

## **Discipline, discourse and new viewpoints. A Report on the *Coruña Corpus***

Isabel Moskowich, Begoña Crespo, Luis Puente-Castelo and Leida  
Maria Monaco (University of A Coruña, Spain)

The *Coruña Corpus of English Scientific Writing* (henceforth, *Coruña Corpus* or *CC*) is a corpus of late Modern English scientific writing, covering the period between 1700 and 1900.

When the project started back in 2003, the *Coruña Corpus* was influenced by a growing interest and gradual increase in studies in both the vernacularisation of Science (Crossgrove 2000) and genre conventions (Taavitsainen 2004). With this idea in mind, the *Coruña Corpus* has been designed as a tool for the study of language change in English scientific writing in general and within each different scientific discipline in particular.

The *CC* has been conceived of as a collection of several twin corpora, one for each different discipline, but all sharing the same structure, organisation, mark-up, and principles of compilation. Each subcorpus contains samples from a given discipline (delimited by adapting UNESCO's classification of Sciences to the reality of the period and with an inclusive view) at a rate of two c.10,000-word samples per decade, thus adding up some 400,000 words per subcorpus. Each sample is encoded in XML following TEI conventions, respecting the original layout and spelling variables as much as it is feasible. Each sample is accompanied by a metadata file containing information about the author of the text as well as about the text itself.

The *Coruña Corpus* is exploited with a specifically designed information retrieval application, the *Coruña Corpus Tool (CCT)*, which works as most concordancers but offers the possibility of searching old characters, spelling variants, and tags; as well as selecting samples on account of metadata parameters.

As of now, the subcorpora on Astronomy (Moskowich & Crespo 2012) and Philosophy (Moskowich *et al.* 2016) have been finished and published, and the subcorpora on Life Sciences, History and Chemistry are in several stages of beta development, with compilation of the samples having finished but work on metadata still in process. Work is now underway in the compilation of the subcorpus on language and linguistics. The fact that disciplines pertaining to the traditionally labelled hard and soft sciences are compiled allows also comparative studies regarding the evolution of different discursive trends.

This poster presents the latest developments in the project, with special attention being paid to the corpora on history, chemistry, and linguistics.

### **References**

Crossgrove, W. (2000). The Vernacularization of Science, Medicine, and

- Technology in Late Medieval Europe: Broadening Our Perspectives. *Early Science and Medicine*, 5 (1), 47-63.
- Moskowich, I. & Crespo, B. (Eds.). (2012). *Astronomy 'playne and simple': The writing of science between 1700 and 1900*. Amsterdam/Philadelphia: John Benjamins.
- Moskowich, I., Camiña-Rioboo, G., Lareo, I. & Crespo, B. (Eds.). (2016). *The Conditioned and the Unconditioned: Late Modern English Texts on Philosophy*. Amsterdam: John Benjamins.
- Taavitsainen, I. (2004). Transferring classical discourse conventions into the vernacular. In I. Taavitsainen & P. Pahta (Eds.), *Medical and scientific writing in Late Medieval English* 37-72. Cambridge: Cambridge University Press.