

Discourse Structuring Devices on Twitter

Tatjana Scheffler, Rike Schlüter, Manfred Stede
Universität Potsdam, Germany
tatjana.scheffler@uni-potsdam.de
@tschfflr



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?

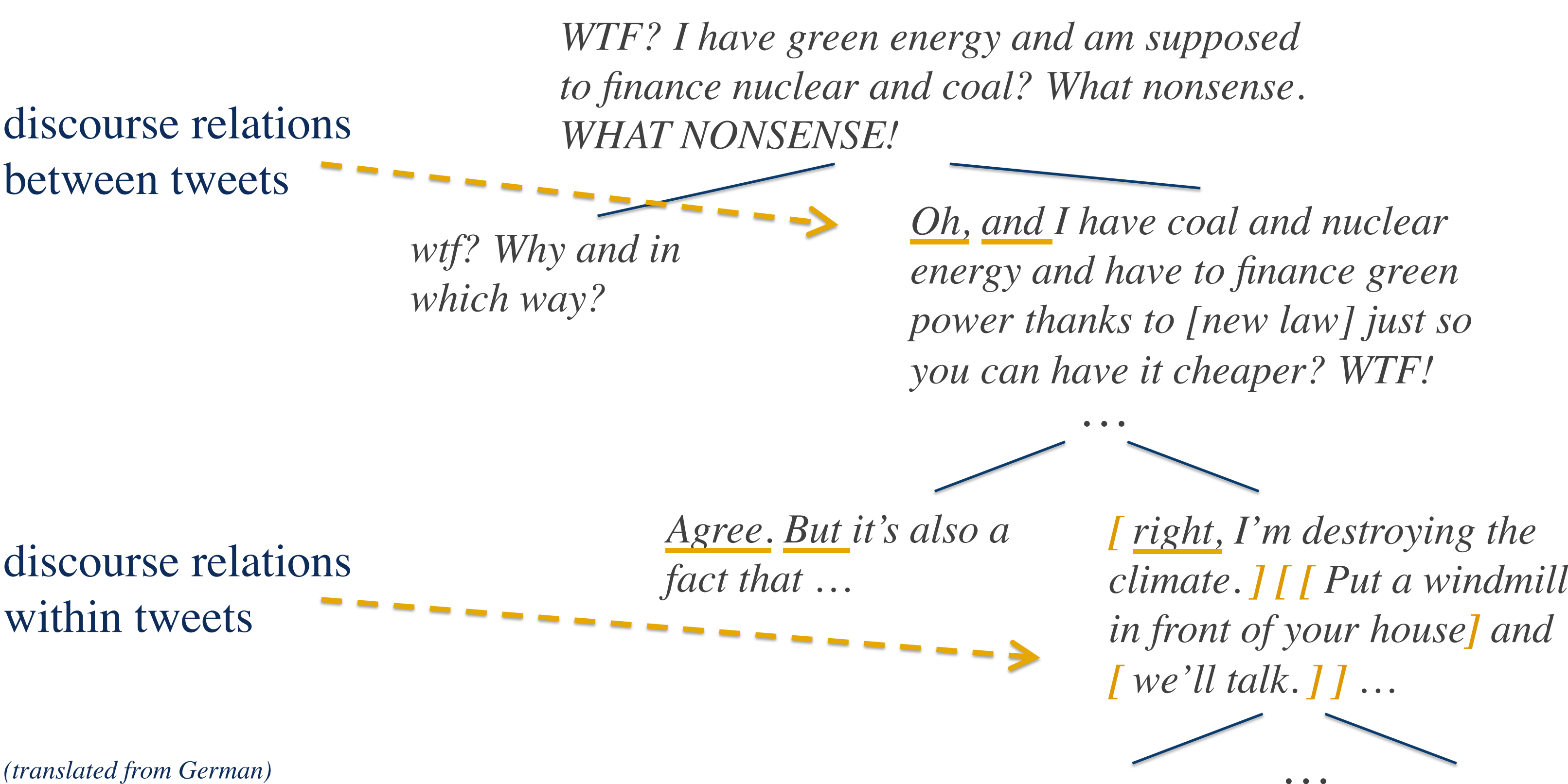
discourse structuring devices

- multi-participant dialog structure
- connectives
- rhetorical structure
- other coherence devices

Question: How is coherence achieved in social media conversations?

Twitter Conversations

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



Connectives – Twitter vs. Newspaper Text

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

Connective (Twitter)	Twitter (total)	Twitter (relative)	PCC (newspaper)	PCC (relative)
und ('and')	43	20.8%	und ('and')	21.1%
aber ('but')	39	18.8%	aber ('but')	8.2%
wenn ('if/when')	37	17.9%	wenn ('if/when')	8.2%
dann ('then')	29	14.0%	doch ('but')	7.7%
also ('so')	8	3.9%	denn ('because')	4.3%
oder ('or')	8	3.9%	dann ('then')	3.6%

