AG 11: Web Data as a Challenge for Theoretical Linguistics and Corpus Design
Koordination: Felix Bildhauer, Roland Schäfer

######## Zeitplan ###########

Wednesday, 5 March

14:00
Felix Bildhauer & Roland Schäfer (Freie Universität Berlin)
Web data as a challenge for theoretical linguistics and corpus design (Introduction)

14:30
Sonja Müller (Universität Bielefeld)
How webdata can challenge traditional generalizations: a case study of the order of modal particles in German

15:00
Susanne Flach (Freie Universität Berlin)
Solving the rare phenomenon problem? 'Quasi-serial' verb constructions in English

15:30
Dirk Goldhahn & Uwe Quasthoff (Universität Leipzig)
Using corpus-based statistics for linguistic typology

16:00 Coffe break

16:30
For a few points more: improving decision processes in web corpus construction
Adrien Barbaresi (ENS Lyon)

17:00
Lea Helmers (Freie Universität Berlin)
Named entity recognition on German web corpora

17:30
Vladimír Benko (Slovak Academy of Sciences)
Near-duplicate data in web corpora

18:00
Jack Grieve (Aston University), Asnaghi Costanza (Università Cattolica del Sacro Cuore) & Tom Ruette (Humboldt-Universität zu Berlin)
Googleology is good science

Thursday, 6 March

9:00
Stefan Evert (invited talk)
An NLP approach to the evaluation of web corpora

10:00
Ines Rehbein (Universität Potsdam)
Using Twitter for linguistic purposes – three case studies

10:30
Kazuya Abe (Atomi University)
Twitter corpus and collection of German phrases

11:00 Coffee break

11:30
Tom Ruette (Humboldt-Universität zu Berlin) & Jack Grieve (Aston University)
Cognitive sociolinguistics with Twitter: why do the Dutch swear with diseases?

12:00
Peter Grube (Martin-Luther-Universität Halle-Wittenberg)
A diachronic corpus of personal weblogs: possibilities and current constraints