Joint Parsing, Named Entity Recognition, (and Semantic Role Labeling)

Jenny Finkel
(joint work with Chris Manning)
GALE Banks Workshop
February 24, 2009

Motivation

- Core NLP tasks: part of speech tagging, named entity recognition, parsing, word sense disambiguation, semantic role labeling, coreference resolution

- Higher level NLP tasks: question answering, machine translation, textual entailment, semantics

- Goal: to bridge this divide

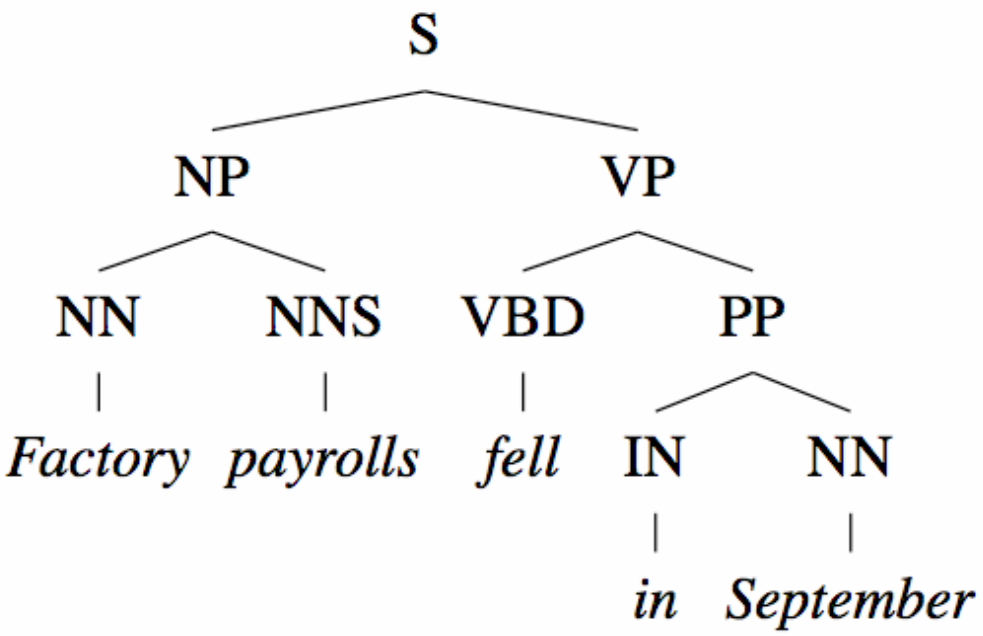
Motivation

- Different types of annotations should be able to help each other.
  - Named entities should correspond to constituents in the parse (or at least not have crossing brackets)
  - Named entities have distinct internal structure
  - Named entities are more likely to appear in certain structural contexts.
  - Named entities are more likely to be semantic arguments. Different kinds of entities are more likely to be different kinds of arguments
  - Semantic structure and syntactic structure are highly correlated and should be able to inform one another

Talk Outline

1. Motivation
2. Discriminative Parsing
3. Adding Named Entity Recognition
4. Adding Semantic Role Labeling (work in progress)
5. Conclusions / Future Work

