PALA 2014 Summer School: *Corpus and Data-driven Stylistics*

The PALA 2014 Summer School entitled **Corpus and Data-driven Stylistics** will take place prior to the PALA 2014 annual conference at the University of Maribor in Slovenia. It will start on Monday morning, 14 July, and end on Wednesday, 16 July, around noon. (A detailed time schedule will be announced in the weeks before the conference.)

Total number of participants: 20 (registration on a first-come-first-served basis)
Summer School registration fee: 100€

The program will comprise a series of lectures and workshops divided into the following two successive sections:

Section 1: Corpus Stylistics (Dan McIntyre and Brian Walker)

1)   Lecture #1: Introduction to corpus linguistics
In this lecture we will introduce some of the important concepts, methodologies and terminology used in corpus linguistics. This lecture provides a foundation for the lecture and workshops that follow.

2)   Lecture #2: Combining Corpus linguistics and stylistics
The combination of corpus linguistics and stylistics is often referred to as *corpus stylistics*. In this lecture we will discuss how methods, approaches and theories from corpus linguistics can be combined with stylistics in order to enrich analyses of texts, as well as to perform analyses that would be extremely difficult (or near impossible) if carried out manually. We will discuss the advantages and issues associated with corpus stylistics with reference to examples from existing research.
We will also use this lecture to introduce *WMatrix* (Rayson 2009), which is a web interface to the USAS (UCREL semantic analysis system), and CLAWS (Constituent Likelihood Automatic Word-tagging System) corpus annotation tools. Processing a text using *WMatrix* produces frequency lists at the word level, grammatical level, and semantic level. These frequency lists can be compared to frequency lists generated by *WMatrix* for other texts in order to generate key words, key word-classes, and key semantic-categories.
We will be showing participants how to use this corpus tool in the workshops that follow.

3)   Workshop#1: Introducing the basic functionality of *WMatrix*
This workshop will give participants the opportunity to become familiar with *WMatrix* and how it can be used. We will start by briefly talking about how *WMatrix* works, after which participants will have the opportunity to familiarise themselves generally with the *WMatrix* site. The main part of the workshop will give practice in: (i) uploading documents to *WMatrix*; (ii) using *WMatrix* to compare texts with a large corpus (the BNC sampler); and (iii) using *WMatrix* to compare a particular text with another text. We will use these comparisons to look at word lists and semantic groupings. Throughout the workshop there will be opportunities to discuss the findings that participants have made.

4)   Workshop#2: *WMatrix* – advanced features
This workshop will give participants the opportunity to become familiar with some of the more advanced features of *WMatrix*. In particular we will be demonstrating in more detail the stages involved in preparing texts for upload into the *WMatrix* environment. We will also give participants the opportunity to analyse a novel and a film script using *WMatrix* in order to explore how this tool can be used in the stylistic analysis of a larger texts.

5)   Lecture #3: *WMatrix* in practice – two case studies showing how WMatrix can be used for stylistic analysis.
In this lecture, we will focus on the semantic output produced by *WMatrix* and show how it can be used as part of an analysis of the main characters in the novel, and of the characters in a film script. An important point which will emerge is that this corpus tool, like other such tools, does not give us automatic answers, but points out potentially interesting parts of the text that might be examined in more detail.

6)   Workshop #3: *AntConc* – introduction
This workshop will introduce a useful corpus tool called *AntConc*. This is freeware that can be easily downloaded from the web. We will demonstrate the main features (wordlists, collocates, keywords) of this handy and extremely portable software.
Because *AntConc* is free and portable, participants can take this software away with them and use it immediately (if they want to) for their own research.

7)   Workshop #4: *AntConc* – intermediate
In this workshop we use *AntConc* to investigate a small corpus of email messages containing approximately 12000 words. The investigation will involve comparing the email corpus with a corpus of contemporary written British English called HUBE12. Delegates will use *AntConc* to investigate whether there are any linguistic features peculiar to emails (as represented by the corpus) when compared with written British English (as represented by HUBE12). There will be time at the end of the workshop for delegates to discuss their findings, and any issues with this investigation.

8)   Lecture #4: AntConc in practice – case studies showing how AntConc can be used to analyse newspaper report and political texts.
In this lecture we demonstrate how AntConc can be practically applied to the analysis of corpora by giving an overview of three recent research projects carried out at the University of Huddersfield. We focus on the methodological issues associated with moving from quantitative to qualitative analysis.

