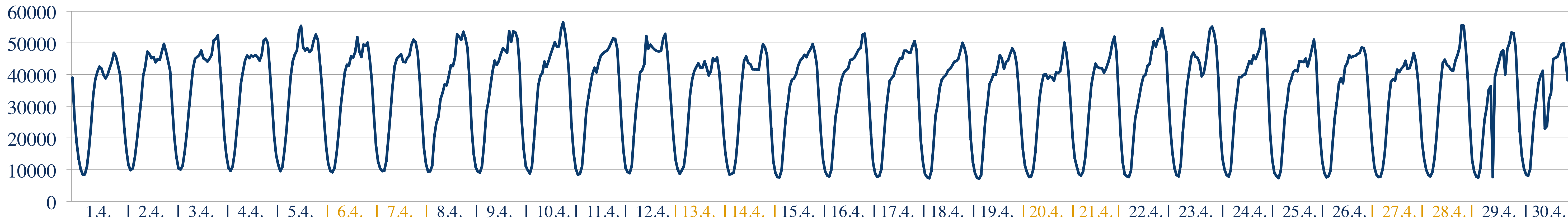


Conversations on German Twitter

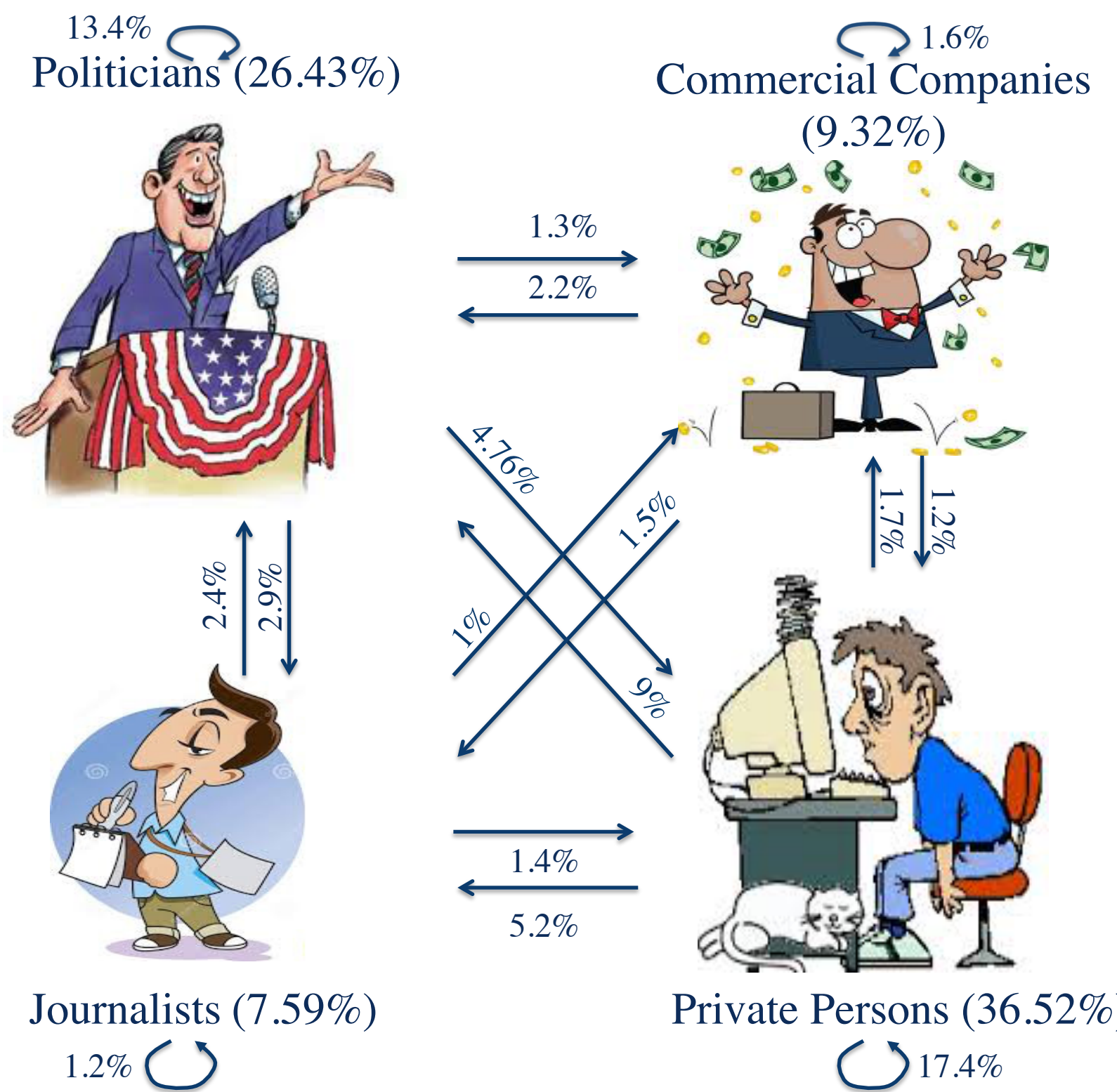
Tatjana Scheffler, Wladimir Sidorenko, Manfred Stede
Universität Potsdam
tatjana.scheffler@uni-potsdam.de



