Software for translators

Laura Carolina Collada Ali[1], Laurence Auffret[2]

[1]*OnTranslation, Italy*
[2]*Cinetique Translations, UK*

Correspondence to:

Laura Carolina Collada Ali
OnTranslation,
laura.collada@ontranslation.it

Abstract

In recent years, a myriad of software has revolutionised the translation sector. This article presents a series of translation tools commonly used in the translation industry, depicting their main features and the way they have transformed the translation process from project assessment to delivering client-ready documentation. This article is written from a translator's perspective and aims at explaining how this software is included in the intricacy of a translation project.

Keywords: Translation software, Translation tools, Computer-assisted translation

In the past 15 years, the increasing use of Computer-assisted translation (CAT) tools has radically changed the translation process. With an ever increasing volume of software on the market, sometimes claiming to be 'the solution' to all translation requirements, it can be challenging to understand the way CAT tools are integrated in the translation workflow and what their use means for translators, agencies, and clients alike.

This article presents the main features of CAT tools frequently used by translators, their role in the translation process, and their major advantages.

Counting content

Freelance translators, as well as translation managers and agencies, often issue quotations and invoices based on document text analysis, of which length – measured in terms of characters, words, lines, pages, and sometimes including repetitions – is one of the most important characteristics. It may seem simple to measure the number of words and pages in a Word document, but in real life, translators work with a myriad of different formats. Thus, software that is able to measure content no matter what the format is essential. Let us see two examples of the available tools in the following sections.

CATCount
Translation agencies often provide translators with a rate scheme using the volume of recurrent content in the source text(s); CAT tools are able to memorise this in the translation memory, as we will see later.[1] Professional translators are not always happy with this approach, because repetitions in a given target text may still need to be translated differently according to the context and, thus, they can take as much time to translate as the non-recurrent content. However, a tool such as CATCount may help in calculating the final rate, as it converts complex CAT schemes into one easy number.[2] In addition, it provides great features for the printing and export of used CAT schemes together with the original word counts, and the resulting CATCount.

The concept of using a single number of words in a given text is not new, and has been used extensively by many translation agencies worldwide under various names: leveraged word count, weighted word count, equivalent word count, etc.

CATCount is a freeware product.

FineCount

FineCount analyses documents in a vast array of formats and provides advanced text statistics. There is a free version, which is available only for single-computer and non-commercial use, and a professional version,[3] which is also able to count the content in websites, export text, and prepare invoices and reports.

Computer-assisted translation

Translators use CAT tools to support, improve, and facilitate the translation process. 'CAT tool' is a broad and imprecise term covering a range of software, from the fairly simple to the more complicated. These tools are not to be confused with machine automatic translation, which is a sub-field of computational linguistics that uses software to translate text from one language to another. A very popular software in this field, although its results are far from satisfying, is GoogleTranslate.[4]

These tools can include:

- Project management software that allows linguists to structure complex translation projects,

assign the various tasks to different linguists, and track the progress of each of these tasks.

- Grammar checkers, either built into word processing software or as add-on programs.
- Electronic dictionaries, monolingual or bilingual.
- Terminology databases, either on the host computer or accessible through the Internet, such as TERMIUM Plus[5] or Grand dictionaire terminologique from the *Office québécois de la langue française.*[6]
- Terminology managers, which allow the translator to manage their own terminology database in an electronic form. This can range from a simple spreadsheet, to a database created in a program such as FileMaker Pro[7] or, for more robust (and more expensive) solutions, specialised software such as SDL MultiTerm,[8] LogiTerm,[9] Termex,[10] etc.
- Translation memory (TM) tools, consisting of a database of text segments in a source language and their translations in one or more target languages.
- Full-text search tools (or indexers), which allow the user to query already translated texts or reference documents of various kinds. Examples of indexers are ISYS Search Software[11] and dtSearch Desktop,[12] among others.
- Concordancers, which are programs that retrieve instances of a word or an expression and their respective context in a monolingual, bilingual, or multilingual corpus, such as a bitext or a TM.
- Bitext aligners, which are tools that establish an in-depth comparison and match from a source text and its translation; the resulting database can then be analysed by using a full-text search tool or a concordance.

