integration into the text as a whole. In this way, a translator feels like he is making his own decisions, rather than having someone else's decisions forced upon him.

The very fact that there are multiple ways to render a given passage in another language may also be a reason why some translators are unhappy about using a TM. Merkel (1998:148) notes that as part of his survey, translators were presented with several different options as translations of a given passage. The choice of "best translation option" varied widely among translators, which leads him to believe that it may be difficult to encourage translators to accept suggestions from TMs.

A related problem that has to do with different working styles of translators is described by Lanctôt (2001:30). When multiple translators are sharing a single TM over a network, it may be that translator A, for example, works by ploughing through a text to complete a full rough draft, and he then goes back over the text a second and third time to clean up any outstanding problems (e.g.

terminological, stylistic). In contrast, translator B's approach is to go more slowly, doing terminological research and addressing stylistic concerns as he goes along. In Lanctôt's scenario, translator B is frustrated by the suggestions proposed by the TM – many of which were produced as part of translator A's first rough draft.

5 Concluding remarks

The aim of this paper has been to introduce and present an analysis of some of the strengths and weaknesses of two categories of tool: BCs and TMs. As noted in section 3, although TMs are widely promoted in the translation industry, BCs are less well known and, in some cases, translators who are vaguely aware of them may erroneously believe that such tools have been completely superseded by TMs and therefore have no interest for the translation community.

It is not our intention to promote one type of tool over the other. Instead, we feel that the two technologies may be considered complementary, rather than competing, in the sense that one may be preferred in certain circumstances, while the other may be favoured in a different situation. Basically, it comes down to a translator being aware of how the two types of tool work and the potential advantages that each offers. The translator must then be able to choose the right tool for the job at hand. What follows are some possible considerations that a translator might take into account when deciding which tool to use.

One critical factor that comes into play when choosing which tool to use is the nature of the job itself. Not all translation jobs are equal, and they will not necessarily all benefit from the same technology. Part of the frustration experienced by some translators using translation tools may result from them applying the tool in an inappropriate situation. Sometimes it may be the client who insists that a particular tool be used without really understanding that it may not be suitable, whereas in other cases, it may be the translator who is not aware that another more appropriate tool exists.

Another consideration might be the size of the job. In many cases, a translation job amounts to just a few thousand words, which typically comes with a short deadline. And since each job is different, it may not be possible to use any tool without making some adaptation to either the tool or the corpus it will be used to process. As pointed out by Bédard (1995:28), by the time the tool is made operational, the deadline may be fast approaching and the cost of getting the tool to work may have exceeded the value of the job. As noted in section 4.1, TMs typically require more in terms of a learning curve and data preparation than

do BCs, so it may be that while a TM could provide a good return on investment for a large job, a BC might be a better choice for a small job.

Text type is also an important factor to consider. There are certain types of texts and writing styles that are highly conducive to being processed with a TM. In particular, texts that are a revision of a previous document (e.g. an updated version of a user manual, a re-negotiated collective agreement), are good candidates for translation with a TM because they will contain many repetitions at the sentence (or even paragraph) level. Another good candidate for use with a TM is a text where the repetitive sentences are varied (i.e., many sentences with few occurrences of each) and scattered throughout the document. However, such documents are not the only type that translators work with. Many translators are faced with texts that contain repetition primarily at the sub-sentence level. In such a case, since the manual searches initiated by the translator using a BC may be more flexible and productive than the auto-concordance search in a TM, a BC may be preferable.

The choice may also be motivated by whether the work is being done for a regular client or for a new client. If a translator works regularly for a particular client and has a corpus consisting exclusively or primarily of similar types of texts translated for that client, it may be reasonable to use a TM since presumably the "sentence salad" effect will be lessened by the fact that the documents will all contain similar terminological and stylistic preferences. In contrast, if the job is for a new client and the corpus does not contain previous work done for that client, perhaps a BC would be a better choice since the translator could consult it merely for inspiration without feeling constrained by choices made previously to suit other clients or text types.

The decision of whether to use a TM or a BC may also depend on the translator's preferred working style. Just as some drivers prefer driving a car with a manual transmission over one with an automatic transmission, some translators may favour a system that does a greater degree of automatic text processing (e.g. TM), while others may opt for one that does less (e.g. BC).

Another relevant issue may be the amount of experience the translator has. A translator who is very experienced may prefer the flexibility offered by a BC, which allows him to look up only those expressions for which he needs help. In contrast, a translator who is just embarking on his career may value the fact that a TM automatically makes suggestions for all types of text strings.

A final factor that may come into play could be cost. A single licence for a BC typically costs less

than \$200 (US), whereas a single licence for a limited version[6] of a TM retails for closer to \$1000 (US). It is true that there are usually additional features present with TM software (e.g. a termbase), and if these features will be used, then the additional cost may be worthwhile. However, if a translator intends to use mainly the concordancing feature of a tool, then it may be preferable to purchase a more modestly priced BC.

