

Paraconc and Wordfast: Improving Productivity

By Emily Lemon

With the rapid development of the digital world, the busywork of translation is decreased, allowing more time for the exploration of translation as an art form. Repetitive phrases are inserted automatically, dictionaries are accessed with the click of a mouse, and content specific terminology is cross-referenced to insure accuracy. The question is, with so many positives, where are the negatives? And is it realistic that a freelance translator can compete in today's job market without the help of translation software? In the following pages the advantages and disadvantages of the translation memory programme Wordfast, created by Yves Champollion, will be compared with those of the parallel concordancer, Paraconc, specifically the free beta version created by Michael Barlow. The two programmes will be examined on many levels, some very practical, such as usability, efficiency, and cost effectiveness, and some more theoretical, such as their impact on the work and language comprehension level of the translator.

Before getting into the strengths and weaknesses of Paraconc and Wordfast, I will address the question of necessity. The move toward computer assisted translation tools has been picking up speed with the advent of low cost or free software. With so many translators taking advantage of these offers and thereby increasing their productivity, those who have not taken this step must ask themselves if they will be left behind. With a glance through online job postings for freelancers, it leads one to believe that there is only work available for those using translation memory. Most agencies are more likely to give a job to someone who can do it in less time and with more consistency. The pressure to use TM might not affect the freelancers who have established direct clients providing regular work and trust only them to do it. The general consensus seems to be that without these tools the agency work will be scarce¹.

The first issue that must be addressed is that of building a parallel corpus. The most important stage in any construction is building a strong foundation. In the case of translation software, this is the construction of the corpus. This begins with locating parallel texts, texts that have already been translated and are available in both the source language and the target language. Depending on the subject area and the language pair, these parallel texts may be difficult to come by. As Maeve Olohan points out in her book *Introducing Corpora in Translation Studies*, the Internet offers some extensive sources for building a corpus but when they are published by international organisations and written by several authors it can be difficult to ascertain which one is the source text.² Being that texts posted on the Internet are not necessarily edited or proofread before being made public, the source texts and their translations must be scrutinised for accuracy. The availability of relevant material can pose problems because the nature of texts varies greatly depending on the language pair. Texts translated

¹ <http://www.proz.com/post/296196#296196>

² Maeve Olohan, *Introducing Corpora in Translation Studies* (Oxfordshire: Routledge, 2004), p. 25.

from Italian to English are usually written at a higher level, whereas those from English to Italian are more often of the best-seller variety.³

The next step in building the corpus is alignment of the source and target texts. Precision is essential here; misalignment due to an extra paragraph break or punctuation that splits a sentence in the wrong place throws off the entire process and makes it impossible to match units. The translation memory database is dependent on accurate matching. Due to the inherent characteristics of the language pair that I work in, German and English, there are frequent occurrences of misalignment. The style of written German often includes longer sentences than in English. Because of this, it is recommended in the Paraconc instructions that the shorter unit, usually the English version, be merged with the next unit rather than breaking up the longer, usually German, text unit. Another issue that leads to misalignment is abbreviations. German and English do not necessarily have corresponding abbreviated forms. Being that an automatic alignment breaks up units at every full stop, the English abbreviation *ex.*, for 'example', would break off after the full stop whereas the German equivalent, *z.B.*, for '*zum Beispiel*', would break off twice. Another example that became a problem for me was the way the date is written. In German there is a full stop after the day and month whereas English uses the slash to separate these. When aligning a large text, these details can add up to an enormous amount of correction.

Depending on the software used, aligning the texts can take days or hours. Using Paraconc, the translator must complete all the steps manually. This is sped up with Microsoft Word's "find and replace" tool, but is quite time-consuming in spite of this. The source text and target text must be divided into units by inserting paragraph marks at the end of each sentence. This is done by finding and replacing each full stop, exclamation point, or question mark with the same mark followed by a paragraph mark. The problem is, depending on the texts being worked with, full stops can be scattered generously throughout the text, not necessarily denoting the end of a sentence. This is, as mentioned above, common in German. With this being the case, paragraph marks have to be inserted manually rather than automatically. Microsoft Word then numbers each unit and the text is viewed as a long list of units, during which the alignment must be checked and edited. The next steps are to remove the numbers, spell check the documents, and save them as text files.

