Enough is Enough! A Case Study on the Effect of Data Size for Evaluation Using Universal Dependencies

Rob van der Goot   Zoey Liu   Max Müller-Eberstein

Small
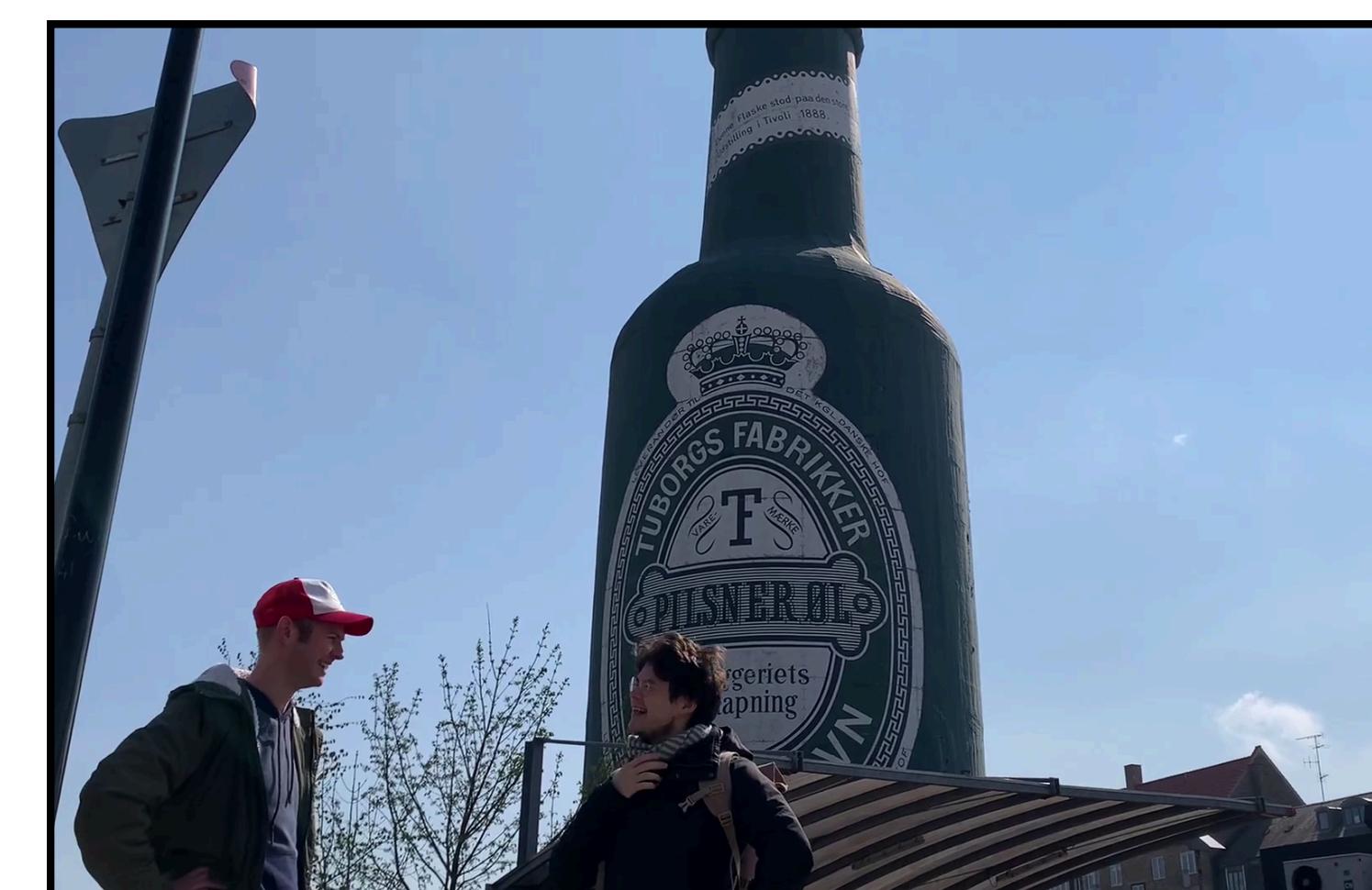
Medium

Large

Too large..