References

J. Allen. 2003. *Post-editing*. In "Computers and Translation: a Translator's Guide", H. Somers, ed., pages 297-317, John Benjamins, Amsterdam.

C. Arrouart and C. Bédard. 2001. Éloge du bitexte. *Circuit* 73:30.

F. Austermühl. 2001. *Electronic Tools for Translators*. St. Jerome Publishing, Manchester.

M. Barlow. 2000. *Parallel Texts in Language Teaching*. In "Multilingual Corpora in Teaching and Research", S. Botley, T. McEnery & A. Wilson, ed., pages 106-115, Rodopi, Amsterdam

C. Bédard. 1995.L'automatisation: faut-il y croire ? *Circuit* 48:28.

C. Bédard. 1998. Ce qu'il faut savoir sur les mémoires de traduction. *Circuit* 60:25.

C. Bédard. 2000. Mémoire de traduction cherche traducteur de phrases… *Traduire* 186.

S. Bernardini. 2002. *Educating Translators for the Challenges of the New Millenium: The Potential of Parallel Bi-directional Corpora*. In "Training the Language Services Provider for the New Millenium", B. Maia, J. Haller & M. Ulrych, ed., pages 173-186, Faculdade de Letras da Universidade do Porto.

L. Bowker. 2002. *Computer-Assisted Translation Technology: A Practical Introduction*. University of Ottawa Press, Ottawa.

J. DeCesaris. 1996. *Computerized Translation Managers as Teaching Aids*. In "Teaching Translation and Interpreting 3: New Horizons", C. Dollerup & V. Appel, ed., pages 263-269, John Benjamins, Amsterdam.

O. Evans. 2002. ATIO Offers Members Professional Development on Concordancing Tools. *Newsletter of the Association of Translators and Interpreters of Ontario* 31(2):7.

W.A. Gale and K.W. Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics* 19:75-102.

S. Hansen and E. Teich. 2002. *The Creation and Exploitation of a Translation Reference Corpus*. In "Proceedings of the Workshop on Language Resources in Translation Work and Research", E. Yuste-Rodrigo, ed., pages 1-4. European Language Resources Association (ELRA), Paris.

M. Heyn. 1998. *Translation Memories: Insights and Prospects*. In "Unity in Diversity? Current Trends in Translation Studies", L. Bowker, M. Cronin, D. Kenny & J. Pearson, ed., pages 123-136, St. Jerome Publishing, Manchester.

T. Johns. 1986. Microconcord: A Language Learner's Research Tool. *System* 14(2):151-162.

D. Kenny. 1999. CAT Tools in an Academic Environment. *Target* 11 (1):65-82.

F. Lanctôt. 2001. Splendeurs et petites misères… des mémoires de traduction. *Circuit* 72: 30.

M.-C. L'Homme. 1999. *Initiation à la traductique*. Linguatech, Brossard, Quebec.

M. Merkel. 1998. *Consistency and Variation in Technical Translation: A Study of Translators' Attitudes*. In "Unity in Diversity? Current Trends in Translation Studies", L. Bowker, M. Cronin, D. Kenny & J. Pearson, ed., pages 137-149, St. Jerome Publishing, Manchester.

D. Mindt. 1986. *Corpus, Grammar and Teaching English as a Foreign Language*. In "The English Reference Grammar: Language and Linguistics, Writers and Readers", G. Leitner, ed., pages 125-139, Niemeyer, Tübingen.

G. Palumbo. 2002. *The Use of Phraseology for Training and Research in the Translation of LSP Texts*. In "Training the Language Services Provider for the New Millenium", B. Maia, J. Haller & M. Ulrych, ed., pages 199-212, Faculdade de Letras da Universidade do Porto.

J. Pearson. 2000. Une tentative d'exploitation bi-directionnelle d'un corpus bilingue. *Cahiers de Grammaire* 25:53-69.

D. Shadbolt. 2002. The Translation Industry in Canada. *Multilingual Computing and Technology* 13(2):30-34.

R.C. Sprung. 2000. *Introduction*. In "Translating into Success: Cutting-edge strategies for going multilingual in a global age", R.C. Sprung, ed., pages ix-xxii, John Benjamins, Amsterdam.

S.E.O. Tagnin. 2002. *Corpora and the Innocent Translator: How Can They Help Him?* In "Translation and Meaning, Part 6", B. Lewandowska-Tomaszczyk & M. Thelen, ed., pages 489-496, Hogeschool Zuyd, Maastricht.

F. Zanettin. 1998. Bilingual Comparable Corpora and the Training of Translators. *Meta* 43(4):616-630.

---

[6] "Freelance" or "lite" versions may restrict database or termbase size or may lack network capabilities. "Professional" versions may cost several \$1000 (US).