Wordfast, on the other hand, has a set of tools called PlusTools that offers a quicker and more efficient method of alignment. Compared to Paraconc, this is a relatively simple and straightforward procedure that lets the computer do the work. After opening Wordfast, multiple source files can be selected and the units extracted simultaneously. After extraction the translation units are counted saved; the same is done with the target texts. Following this, PlusTools is opened and by clicking on 'Align', the source and target texts will be processed and appear either side by side in two separate documents or as one document in

³ Federico Zanettin, 'Parallel corpora in translation studies: Issues in corpus design', in *Intercultural Faultlines. Research methods in translation studies I: Textual and cognitive aspects*, ed. by M. Olohan (Manchester: St Jerome), pp. 105-108

table format. The texts can then be manually modified and verified using keyboard shortcuts. During the exploration of these shortcuts I did encounter some problems; Alt M (merge), Alt W (delete cell), and Alt I (insert cell) did not work, therefore I used the mouse and the menu bar. The 'find and replace' tool in MS Word can also be used to correct repeated causes of misalignment. The documentation that accompanies the PlusTools download offers tips for the alignment process.

As the work of the translator changes with time, it is of interest to consider the impact of these programmes on this work. The quality of a translator's work depends on several factors, one of the most important being the depth of his or her understanding of the languages worked with. It is necessary to have extensive resources accessible, with which one can confirm questionable phrases or terminology. In the past these resources were made up of hard-copy dictionaries or personal glossaries. With the development of parallel concordancing software, the translator is now provided with a tool that is as dynamic as language itself. Texts relevant to the theme of the translation are aligned and can be used as an electronic resource; terms can be searched and viewed contextually, providing a more comprehensive explanation of use. Parallel concordancing software is more often used as a teaching or learning tool in language development than as a translation tool. As Olohan points out, professional translators rarely allow themselves the time to become familiar with software that will not directly speed up their output. Parallel corpora are useful in exploring specific phraseology in depth; this exploration takes time, which is not always available. Translators are more likely to invest time in learning a translation memory programme.⁴ Another helpful application of corpora is in the acquisition of LSP, or languages for specific purposes. Parallel texts can be loaded and used as a resource to develop a more extensive vocabulary when working in specific fields, such as technical, legal, or medical. The point where parallel concordancing may have an advantage over TM is in the translation of more creative texts. There is little terminology repetition and a greater focus on writing style. Here, the corpus is searched for patterns or oddities that distinguish one character from another or make a writer's work their own.⁵ When a translator takes the time to learn and make use of parallel corpora, this depth of research and precision will surely be reflected in the equivalence level of the translation and the "native-sound".

Translation memory programmes are used widely by freelance translators. They provide instant results that are visible in the increase of productivity. This depends a lot on the field that the translator works in but when there is a demand for high output, TM can really help. If the texts are of a more technical nature or have large amount of repeated phrases or terminology, the TM will automatically insert them into the translation. Wordfast, like many other TM programmes, leads the translator through the source text unit by unit. This insures that every unit is translated and nothing is missed in the rush to meet deadlines. As Frank Austermühl points out in his book *Electronic Tools for Translators*, three important factors should be evaluated before deciding on the

⁴ Olohan, p. 176.

⁵ Ibid, p. 180.

usefulness of TM. These are as follows: text type, TM is better suited to technical texts, re-usability, or how much repetitive content exists, and thirdly, volume, longer texts are more likely to have repetitive content.⁶ Another enormous benefit of a TM involves revisions. In her master's thesis on translation memory, Lynn Webb talks about the amount of unnecessary work that can be avoided when a client sends a revised version of a text that the translator has already begun working on. Rather than having to search for the alterations or additions, the revised version can be run through the TM and the software finds changes immediately.⁷

Throughout the development of computer aided translation software, the factors that must be weighed have been changing. Until a few years ago the question of cost was a significant one. The market leader for TM software was Trados, which was only available at a very high cost. Since many freelance translators work on a tight budget, it was necessary to evaluate whether the tool would be useful for their field before committing to its purchase. The price of Trados 7 Freelance now runs at approximately £500, whereas Paraconc costs about £55. Although the capabilities of Trados vary greatly from those of Paraconc, the difference in price often tipped the scales. In 1999 the idea of Wordfast was born.⁸ Providing an affordable TM programme that was compatible with different platforms and could be networked has had an enormous effect on the way freelance translators work. A demo version of Wordfast can be downloaded and used for an unlimited time; the number of translation units is limited to 500. This is big enough for small to medium jobs and when the TM is full, another can be created. Buying a licensed version of the software costs £125. Wordfast offers PlusTools as a free add-on alignment tool, as well as an online sharable TM called VLTm (very large translation memory) for free. By now, others have developed TM software at a low price, such as Metatexis or Star Transit. With this variety available, the question of cost has become almost irrelevant. This leads us to the other factors that have begun to take precedence in the decision making process.

