

===

Sunday 16 September, 10:00-13:00

Workshop 1 (PC room 1: 40 people)

Stefan Evert (web: [www.stefan-evert.de](http://www.stefan-evert.de))

FAU Erlangen-Nürnberg

### **Multivariate Analysis (with R)**

Multivariate analysis exploits similarities and correlations between large sets of quantitative features in order to identify underlying categories or dimensions of linguistic variation. In addition to Doug Biber's seminal work on the dimensions of register variation, multivariate techniques have been applied in literary stylometry, dialectology, sociology, historical linguistics, and many other fields of research.

This workshop introduces the basic mathematical concepts and classic techniques of multivariate analysis, such as clustering, principal component analysis (PCA), factor analysis (FA), correspondence analysis (CA) and linear discriminant analysis (LDA). I will show how to apply such techniques as well as visualize and interpret analysis results using the statistical software package R. Participants will carry out hands-on exercises based on examples from register variation, translation studies and authorship attribution.

===

Sunday 16 September, 10:00-13:00

Workshop 2 (PC room 2: 15 people)

Michael Barlow

University of Auckland

### **WordSkew: Text and Corpus**

In the workshop we will use WordSkew to investigate text structure within a corpus. We will examine how the clustering of words relates to positions in particular units in a text (e.g., sentences, paragraphs and sections). We can examine a variety of research questions: (i) how a particular word or phrase (or other abstract category such as part-of-speech class) is distributed with respect to different positions in the units of discourse and (ii) how the meaning or function of a word or phrase varies as we look at different parts of a discourse. These approaches may be of interest in both grammatical and literary investigations.

We will cover the basics of the operation of WordSkew and carry out some illustrative investigations.

===

Sunday 16 September, 14:00-17:00

Workshop 3 (PC room 1: 40 people)

Mike Scott

Lexical Analysis Software and Aston University

### **Exploring Features of WordSmith 7.0**

Outline

This workshop is intended for participants who have previously used at least one previous version of WordSmith Tools, and assumes familiarity with basic functions. Materials we will use will be in English and supplied at the workshop, but the principles we explore will apply to Japanese, Korean, Chinese etc too.

Participants will be using the latest version of WordSmith 7.0, hands-on. The workshop will be conducted using a mix of demonstration via a display projector, individual or pair hands-on experimentation and activity, general feedback and discussion.

Topics the workshop may cover include

- detailed consistency
- boilerplate text
- relevance checking
- colour categories and breakdowns
- TED talks and video concordancing (requires VLC Media player)
- collocation and collocate statistics
- Text Converter functions
- Corpus clean-up techniques

==

Sunday 16 September, 14:00-17:00

Workshop 4 (PC room 2: 15 people)

Stephen Jeaco

Xi'an Jiaotong-Liverpool University, Suzhou, China

### **The Prime Machine – a tool to help English language learners get started with the exploration of English language and literature.**

The Prime Machine was developed to make it easy for English language learners at my institution to notice and explore patterns of language use in corpora. Version 3 of The Prime Machine has very recently been made publicly available, with many of its features

and corpora open for the first time to users from outside my institution. Some of the principles guiding its development were:

- The design should help language learners explore differences between words and phrases;
- The design of the display for concordance lines should help language learners notice textual colligation, co-text and contexts;
- The design should help language learners notice features in the patterning of words and phrases. (Claims as presented in Jeaco, 2017, pp. 277, 282 & 284)

Through working with several cohorts of undergraduate English majors at a Sino-British institution in China, the software has been further developed and ways of presenting its features to students have been fine-tuned. This workshop will provide a platform on which to introduce the software to a wider audience and to see how it could perhaps form a useful stepping stone for English language students on their way to more expert and sophisticated corpus analysis.

Specifically, the participants will see and try out:

- Search support features
- Side by side comparisons
- Methods for sorting concordance lines
- Graphs and filtering for different priming features
- tPM's Research Tools

The workshop will be designed for anyone who is interested in introducing a learner-friendly corpus tool to English language learners with a view to helping them with different aspects of the writing process, or to help them explore features of a text or a small collection of texts. As well as covering the practicalities of hands-on software use, the workshop will discuss some teaching points and pedagogical aims to be considered when designing class sessions. The software hosted on a server in Suzhou, China and is available from [www.theprimemachine.net](http://www.theprimemachine.net).

## References

Jeaco, S. (2017). Concordancing Lexical Primings. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical Priming: Applications and Advances* (pp. 273-296). Amsterdam: John Benjamins.