German tweets – April 2013



Communication of Social Groups



Corpus:

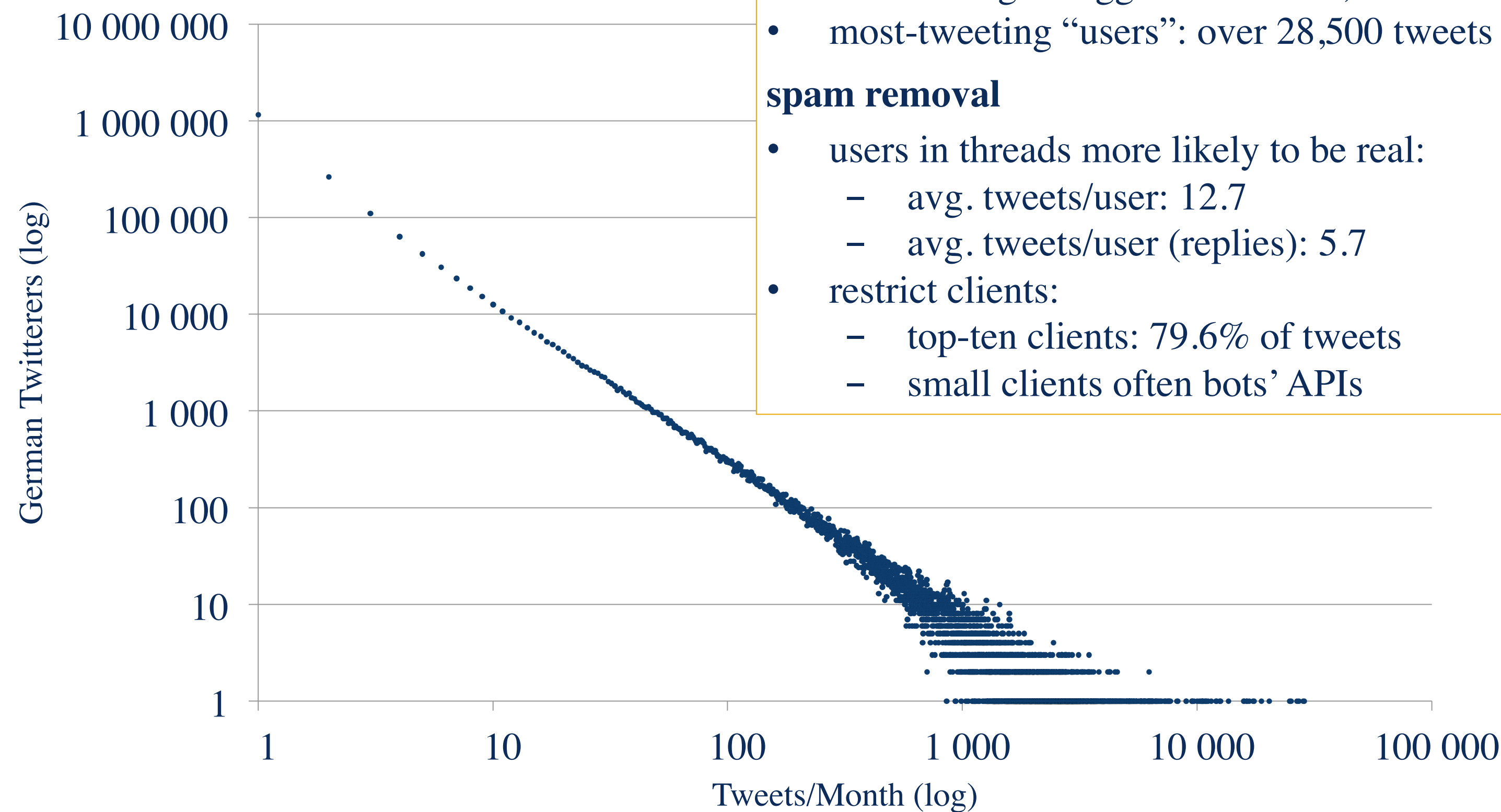
- 56,649 tweets (3,101 discussions) about energy turnaround in Germany (August–November 2013);
- 2,655 tweets (729 discussions) manually annotated with sentiments, social group of authors and addressees;