Since the number of translators using CAT is growing daily, there is a huge demand for support during the learning process. Because the Paraconc community is focused more on the language development benefits of the programme, there is not very much information available online for those using it actively in translation. In fact, I was unable to find any resources online besides the main Paraconc site, <http://www.athel.com/para.html>. Here one can find informative articles about different uses for the programme, but a user-based forum would provide a more lively discussion. The documentation that is available to download is extensive but demands a significant amount of time to sift through and understand. I found the 90 pages of Paraconc instructions rather daunting. On top of this virtual stack of paper, the fact that there was no forum upon which questions can be posted leaves one feeling frustrated and overwhelmed.

⁶ Frank Austermtühl, *Electronic Tools for Translators* (Manchester: St. Jerome Publishing, 2001), p. 138.

⁷ Lynn E. Webb, 'Advantages and Disadvantages of Translation Memory: A Cost Benefit Analysis' (master's thesis, Monterey Institute of International Studies, 1998-2000), p. 14.

⁸ Logos Spa & Y. Champollion, 'Previous Versions', <http://www.wordfast.org/site/version.html> [accessed 6 January 2006]

In general TM software has a larger user-base; this has led to a greater amount of support, not only from the producers of the software, but also a network between the users. Wordfast is the most widely used low-cost TM software and therefore has a large number of users to be queried. There are numerous forums on translation websites, such as www.proz.com or www.translatorscafe.com, as well as Wordfast's own mailing list or peer-to-peer groups, which are available in nine different languages. The user manual is rather large at 100 pages, but also clearly written and with a linked table of contents.

The ease of use of these two programmes must also be examined. Paraconc is an independent programme that runs separately from the word processing programme used to write the translation. The one great challenge that I encountered with the programme was that I could only find the Windows version. From what I have been able to find out, the original version was written for Mac and Windows, but all links leading to where the Mac version should be are dead. The user manual only makes reference to compatible versions of Windows. Due to this fact, I have been running the Windows version of Paraconc through Virtual PC on my Mac, which of course, slows everything down. Working on two platforms simultaneously made me aware of the need to convert documents saved as 'text only' in Word for Mac into 'text only' in Word for Windows. Without this step, the German characters cannot be read properly when opened in Windows. Having worked mainly on Macs for the past 20 years, the lack of software availability is a familiar problem. These extra steps needed to make Paraconc work did add another hurdle though. That being said, using the programme itself is relatively simple. The interface is clear and navigable.

Because Wordfast is a template of Microsoft Word, if the user is familiar with Word, learning Wordfast comes naturally. As it is integrated with the programme that most translators use to do their translation, it is fast and responsive and does not use much RAM. The fact that the Mac version is as accessible as the Windows version makes Mac users that much happier. Wordfast is, as far as I could find, the only TM software that works on both Mac and Windows. In terms of using the application, the menu bar leaves something to be desired. Visually it looks outdated, but in spite of this it functions at a high level. Wordfast follows the format of Word's toolbars and lies quietly alongside the text, waiting to be used. PlusTools, the free add-on mentioned above completes the picture by providing an automatic alignment tool.

