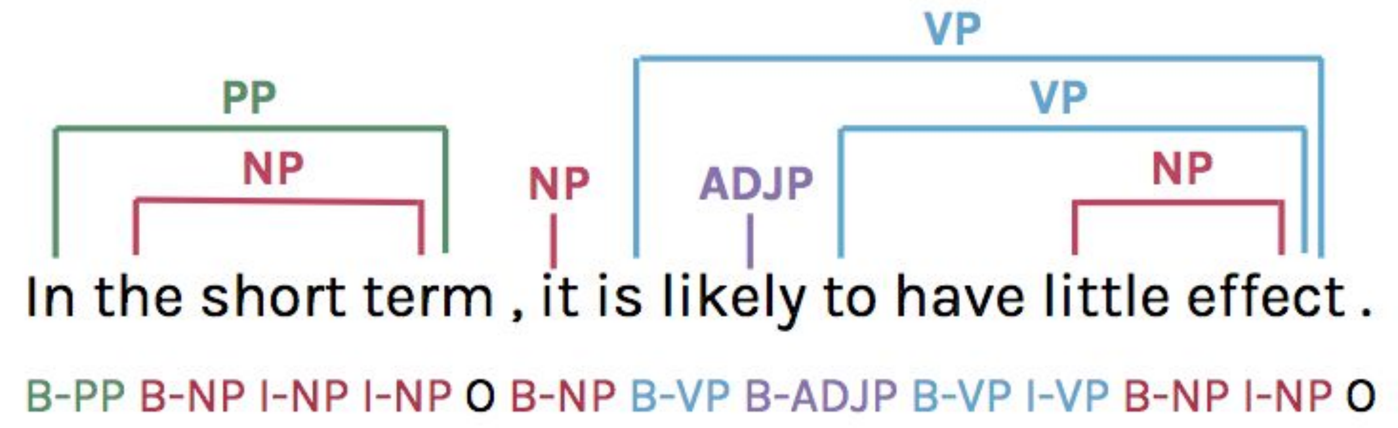
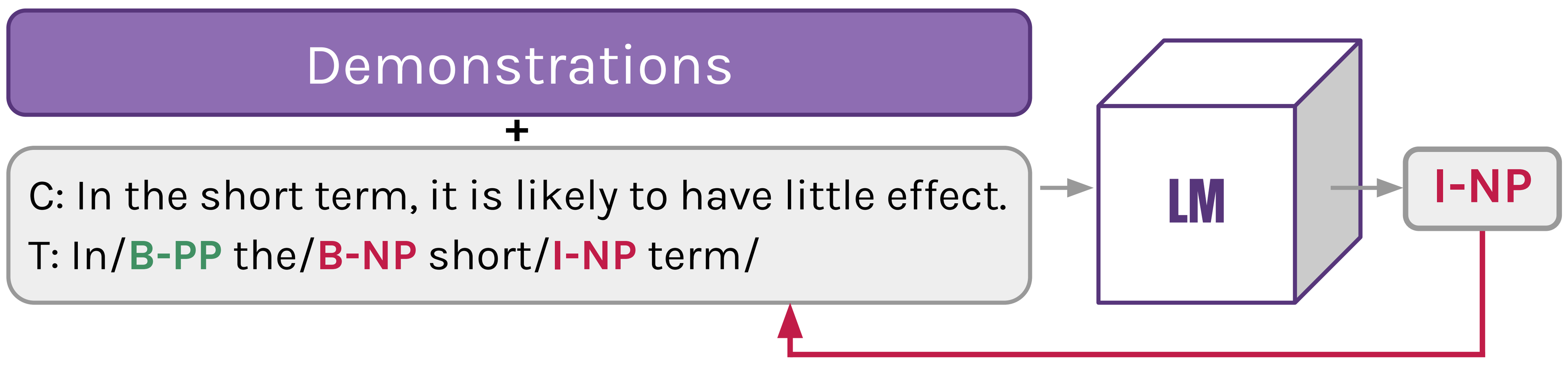
# Prompting Language Models for Linguistic Structure

