

- For You
- Following
- Explore
- LIVE

Log in to follow creators, like videos, and view comments.

Log in

Suggested accounts

ITU
IT University of Copenhagen

AiCentreDK
Pioneer Centre for AI CPH

POS tagging Danish transfer learning cross-lingual cross-domain social media

POS-TAGGING

wauw	INTJ
😊	SYM
Du	PRON
er	VERB
dygtig	ADJ
🙏	SYM
😊	SYM
Han	PRON
er	VERB
f*ndme	ADV
den	DET
beste	ADJ
😓	SYM
😂	SYM

We annotate 8,028 tokens with Universal POS tags + three additional layers #UD

Trine Naja Eriksen

CODE-SWITCHING

INTRA	MIXED	INTER
🇬🇧	🇬🇧🇩🇰	🇬🇧🇬🇧

@username de er dem der altid **roaster** hinanden 🇩🇰

I ligner **lowkey** hinanden 🇬🇧

Min fjernbetjening bruger batterier ... **get it ?** 🇬🇧🇬🇧

ej hvor kan jeg da **relate** xD 🇩🇰

@username at jeg **simpede** for det der før lol 🇩🇰🇩🇰

Klassisk Københavner **sorry to say** 🇩🇰🇩🇰

#Code-switched words are exclusively English and are harder to tag for our models

Cathrine Damgaard

NORMALIZATION

Dt r fkn sjovt
PRON AUX ADV ADJ

Det er fucking sjovt
PRON AUX ADV ADJ

#Normalization is important on social media and improves the performance of #POSTaggers

Kia Kirstein Hansen

Stratified Class	Ælæctra(-D+L)	TwHIN(+D+L)
POS Certainty (N=203)	-25	-15
Normalized (N=3338)	15	10
In Fine-tune Vocab (N=1234)	-5	-10
Code-switched (N=2808)	15	5

er det ik hestennettet længere

vent

Our extra annotation layers show which features connect to #syntax on contemporary social media

Rob van der Goot

er det ik hestennettet længere

vent

Words for which annotators were #uncertain also posed the largest problems for automatic taggers

Maria Barrett

Language Model	Training Data	Accuracy (%)
BERTweet-B (+D-L)	LinES (-D-L)	49.50
Ælæctra (-D+L)	TwB (+D-L)	60.82
TwHIN (+D+L)	DDT (-D+L)	77.96
	TwB (+D+L)	85.35
	DDT (-D+L)	84.19
	TwB (+D+L)	85.15
	DDT (-D+L)	85.35

Domain beats language?!

#Multilingual models transfer better w/ in-domain, English than out-of-domain, Danish data

Max Müller-Eberstein