Deciding which programme to use as a translation aid depends largely on what kind of texts are being translated. Wordfast may greatly increase the productivity level of someone working with texts having a high level of repetition but may not affect the output of a translator working on literary texts. The topic of my sample parallel corpus is film. The corpus is made up of reviews, critiques, and film festival press releases from Germany and Austria. Due to the nature of the texts, the repetition rate is rather low. After the corpus contained over 10,000 words, I attempted to use it to translate another film review from one of the sites that provided many texts for my corpus. What I discovered rather quickly was that Wordfast had no suggestions for me at all. Initially this came as a surprise to me; I was under the false impression that if vocabulary repeated itself it would also be suggested. I quickly realised that the entire TU had to be a closer match in

order to even appear as a fuzzy match. In order to test the functionality of the matching process, I took one of the original sections of my corpus, changed some sentences, and ran it through Wordfast to see what the matching rate was. This gave me the results that I have been reading about. The text was quickly translated, leaving un-translated only the sentences that I had changed. It seems that if the corpus were bigger and the translation aid had more units to choose from, there would be a greater chance of finding matching units. In general though, the TM function of Wordfast is not very well suited to this text type. Other features of the software are of great assistance. For example, the context search allows the translator to highlight a word in the source or target text and perform a search of both the background TM and the regular TM for it. This informs the translator whether the term is suitable or not for the given context. Other search functions, such as the dictionary and glossary look-up as well as the reference search provide useful tools during the translation. Translators also use Wordfast for non-repetitive texts, not for the TM but rather for these other features. Guiding the translator through the text one segment at a time eases the eyestrain and improves the end product.⁹

Determining the usefulness of Paraconc initially presented more of a challenge. The programme is frequently used for language acquisition, but as pointed out earlier in this text, due to time restrictions freelance translators are more likely to teach themselves how to use a TM programme than a parallel concordancing programme. In spite of this fact, there are many uses for this type of software for the translator. The greatest of which is using it as a replacement for the dictionary. Michael Wilkinson describes in detail the process of narrowing down the most suitable collocation in a given context. When the translator has an idea of what he or she is looking for, searches are performed for terms in order to establish whether the collocation is natural. One can determine this by how many hits come up and through sorting the resulting collocations.¹⁰ Following Wilkinson's model, I performed a similar search and the main problem that I continuously ran into was that there was an insignificant amount of hits to draw any conclusions. This confirmed my suspicion that a much larger corpus is necessary when working with non-technical texts. Since my corpus field is film, I settled on the German word *Kino*, most often translated as 'cinema', in order to get a large number of hits. Paraconc came up with 30, most of which were indeed translated as 'cinema'. After sorting the hits and comparing them with their matching segments in English it became clear that there were some interesting results. In one example, *Kino der Gegenwart* is nicely rendered with 'contemporary film'. This is a natural collocation in English that, when appearing in a Paraconc search, would clearly assist the translator in making the target text smooth. Another common collocation appears with the translation of the unit *Er besucht das Kino* as 'He goes to the movies'. Although the Langenscheidts dictionary definition of *Kino* includes variations such as 'go to the cinema

⁹ Andrei Gerasimov, *Ph.D.*, 'An Effective and Inexpensive Translation Memory Tool' *Translation Journal*, Vol. 5, No. 3, (July 2001) <http://accurapid.com/journal/17wordfast.htm> [accessed 7 January 2006] (para. 6 of 11)

¹⁰ Michael Wilkinson, 'Discovering Translation Equivalents in a Tourism Corpus by Means of Fuzzy Searching' *Translation Journal*, Vol. 9, No. 4, (October 2005) <http://accurapid.com/journal/34corpus.htm> [accessed 7 January 2006]

(*Am. movies*)', the key term 'contemporary film' does not appear¹¹. This proves the point that Paraconc provides useful examples of language in use and their naturally occurring collocations that may not be found in a dictionary.

In conclusion, weighing the time needed to thoroughly familiarise oneself with the functions of Paraconc or the Wordfast template against what the two programmes have to offer the freelance translator, the translation memory software is clearly more efficient. Being that cost is not the issue that it once was and that Wordfast offers so many other useful features such as automatic alignment, a reference and context search, and guided segmentation during the translation, even if the translator is working with a small corpus and a text with little repetitive content there is a plethora of reasons to use such a TM programme. Although Paraconc has some useful features to aid the linguist in his LSP development and is a helpful tool for finding natural expressions while translating, the fact that it runs separately from the word processing program, that the corpus has to be manually aligned, and that it serves mainly as a dynamic reference leads me to question why one would choose to invest time such a programme when software like Wordfast exists, which offers the same features and more for approximately the same price.

Emily Lemon M.A.

www.lemontranslation.com

¹¹ *Langenscheidts Großes Schulwörterbuch Deutsch-Englisch*, 1996, s.v. "Kino"