Corpora & Language

Workshop   and   Conference
March 27-29, 2017   and   **March 29-31, 2017**

Building J, Paderborn University

The conference ''Corpora and Language'' / ''Korpora und Sprache'' will be hosted by the chair of English Linguistics at the Department of English and American Studies at Paderborn University, Germany on March 29-31, 2017. The conference is a platform for researchers from the field of corpus linguistics to discuss current topics and present their projects and latest results. The languages of the conference are English and German. Workshops on corpora and language will take place from March 27-29, 2017.

Confirmed Speakers

Dr. Sabine Bartsch (TU Darmstadt)
Prof. Dr. Klaus P. Schneider (Universität Bonn)
Prof. em. Dr. Dr. h.c. mult. Norbert Richard Wolf (Emeritus Professor, Universität Würzburg)

Organiser

Prof. Dr. Ilka Mindt (Paderborn University)

PADERBORN UNIVERSITY
The University for the Information Society

Workshops

Monday, 27.03.2017, 9am-4pm, Dr. Markus Freudinger

In this workshop, we will be dealing with some basic questions: what is a corpus? What is corpus linguistics? Which chances and limitations are there in corpus linguistic studies? What do we need to take into account for corpus-design? How does corpus analysis actually work?

We will be working with different English and German corpora (e.g. BNC, COCA, Cosmas…), corpus tools (e.g. Antconc) and general data processing software (e.g. Excel).

In total, this workshop aims at enabling students to successfully work on their PhD/Master's thesis and to provide them with the necessary set of tools. For this reason, there will be many hands-on activities with corpora and respective software.

Tuesday, 28.03.2017, 9am-4pm, Prof. Dr. Christian Langstrof

This workshop familiarises participants with a range of approaches to analysing speech using the freely available program PRAAT by Paul Boersma and David Weenink (www.praat.org). In terms of thematic depth, the workshop mainly aims at participants with little or no prior experience with PRAAT. Introductory knowledge of phonetics on the level of an introductory linguistics class is assumed, but no further experience with acoustic phonetics is required.

We will approach the topic by looking at a range of well-known linguistic variation phenomena including, *inter alia*, vowel length in Scottish English vis-a-vis RP and vowel quality differences between major varieties of English. The data we will be looking at comes from the International Dialects of English Archive (http://www.dialectsarchive.com/). The workshop focuses on four interwoven aspects in basic speech analysis: PRAAT – the basics, Spectrogram analysis, Annotation and Automatised measurements

Wednesday, 29.03.2017, 9am-1pm, Dr. Sabine Bartsch

Annotation and query are a mainstay of corpus linguistic research. Corpus studies can benefit greatly from high-quality annotation. It is a means of making linguistic expertise and intuitions explicit and thus accessible to the analyst and to further steps in the research process. Yet, in order to harness the linguistic knowledge provided by corpora and annotations, linguists must also be able to formulate or operationalise their linguistic questions and address them to the corpus.

This workshop offers students of linguistics and other philologies a hands-on tutorial in annotation and query techniques for the analysis of digital corpora. In the workshop, we are going explore techniques of automatic and semi-automatic annotation together with matching techniques for querying corpus data. Students are going to learn to employ the Stanford NLP Part of Speech Tagger as an example of an automatic annotation process they can use to annotate their own corpora. We are furthermore going to study the structure of linguistic data in plain text and annotated corpora and the resulting affordances for operationalisations of linguistic research questions in the form of queries. Query techniques to be explored in the workshop include the integrated search for linguistic forms and annotations based on the use of regular expressions as implemented in standard tools such as text editors (e.g. Notepad++) and the widely used concordancing software such as AntConc. Working examples in the workshop are from a corpus of political speeches.

Programme

Monday, 27.03.2017 (Pre-conference workshop)

Basic Concepts of Corpus Linguistics

9am - 12noon and 1pm - 4pm

Dr. Markus Freudinger

Tuesday, 28.03.2017 (Pre-conference workshop)

Analysing Speech with PRAAT

9am – 12noon and 2pm – 4pm

Prof. Dr. Christian Langstrof

Wednesday, 29.03.2017 (Pre-conference workshop)

Workshop on Annotation and Query in Corpus Linguistics (AnQCor)

9am – 1pm

Dr. Sabine Bartsch

7pm: Conference Warming

Thursday, 30.03.2017

9am: Opening of Conference

9am - 6pm: Presentations

Friday, 31.03.2017

9am – 12 noon: Presentations

12 noon: Closing of Conference

Registration for Workshops and Conference

Registration for Workshops:

The registration for one or more of the pre-conference workshops is mandatory and open until 20th March 2017. Please register by sending an e-mail to korpora2017@upb.de. The places in the workshops are limited, so we recommend registering early.

Registration for Conference:

The registration for attending the conference is open until 20th March 2017. Please register by sending an e-mail to korpora2017@upb.de.

For more information, visit our website
http://kw.uni-paderborn.de/en/anglistik-amerikanistik/korpora2017/workshop/

Interested students and staff are welcome!

Conference Venue

The conference will be held in building J at Paderborn University,
Warburger Str. 100, 33098 Paderborn, Germany.

For a detailed description, visit our website
http://kw.uni-paderborn.de/en/anglistik-amerikanistik/korpora2017/

Contact details

Prof. Dr. Ilka Mindt   Universität Paderborn
Englische Sprachwissenschaft   Warburger Straße 100
Institut für Anglistik und Amerikanistik   33098 Paderborn

korpora2017@upb.de