Conversation Statistics:

Author	Person	Polit.	Journ.	.com	.org
Person	270	140	81	27	50
Politician	74	208	45	20	64
Journalist	22	38	19	16	23
.com	18	34	23	25	45
.org	31	92	52	41	97

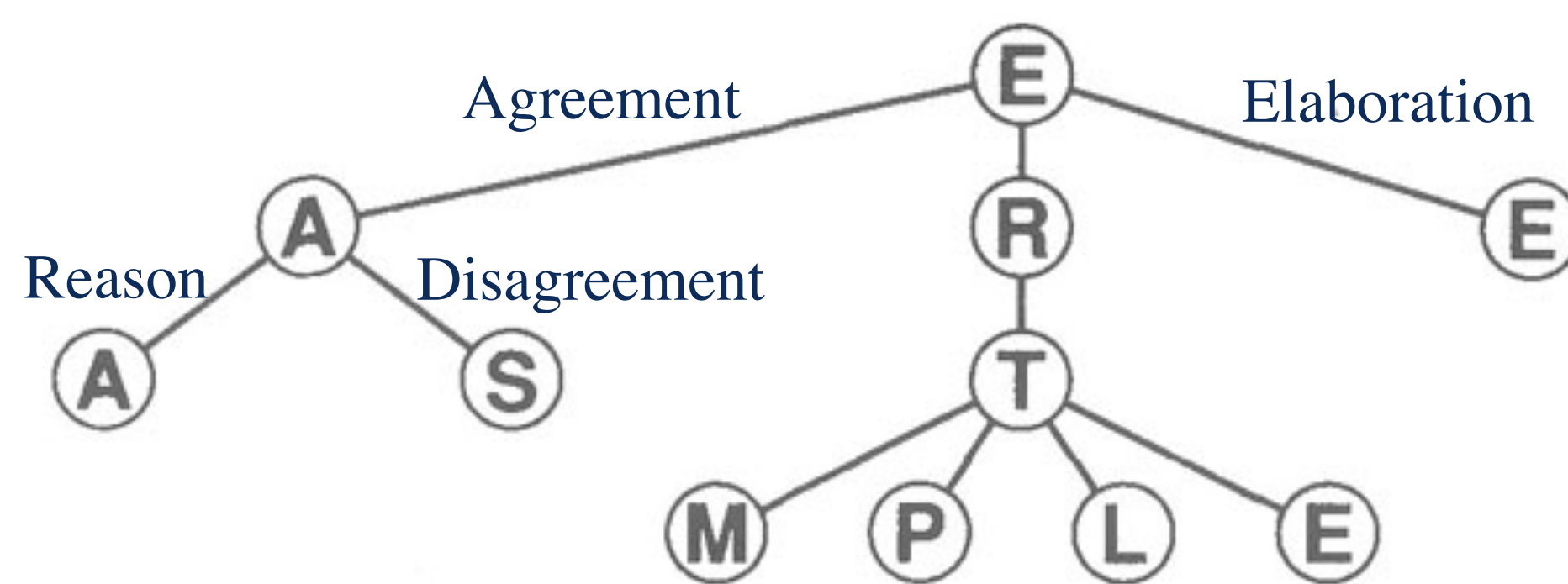
(This data was contributed by our project partners at LMU München: Prof. Dr. C. Neuberger / Dr. I. Engelmann.)

