Moving on from OntoNotes: Coreference Model Transfer
Patrick Xia and Benjamin Van Durme

Motivation
Coref: cluster spans of text that refer to the same entity

Poor coref transfer across:
• Annotation standards – include singleton clusters?
• Domain – which entity types are annotated?
• Language – is there any cross-lingual transfer?

Results

- Transfer models usually outperform randomly initialized models
- PreCo is as effective as OntoNotes
- PreCo is better with gold mention boundaries
- Continued training of small (publicly available) encoders is effective with low # training docs

Methods
Three model initialization methods:

- 1. Pretrained encoder only
- 2. Finetuned encoder on coref
- 3. Fully trained on source

Source datasets: OntoNotes, PreCo
Target datasets:
English: PreCo, LitBank (books), ARRAU (news), SARA (legal), QBCoref (quiz)
Other Languages: OntoNotes (zh, ar), SemEval (ca, es, it, nl)
Encoders: SpanBERT, XLM-R

Additional Findings
1. Continued training is effective cross-lingually
   New baseline numbers on several datasets

2. Allocate few docs for model selection
   84.6 vs. 84.9 F1 with 5 vs. 500 dev docs

3. There is still catastrophic forgetting:
   • Larger drops across annotation guideline changes
   • Smaller drops across domain or language

4. Finetune only top encoder layers with continued training

Code and models available at:
https://nlp.jhu.edu/coref-transfer