

Cork workshop on Free Tools for Multi-Re-Translation Studies

Advance notes

It will be interesting if at least some (two or three) participants/groups can prepare their own *multi-re-translation corpus* to experiment with in the workshop.

This means a folder which contains:

- A Source Text, in 2 file formats: a Word .doc/docx file, and a UTF8 .txt file. UTF8 is a kind of plain text (txt) format some programmes require. You can save from Word to that format (it's an 'Other encoding').
- Target Text subfolders: two of them: each contains a set of differing Target Texts; one folder contains Word .doc/docx files and the other contains UTF8 .txt files.

Note:

- Word files should be 'cleaned' of extraneous formatting – e.g. copy/paste into Notepad and copy/paste back into Word; or open the .txt file, then save it as a Word .doc file.
- Paragraphs (or lines of verse) are best marked with 2 paragraph break codes: that will make some processing easier.
- For some purposes it's best if you mark sentences as paragraphs that way, too.
- Name files in a simple, consistent way, e.g. with the name of the creator/translator (or a student number).
- The ST can be any kind of text; any length. TTs can be close translations or free versions, by professionals, or students, or anyone. But the more closely a TT is aligned with the ST, the easier it is to work with.
- Any language pair and direction is fine. More tools work with English than other language. And if Target Texts are in English, more people can read the differences.

We will be playing with the following, so feel free to check them out in advance:

- **LF Aligner** – freeware for aligning segments in as many parallel texts as you want. Download: <https://sourceforge.net/projects/aligner/>
- **Version Variation Visualisation (VVV)**. A platform (made in Swansea) for viewing and running certain kinds of analysis of multi-re-translation corpora. 'Guest' access at www.delightedbeauty.org/vvvdraft (latest installation, with my current Bovary/English project) or www.delightedbeauty.org/vvvclosed (includes my Othello/German project)
- **Word2Dream** by Eran Hadas and Eyal Gruss – <http://eranhadas.com/word2dream/> (2015). This is two poet-programmers' critical response to Google's **Word2Vec** (2013) - a statistical modelling of language which helps underpin the rapid recent advance of MT. See Word2Vec visualised at **Word2Viz**: <https://lamyowce.github.io/word2viz/> (by Julia Bazińska with Piotr Migdał, 2017)

We might also have time to look at things like

- **Voyant** - <https://voyant-tools.org/>
- **Textalyser** <http://textalyser.net/index.php?lang=en#analysis>
- **MALLET** Topic Modelling tool: <https://code.google.com/archive/p/topic-modeling-tool/> (download under 'Source')