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)

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00–15:45</td>
<td><strong>Technical Aspects in Harvesting Data from Social Network Sites</strong>&lt;br&gt;Aivars Glaznieks &amp; Egon Stemle (European Academy of Bozen)&lt;br&gt;<strong>Spelling Variation in Social Media</strong>&lt;br&gt;Nelleke Oostdijk &amp; Eric Sanders (Radboud University Nijmegen)&lt;br&gt;<strong>Challenges and Solutions in Automatically Annotating CMC Data</strong>&lt;br&gt;Torsten Zesch (TU Darmstadt)</td>
</tr>
<tr>
<td>15:45–16:15</td>
<td>Coffee break</td>
</tr>
<tr>
<td>16:15–18:00</td>
<td><strong>Anonymising CMC Corpora: A Reasonable Way to Share Corpora</strong>&lt;br&gt;Christophe Reffay (UMR Sciences Techniques Education, ENS-Cachan)&lt;br&gt;<strong>A TEI Schema for the Annotation of CMC Genres</strong>&lt;br&gt;Michael Beißwenger (TU Dortmund), Alexander Geyken, Maria Ermakova, Lothar Lemnitzer (Berlin-Brandenburg Academy of Sciences), Angelika Storrer (TU Dortmund)&lt;br&gt;<strong>Tackling Diachronic Network Representation of Wikipedia and Preprocessing</strong>&lt;br&gt;Rüdiger Gleim &amp; Alexander Mehler (Goethe-Universität Frankfurt)</td>
</tr>
<tr>
<td>19:00–</td>
<td>Dinner at Tryp Hotel</td>
</tr>
</tbody>
</table>

**Friday 15 February**

## Part II, continued

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00–10:45</td>
<td><strong>NCAT (Net Corpora Administration Tool):</strong>&lt;br&gt;Building a System for Exploring Web Forum Data&lt;br&gt;Daniel Alcón Lopez (Universität Freiburg)&lt;br&gt;<strong>Web Forum Corpora on Globalized Varieties:</strong>&lt;br&gt;Notes from the User Perspective&lt;br&gt;Theresa Heyd (Universität Freiburg)&lt;br&gt;<strong>Experiments with Tokenization and Part-of-speech Tagging for German CMC Discourse</strong>&lt;br&gt;Thomas Bartz &amp; Michael Beißwenger (TU Dortmund)</td>
</tr>
<tr>
<td>10:45–11:15</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:15–13:00</td>
<td><strong>Discussion: issues of joint interest / perspectives for further cooperation</strong></td>
</tr>
<tr>
<td>13:00</td>
<td>End of workshop</td>
</tr>
</tbody>
</table>

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

Financial support for the workshop is granted as part of an individual funding by the Mercator Research Center Ruhr (http://www.mercur-research.de).