Terra Blevins Hila Gonen Luke Zettlemoyer

University of Washington

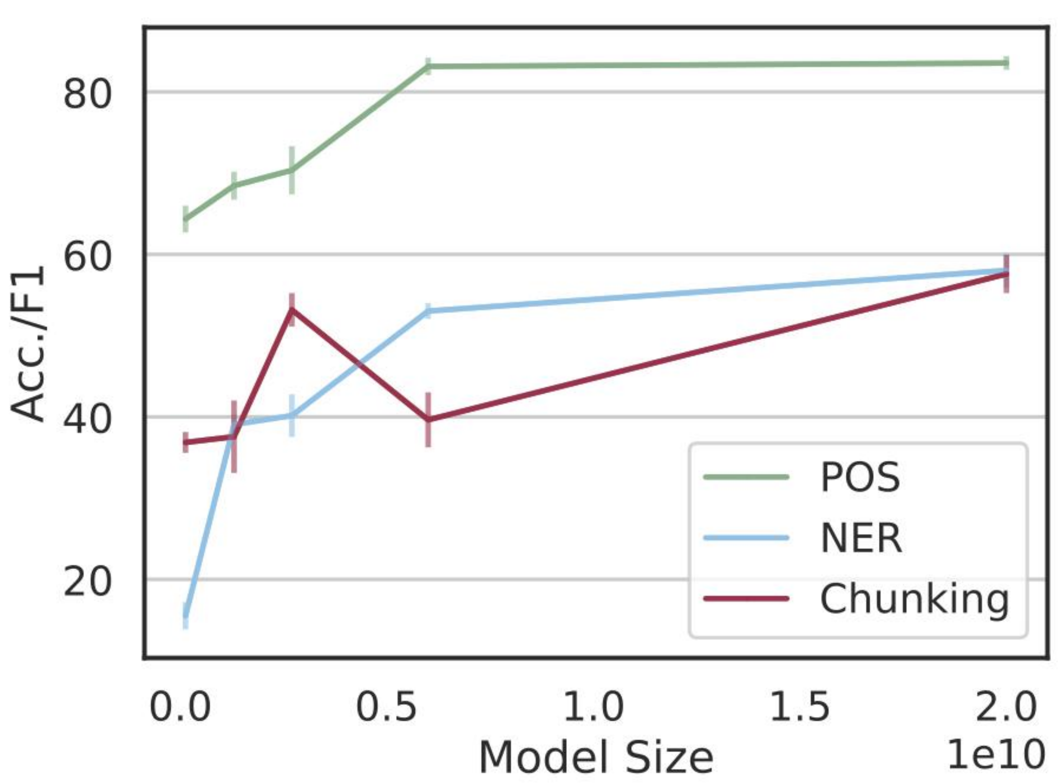



## Key Findings

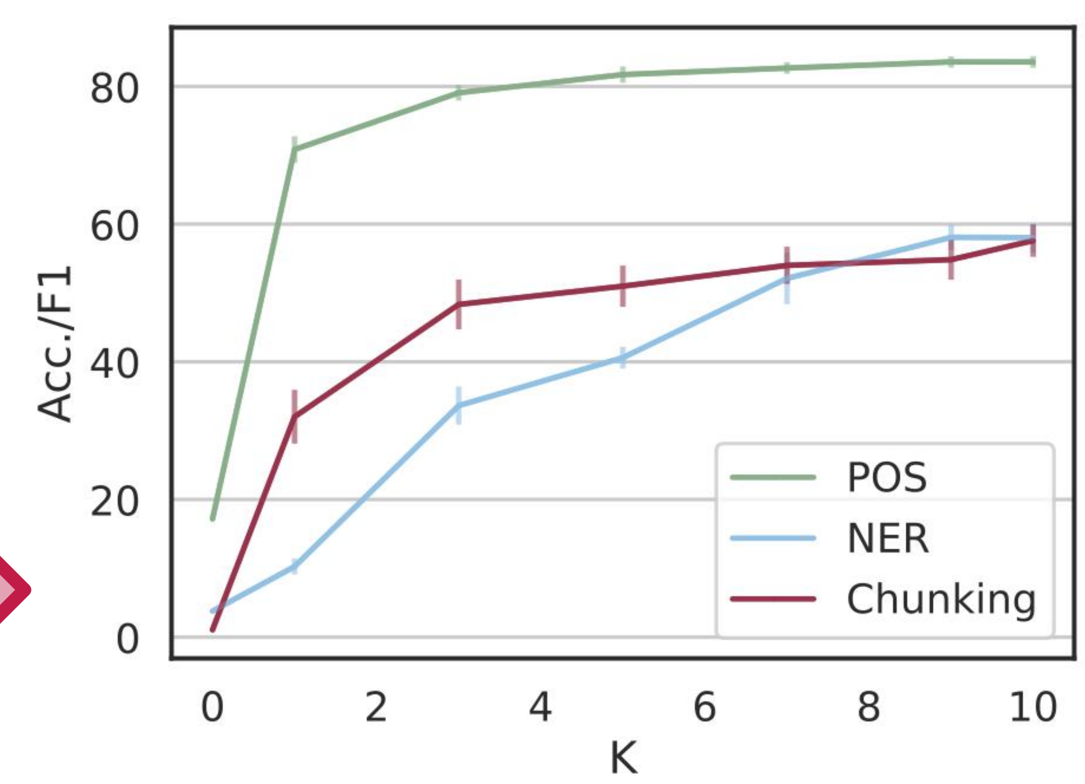
- Structured prompting tests LMs for linguistic knowledge framed as sequence tagging tasks
- LMs have strong priors over label meaning, likely due to pretraining on labeled task data
- But LMs can also in-context learn without priors by using unseen, informative labels!

