

# Corpus Approaches to the Language of Literature

---

Martin Wynne<sup>1</sup> and Ylva Berglund Prytz<sup>1</sup>

## Abstract

Work in stylistics relies on the evidence of the language of literature. Corpus linguistics is also an empirical approach to linguistic description, relying on the evidence of language usage as collected and analysed in corpora. As linguists and stylisticians have become more aware of the possibilities offered by corpus resources and techniques, then increasingly it is pointed out that the coming together of these fields could be fruitful. But there is as yet little actual research in the area. This colloquium wants to offer an opportunity to disseminate and discuss examples of successful research which has shed new light on literary texts through the techniques of corpus linguistics. Furthermore, it will point to ways forward in demonstrating the resources and techniques necessary for such work in the future.

The session will consist of an introductory presentation, providing an introduction to the topic, an outline of the current state of the relatively new area of corpus stylistics and a summary of previous, recent research and activities (15 min). It will be followed by three papers, each offering excellent examples of current trends (3\*20 min). The colloquium will be concluded by a general discussion, which is expected to take the form of a discussion of what is meant by corpus stylistics and where the field is moving now (30 min). (The time allocations allow time for switching between speakers, spontaneous questions and possible delays and interruptions. If not used, the discussion session will be extended).

## PAPER ABSTRACTS

### *Language Norms, Creativity and Local Textual Functions*

Michaela Mahlberg<sup>2</sup>

This paper will look at examples of how corpus linguistic methodology can aid the literary analysis of texts with the help of collocations, clusters and keywords, and the paper will also discuss theoretical questions about the relationship between literary and linguistic categories of description. The concept of local textual functions (cf. Mahlberg 2005, and forthcoming) will be suggested as a tool to capture internal norms of texts that play a part in the description of characters and the creation of point of view. Local textual functions characterise patterns of lexical items in a particular text or in texts that share functional features. The examples that are discussed in the paper will mainly be derived from a corpus of 19<sup>th</sup> century fiction.

---

<sup>1</sup> Convenors' contact details:

Martin Wynne, Oxford Text Archive, Oxford University

*e-mail:* martin.wynne@oucs.ox.ac.uk

Ylva Berglund Prytz, Research Technologies Service, Oxford University

*e-mail:* ylva.berglund@oucs.ox.ac.uk

<sup>2</sup> University of Liverpool, *e-mail:* M.Mahlberg@liverpool.ac.uk

*Speech and Thought Presentation in a Corpus of Nineteenth-Century  
Autobiographies and Fictional Texts*

Beatrix Busse<sup>3</sup>

This paper will present quantitative and qualitative information about the distribution and (communicative) functions of various types of speech and thought presentation in a corpus of nineteenth-century narrative fiction and autobiography. It is concerned to a large extent with the representation of consciousness on various levels – authorial, narratological, psychological, cognitive, and more.

Another aim of this paper is to offer a selective diachronic comparison: the results gained from the analysis described above and those based on twentieth-century narratives (as gained by Semino and Short [2004]) will be assessed in terms of corpus design, annotation, similarities and changes. This will especially be pursued because existing studies on the topic are often synchronically oriented and do not draw on larger corpora.

*Lexical Patterns as a Means of Text Segmentation*

Bettina Fischer-Starcke<sup>4</sup>

Segmenting a text into its constituent parts is useful for example for summarizing a text or to better understand text progression. The latter is of particular importance for the interpretation of literary texts, where plot development may be crucial for their interpretation. Based on the concept of cohesion, the occurrence of lexis from one semantic field creates coherence and textual unity. We can therefore assume that parts of text are characterized by the occurrence of similar lexis as in words belonging to the same semantic field.

In this paper, I demonstrate two complementary analyses to identify the different parts of a text based on its lexis. The textual basis for this is Austen's novel *Northanger Abbey*. I compare my findings to those of literary critics and differences between the linguistic and the literary segmentations are discussed.

**References** (joint)

- Adolphs, S. and R. Carter. 2002. "Point of view and Semantic Prosodies in Virginia Woolf's *To the Lighthouse*", *Poetica*, 7–20.
- Carter, R. 2004. *Language and creativity. The art of common talk*. London: Routledge.
- Fludernik, M. (1993). *The Fictions of Language and the Languages of Fictions*. London and New York: Routledge.
- Leech, G.N. and M. Short (1981). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. London and New York: Longman.
- Mahlberg, M. 2005. *English General Nouns. A Corpus Theoretical Approach*. Amsterdam: Benjamins.

---

<sup>3</sup> Westfälische Wilhelms-Universität Münster, e-mail: busseb@uni-muenster.de

<sup>4</sup> University of Trier, e-mail: starcke@uni-trier.de

- Mahlberg, M. (forthcoming). "Corpus stylistics: bridging the gap between linguistic and literary studies". In: M. Hoey, M. Mahlberg, M. Stubbs, W. Teubert. 2007. *Text, Discourse and Corpora*. London: Continuum.
- Scott, M. (1999). *WordSmith Tools*. 3.0. OUP. [Computer Software]
- Semino, E. and M. Short. 2004. *Corpus Stylistics. Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- Sinclair, J. 2004. *Trust the Text. Language, corpus and discourse*. London: Routledge.
- Stubbs, M. 2005. "Conrad in the computer: examples of quantitative stylistic methods". *Language and Literature*, 14 (1): 5–24.
- Trudgill, P. and R. Watts (2002). "Introduction. In the year 2525." *Alternative Histories of English*. Ed. Peter Trudgill and Richard Watts. London and New York: Routledge, pp. 1–3.
- Youmans, G. (2001). *Vocabulary Management Profiles*. [Computer Software]  
<http://web.missouri.edu/~youmansc/vmp/index.shtml> 1 December 2006
- Youmans, G. (1991). A new tool for discourse analysis: the Vocabulary-Management-Profile. *Language*. 67(4). 763–89.  
<http://www.missouri.edu/~youmansc/vmp/help/Youmans-NewTool.pdf> 1 December 2006