Monthly Users

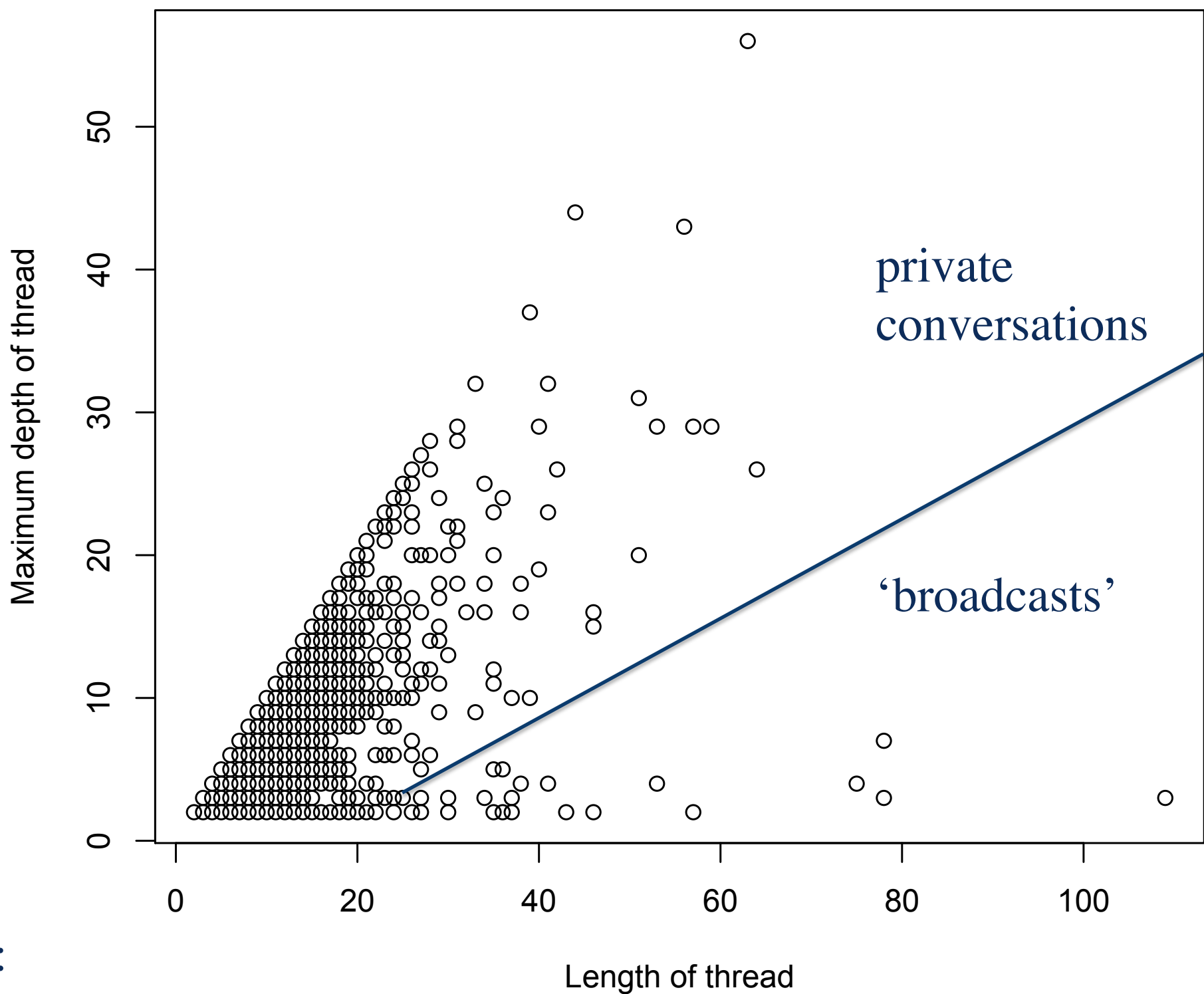


Twitter Threads

- over 30% of tweets are part of a conversation
- in_reply_to_id creates discussion trees:



Depth vs. length of all threads on April 1, 2013:



