

International workshop:

Building Corpora of Computer-Mediated Communication: Issues, Challenges, and Perspectives

Feb 13–15, 2013, TU Dortmund University, Erich-Brost-Haus



Organizers:

PD Dr. Michael Beißwenger – Prof. Dr. Angelika Storrer
TU Dortmund University, Faculty of Culture Studies
Department of German Language and Literature

Wednesday 13 February

19:00– Pre-workshop warm-up (*Tryp Hotel, Restaurant „Bodega del Sol“*)

Thursday 14 February

08:00 Registration

09:00 Opening / introduction

Part I: Overview of CMC Corpus Projects

(15 min. talk + 5 min. for questions/discussion)

09:10–10:30 **Computer-Mediated Communication in SoNaR: Design & Collection**
Nelleke Oostdijk & Eric Sanders (Radboud University Nijmegen)

CMC Data in Learning and Teaching (LETEC) Corpora

Thierry Chanier (Université Blaise Pascal, Clermont-Ferrand)

The Project DIDI (*Digital Natives and Digital Immigrants*): Writing on Social Network Sites – A Corpus-based Observation of the Current Language Use in South Tyrol, with Particular Consideration of the Writers' Age

Aivars Glaznieks (European Academy of Bozen)

Project *Cybercreole/RomWeb*: An Overview

Daniel Alcón Lopez & Theresa Heyd (Universität Freiburg)

10:30–11:00 Coffee break

11:00–12:20 **Web2Corpus_it: A Balanced Pilot Corpus of Conversational
Computer-Mediated Communication**

Isabella Chiari (Università La Sapienza di Roma)

Building a Reference Corpus of German Computer-Mediated Communication: Introducing the *DeRiK* project

Michael Beißwenger (TU Dortmund), Alexander Geyken, Maria Ermakova,
Lothar Lemnitzer (Berlin-Brandenburg Academy of Sciences), Angelika
Storrer (TU Dortmund)

Building and Annotating Corpora of Collaborative Authoring in *Wikipedia*

Johannes Daxenberger & Torsten Zesch (TU Darmstadt)

***Wikipedia* as a Linguistic Resource**

Eliza Margaretha (Institut für deutsche Sprache, Mannheim)

12:20–14:00 Lunch

Part II: Special Topics in Building CMC Corpora

(25 min. talk + 10 min. for questions/discussion)

- 14:00–15:45 **Technical Aspects in Harvesting Data from Social Network Sites**
Aivars Glaznieks & Egon Stemle (European Academy of Bozen)
- Anonymising CMC Corpora: A Reasonable Way to Share Corpora**
Christophe Reffay (UMR Sciences Techniques Education, ENS-Cachan)
- Challenges and Solutions in Automatically Annotating CMC Data**
Torsten Zesch (TU Darmstadt)
- 15:45–16:15 Coffee break
- 16:15–18:00 **A TEI Schema for the Annotation of CMC Genres**
Michael Beißwenger (TU Dortmund), Alexander Geyken, Maria Ermakova, Lothar Lemnitzer (Berlin-Brandenburg Academy of Sciences), Angelika Storrer (TU Dortmund)
- NCAT (Net Corpora Administration Tool): Building a System for Exploring Web Forum Data**
Daniel Alcón Lopez (Universität Freiburg)
- Web Forum Corpora on Globalized Varieties: Notes from the User Perspective**
Theresa Heyd (Universität Freiburg)
- 19:00– Dinner at *Tryp Hotel*

Friday 15 February

Part II, continued

- 09:00–10:45 **Spelling Variation in Social Media**
Nelleke Oostdijk & Eric Sanders (Radboud University Nijmegen)
- Experiments with Tokenization and Part-of-speech Tagging for German CMC Discourse**
Thomas Bartz, Michael Beißwenger & Angelika Storrer (TU Dortmund)
- Tackling Diachronic Network Representation of *Wikipedia* and Preprocessing**
Rüdiger Gleim & Alexander Mehler (Goethe-Universität Frankfurt)
- 10:45–11:15 Coffee break
- 11:15–13:00 **Discussion: issues of joint interest / perspectives for further cooperation**
- 13:00 End of workshop

Contact:

Michael Beißwenger, Technische Universität Dortmund, Institut für deutsche Sprache und Literatur, D-44221 Dortmund, michael.beisswenger@tu-dortmund.de