

## Abbreviated (and other OOV) words in Dutch tweets

Er zijn horrorfilms bezig tot 6 uur **lkk**  
Je kan me **mss** gek verklaren, ik heb er voor over  
**gwn** zoe zei vraag maksiem **drm** moest ik lachen

### Task

- Given an out-of-vocabulary word
- Find the corresponding in-vocabulary word

### Hypothesis

Abbreviated form is distributionally similar to the canonical form  
Gouws, Hovy, Metzler, *Unsupervised mining of Lexical Variants from Noisy Text*, EMNLP 2011

# Abbreviated (and other OOV) words in Dutch tweets

## Method

- A large Twitter corpus ✓
- A large corpus with standard vocabulary ✓
- Similarity scores for words in both corpora (**word2vec**!!)
- Evaluation data