**Research question:
How large is too large?**

**Problems with too
large evaluation sets:**
- Higher costs 💸
- Less training data 🏃🏽‍♀️
- Less coverage 👴🏻👴🏻👴🏻
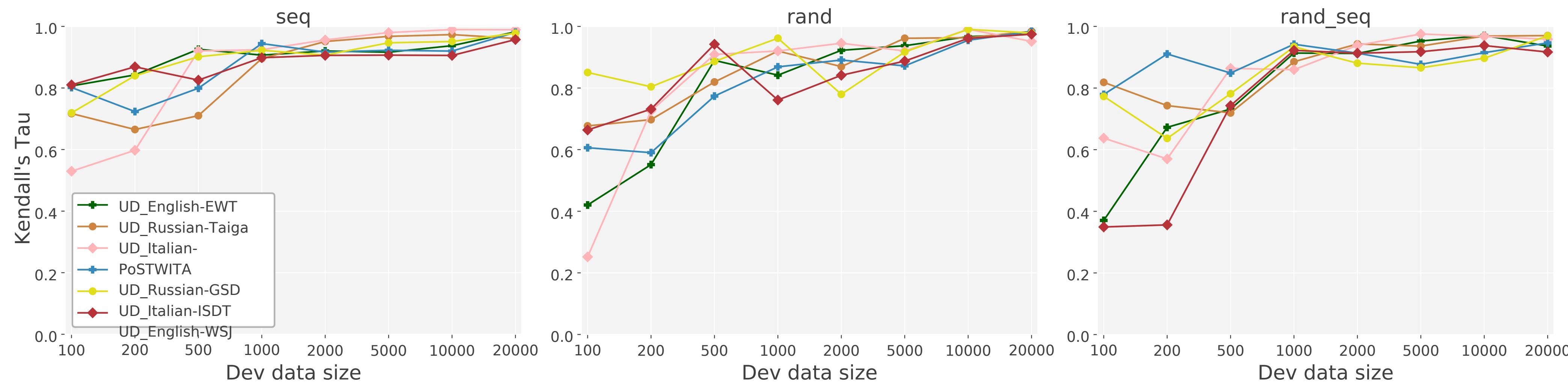
Setup
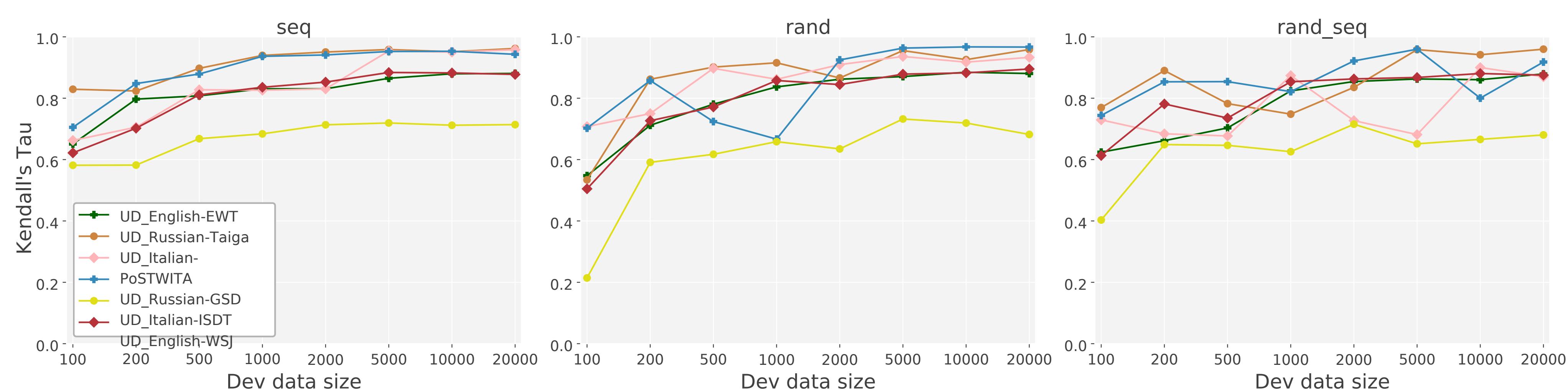
We resplit existing datasets in
new train (T) and dev (D)

Seq	T	T	T				D	D
Rand	T	T	T	D		D		
Rand_seq	T	T	T		D	D		
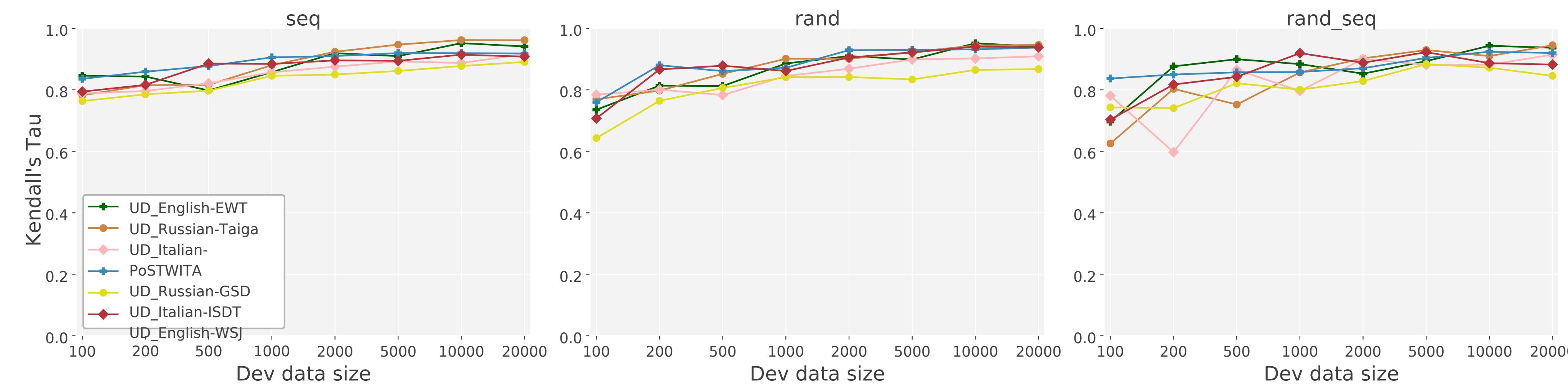
In-treebank

Cross-domain

Cross-lingual