



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Benjamin Milde, Prof. Dr. Chris Biemann

---

**DEEP LEARNING FOR LANGUAGE AND  
SPEECH - SEMINAR PROJECTS**

# Projects

# #1 GermaNER tagger (1)

- Alex Tronje
- German named entity recognition (organizations, names, places) with a twist - can you make the model compact?
- Dataset: <https://github.com/tudarmstadt-lt/GermaNER>
- Model accuracy vs. model size
- Many OOVs, will need a character model
- Might be used by LT projects e.g. new/s/leak (NetWork of Searchable Leaks), <https://www.inf.uni-hamburg.de/en/inst/ab/lt/resources/demos/new-s-leak.html>

# #1 GermaNER tagger (2)

Schartau B-PER  
sagte 0  
dem 0  
" 0  
Tagesspiegel B-ORG  
" 0  
vom 0  
Freitag 0  
,0  
Fischer B-PER  
sei 0  
" 0  
... 0

Firmengründer 0  
Wolf B-PER  
Peter I-PER  
Bree I-PER  
arbeitete 0  
Anfang 0  
der 0

## #2 Assessing Humor

- Elnur, Fynn, Vinh
- Competition with international teams!
- <https://competitions.codalab.org/competitions/20970>
- 15095 headlines with simple edits designed to make them funny
- <https://www.aclweb.org/anthology/N19-1012.pdf>

Original Headline	Substitute	Grade
Kushner to visit <b>Mexico</b> following latest Trump tirades	therapist	2.8
Hillary Clinton Staffers Considered Campaign Slogan 'Because It's Her <b>Turn</b> '	fault	2.8
The Latest: BBC cuts <b>ties</b> with Myanmar TV station	pies	1.8
Oklahoma isn't <b>working</b> . Can anyone fix this failing American state?	okay	0.0
4 <b>soldiers</b> killed in Nagorno-Karabakh fighting: Officials	rabbits	0.0

# #3 Detection of Propaganda Techniques in News Articles

- Ronja, Felix, Henri
- Competition with training data of annotated news articles
- <https://propaganda.qcri.org/semEval2020-task11/>
- Name calling, labeling, loaded language, exagg., ...

Manchin says Democrats acted like <sup>34</sup>babies<sup>40</sup> at the SOTU (video) Personal Liberty Poll Exercise your right to vote.

Democrat West Virginia Sen. Joe Manchin says his colleagues' refusal to stand or applaud during President Donald Trump's State of the Union speech was disrespectful and a signal that <sup>299</sup>the party is more concerned with obstruction than it is with progress<sup>368</sup>.

In a glaring sign of just how <sup>400</sup>stupid and petty<sup>416</sup> things have become in Washington these days, Manchin was invited on Fox News Tuesday morning to discuss how he was one of the only Democrats in the chamber for the State of the Union speech <sup>607</sup>not looking as though Trump <sup>635</sup>killed his grandma<sup>653</sup>.

When others in his party declined to applaud even for the most uncontroversial of the president's remarks, Manchin did.

He even stood for the president when Trump entered the room, a customary show of respect for the office in which his colleagues declined to participate.

file: article123456.txt

## #4 Counterfactual Statements

- Fabian, Christian, Jose
- Competition: <https://competitions.codalab.org/competitions/21691>
- Counterfactual statements describe events that did not actually happen or cannot happen, as well as the possible consequence if the events have had happened.
- Examples:

Her post-traumatic stress could have been avoided if a combination of paroxetine and exposure therapy had been prescribed two months earlier.

Finance Minister Jose Antonio Meade noted that if a jump in tomato prices had been factored out, inflation would have begun to drop.

## #5 One Million Posts Corpus

- Oliver, Julius, Jens
- DER STANDARD is an Austrian daily broadsheet newspaper.
- The data set contains a selection of user posts from the 12 month time span from 2015-06-01 to 2016-05-31.
- There are 11,773 labeled and 1,000,000 unlabeled posts in the data set.
- Labels: **Sentiment** (negative/neutral/positive), **Off-Topic** (yes/no), **Inappropriate** (yes/no), **Discriminating** (yes/no)



## #6 Common voice

- Recently released German **speech dataset** from Mozilla
- Some information about age, gender and accent available
- Predict those attributes (separately or together)
- Try transfer or multitask learning!

## #7 Who is the Better Speaker?

- Learning to Determine Who is the Better Speaker
- **Speech dataset** with ranked spoken (German) wikipedia articles
- From the Spoken Wikipedia Corpus
- People were asked which speaker "reads better"

## #8 Your own idea here

- As long as it is about text/speech
- And doable in a short time

## #9 Extra: Irony detection

- Imon Elham
- SemEval-2018 task 3 competition
- [competitions.codalab.org/competitions/17468](https://competitions.codalab.org/competitions/17468)
- Task a) binary classification:

I just love when you test my patience!! #not  
Had no sleep and have got school now #not happy

- Task b)
- i) verbal irony realized through a polarity contrast, ii) verbal irony without such a polarity contrast (i.e., other verbal irony), iii) descriptions of situational irony, iv) non-irony.