

Lexical Normalization for Dutch

Task:

nee ! :-D kzal no es vriendelijk doen lol
nee ! :-D ik zal nog eens vriendelijk doen lol

Lexical Normalization for Dutch

Problem existing corpus:

- Small
- Flemish
- Collected in 1 evening
- Evaluation
- Capitalization?

Lexical Normalization for Dutch

Evaluation problem:

- Bleu
- WER
- CER

Lexical Normalization for Dutch

Evaluation problem:

Norm.	Tok.
708	1,322

Lexical Normalization for Dutch

Evaluation solution:

- Use categories proposed by van der Goot et al(2018)
- Add tokenization categories

Lexical Normalization for Dutch

Problems

- Data collection
- Tokenization
- Predict normalization categories
- Effect on end-tasks (NE, POS, etc.)
- (Unsupervised ranking of candidates)