Abs-163
Catherine Smith and Michaela Mahlberg (University of Nottingham)
CLiC (Corpus Linguistics in Cheshire): Developing a tool for corpus stylistics

The term 'corpus stylistics' is now commonly used to refer to work that combines literary and corpus studies. However the qualities of literary texts and the questions that stylisticians want to ask differ to some extent from research questions in mainstream corpus linguistics. Therefore literature specific corpus tools are needed if corpus stylistics is to realise its potential.

This paper will report on the development of CLiC (Corpus Linguistics in Cheshire), a web-based search tool and automated corpus annotation scripts to support research in corpus stylistics. We have so far developed a prototype of the tool that allows us to explore options for text searches. Our main interest is in the association of lexico-grammatical patterns with positions in the text. Research on text types other than literature has started to show that lexical items have a tendency to occur at particular places in a text, i.e. the beginning or end of texts, paragraphs or sentences (e.g. Scott and Tribble 2006; Hoey and O'Donnell 2008; Mahlberg 2009; Csomay and Cortes 2010; Römer 2010). For literary texts other relevant textual positions are stretches of text in relation to quoted speech, e.g. words and phrases in direct speech may differ from the language in the narrator text or reporting clauses that accompany speech have preferences for certain reporting verbs (c.f. Semino & Short 2004). We use texts from project Gutenberg to annotate them in XML using a series of scripts to add chapter, paragraph and sentence structures. On top of this basic structure further scripts are used to annotate sections of the text between quotation marks, these being roughly equivalent to direct speech or thought. CLiC is based on an XML search engine called Cheshire3 which allows searches based on the XML structure in the texts. We want to develop the CLiC interface so as to allow complex context based searching of the XML. As well as more basic searches that are similar to standard corpus tools, CLiC allows the user to search for two or more words in the same sentence, paragraph or speech instance. Results can be viewed highlighted in full chapters of the text to study context or in the more traditional corpus style key word in context view.

At the same time as demonstrating initial functionalities of the prototype, the paper will look at possibilities for stylistic research that arise through such search options and outline the direction into which we aim to develop CLiC.

References

Csomay, Eniko & Viviana Cortes. 2010. Lexical bundle distribution in university lectures in Stefan Th. Gries, Stefanie Wulff & Mark Davies (eds.) Corpus linguistic applications: Current studies, new directions. Amsterdam: Rodopi.

Hoey, Michael & Matthew Brook O'Donnell. 2008. Lexicography, grammar, and textual position. International Journal of Lexicography 2008 21(3): 293-309

Mahlberg, Michaela. 2009. Local textual functions of move in newspaper story patterns. In Römer & Schulze: 265–287.

Scott, Mike & Chris Tribble. 2006. Textual Patterns: Key words and corpus analysis in language education. Amsterdam: John Benjamins.

Semino, E. and Short, M. 2004. Corpus Stylistics. Speech, Writing and Thought Presentation in a Corpus of English Writing. London: Routledge.

Römer, Ute. 2010. Establishing the phraseological profile of a text type: The construction of meaning

in academic book reviews. English Text Construction 3(1): 95-119.