Other Mode-Specific Phenomena

hashtags

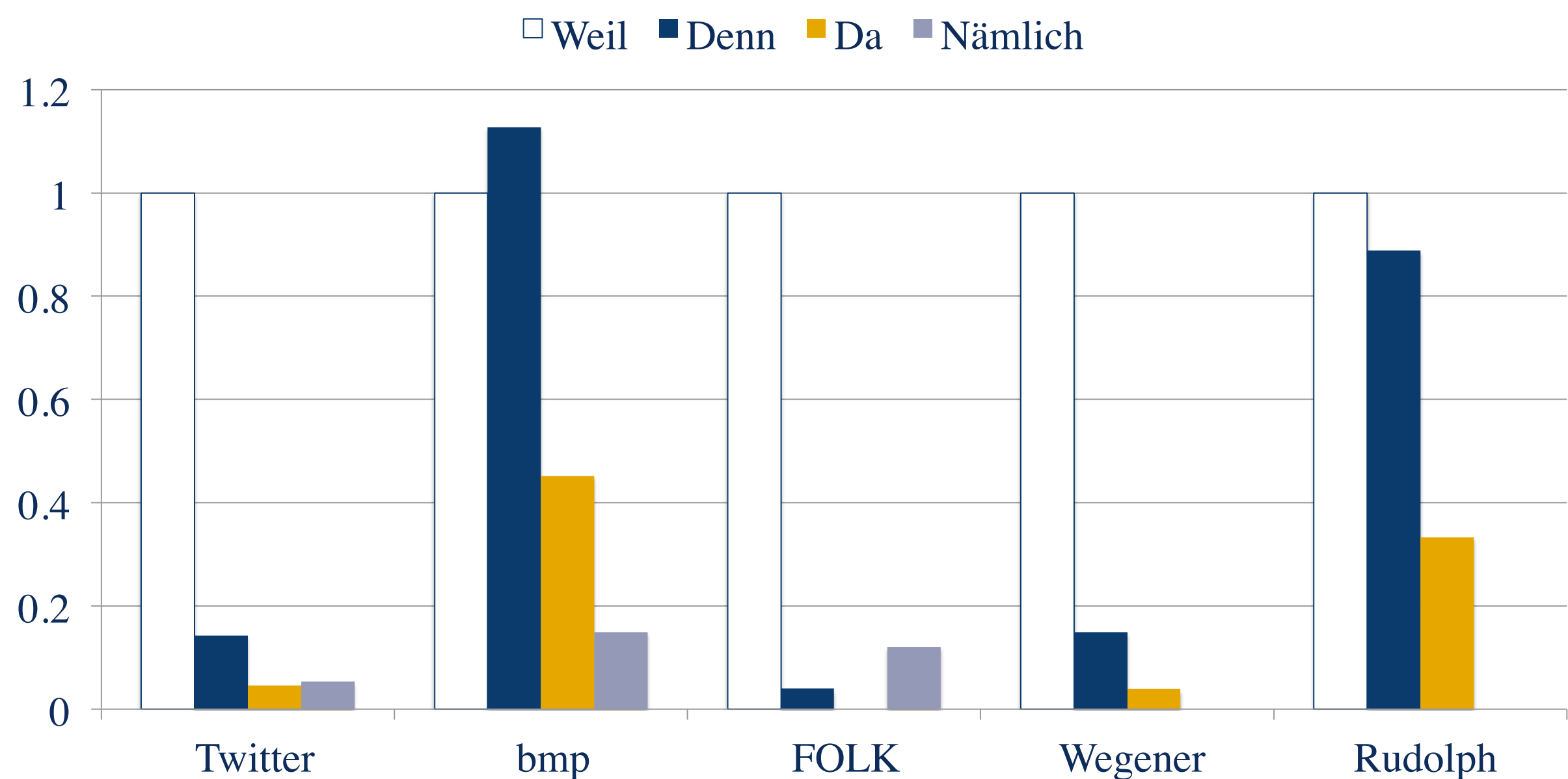
- reference resolution, search, topic (#panamaleaks, #Germany)
- evaluation (#fail, #sarcasm, #yay)
- contributing content (#onemoresleep, #ihateitwhen)
- memes (#bestdayofmylife)

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 – Types

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



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

Twitter = Wulff-corpus; 253,172 German tweets about the Wulff-scandal // **bmp** = Berliner Morgenpost/COSMAS II (daily newspaper) // **FOLK** = Forschungs- und Lehrkorpus Gesprochenes Deutsch; dialogs // **Wegener** = spoken corpora 1980-1999 from (Wegener 1999, Tab. 1) // **Rudolph** = written texts (Rudolph 1982) cit. in (Wegener 1999)

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

Causal Connectives – Semantic Levels

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

classification with Sweetser’s (1990) semantic levels

- most uses are on the propositional level
- in cross-tweet cases, slightly more epistemic uses (19%), because speakers are often asked for justifications:

A: *Why do you think that?*

B: *Because the FDP [pol. party] was economically liberal in many ways.*

s	tweets (<i>weil</i> only)		spoken (<i>weil, da, denn, nämlich</i>)
	total	percent	
propositional level	159	79.5%	34.7%
epistemic level	28	14.0%	34.9%
speech act level	13	6.5%	28.0%