

SentiAnno

Building a Sentiment-Annotated, Topic-Specific Corpus of Austrian Historical Newspapers

CORPUS AND CODE

<https://github.com/lucijakrusic/SentiAnno>



BASE CORPUS

Focus: Migration & minorities in Austrian historical newspapers

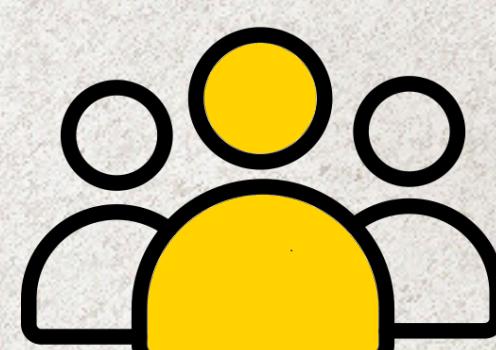
Scope: 1005 sentences from Neue Freie Presse & Das Vaterland (1850–1900)

Method: BERTopic with seed words (migration, labour, Czechs, Jews, Croats...)

ANNOTATION PROCESS

Annotators

3 semi-experts
(Master's students)



Categories
positive
negative
neutral
mixed



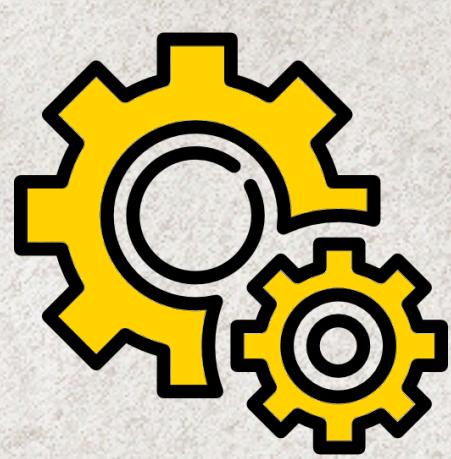
Perspective
as intended by
newspaper



Unit
sentence



Tools
Google Forms
Google Sheets
Doccano



INTER-RATER AGREEMENT

Fleiss' kappa
0.405 (fair, verging on
moderate)

APA
two-out-of-three: 93%

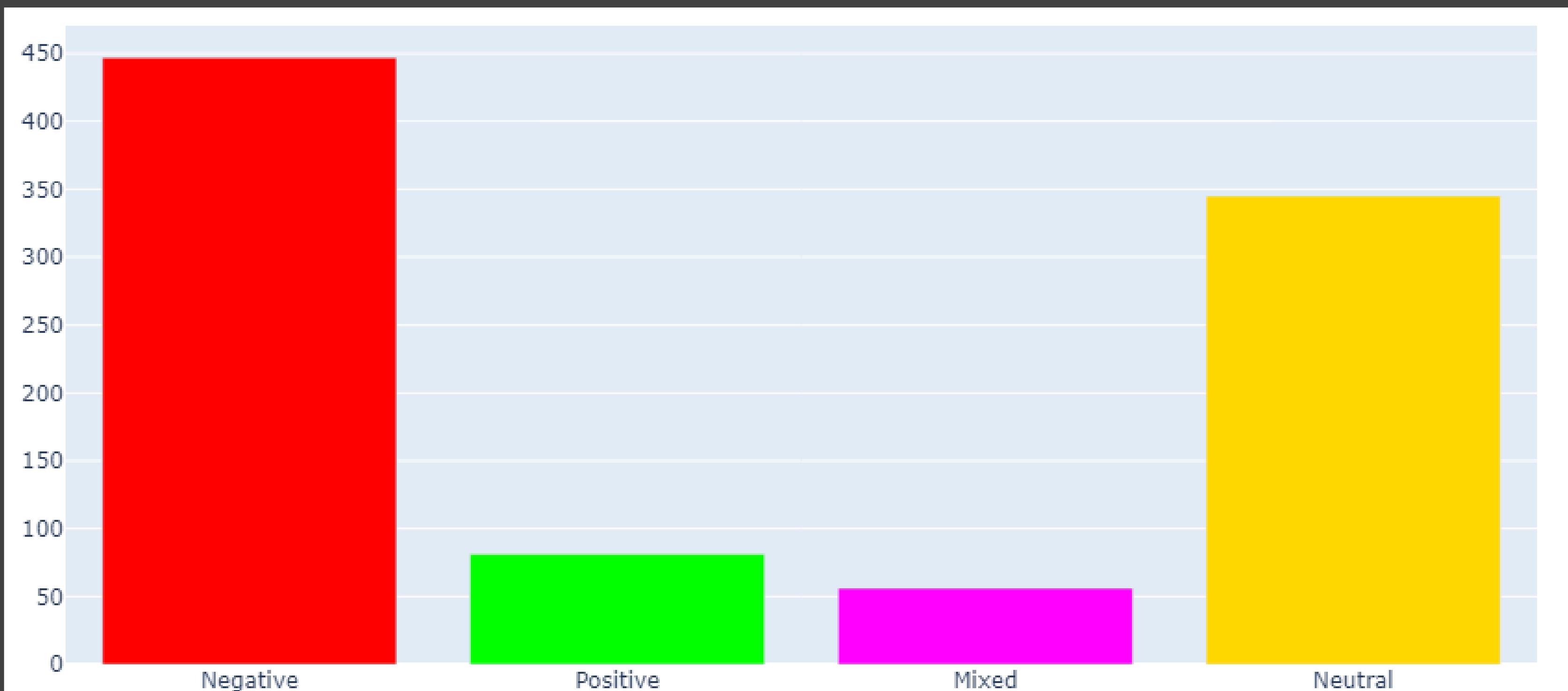
Complete agreement

Even if necessity drives a Serb or Wallachian to work, they do not exert themselves at all.

Complete disagreement

Despite the many mutual assurances of fraternal friendship, relations between Poland and Czechs have recently become strained - as is almost regularly the case before every reconvening of the Imperial Council.

DISTRIBUTION OF ANNOTATIONS



Lucija Krušić
lucija.krusic@uni-graz.at

Melanie Frauendorfer
melanie.frauendorfer@uni-graz.at

Clara Hochreiter
clara.hochreiter@edu.uni-graz.at

