

Computer-aided methods of textual analysis

JMCE / RECON Workshop
WP 6 – The Foreign and Security Dimension

Berlin, May 27-28 2010.

The joint workshop of the Jean Monnet Centre of Excellence (JMCE) and RECON is organised by the Otto-Suhr-Institute of Political Science, Free University Berlin, under work package 6 of the RECON project. WP 6 – The Foreign and Security Dimension – aims to assess the status as well as the prospect for democracy within the field of foreign and security policy in Europe.

In the digital era, large bodies of electronic data hold great promise for the exploration of researchers' questions. Social scientists have started to use computer-aided and linguistic methods for the investigation of social and political issues, drawing on large text corpora. In the process, they have invariably been confronted with a host of problems related to the compilation of an issue-specific corpus; the retrieval of content-related information and its frequency-based generalisation; the down-sizing of corpora; the detailed qualitative-linguistic analysis of smaller sub-corpora; and the triangulation of text analytical methods for the qualitative and quantitative exploration of texts. At the same time, computational and corpus linguists have developed and applied new tools for the computational retrieval, annotation, and representation of linguistic information in large text corpora. However, they are faced with the challenge of communicating these innovations to a broader scientific audience that has an increasing interest in these tools of analysis, but often operates with different (content-related) categories.

The objective of this workshop is to explore the common methodological interests of computational and corpus linguists working on computer-aided linguistic analysis, on the one hand, and social scientists, who have used computational and corpus-linguistic tools to advance text analysis in the social sciences, on the other. We hope to foster interdisciplinary cooperation focusing on method-related issues:

- Innovations in computational and corpus linguistics: How to retrieve, annotate, and represent linguistic information and how to bring these insights to fruition across disciplines. Typical wording as a key to meaning. Innovative approaches to tracing complex structures of meaning.
- Method-mix: How to combine multiple computer-aided text analytical methods on the threshold between quantitative / quantifying and qualitative analysis and between different text analytical approaches (content analysis, corpus analysis, discourse analysis).

The workshop is designed as an 'authors' workshop.' It provides a forum for the discussion of results gained in ongoing research projects with the aim of compiling these results in a journal special issue on interdisciplinary methods. The workshop is meant to strengthen existing research collaboration in this field and to create new avenues for trans-disciplinary research.

Programme

Thursday, 27 May 2010:

How to make use of computer- and corpus-linguistic methods for large-n text analysis in the social sciences

12:30-13:00 **Arrival, registration and coffee**

13:00-13:15 **Welcome and opening**

Thomas Risse

Free University Berlin

13:15-13:30 **Introduction: How to make use of computer- and corpus-linguistic methods for large-n text analysis in the social sciences**

Amelie Kutter and Cathleen Kantner

Free University Berlin

The preparation of large multi-lingual corpora

13:30-14:00 **Computing in the humanities by means of the eHumanities desktop**

Alexander Mehler and Rüdiger Gleim

University of Bielefeld and Goethe University Frankfurt

14:00-14:15 **Discussion**

14:15-14:45 **How to get rid of the noise in the corpus: cleansing large samples of digital newspaper texts**

Amelie Kutter and Cathleen Kantner

Free University Berlin

14:45-15:00 **Discussion**

15:00-15:30 *Coffee break*

Text analysis between human coding/annotation and automatisisation

15:30-16:00 **Toward identifying arguments automatically**

Manfred Stede

University of Postdam

16:00-16:15 **Discussion**

16:15-16:45 **Improving manual annotation with usable software tools**

Manuel Burghardt and Christian Wolff

University of Regensburg

16:45-17:00 **Discussion**

17:00-17:30 **Electoral campaigns, relation mining, and dependency parsing: Extracting semantic network data from Swiss newspaper articles**

Bruno Wüest, Simon Clematide and Daniel Laupper

University of Zürich

17:30-17:45 **Discussion**

17:45-18:00 **Final debate and conclusions**

19:30 **Dinner, Jules Verne Restaurant**

Schlüterstraße 61, S-Bhf Savignyplatz

Friday, 28 May 2010

How to analyse millions of texts and still be home for tea? Innovations in computer-aided linguistics and the social sciences

10:00-10:15 *Coffee*

Typical wording as a key to meaning

10:15-10:45 **Avoiding the hammer-and-nails problem. A theory driven approach to computer-aided political keywords analysis**

Michal Bukowski

Jagiellonian University Krakow

10:45-11:15 **Automatic meaning? Large text corpora as the source of social knowledge**

Jacek Kolodziej

Jagiellonian University Krakow

11:15-11:30 **Discussion**

11:30-12:00 **Corpus-based extraction of domain-specific terminology**

Fabienne Fritzing

University of Stuttgart

12:00-12:15 **Discussion**

12:15-14:00

Lunch: *Harnack-Haus*

Imnestraße 16-20, U-Bhf Thielplatz

Typical wording as a key to meaning - continued

14:00-14:30 **Semantic fields and conceptual categories: exploring the synergy of corpus and content analysis**

Cathleen Kantner and Amelie Kutter

Free University Berlin

14:30-14:45 **Discussion**

14:45-15:00 *Coffee*

Approaches to tracing complex structures of meaning

15:00-15:30 **A text-mining system for content analysis of large text corpora**

Peter Kolb

University of California, San Francisco

15:30-15:45 **Discussion**

15:45-16:15 **Speech act theory and grammar-based text analysis**

Brit Helle Aarskog

University of Bergen

16:15-16:30 **Discussion**

16:30-16:45 *Coffee*

Open discussion on problem-solutions and recommendations and concluding remarks

16:45-17:30 **Presenters and other experts:**

Brit Helle Aarskog, *University of Bergen*

Michal Bukowski, *Jagiellonian University Krakow*

Manuel Burghardt, *University of Regensburg*

Fabienne Fritzing, *University of Stuttgart*
Rüdiger Gleim, *Goethe University Frankfurt*
Cathleen Kantner, *Free University Berlin*
Peter Kolb, *University of California, San Francisco*
Jacek Kolodziej, *Jagiellonian University Krakow*
Amelie Kutter, *Free University Berlin*
Alexander Mehler, *University of Bielefeld*
Thomas Risse, *Free University Berlin*
Manfred Stede, *University of Postdam*
Christian Wolff, *University of Regensburg*

19:30

Dinner: *Anda Lucia, Restaurant & Tapasbar*
Savignyplatz 2, S-Bhf Savignyplatz
