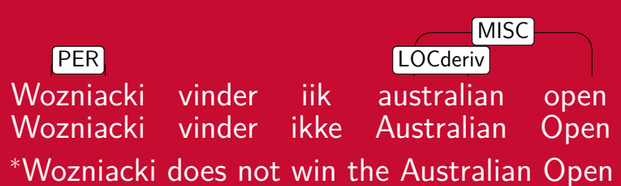
DaN+: Danish Nested Named Entities and Lexical Normalization

Barbara Plank, Kristian Nørgaard Jensen and Rob van der Goot

Task

- Danish NER for multiple domains
- Nested entities
- Lexical normalization layer
- Evaluate transfer from German



Results

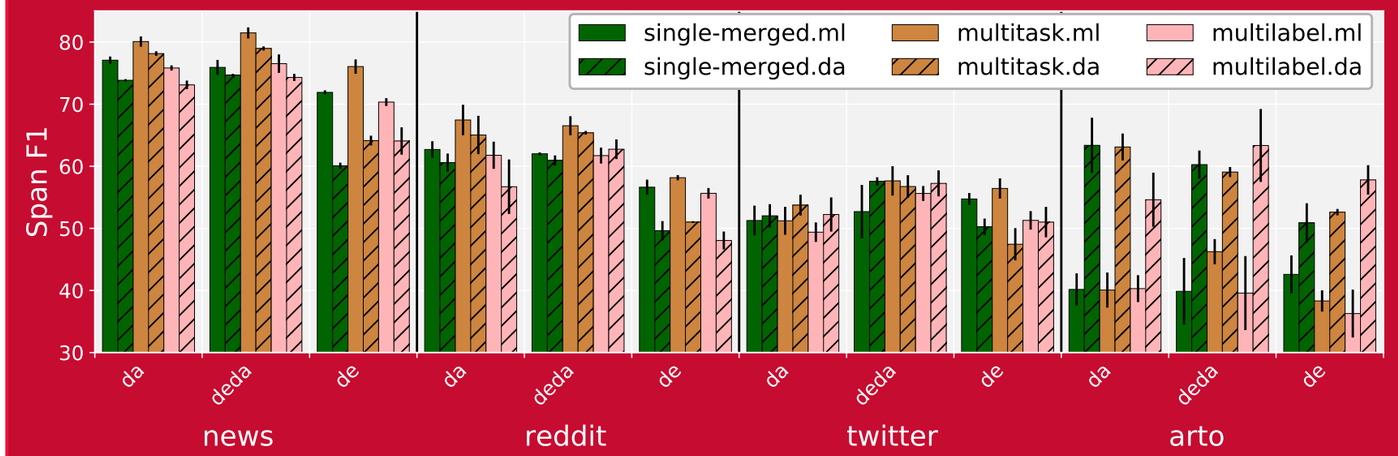


Figure 1: Nested NER results. Models trained on German (de), Danish (da) or both (de+da) and with mlBERT (ml) vs danishBERT (da). Average over 3 runs. Std. dev. indicated as error line.

DaN+ Dataset

Variety	Sentences	Tokens	Sent./NE	1st NE	2nd NE	%normed
News (DE)						
train	24,002	452,853	59%	29,078	2,467	—
News (DA)						
train	4,383	80,378	45%	3,800	235	—
dev	564	10,332	47%	468	36	—
test	565	10,023	48%	525	41	—
Reddit						
dev	326	4,547	60%	319	36	—
test	126	4,497	56%	128	20	—
Twitter						
dev	120	5,347	80%	279	13	3.5%
test	110	5,086	77%	284	32	2.3%
Arto						
dev	336	5,496	21%	93	1	16.7%
test	337	4,389	20%	103	12	15.2%

Table 1: Overview of DaN+: **Da**nish **N**ested **N**amed entities and lexical **N**ormalization, which includes news and social media varieties (Reddit, Twitter, Arto). First column: GermEval (Benikova et al., 2014).

Analysis

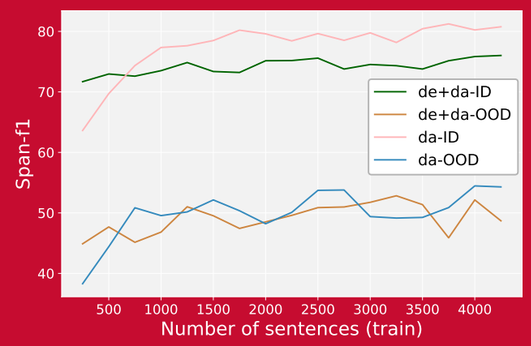


Figure 2: Learning curve for multi-task, mlBERT on in-domain (ID) news and average over all out-of-domain datasets (OOD).

	Normalization		NE tagging	
	Twitter	Arto	Twitter	Arto
Baseline	97.17	83.93	57.65	46.25
MoNoise	97.17	92.52	58.59	55.83
Gold	100.00	100.00	59.18	65.71

Table 2: Normalization accuracy, and its downstream effect on NER. For NER, multitask, mlBERT trained on de+da is used

	German	News	Reddit	Twitter	Arto
boundary-aware	57.89	56.89	16.48	21.37	13.77
Raw (ml)	83.31	80.73	57.99	60.87	54.88
Norm'ed (ml)	—	—	—	61.74	56.38
Raw (da)	72.80	80.13	50.99	62.88	55.48
Norm'ed (da)	—	—	—	62.91	56.59

Table 3: Nested NER F1 score on the test sets for models with mlBERT (ml) vs danishBERT (da).