

INTRODUCTION TO CAT TOOLS

Workshop presented by Eduardo Berinstein at the 2018 Annual NETA Conference

Welcome. Thank you to organizers of this wonderful yearly conference, that allows us to get together, socialize and exchange valuable information.

Scope: Quick glimpse into the state-of-the-art technology available to translators

Disclaimers: I will show you the tools that I use, but there are many different brands that do exactly the same thing. I do not represent any of these companies, nor do I have a vested interest in any of them.

The programs I use are:

IntelliWebSearch

GT4T

SDL Trados 2017 Freelance Plus

Dragon Professional Individual

I) Project management

In project with multiple files, CAT tools can analyze the content and generate reports that show

- Full and fuzzy match analysis
- Intra-and cross file repetitions
- Translation memory leverage
- List of files with individual word counts

Example: a project with 47 files

Analyze Files Report

Summary

Task:

Analyze Files

Totals

Total

Files:47

Chars/Word:5.93

Type	Segments	Words		
PerfectMatch	0	0		
Context Match	4	4		
Repetitions	312	1958		
Cross-file Repetitions	579	5393		
100%	84	110		
95% - 99%	99	118		
85% - 94%	18	56		
75% - 84%	37	158		
50% - 74%	32	134		
Total	3417	42002		

File Details

Think about how long it would take to analyze this by hand!

II) Translating different file types

- SDL Trados 2017 Freelance (and many other similar CAT tools) can handle
 - Multiple file formats: all Microsoft Office line, InDesign, HTML and more.
They are all converted to XLIFF

XLIFF means “XML Localization Interchange File Format”

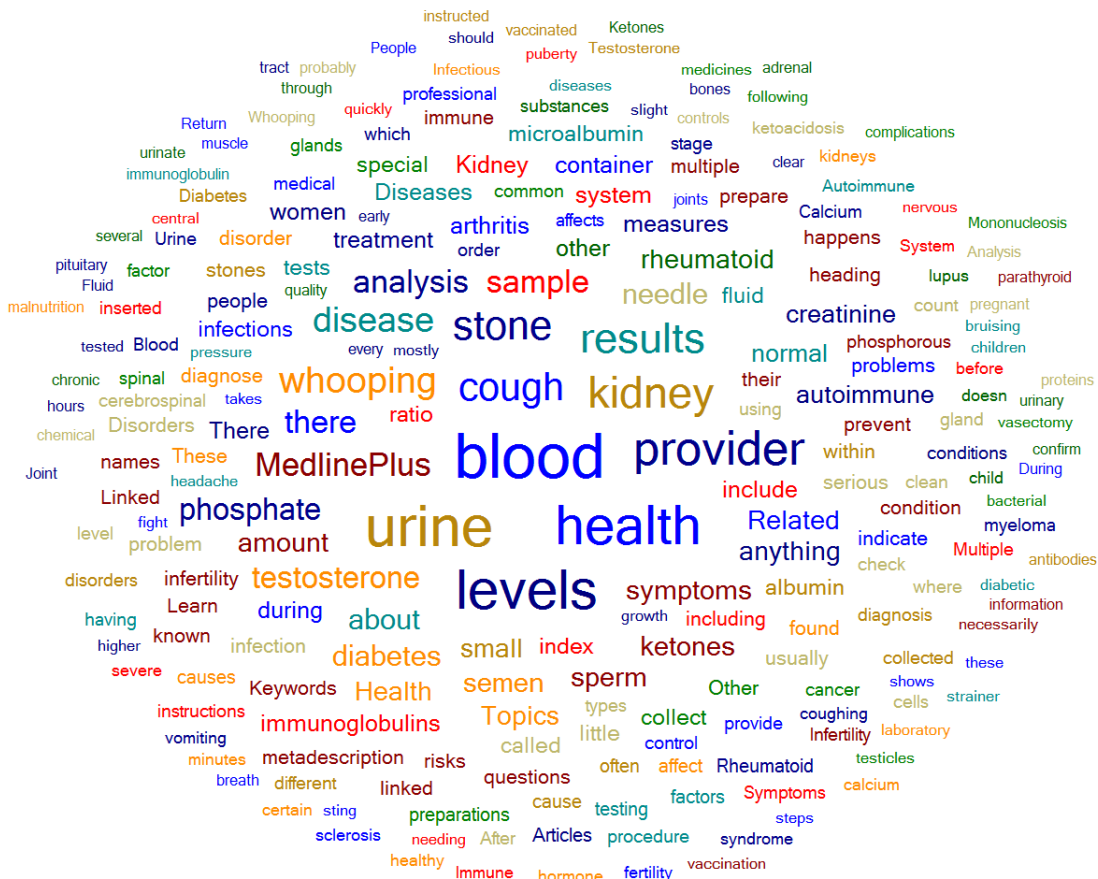
After translation, XLIFFs can be exported back to the native file format.

