



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG

PHILOSOPHISCHE FAKULTÄT  
UND FACHBEREICH THEOLOGIE

# Reconstructing Twitter arguments with corpus linguistics

Natalie Dykes, Philipp Heinrich, Stefan Evert  
FAU Erlangen-Nürnberg

[www.linguistik.fau.de](http://www.linguistik.fau.de)

# Introduction

## Argument mining

- relatively new field in NLP and CSS
- goal: automatic extraction and representation of arguments from texts
- classic argumentation schemes:
  - Modus Ponens ( $X \rightarrow Y, X \mid Y$ )
  - Modus Tollens ( $X \rightarrow Y, \neg Y \mid \neg X$ )
- task: identify premises and conclusions

# Related work

Challenges when looking at real text:

- premises (or even conclusion) are often left implicit (Bosc et al. 2016)
- non-traditional forms of argumentation (“defeasible” arguments, see Walton et al. 2008)
- persuasion is achieved through rhetorical strategies, particularly selection, arrangement, phrasing of argumentative units (Wachsmuth et al. 2018)

# Related work

Finding and classifying arguments automatically in social media is hard (see e.g. Goudas et al. 2014):

- detection of argumentative sentences: ca. 77% F1
- splitting claims and premises: ca. 40% F1
- features for MLs include number of comma tokens, verbs in passive voice, cue words, mean word length, etc.

# Related work

## Challenges on social media (Goudas et al. 2014)

- defeasible arguments (ad hominem in different flavours – Dykes et al. forthcoming)
- implicitness (restriction to 140 characters on Twitter)
- unmediated environment (no structure or guidelines, multimodality)
- non-standard language

# Argument Mining and Social Media

*Ad hominem*: the opponent is discredited to strengthen the speaker's own stance

- 1) If you are too lazy to vote then you have no right to complain
- 2) donald trump is too stupid to know the significance of brexit

(source)  
JunckerEU at 2014-05-20 20:16:40:  
  
In the very long run, we will need a European army.  
Because we have to be credible when it comes to foreign policy  
#wahlarena #withJuncker

quotes

(A1)  
sandieshos at 2016-05-28 10:24:45:  
  
@JunckerEU  
What an awful little man you are. The epitome of a dictator.  
The UK will reject you & your vision for Europe.  
#Brexit

(A2)  
DevonBloke at 2016-06-09 18:20:39:  
  
@sandieshoes @TeddyUKIPClark @JunckerEU  
The 'army' also needed to subdue 'disruptive' EU states. #Brexit

linear sub-thread

(B1)  
SocialistVoice at 2016-05-28 19:57:11:  
  
EU President @JunckerEU wants a European Army.  
Now, if I ever knew of a good reason to vote #Brexit

(B2)  
scotpolitik at 2016-05-28 20:06:53:  
  
@SocialistVoice @JunckerEU  
#brexit <https://t.co/CJhhIrVOS>

linear sub-thread

(C1)  
LouiseMensch at 2016-05-28 14:37:45:  
  
.@JunckerEU Nobody is #WithJuncker  
fuck off - no EU Army. #Brexit #voteleave

(C24)  
scotpolitik  
2016-05-28 18:01:36

(C23)  
Cllrporter  
2016-05-28 14:39:14

(C22)  
cheekylatte  
2016-05-28 14:44:42

(C21)  
JOHN\_POULTER  
2016-05-28 14:58:40

independent responses to a first-order response

# Corpus

- 6 million tweets containing “Brexit” collected by Milajevs between 05/05-24/08/2016 (<https://zenodo.org/record/263584/>)
- Pre-referendum only (higher consistency in argumentation)
- Preprocessing
  - Off-the-shelf algorithms for tokenisation and tagging (Owoputi et al. 2013)
  - Custom lemmatiser based on Minnen et al. (2001)
  - Removed near-duplicate tweets generated by social bots (Schäfer et al. 2017).
  - (NER and phrase chunking)

# Mapping queries to formulae

$\forall x \{x : \text{entity}\} \in \{y : \text{entity}\} : \{z : \text{property}\}(x)$   
'all entities in entity 1 have property 2 e.g. being an idiot'

...or, on a more positive note:

- Common Folks *ad populum* (Walton et al. 2008)
  - Premise: I (the speaker) am an ordinary person, that is, I share a common background with you (the audience).
  - Conclusion: Therefore, you ought to take what I say as being more credible or acceptable
  
- Position to Know *ad populum* (Walton et al. 2008)
  - Premise 1: Everybody in this group G accepts A.
  - Premise 2: This group is in a special position to know that A is true.
  - Conclusion: Therefore, A is (plausibly) true.

