

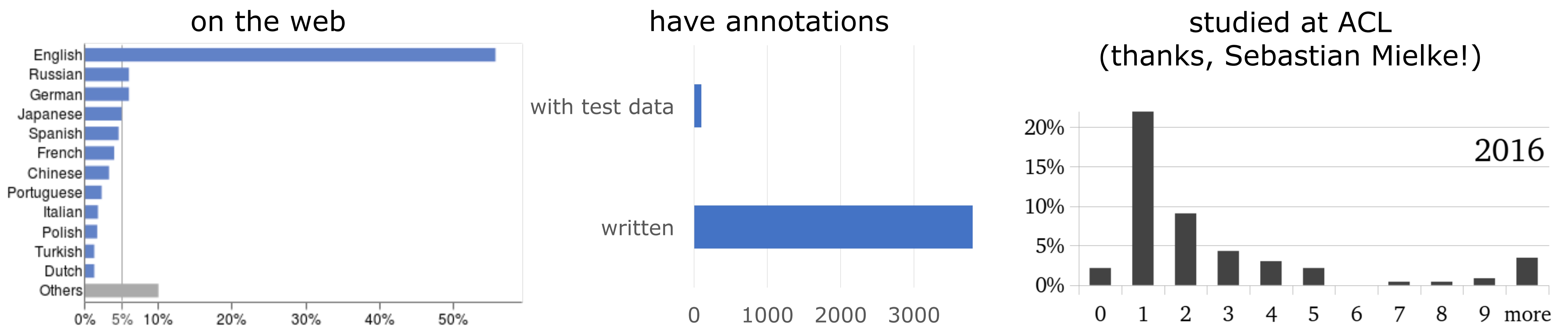
Cross-lingual tagger evaluation without test data

Željko Agić
zeag@itu.dk

Barbara Plank
b.plank@rug.nl

Anders Søgaard
soegaard@di.ku.dk

Three views on languages

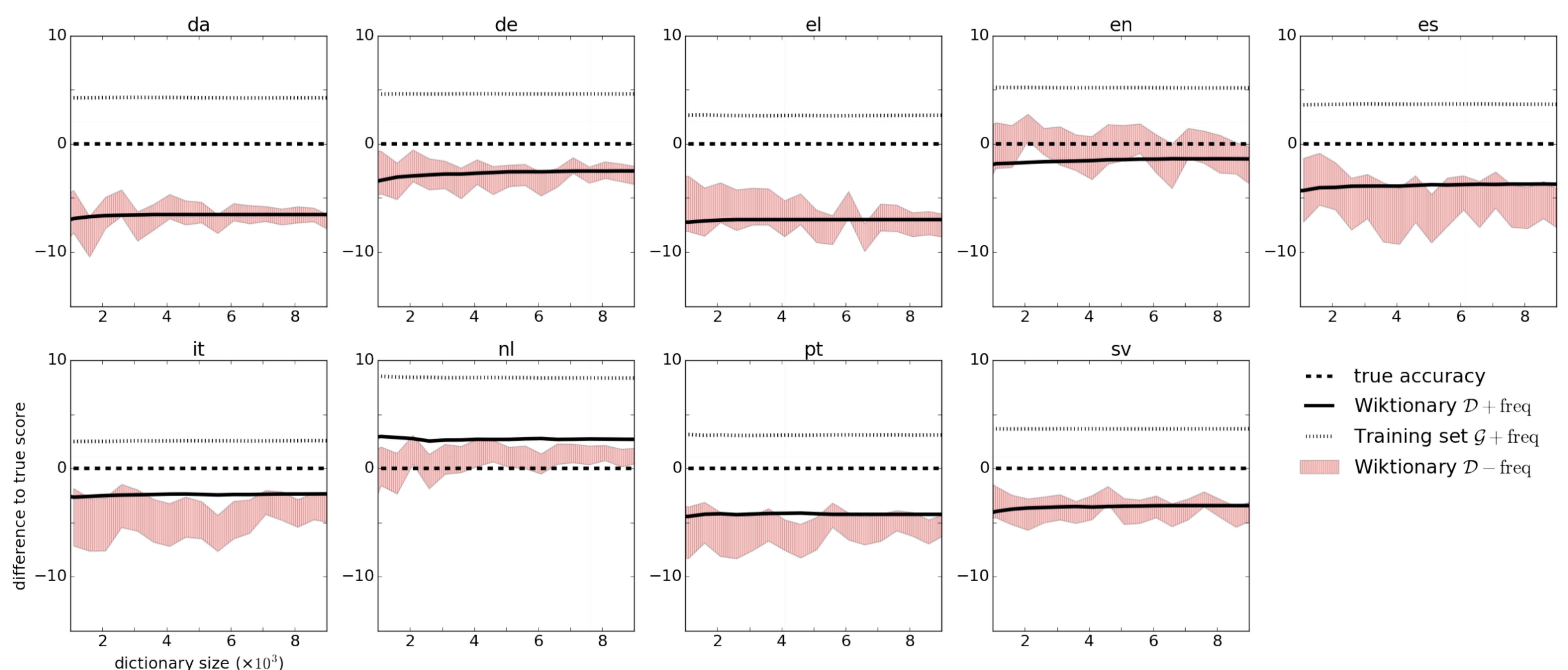


An evaluation impasse

“What we create is affected by what we can evaluate.”

—Winston Churchill 😊

Dictionaries to the rescue!



Wiktionary quality

	P	R
D	86.81	18.38
D_{trans}	58.09	11.01

Metrics evaluation

	acc diff	P@1	T _b
A_{soft}	-11,98	14/25	0.42
A_{trans}	-8,85	11/20	0.27