This is very convenient. For example, in InDesign documents, you don’t need to extract the copy to translate it, for the designer to later on place it back manually in the publication. You can translate InDesign documents **directly**, and when you’re done you can export them back to InDesign, and voila! The source language text is replaced by the translation. This saves work on both ends: the translator does not need to extract text from a PDF, and the designer does not need to place back the translated text back manually (which often times opens the door for errors to creep in).

- PDF conversion (“Pretty Darn Frustrating” no longer): SDL Trados 2017 can convert PDFs into fully editable Microsoft Word documents automatically in the process of creating a project. Prior to this, separate programs—which were pretty costly by the way—were required to accomplish this (for example, OmniPage, ABBYY FineReader, etc.)
- SDL Trados 2017 has a built-in OCR function. It is no longer necessary to have a separate program OCR program. In the process of creating a project, text embedded as an image is automatically “read” and converted into fully editable text.

III) Terminology extraction and glossary creation

There is an app for SDL Trados 2017 called **projectTermExtract** that, as its name indicates, extracts the terminology from projects and creates a “word cloud”:



The translator can set criteria for the terminology extraction, for example, terms must have a minimum number of characters, or appear a certain number of times. The extracted terminology list can be added to the project as a separate file, translated, and converted into a term base. This can be useful for working out the terminology before actually translating, or for standardizing terminology when working with a team of translators.

- Trivia: In Spanish, “Computer Science” is called “Informática” (information manipulation)

You can see a demo of how this works in the YouTube video posted in this [webpage](#).

Other useful programs and apps for terminology management

- SDL MultiTerm (other CAT tools have equivalent terminology management programs)
- SDL Trados app Glossary Converter: Drag-and-drop an Excel file and it will be converted into a term base, and vice versa. Very easy to use!
- QA: Making sure the desired terminology has been followed

IV) Alignment tools

These tools allow you to create translation memories from bilingual corpora or legacy materials. For example, you are asked to translate a website that has already been partially translated by someone else, and you must maintain consistency.

- <http://www.youalign.com/>
 - Has paid and unpaid versions. Pros and cons: Unpaid version raises confidentiality issues because you have to upload files to their server. Paid version can be local in your computer.
- After text is aligned, it can be imported into a translation memory

V) Electronic reference materials (dictionaries, thesauri, encyclopedias)

VI) Machine translation

Machine translation is here to stay, and it is getting better by the day. It can be a useful way of creating a first draft to be further edited.

Issues with machine translation

1. Confidentiality (unless it is used **selectively**, anything that is machine-translated is sent to an outside server, and therefore the confidentiality goes out the window). Cannot be used with certain types of documents.
2. Machine translation can be a double-edged sword in the hands of the inexperienced.

If you are a **good** translator, with MT you can produce **good** (or maybe even better) translations, twice as fast, with a fraction of the effort. If you are a **bad** translator, you will produce **bad** translations, twice as fast, with a fraction of the effort.

A little translation exercise: Howard Zinn book title example.

A People's History of the United States: 1492 to present (a book of American history told from the point of view of the oppressed, slaves, American Indians, woman, men who didn't own property and therefore initially did not have the right to vote, etc.)

