

Ideas for master projects

Malvina Nissim
m.nissim@rug.nl
room 1311.421

IK meeting, 9 November 2018

Project 1: (Non-lexicalised) Gender Detection

(Non-lexicalised) Gender Detection

other people involved: Rob, Hessel, Masha

- cross-dataset (in two different languages, possibly adding English, too)
- cross-language
- human experiments
- self-reported gender for more accurate classification (beyond F/M, and $M \rightarrow F$, $F \rightarrow M$)

Project 2: Politicians between News and Twitter

Politicians between News and Twitter

other people involved: Marcel Broersma (media/journalism)

Data:

- datasets of News and Twitter (politicians in 2017 elections)
- metadata including gender, party, incumbency, etc

Possible tasks/questions:

- quantitative - model different targets: on what grounds are texts more similar?
 - gender
 - political party
 - incumbency
 - (polls)
 - ...
- qualitative - compare mentions of politicians in news with tweets (self-reporting)

Project 3: Embeddings

Label-specific Embeddings

- lookup in different spaces for different labels
- possible applications: profiling (space for F vs space for M), hate speech (space for hate, space for non-hate), hyperpartisan, ...
- WSD literature
- comparison with generic embeddings

Project 4: Hyperpartisan project: extensions