Biographical notes.
Dan McIntyre is Professor of English Language and Linguistics in the Department of Linguistics and Modern Languages at the University of Huddersfield. He began his career teaching English as a foreign language in Italy and has a long-standing interest in pedagogical stylistics. Among his publications on this topic is the edited collection *Teaching Stylistics* (Palgrave, 2011). His other major publications are *Stylistics* (Cambridge University Press, 2010; with Lesley Jeffries), *Language and Style* (Palgrave, 2010; co-edited with Beatrix Busse), *History of English* (Routledge, 2009) and *Point of View in Plays* (John Benjamins, 2006). Dan is series editor of *Advances in Stylistics* (Bloomsbury) and *Perspectives on the English Language* (Palgrave). He is also co-editor of *Babel: The Language Magazine* (www.babelzine.com).

Brian Walker is Senior Lecturer in Corpus Stylistics in the Department of Linguistics and Modern Languages at the University of Huddersfield. His PhD, which he received from Lancaster University (UK), combined corpus linguistics and stylistics to investigate the construction of characters in Julian Barnes' novel *Talking It Over*. This research explored the potential connection between statistical salience and interpretative relevance, and the combination of corpus-led quantitative analysis and qualitative analysis. His published research since then includes an investigation speech, writing and thought presentation in Early Modern English (1500-1700) texts, and a corpus-driven analysis of the cultural keywords in the English press during the period when Tony Blair was prime minister. His research interests are rooted in Stylistics and Corpus Linguistics.

Linguistics and Modern Languages
School of Music, Humanities and Media
University of Huddersfield
Queensgate
Huddersfield
HD1 3DH
UK
Tel: +44 (0)1484 471365
Email: d.mcintyre@hud.ac.uk and b.d.walker@hud.ac.uk

Section 2: How to Handle Quantitative Data? (Willie van Peer)

Good research requires either genius or data. Because most of us aren't geniuses, we must rely on data. Good data. Without data, no real research. First, one has to collect data, of course. This workshop works from the premise that you already have done so. But now the data are there: what to do? That is the focus of the present workshop.

In the morning session we will confront basic concepts in analyzing data, which we will apply in the afternoon session. It will be a hands-on approach, without much jargon, aimed at allowing you to do such analyses yourself. Therefore, we will work at the computer with the program SPSS, which you will be using yourself in simulations, so that we can all go home afterwards and run analyses of your data.

1. Lecture - *Descriptive Data* (Averages, Standard Deviation, Normal Distribution, Graphs)

2. Lecture - *Inferences from Data* (Probability, errors, significance value, decision matrix for tests)

3. Workshop - How to obtain descriptive tests and graphs, how to incorporate graphs into your report, how to investigate the nature of your data.

4. Workshop - How to run inference tests: T-test, ANOVA, GLM, non-parametric tests, how to draw conclusions from the tests for your research.

Biographical notes

Willie van Peer (Ph.D. Lancaster University, 1980) is Professor of Literary Studies and Intercultural Hermeneutics at the University of Munich, former President of IGEL (International Society for the Empirical Study of Literature), former Chair of PALA, and until recently the General Editor of the journal *Scientific Study of Literature* (see: http://benjamins.com/#catalog/journals/ssol/main). He has been Visiting Scholar at Stanford and Princeton University, at the University of Alberta, and in the Program of Cognitive Psychology at the University of Memphis. His publications include *Stylistics and Psychology: Investigations of Foregrounding* (Croom Helm, 1986), *The Taming of the Text* (Routledge, 1988), *New Perspectives on Narrative Perspective* (SUNY Press, 2001; co-edited with Seymour Chatman), *Thematics. Interdisciplinary Studies* (Benjamins, 2002; co-edited with Max Louwerse), *The Quality of Literature* (Benjamins, 2008), *New Beginnings in the Study of Literature* (Cambridge Scholars Publications, 2008; co-edited with Jan Auracher) and *Scientific Methods for the Humanities* (Benjamins, 2012; co-authored with Frank Hakemulder and Sonia Zyngier). He has co-founded the Project Mali-ka-di (see www.mali-ka-di-org) for development projects in Mali.

Prof. Willie van Peer, Ph.D.
Professor of Intercultural Hermeneutics
Ludwig Maximilian University, Munich
Email: w.vanpeer@gmail.com