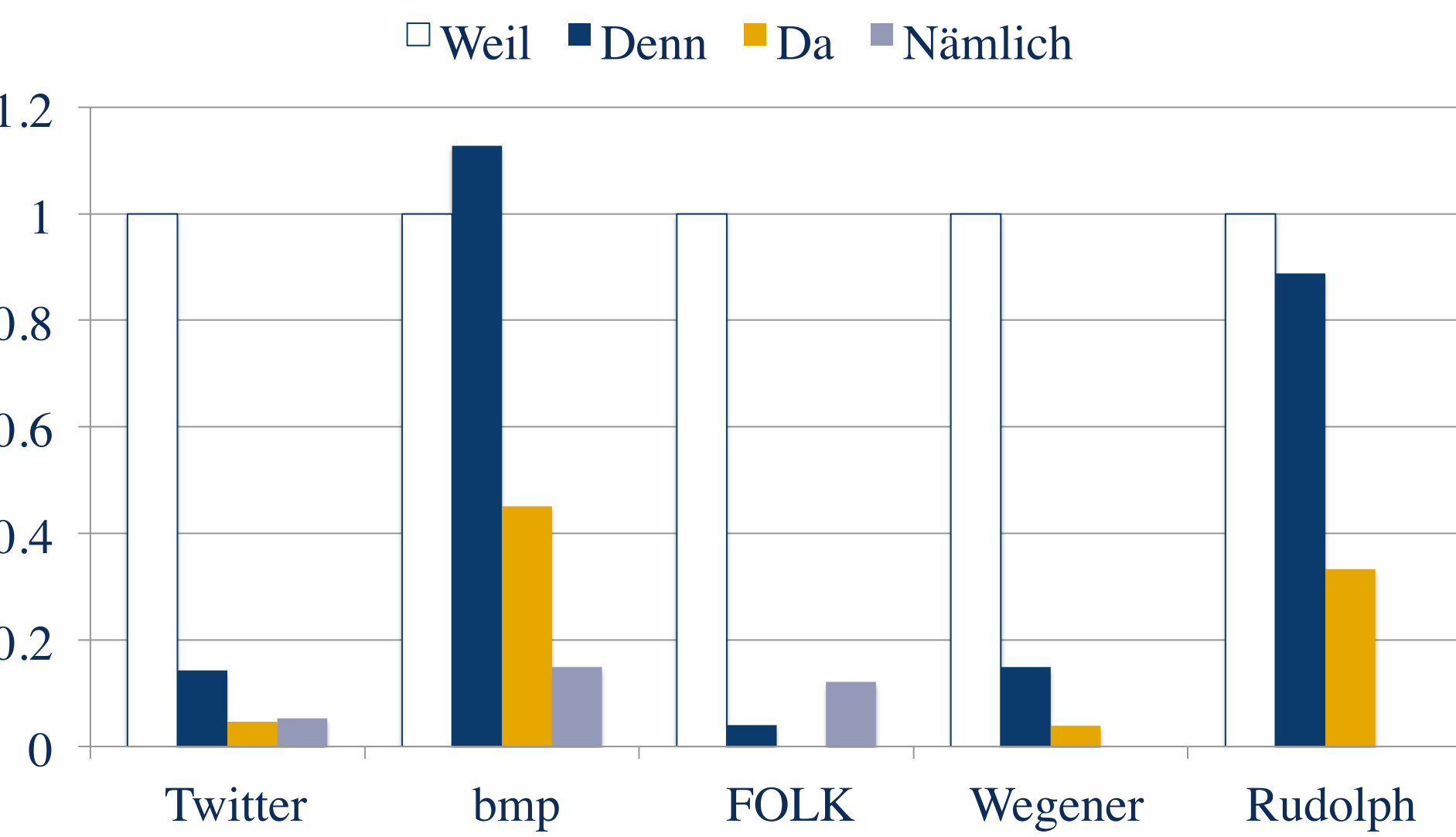
Discourse Connectives

- Discourse connectives are common in Twitter conversations

Corpus: 27.048.887 Tweets (April 2013)
reply n/y: 20.256.317 6.792.570
with connective: 23.61% 33.41%

Top ten connectives in German:
und – aber – dann – da – oder – doch – weil – denn – also

- Causal connectives on Twitter:**
- 1.7% of tweets / 2.6% of replies
 - “spoken”/informal style of justification