Workflow	Activity	CAT tools
File submission	Project analysis	Wordcount
	Workload estimation pricing	Analysis of recurrent segments
Project start	Organisation of reference materials	Alignment Glossaries
	Linguistic guidelines	Concordancers
Translation	Translation	TM
Editing	Editing	Spellcheckers
Proofing	Proofing	
Quality assurance (QA)	QA	TM Linguistic guidelines
Project close	Archive new translation in the TM	TM*

*This updated TM can then be re-used to analyse, translate, and proof future projects.

We find the following among the most frequently used features of this kind of software.

Translation memory

TMs are databases that store source sentences and their translations into manageable units known as 'segment pairs'. The TM 'remembers' each source–target segment that is translated, and as you work through a document, if the program finds an identical or a similar matching source segment in its database it provides a previous translation for re-use, and thus, it does not need to be translated from scratch. If the program does not find an identical or a similar matching source segment, it allows you to enter a translation for the new segment. After the translation is completed, the program stores the new translation and moves on to the next segment. This technology offers three main advantages:

- Faster turnaround times and cost-savings than when the TMs are not in use, as repetitive content can be translated quickly.
- Better consistency and readability, as you can make sure that the program does not translate the repeated occurrences of the same sentence differently, which can confuse the readers.
- A CAT system can automatically scan the documents and identify places where numbers have been mistranslated, simple typos, or incorrect punctuation.

Alignment

When you start using a TM tool, you will have to fill it with your own translations, as the database is empty at the beginning of the project. Alignment programs take completed translations, divide both the source and the target texts into segments, and attempt to determine which source and target segments can be matched in order to build a TM database with the given content. These segment pairs can then be imported into your TM database. This technology is invaluable to easily re-use your previous translations.

Terminology databases

Terminology databases are similar to electronic dictionaries. Rather than storing whole source–target segments as translation memories do, they contain single words or expressions. Creating and maintaining terminology databases is useful for storing organisation-specific, customer-specific, or product-specific terms, which cannot be found in standard dictionaries.

When translating a document, the source sentences are automatically searched for relevant terminology in the background. The corresponding target terms can then easily be inserted into the translation.

This saves you the trouble of having to open a separate application in order to do time-consuming research work. Some of the more advanced systems enable the translators to check – interactively or in batch mode – if the correct source–target term combination has been used within the TM segments in a given project. Terminology databases offer the following main advantages:

- Cost and time savings, as you avoid extra work such as researching and typing long-winded expressions.
- Readability improves, as the same expressions are used consistently throughout a document. The terminology databases help you to effectively avoid unnecessary variations, which can confuse the readers.
- You can create and manage terminology that is customer- or product-specific and that cannot be found in any standard dictionary.

Some notable CAT tools

The list below includes some of the most frequently used tools available on the market, although several relevant tools are missing.[13]

Tool	Supported file formats	Operating system	Licence
MetaTexis	Microsoft Word, Excel, and PowerPoint, and all kinds of text formats	Microsoft Office Word add-in	Proprietary
MemoQ	Microsoft Office files, HTML, Adobe FrameMaker, Adobe InDesign, XLIFF, XML, Scalable Vector Graphics (SVG), Typo3, and many others	Windows	Proprietary
OmegaT	Microsoft Word, Excel, PowerPoint, XHTML, and HTML, Open Document Format, MediaWiki, Plain text, and many others	Cross-platform (Java)	GNU GPL
SDL Trados	Microsoft Word, Excel, PowerPoint, Open Office, InDesign, QuarkXPress, PageMaker, Interleaf, FrameMaker, HTML, SGML, XML, SVG	Windows	Proprietary
Wordfast	Microsoft Word, Excel, PowerPoint, PDF, SGML, HTML, XML, InDesign, FrameMaker, tagged documents, XLIFF	Cross-platform (Java)	Proprietary

According to a 2013 survey undertaken by Jared Tabor of over 3000 translation professionals from different countries, primary tool usage was reported as follows: SDL Trados (43.2%), Wordfast (11.5%), MemoQ (8%), and other softwares (37.3%).[14]

Voice recognition

Typing is considered to be an 'old activity' and some still seem to be reticent about abandoning it, yet it strains the fingers, the eyes, the wrists, and the back. Translators may type from 2000 words a day – if the source texts are rather technical or research in the field is needed, to 3000 words a day – if the subject matter corresponds to their specialisation domain. Typing may be slow and exhausting.

