

# MACHINE LEARNING AND GENDER NEUTRAL LANGUAGE

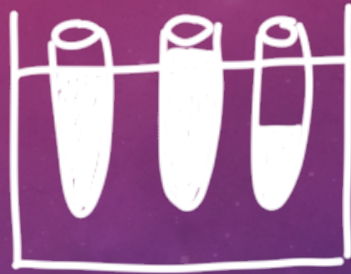
IN GERMAN

LARISSA HAAS

@L\_R\_SS\_



Motivation



Experiment 1:  
Preprocessing



Experiment 2:  
Word Embeddings





# ENTWICKLER

- Entwickler/Entwicklerinnen
  - Entwickler/innen
  - EntwicklerInnen
  - Entwickler(innen)
- Entwickler:innen
  - Entwickler\*innen
  - Entwickler\_innen





**Elea Brandt, queerfeministische Shitstormtrooperin**  
@EleaBrandt

Das mit der Aussprache ist echt hart, ich weiß.

Bei der Gesellschaft deutscher Sprache werden Leute auch nie beurlaubt, kochen nie Spiegelei und kreieren keine schönen Dinge.

Weil das kann einfach NIEMAND aussprechen!



**MDR AKTUELL** ✓ @MDRAktuell · 22 Std.

Die Gesellschaft für deutsche Sprache rät von der Nutzung des Gendersternchens ab. Der Verein teilte mit, das Sternchen sei mit Grammatik und Rechtschreibung nicht vereinbar. Zudem sei die Aussprache unklar.

2:59 nachm. · 13. Aug. 2020 · [Twitter Web App](#)

192 Retweets 685 „Gefällt mir“-Angaben 11 Zitate



**James A. Sullivan**  
@fantasyautor

Die Gesellschaft für deutsche Sprache hat also was gegen das Gendersternchen. 🙄

Die Positionen, die da vertreten werden, zeigen nicht gerade ihre Aufgeschlossenheit. Dort scheint kaum ein Interesse an einer lebendigen Sprache vorhanden zu sein.

[gfds.de/gendersternche...](https://gfds.de/gendersternche...)

9:57 nachm. · 13. Aug. 2020 · [Twitter for iPhone](#)

45 Retweets 177 „Gefällt mir“-Angaben



**Roxane Bicker | Autorin**  
@roxane\_bicker

Zwecks "kann man ja nicht aussprechen" vom Genderstern, ne?

Vielleicht sollten sich alle mal etwas mehr mit sprachlichen Hintergründen beschäftigen, Sprache ist nämlich sowas von inkonsequent 😂  
1/?



**Elea Brandt, queerfeministische Shitstormtrooperin** @EleaBr... · 20 Std.  
Das mit der Aussprache ist echt hart, ich weiß.

Bei der Gesellschaft deutscher Sprache werden Leute auch nie beurlaubt, kochen nie Spiegelei und kreieren keine schönen Dinge.

Weil das kann einfach NIEMAND aussprechen! [twitter.com/MDRAktuell/sta...](https://twitter.com/MDRAktuell/status/1291111111)  
[Diesen Thread anzeigen](#)

7:39 vorm. · 14. Aug. 2020 · [Twitter for Android](#)

15 Retweets 37 „Gefällt mir“-Angaben 1 Zitat

Google





autor



Alle



Bilder



News



Bücher



Videos



Ungefähr 1.070.000.000 Ergebnisse (0,57 Sekunden)

de.wikipedia.org › wiki › Autor ▼

## Autor – Wikipedia

**Autor**, weiblich **Autorin**, von lateinisch auctor „Urheber, Schöpfer, Förderer“ und englisch author, seit dem 17. Jahrhundert auch Verfasser ...

[Geschichte](#) · [Konnotationen des Begriffs](#) · [Publikationsmöglichkeiten](#)

autor\*in

de.wikipedia.org › wiki › Autor ▼

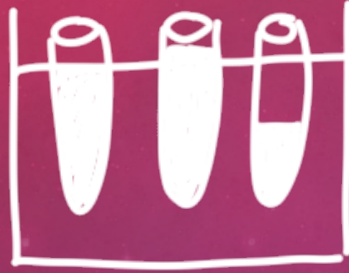
## Autor – Wikipedia

**Autor**, weiblich **Autorin**, von lateinisch auctor „Urheber, Schöpfer, Förderer“ und englisch author, seit dem 17. Jahrhundert auch Verfasser ...

[Geschichte](#) · [Konnotationen des Begriffs](#) · [Publikationsmöglichkeiten](#) ...