## Structured Prompting extends in-context learning (ICL) to sequence tagging tasks



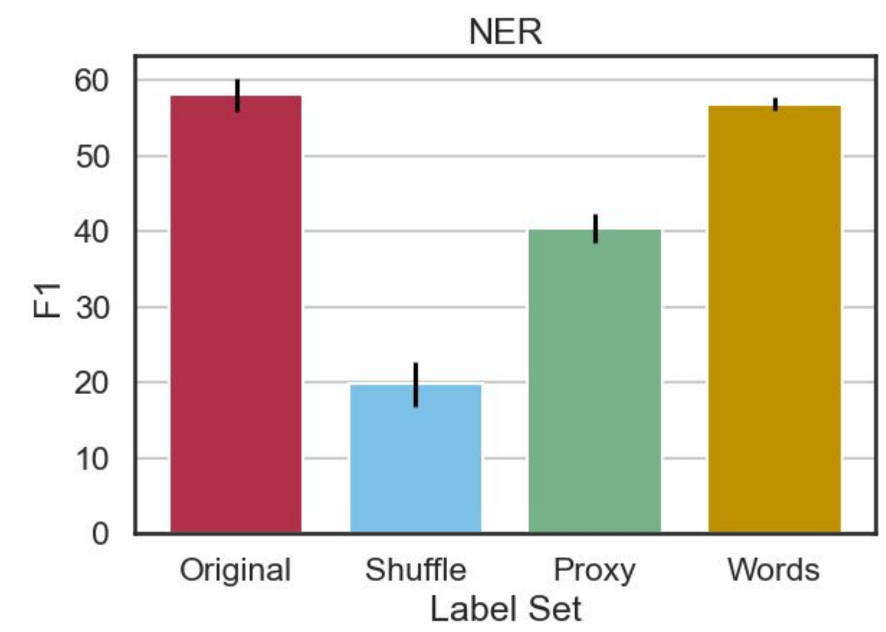

Performance across model sizes

Performance with different # of examples

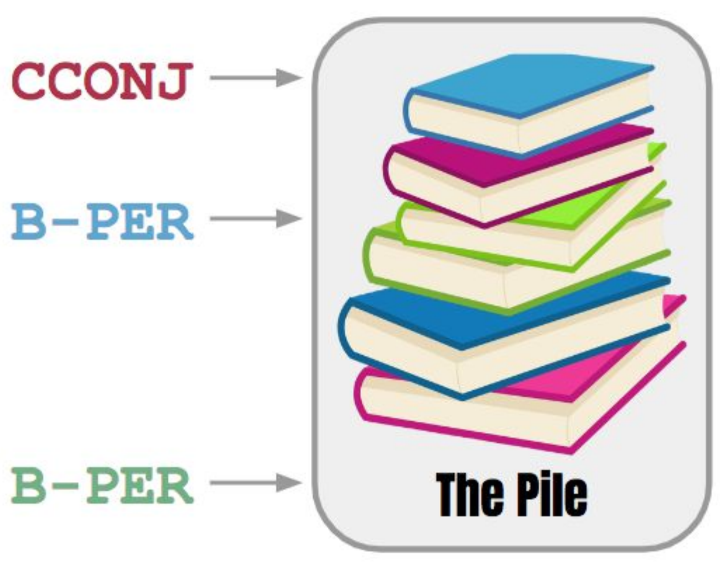



## How does label choice affect structured prompting?

| | The | cat | is | a | ... |
|---|---|---|---|---|---|
| Original Labels | DET | NOUN | AUX | DET | |
| Shuffled Labels | PUNCT | ADV | PROPN | PUNCT | |
| Proxy Labels | 13 | 18 | 26 | 13 | |
| Words as Labels | determiner | noun | auxiliary | determiner | |


## How does the pretraining data affect structured prompting?

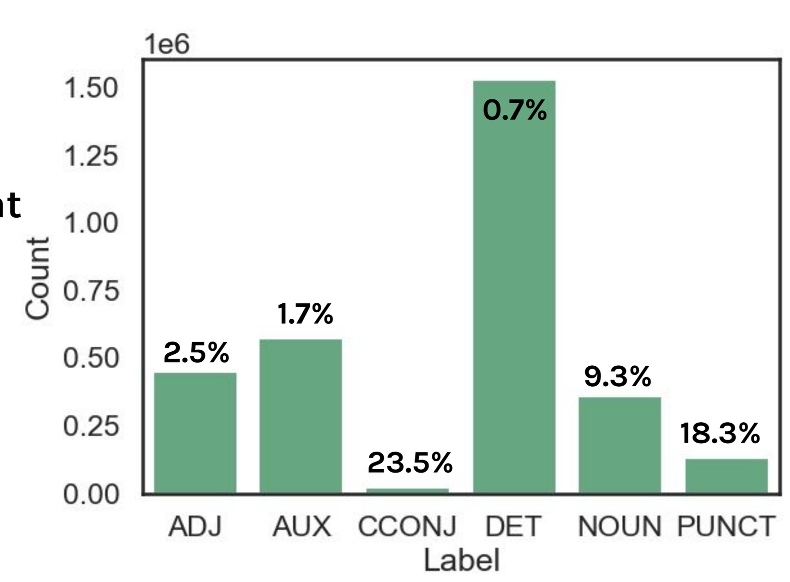



- **Labeled data:** `13 \t und \t und \t CCONJ \t KON \t _ \t 14 \t cc`
- **Task descriptions:** `*I-PER* label usually follows *B-PER* and *I-PER*, but it cannot follow *B-ORG* or *I-ORG*.`
- **Unrelated contexts:** `Bacterial pellets were lysed in 10 ml B-PER Bacterial Protein Extraction Reagent`