Machine translation: Una historia del pueblo de los Estados Unidos: Desde 1492 hasta el presente (even though this is perfectly correct, to most Spanish readers this would look like **another** history book)

Human translation (published title of the book in Spanish): La **otra** historia de los Estados Unidos: Desde 1492 hasta el presente. (Bold and underline are mine, the back translation would be “The Other History of the United States of America”) The human translation succeeds in conveying the idea that this is a different type of history book. But the translator obviously took some “poetic license.”

This is where humans can outperform machines! Machines tend to err on the side of being too literal. Good translators, instead, understand that their job is not just to transpose words or phrases from one language to another, but to distill the idea the author meant to communicate, and to re-express it in another language so that it is **understood**. This often means rewriting and departing from the actual **form** of the message. Some translation scholars have called this interlingual intercultural mediation (rather than just translation).

- Concept of FAUT (Fully Automated Useful Translation) and FAUHQT (Fully Automated Useful High-Quality Translation, the “Holy Grail”). Ray Kurzweil predicted that FAUHQT will be available in 2029 (9 years from now!)
- Is output useful? Depends on the POV (point of view). Useful for what? For the translator, the usefulness could be a decent first draft that can be edited and improved **faster** than translating it from scratch by typing or dictating. The usefulness could also be that it offers an easy and fast way to consider other translation options of words, phrases or sentences, or even to research terminology.
- Quality indicator of MT: Can output be understood without looking at the source text?
- A quick history
 - Rules-based MT
 - Statistical Phrase Based MT
 - Neural Machine Translation
 - Hybrid

GT4T

- What does GT4T do? See this short demo: <https://gt4t.net/en/documentation/>
- Stand-alone program independent of any CATs or programs; works everywhere you can type
- Allows control of lexicon via dictionary
- Allows for concordance searches within a document
- Allows for selective use of MT
- Draws from multiple sources, including DeepL alternative translations
- Very flexible pricing (can be purchased temporarily for a project)

Other machine translation options

- SDL Trados Adaptive Machine Translation (machine translation engine that is personalized and learns from your corrections and edits)
- DeepL and DeepL Pro (interactive machine translation)

VII) Internet research tools

IntelliWebSearch: incredibly useful tool to query anything online or selectively fetch GT or other MT
<http://www.intelliwebsearch.com/version-5/>

What does IntelliWebSearch do?

Pressing a convenient customizable shortcut key:

1. copies selected text from your translation environment by simulating the Windows copy function;
2. strips the text of paragraph marks, line breaks, tabs, double spaces, punctuation marks and various other superfluous characters (customizable);
3. opens your default browser (Microsoft Edge, MS Internet Explorer, Firefox, Google Chrome, Opera, etc.) and sends the copied text to one of a virtually unlimited number of user preset search engines, on-line dictionaries or on-line encyclopaedias together with other customizable instructions which limit the search according to the advanced search options available for the chosen web resource. Alternatively it can send the same text and instructions to local dictionaries on CD-ROMs or installed on your hard disk.

Another convenient shortcut key (known as the *Return Key*) copies any text you select in your browser or local dictionary and returns you to your translation environment, where you can choose to paste in the text using your translation environment tool's own paste function

VIII) **Translation Memory** (for leveraging our previous work and “collecting royalties” on it).

What is Translation Memory?

<https://www.youtube.com/watch?v=MmsBe7BitG4>

Now, in SDL Trados, you can leverage not just full sentences but also words and phrases via SDL upLIFT technology, which introduces matching based on fragments.

And upLIFT Fuzzy Repair intelligently uses your own trusted resources to repair fuzzy matches, so you save time and get the best match possible.

Fuzzy Match Repair can draw from a number of translation sources to perform fuzzy match repair:

- Machine Translation
- Termbases
- Legacy and new translation memories.

SDL Trados AutoSuggest: Predictive typing based on your own way of writing, draws from your own Translation Memory

IX) **Speech Recognition (for entering text by speaking)**

X) **Text-to-Speech Readers (for proofreading text by listening instead of reading)**

There are many products in the market, and many different languages.

Example: <https://www.naturalreaders.com/online/>

For more information on CAT tools, I recommend subscribing to

<http://www.internationalwriters.com/toolkit/index.html>