# From schemes to KWIC

## Different linguistic representations

- Common Folks *ad populum*
  - *to stay , because <the average person doesn't need to be left in the hands of the brexit leaders> !! Are ppl really*
  - *@DrAlanGreene <I'm as against #Brexit as the next man> but this is nonsense*
- Position to Know *ad populum*
  - *<. As an Irishman I wouldnt mind erasing that border> lol*
  - *Fortunately most <business people like myself know better>*

# Corpus linguistics for ArgMining

- Linguistic patterns, but not directly tied to word level
- Corpus-linguistic approach: CQP query language (Evert & Hardie 2011)
  - Phrase/ clause structure patterns defined by POS sequences
  - Word lists representing lexico-semantic categories
  - Iterative development informed by regular concordance analysis and consulting with the CS team

```
"common|regular|normal|average|ordinary"  
[lemma=$nouns_person_common] (/vp[] | /pp[] | /np[] |  
/ap[] | ".?["bB]rexit")+;
```

# Syntactic macros

```
## A determiner phrase
MACRO dp(0)
(
  [pos = "D"]?
  (
    /ap[]
  )*
  [pos = "N|Z"]+ #noun or proper name
)
;

MACRO np_base(0)
(
  /dp[]          # full determiner phrase
| /pron[] #pronoun
)
;
```

# Wordlists for semantic grouping

caveman  
cheater  
child  
clown  
contempt  
crap  
cretin  
criminal  
crowd  
cult  
cunt  
dick[a-z]\*  
dictator  
dinosaur  
disdain  
disgrace  
dodo  
douche  
douchebag

\$nouns\_person\_negative (curently at 150 items)

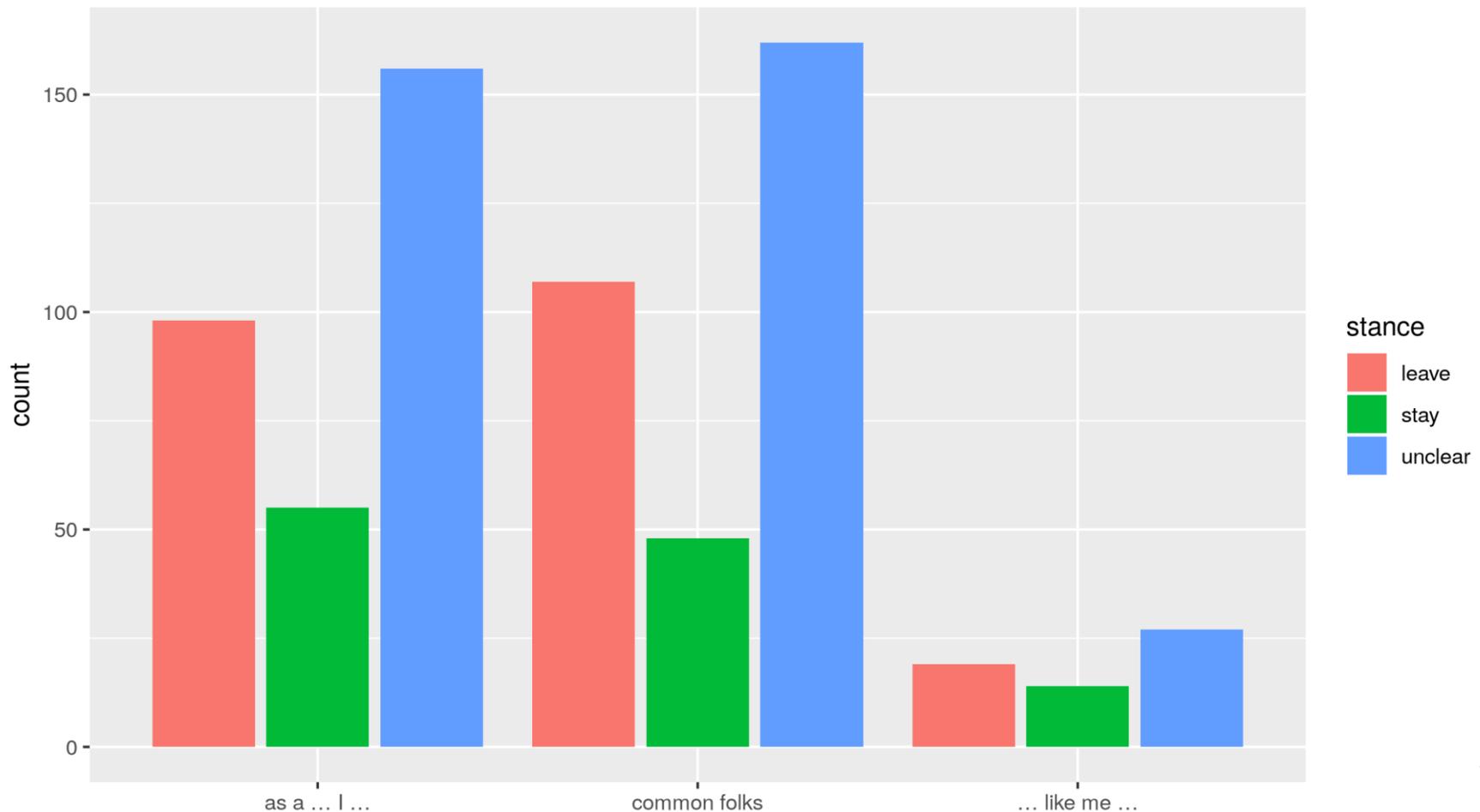
# Case study: Self-identification in the Brexit discourse on Twitter