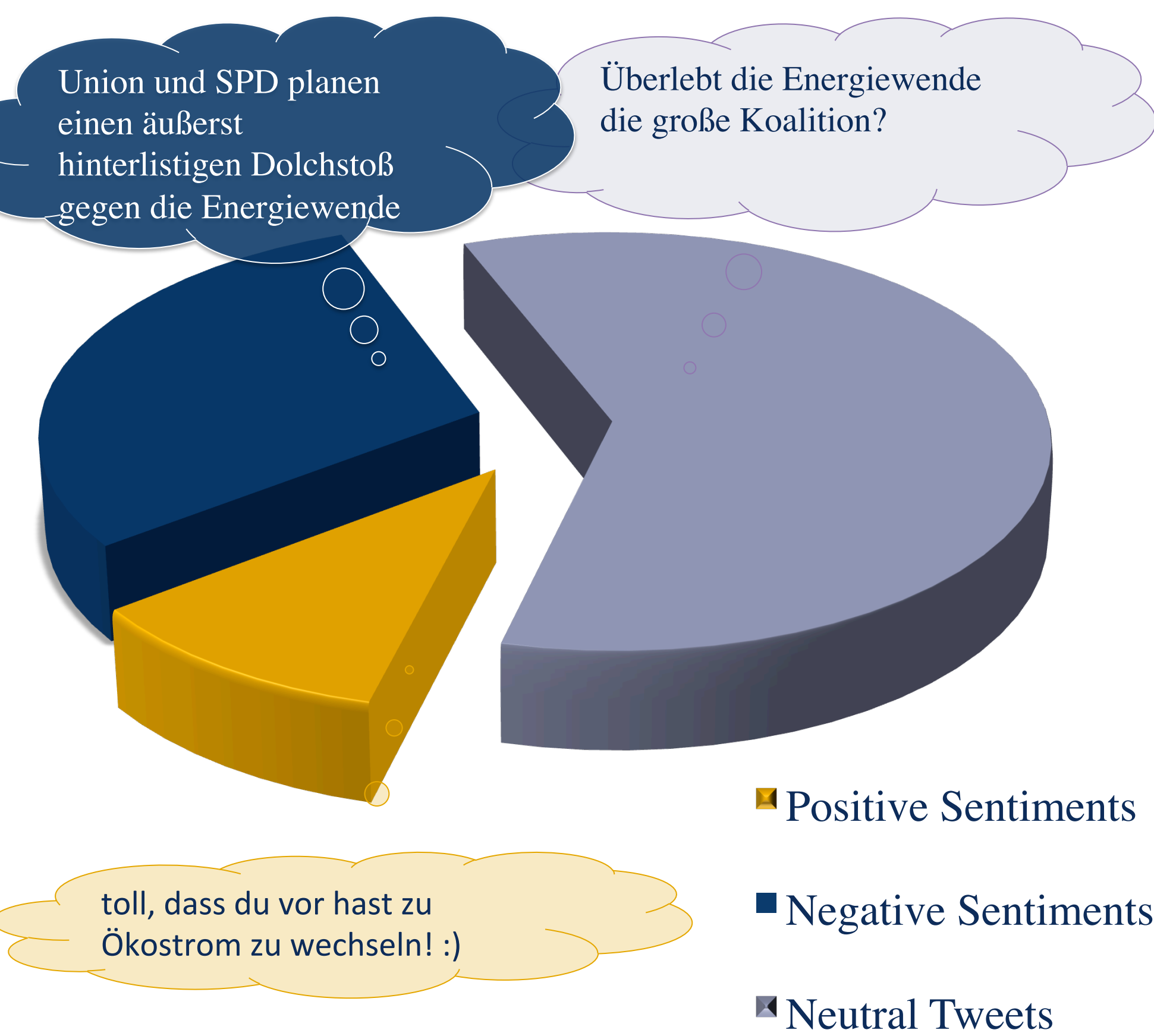
Relative frequencies of connectives ‘denn’, ‘da’, and ‘nämlich’ compared with ‘weil’ (all, ‘because’) in corpora of spoken and written German, and in Twitter.

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) referenced 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

Sentiment

Sentiment Distribution on Twitter



Automatic Sentiment Classification

Classification Algorithm	F-Measure
ZeroR	43.2%
Logistic Regression	57.8%
AdaBoost	58.6%
Näive Bayes	65.5%
Multinomial Näive Bayes	62.8%
LibLinear	63.2%
SMO	66.6%