**Motivation**

Twitter as a medium
- microblogs
- written, but pseudo-realtime
- many unknown participants: lack of context and common ground
- mix of formality levels and styles

**Question:** Where do Twitter conversations fall on the spoken/written, informal/formal continuum?

**Tatjana Scheffler, Rike Schlüter, Manfred Stede**

Universität Potsdam, Germany
tatjana.scheffler@uni-potsdam.de
@tschfflr

**Connectives – Twitter vs. Newspaper Text**

- 100 conversations, 451 tweets
- 207 connective instances, in 165 tweets (37% of tweets)

<table>
<thead>
<tr>
<th>Connective</th>
<th>(Twitter)</th>
<th>Twitter (total)</th>
<th>Twitter (relative)</th>
<th>PCC (newspaper)</th>
<th>PCC (relative)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>und</strong> (‘and’)</td>
<td>43</td>
<td>20.8%</td>
<td><strong>und</strong> (‘and’)</td>
<td>21.1%</td>
<td></td>
</tr>
<tr>
<td><strong>aber</strong> (‘but’)</td>
<td>39</td>
<td>18.8%</td>
<td><strong>aber</strong> (‘but’)</td>
<td>8.2%</td>
<td></td>
</tr>
<tr>
<td><strong>wenn</strong> (‘if/when’)</td>
<td>37</td>
<td>17.9%</td>
<td><strong>wenn</strong> (‘if/when’)</td>
<td>8.2%</td>
<td></td>
</tr>
<tr>
<td><strong>dann</strong> (‘then’)</td>
<td>29</td>
<td>14.0%</td>
<td><strong>doch</strong> (‘but’)</td>
<td>7.7%</td>
<td></td>
</tr>
<tr>
<td><strong>also</strong> (‘so’)</td>
<td>8</td>
<td>3.9%</td>
<td><strong>denn</strong> (‘because’)</td>
<td>4.3%</td>
<td></td>
</tr>
<tr>
<td><strong>oder</strong> (‘or’)</td>
<td>8</td>
<td>3.9%</td>
<td><strong>dann</strong> (‘then’)</td>
<td>3.6%</td>
<td></td>
</tr>
</tbody>
</table>

**Causal Connectives – Types**

- causal connectives are frequent in Twitter:
  - ~1.7% of tweets / 2.6% of replies
  - ‘spoken’/informal style of justification

**Causative Connectives – Semantic Levels**

- analysis of 200 instances of *denn* (‘because’) in German Twitter conversations
- 47/200 instances: external argument (result) of *weil* is expressed in another speaker’s tweet

**Twitter Conversations**

- Up to 40% of German tweets are part of conversations.
- Answer-relation on Twitter creates discourse trees.

**Question:** How is coherence achieved in social media conversations?

**Other Mode-Specific Phenomena**

- hashtags
  - reference resolution, search, topic (#panamaleaks, #Germany)
  - evaluation (#fail, #sarcasm, #yay)
  - contributing content (#onemoresleep, #ihateitwhen)
  - memes (#bestdayofmylife)
  - URLs
    - carry communicative content:
      - Inform, Answer, Opening, Question, Suggestion, …
    - discourse relation to the tweet/conversation is rarely indicated (~20% of cases):
      - Elaboration, Exemplification, Evaluation, …
    - nucleus or satellite of the rhetorical relation

**Causal Connectives – Twitter vs. Newspaper Text**

<table>
<thead>
<tr>
<th>Connective (Twitter)</th>
<th>Twitter (total)</th>
<th>Twitter (relative)</th>
<th>PCC (newspaper)</th>
<th>PCC (relative)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>und</strong> (‘and’)</td>
<td>43</td>
<td>20.8%</td>
<td><strong>und</strong> (‘and’)</td>
<td>21.1%</td>
</tr>
<tr>
<td><strong>aber</strong> (‘but’)</td>
<td>39</td>
<td>18.8%</td>
<td><strong>aber</strong> (‘but’)</td>
<td>8.2%</td>
</tr>
<tr>
<td><strong>wenn</strong> (‘if/when’)</td>
<td>37</td>
<td>17.9%</td>
<td><strong>wenn</strong> (‘if/when’)</td>
<td>8.2%</td>
</tr>
<tr>
<td><strong>dann</strong> (‘then’)</td>
<td>29</td>
<td>14.0%</td>
<td><strong>doch</strong> (‘but’)</td>
<td>7.7%</td>
</tr>
<tr>
<td><strong>also</strong> (‘so’)</td>
<td>8</td>
<td>3.9%</td>
<td><strong>denn</strong> (‘because’)</td>
<td>4.3%</td>
</tr>
<tr>
<td><strong>oder</strong> (‘or’)</td>
<td>8</td>
<td>3.9%</td>
<td><strong>dann</strong> (‘then’)</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

**Relative frequencies of connectives ‘denn’, ‘da’, and ‘nämlich’ compared with ‘weil’ (all, ‘because’) in corpora of spoken and written German, and in Twitter.**


For Twitter and FOLK, the frequencies of causal ‘denn’ and ‘da’ were estimated by manually disambiguating a representative sample of the data. 0 values = no data