



## Goal: Negative Actor Identification

**Claim:** if regular targets of hate speech are turned into perpetrators, then a text is hate speech

- i.e. 'Migrants destroy our democracy'
- actor identification is animacy identification
- 1st step: identification of actorship
- 2nd step: quantification of negative strength

**Research Hypothesis:** Word embeddings are a good basis for a general approach to animacy noun classification

## Animacy Detection for German

### State of the Art

- no corpus of annotated noun phrases available for German
- no contextual features could be used, thus
- insight: animacy bearing words are very often head nouns
- wordnets (GermaNet) allow access to good hyponym candidates
- frequency-based selection overcomes GermaNets over-specific vocabulary: track construction foreman (Gleisbauvorarbeiter)

### Data

- 7,000 nouns hypernyms Mensch (human being) and Gruppe (group)
  - e.g. *milkman, nun, doctor*
- 8,000 nouns hypernym Artefakt (artefact)
  - stone, sun, tree, house
- seed lists are noisy
  - artefact (non-animacy) contains metonymy trigger
  - group (animacy) contains non-animacy nouns (e.g. branch)
- thus: annotation needed

## Corpus and Evaluation

### Corpus Annotations

- 30 students annotated each 600 nouns
- each set has 2 annotators
- thus: 9'000 annotations (15 \* 600)
  - 40/60 distribution:
  - 3,600 animacy
  - 5,400 non-animacy candidates
  - proper names (incl. acronyms) were removed
- final corpus: 3145 animacy and 5512 non-animacy nouns
- IAA: Kappa of 0.82

### Examples

#	noun	type	MT
1	Terrorregime (terror regime)	A	yes
2	Zeitung (newspaper)	A	yes
3	Sünder (sinner)	A	no
4	Minister (minister)	A	no
5	Universität (university)	A	yes
6	Weltcup (world cup)	NA	-
7	Roman (novel)	NA	-
9	Salz (salt)	NA	-
10	Pluralismus (pluralism)	NA	-

### Overall Accuracy

	FastText	BERT
MLP	94.3	92.2
LR	94.1	92.1
LR+CW	94.4	92.2

Logistic regression (LR), Multilayer Preceptron (MLP)

### Class-wise Accuracy

	LR			MLP		
	p	r	f1	p	r	f1
A	92.2	93.0	92.6	91.6	93.3	92.5
NA	95.8	95.2	95.5	95.9	94.8	95.4

Animacy (A), metonymy trigger (MT)

## Results

- data and classifier freely available
- three lists
  - direct animacy denoting nouns
  - metonymy trigger for animate referents
  - non-animacy denoting nouns

### Conclusion

- classifying nouns into (non-)animacy works well
- word embeddings are a good basis
- see poster: Polar Quantification of Actor Noun Phrases for German