Human speech is 5 to 7 times faster than typing; cognitive processing about 10 to 15 times faster, states Hendzel.[15] Nevertheless, our translating brains actually work closer to these speeds – think, for example, about simultaneous interpreting!

In this context, voice recognition may seem a miracle, yet modern voice-recognition software has matured in accuracy to a point that it is difficult to talk faster than the computer processing speed. What are the highlights? This methodology is fast, always available, accurate, and never gets tired. Thus, it enables professional-quality translation at lightning speed.

Professional-quality voice-recognition software must be available in your target language. Dragon Naturally Speaking v. 12 is available in US English, UK English, French, German, Italian, Spanish, Dutch, and Japanese.

However, dictated translations are still drafts that require review and editing, although as you develop and refine your dictation skills you will reduce the volume of editing required, although as in any quality translation process, translations always need to be reviewed.[15] You may also need some interpreter DNA to verbalise translations.

Proofreading

Tracking changes in the translation process is essential for QA. In some regulated industries like clinical trials, all the changes have to be well documented and approved; this means that tracking changes is an everyday task for many translators, reviewers, and project managers.

ApSIC Comparator

ApSIC Comparator is a freeware tool that allows you to create side-by-side reports with all the changes made to a translation by a reviewer in different document formats. These comparison reports can be used both to evaluate the quality of the translation and to provide feedback to the translator.[16]

ChangeTracker

ChangeTracker is a freeware tool for tracking changes in bilingual documents and comparing

translations. This tool generates a report flagging all the changes implemented by a reviewer, in a similar way to Microsoft Word's 'Track Changes' feature. This report can then be edited with the addition of comments and error classification codes. ChangeTracker also allows reports to be exported to Microsoft Excel.[17]

Translation quality assurance

Translation quality assurance (TQA) software compares the source and the target segments of bilingual texts in order to detect translation errors, such as:

- Inconsistencies
- Terms that have not been translated using a project glossary
- Omissions
- Target segments, which are identical to the source segments
- Punctuation, capitalisation, number value, and formatting errors
- Incorrect untranslatables and tags

TQA tools cannot detect mistakes arising from an incorrect understanding of the source text, poor stylistics, or an inappropriate choice of language register. Furthermore, often TQA tools expect the source text to be correct, which is not always the case. Thus, these tools may be very helpful, but they need to be used with care.[18]

One of the most popular ones is ApSIC Xbench, which provides powerful QA features in one single view and supports many different CAT tool formats. Its spellchecker is available in 33 languages. It comes with some predefined quality checks for completeness, consistency, numbers, tags, key terms, and so on, but you can also create your own personal checklists. It is even possible to upload hundreds of files at a time and edit only those that have spelling issues.

In terms of practicality, CAT tools have very clear advantages as they offer translators convenient features such as the possibility of working between several documents without having to look back and forth between them and they allow focusing on the very expression being translated. As the translators save time looking up terminology in hard copies of reference materials, they can dedicate more time to writing and are more productive. CAT tools also eliminate the need for typing several occurrences of the same expression. Above all, they ensure consistent use of terminology and style across all documentation and this has a significant impact when working on large projects.

However, stemming from these advantages, using CAT tools without a clear knowledge of translation issues or adequate training, can also bring a number of problems. First and foremost, focusing on one segment or sentence at a time is detrimental to producing a good overall style for a document. It is essential, therefore, to operate a final proofing stage on the fully translated document, aside from the source. Furthermore, the automatic generation of translation 'segments' carries the risk of disseminating an error. So, it is important to clean up the terminology database prior to starting a new project. Regrettably, this is often overlooked because of the time constraints.