### POS Label Distribution in the Pile

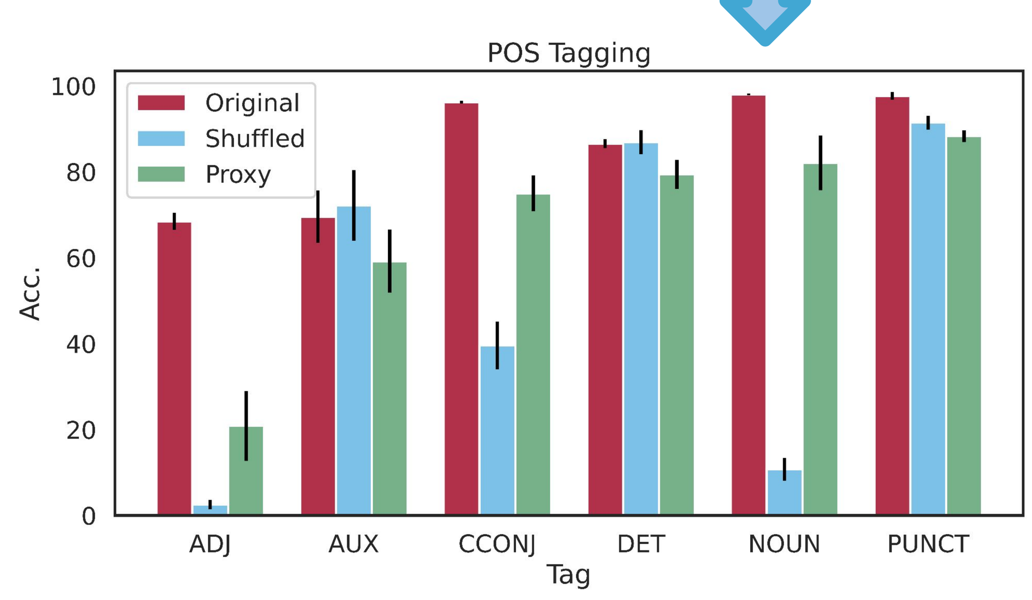
- Many occurrences in labeled UD format
- % of contexts that are task data varies across labels

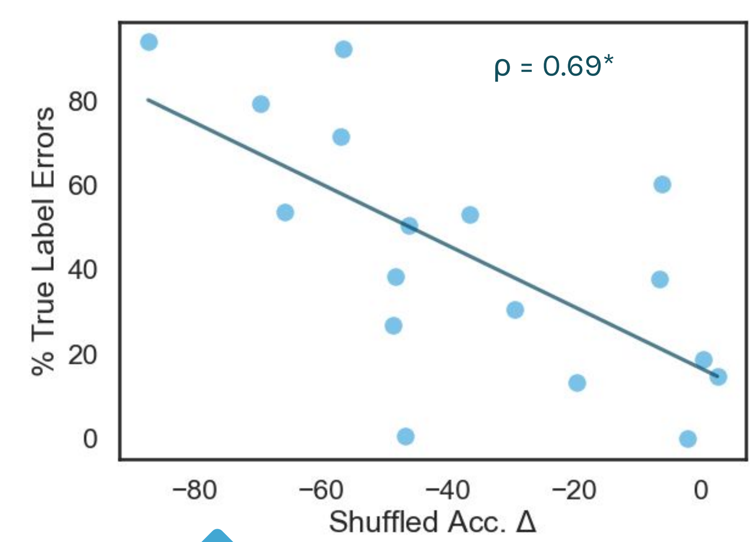
% = fraction of contexts in UD format




## PLMs rely on prior label knowledge for ICL, causing degraded performance in Shuffled setting

61.4% of errors in Shuffled setting are from predicting the Original label instead




Strong ρ in % of true label predictions and shuffled setting POS accuracy Δ

blvns.github.io

@terrablvns

Check out the paper!