`\forall x {?0 : entity} \in {?1 : entity} : {?2 : property}(?0)`

- Three query types:
  - as an X I Y
  - X like me Y
  - {ordinary/normal/common} {people} Y
- Manual categorisation
  - Leave/stay/unclear or NA
  - Group identity statements categorised by domains

# Case study: Self-identification in the Brexit discourse on Twitter

emp. distribution of argumentation schemes s.t. stance

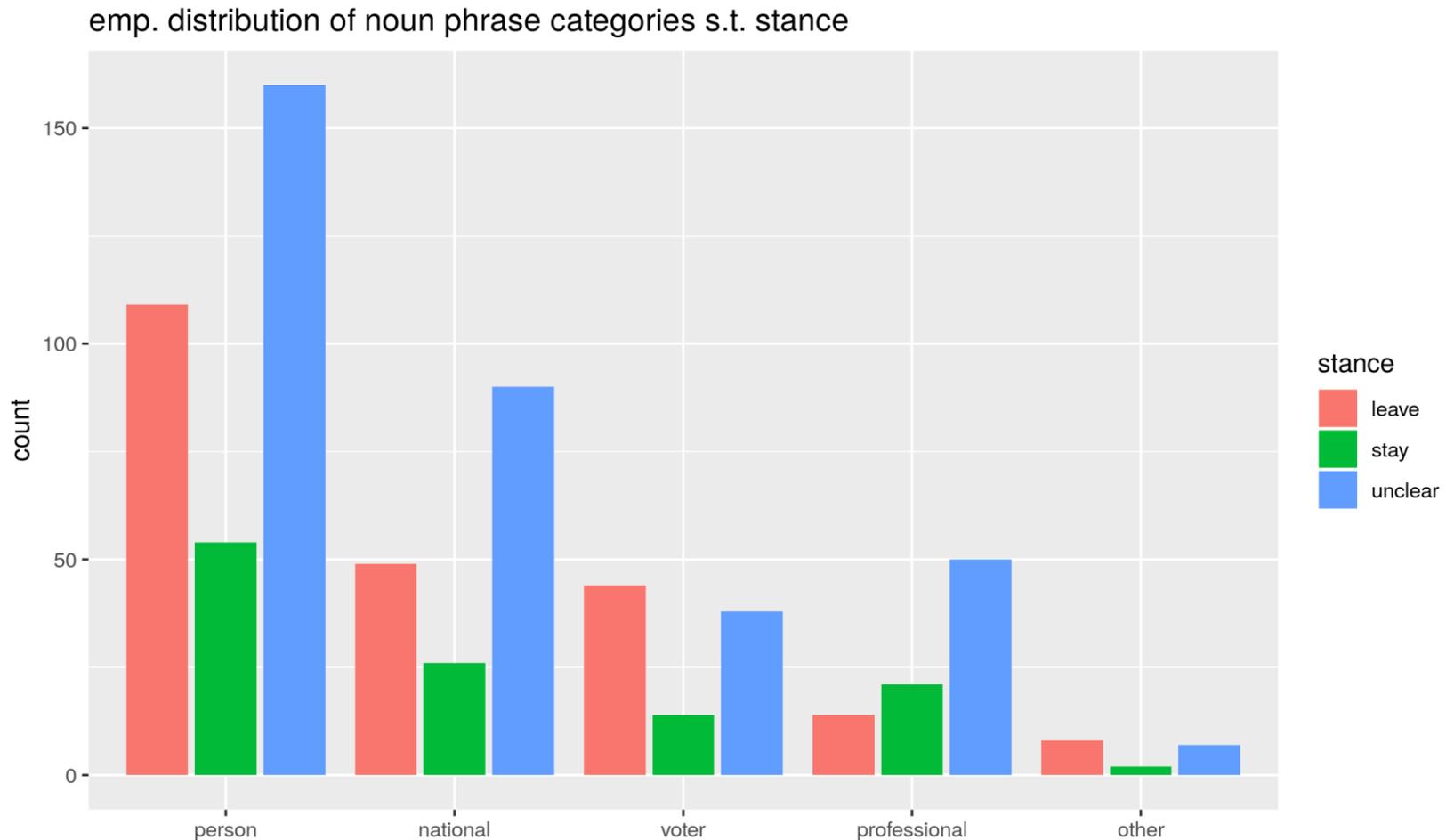


# Case study: Self-identification in the Brexit discourse on Twitter

Majority of hits: no clear stance towards Brexit

- 2282509: zin @FiveRights lol <normal people are can't afford iPhones why> ? b they are f
- 3931632: . BREXIT will be better . <As an American I have> only two things to say :
- 6949948: a Soubry eh .. who thinks <men like me are the problem> . Middle aged male and p

# Case study: Self-identification in the Brexit discourse on Twitter



# Case study: Self-identification in the Brexit discourse on Twitter

Stance may differ within group membership:

- 10364789: . @MyronChristodou @vote\_leave <ordinary folk **will do worst from #Brexit**> - except perhaps t
- 32245155: #brexit . Almost <all ordinary folks I speak to **are voting Leave**> .
- 39052616: emain #brexit #strongerin <As an American I **am hopelessly uninformed about the #Brexit**> , I just hope Britain ca
- 6237133: ty on brexit is a concern . <As an American I **can't understand why UK is in EU**> & London elected a Musli

## Next steps

- Further externalise lexical placeholders into wordlists to improve precision:  
928369: ve read the news recently and <as a result I have mourned> . What should I be scare
- Automated chunks for standard phrases; fallback macros for systematic errors to improve consistency
- Semi-automatic extension of manually compiled wordlists using WordEmbeddings