As professional translators, we could not work without these tools and would like to encourage any aspiring or established translators or proofreaders to invest in these resources and undertake proper training before starting to use them. Being up-to-date in our own field of experience also means being able to implement up-to-date technologies to ensure a better quality for our outputs.

References

1. Translation memory. Wikipedia [cited 2013 Oct 09]. Available from: http://en.wikipedia.org/wiki/Translation_memory.
2. CatCount [cited 2013 Sep 20]. Available from: http://www.catcount.com/.
3. Tilti Systems. FineCount [cited 2013 Sep 20]. Available from: http://www.tilti.com/software-for-translators/finecount/.
4. Google Translate [cited 2013 Oct 09]. Available from: http://translate.google.it/.
5. Public Works and Government Services Canada. Termium [cited 2013 Sep 20]. Available from: http://www.btb.termiumplus.gc.ca/.
6. Office québécois de la langue française. Le grand dictionnaire terminologique (GDT) [cited 2013 Sep 20]. Available from: http://gdt.oqlf.gouv.qc.ca/.
7. File Maker [cited 2013 Sep 20]. Available from: http://www.filemaker.co.uk/.
8. SDL MultiTerm Desktop 2011 [cited 2013 Sep 20]. Available from: http://www.sdl.com/products/sdl-multiterm/desktop.html.
9. Terminotix [cited 2013 Sep 20]. Available from: http://www.terminotix.com/index.asp?content=brand&brand=2&lang=en.
10. Take[lab]; TermX v1.0 [cited 2013 Sep 20]. Available from: http://takelab.fer.hr/termex/.
11. Perceptive software [cited 2013 Sep 20]. Available from: http://www.perceptivesoftware.com/products/perceptive-search.
12. dtSearch [cited 2013 Sep 20]. Available from: http://dtsearch.com/.
13. Computer-assisted translation [cited 2013 Sep 20]. Available from: http://en.wikipedia.org/wiki/Computer-assisted_translation.
14. Tabor J. CAT tool use by translators: what are they using? [cited 2013 Sep 20]. Available from: http://prozcomblog.com/2013/03/28/cat-tool-use-by-translators-what-are-they-using/.

15. Hendzel K. Professional-quality translation at light speed: why voice recognition may well be the most disruptive translation technology you've never heard of [cited 2013 Sep 20]. Available from: http://www.kevinhendzel.com/professional-quality-translation-at-light-speed-why-voice-recognition-may-well-be-the-most-disruptive-translation-technology-youve-never-heard-of/.

16. ApSIC Comparator [cited 2013 Sept 20]. Available from: http://www.apsic.com/en/products_comparator.html.

17. Technolex software. ChangeTracker [cited 2013 Sep 20]. Available from: http://www.change-tracker.com/.

18. Gerasimov A. Review of translation quality assurance software [cited 2013 Sep 20]. Available from: http://www.translationdirectory.com/articles/article1253.htm.

Author information

Laura Carolina Collada Ali, *freelance translator and writer – OnTranslation* She started her career as a medical translator and writer at the European Organisation for Research and Treatment of Cancer (EORTC) in Belgium. She then moved to Italy, where she worked for a not-for-profit research organisation specialising in haematology (GIMEMA). In 2011, she became a freelancer and she now offers translation and editing services.

Laurence Auffret *MSc Bioeng, MA Ling – CINETIQUE Translations* Laurence founded CINETIQUE Translations (ISO 9001:2008) in 2003 after working as a scientist in France and a language lecturer in UK universities (Translation, Computer-assisted learning, Erasmus). CINETIQUE Translations specialises in language services for the life science industry with a strong focus on clinical trial and biotech documentation. Laurence currently manages large translation projects for European- and US-based clients as well as developing her company's activities. In May 2009, Laurence also became part of EMWA's EC and is a past President of the EMWA.