hispaniatranslations

5 Bilingual Data Resources That Work With Your CAT Tool.

By hispaniatranslations 6 Comments Categories: Blog Tags: bilingual data, CAT, corpora, resources, technology, translation memory Tweet

The use of translation memories has become somewhat of a standard in our industry. If you work for agencies they are now expecting that you at least have a command of TRADOS. Although I have used TRADOS, and see that our productivity increases using it, I don't personally own a copy of the software yet. I won't be discussing about TM programs per se, I believe too many articles and blogs have discussed the pro's and con's of it. Rather, I would discuss the resources that can work well with the use of a TM **program or CAT tool** or use as a standalone resource. They are large amounts of bilingual data (corpora) extracted from websites in all particular areas. The reason I love these resources is that they are not the result of machine translation but rather large sources of data that have been translated by translators in a specialized field. So the result is an accurate copy of what you search for. The following list is but a few that I have in

my arsenal of resources and some that I have used so far. They are in no particular order.

Linguee
www.linguee.com

This is the corpus that I use the most and like the most. I do find it the most complete and it is continuously improving its site. It first started with an English German language pairing. It then included Spanish and Portuguese and now French. It gathers large bodies of information from websites and the internet and matches it with your search. It also tells you the link to the website where this information appeared and gives you several contexts in which the query appeared and the website link (This can also be downloaded and aligned to your TM). On the side bar there is a dictionary that will help you to understand the phrase and it also allows you to add information to improve the corpora itself. Of all the TM's I use I find this one the most useful and most complete. You may even download their dictionaries as a GTL file.

My Memory
mymemory.translated.net

Established in 1999, My Memory is another resource that I use extensively. It gives you both, internet searches done by human translators and when none is available it gives you machine translated text (from Google Translate). Like Linguee, it tells you where the source came from but here it allows you to rate and improve the entry. Furthermore, you can also use their existing memory by downloading a TMX file of your document into their system. It searches the memory for you so you can work with it through your CAT tool. You may also contribute your memory to improve the site. They do protect the identity of the material and only use the memory they need. It is available in many language combinations and you can either search by phrase or word. It does not have a dictionary entry like Linguee, but it is also another resource . It is free to join and given that they can provide massive information for you , this is not quite bad at all.

TAUS Data Association
www.tausdata.org

This corpora is a paid subscription that allows you to download the memory directly into your CAT tool. Through their website you can browse their extensive catalog, some are

public and others are for paid members. Categorized by subject, you have access to a vast variety of subjects and through this website www.taustracker.com you will soon be able to have access to directories from a specific translation memory.

Finally, there is always Google, but they to have a particular bilingual engine.

2lingual

2lingual.com

Powered by Google and now Bing search engines this bilingual engine provides side-by-side samples of websites in the language pair that you are looking for by word, phrase, or keyword. Just another free source.

And there you have them, these are just a few that I have found useful to me. There are many others of course , just google "translation memory" or "bilingual data" and hundreds of pages will appear. However, you have to select the ones that are more useful to you. This is my sample.