## Conclusion

- Queries balance grammatical and semantic flexibility in patterns
- Each query: one linguistic instantiation of a given argumentation scheme
- Mapping of schemes to logical formulae (theoretical CS)
- Qualitatively informed approach to handle noisy data



## Questions?

<Presenters like us would like to thank you>

<As FAU researchers we are looking forward to  
your questions>

# References

- Bosc, Tom Elena Cabrio, Serena Villata. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media. 2016. Proceedings of the 6th International Conference on Computational Models of Argument.
- Evert, Stefan & Andrew Hardie. 2011. Twenty-first century corpus workbench: Updating a query architecture for the new millennium. In *Proceedings of CL*.
- Goudas, Theodosios, Christos Louizos, Georgios Petasis & Vangelis Karkaletsis. 2014. Argument Extraction from News, Blogs, and Social Media. In *SETN 2014*.
- Minnen, Guido, John Carroll & Darren Pearce. 2001. Applied morphological processing of English. *Natural Language Engineering* 7(3). 207–223.
- Owoputi, Olutobi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider & Noah Smith. 2013. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. In *Proceedings of NAACL*.
- Schäfer, Fabian, Stefan Evert & Philipp Heinrich. 2017. Japan's 2014 general election: Political bots, right-wing internet activism and PM Abe Shinzō's hidden nationalist agenda 5(4). 294–309.
- Wachsmuth, Stede, El Baff, Al-Khatib, Skeppstedt, Stein. 2018. Argumentation Synthesis following Rhetorical Strategies. In Proceedings of the 27th International Conference on Computational Linguistics. 3753–3765.
- Walton, Douglas, Chris Reed & Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge: Cambridge University Press.