# EXPERIMENT 1

## PREPROCESSING



▶ M↓

```
text = "Autorinnen und Autoren, Autor:innen Autor_innen, AutorInnen Autor/innen, Autor*innen, Autor(innen)"  
pronomen = "jede*r, jede:r, jede_r, jede/r, eine*r, eine:r, eine_r, eine/r"
```



# EXPERIMENT 1

## PREPROCESSING

- Colon is interpreted as separating character, everything else stays
- I know this can be configured – but you need to be aware of these forms to change the behavior from default!
- It's not clear what's the wished behavior

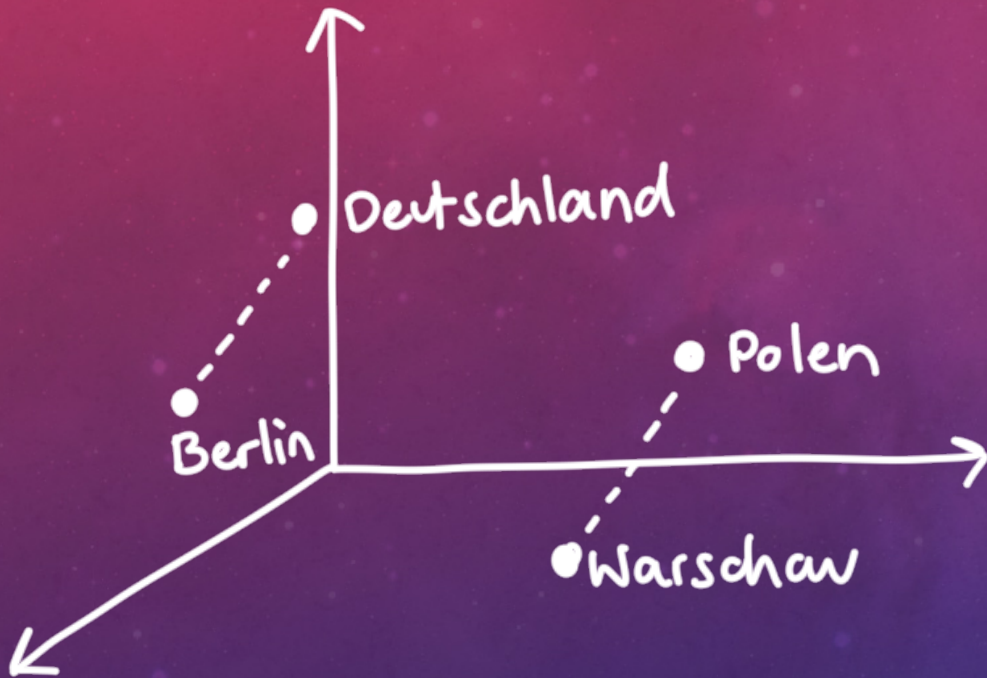
▶ M↓

```
lemmas = []  
for token in doc:  
    if token.is_punct is False and token.is_stop is False:  
        print(token.lemma_)  
        lemmas.append(token.lemma_)
```

```
Autorinnen  
Autoren  
Autor  
innen  
Autor_innen  
AutorInnen  
Autor/innen  
Autor*innen  
Autor  
innen
```

# EXPERIMENT 2

## WORD EMBEDDINGS



[45]

▶ M↓

```
berlin = model.get_vector('Berlin')
deutschland = model.get_vector('Deutschland')
polen = model.get_vector('Polen')
result = berlin - deutschland + polen
```

[47]

▶ M↓

```
model.similar_by_vector(result)
```

```
[('Polen', 0.8874868154525757),
 ('Warschau', 0.7945132851600647),
 ('Berlin', 0.7723734974861145),
 ('Breslau', 0.7452571988105774),
 ('Krakau', 0.7378164529800415),
 ('Stettin', 0.7082524299621582),
```



# EXPERIMENT 2

## WORD EMBEDDINGS

- Though they stay in preprocessing, they cannot be interpreted with word embeddings.

[61]

▶ M↓

```
for w in words:  
    print(w, [w in model.words])
```

```
Autorinnen [True]  
Autoren [True]  
Autorin [True]  
Autor [True]  
Autor:innen [False]  
Autor_innen [False]  
AutorInnen [True]  
Autor/innen [False]  
Autor*innen [False]  
Autor(innen) [False]
```

# REVIEW FROM FIRST TALK

- Define explicit rules how and when to deal with gender neutral language.
- Any ideas from you? Please contact me! 😊



# CONTACT

- Larissa Haas
- Twitter: @L\_R\_SS\_
- You can also find me in the PyData Südwest Telegram Channel or on LinkedIn