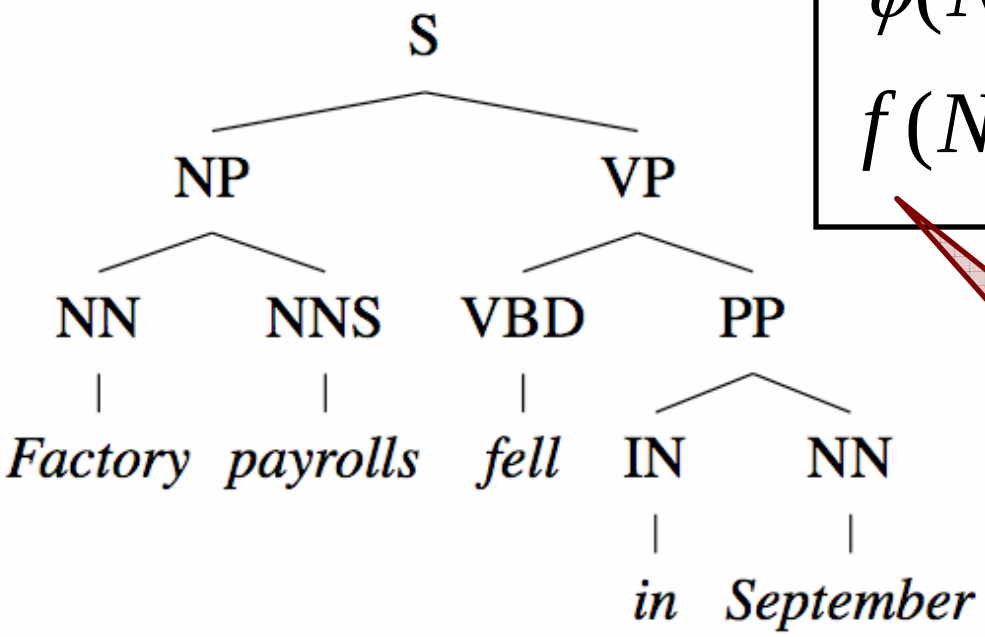
Discriminative Parsing



Generative

$$P(NP \rightarrow NN\ \ NNS) = \frac{\#(NP \rightarrow NN\ \ NNS)}{\#(NP)}$$

Discriminative Parsing



Discriminative

$$\phi(NP_{0,2} \rightarrow NN_{0,1} \;\; NNS_{1,2}) =$$

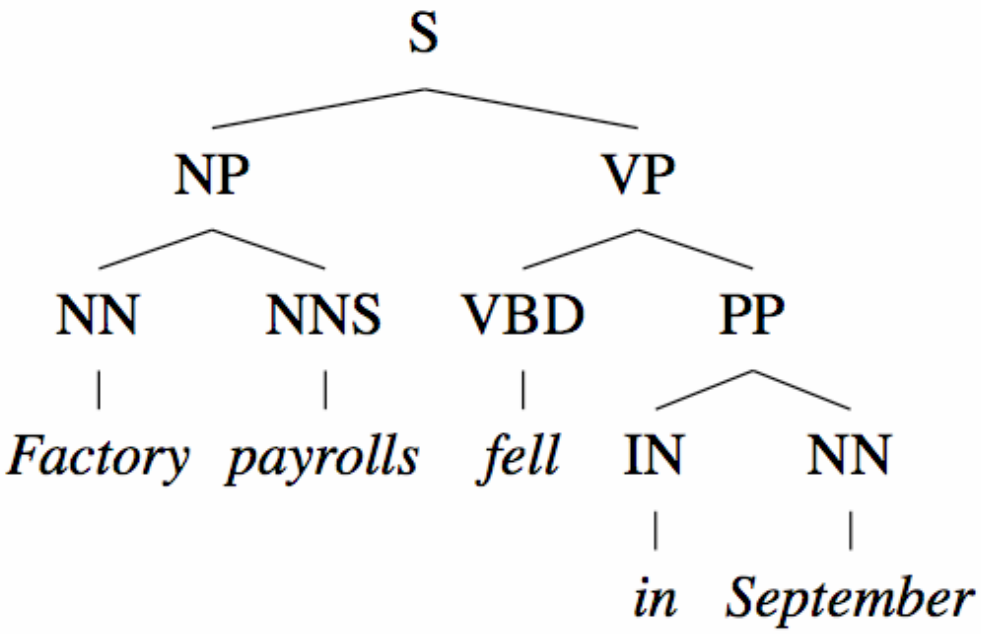
$$f(NP_{0,2} \rightarrow NN_{0,1} \;\; NNS_{1,2}) \cdot \theta$$

feature function

feature weights

Intuition: analogous to changing an HMM into a linear-chain CRF

Discriminative Parsing



$$f(NP_{0,2} \rightarrow NN_{0,1} \;\; NNS_{1,2})$$

label: NP
rule: NP → NN NNS
words: NP/factory
words: NP/payroll
dist sim tags: NP/132
dist sim split: NP/132/45

$$f(NN_{4,5} \rightarrow September)$$

label/word: NN/September
label/dist sim: NN/24
prev word: NN/in

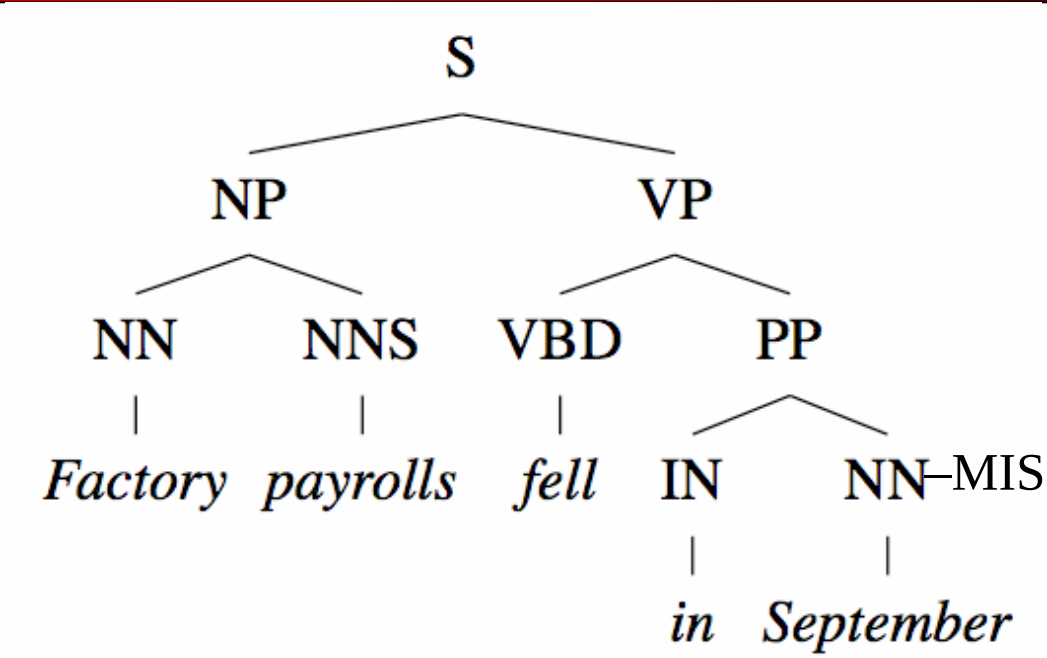
Talk Outline

1. Motivation
2. Discriminative Parsing
3. Adding Named Entity Recognition
4. Adding Semantic Role Labeling (work in progress)
5. Conclusions / Future Work

NER data

- Everything reported is with OntoNotes 2.0
- Only worked with the English data
- Collapsed the entity types down to *person*, *GPE*, *organization*, *misc*
- Made some small changes to the data for the sake of consistency – Chris will talk about this

The Basic Idea

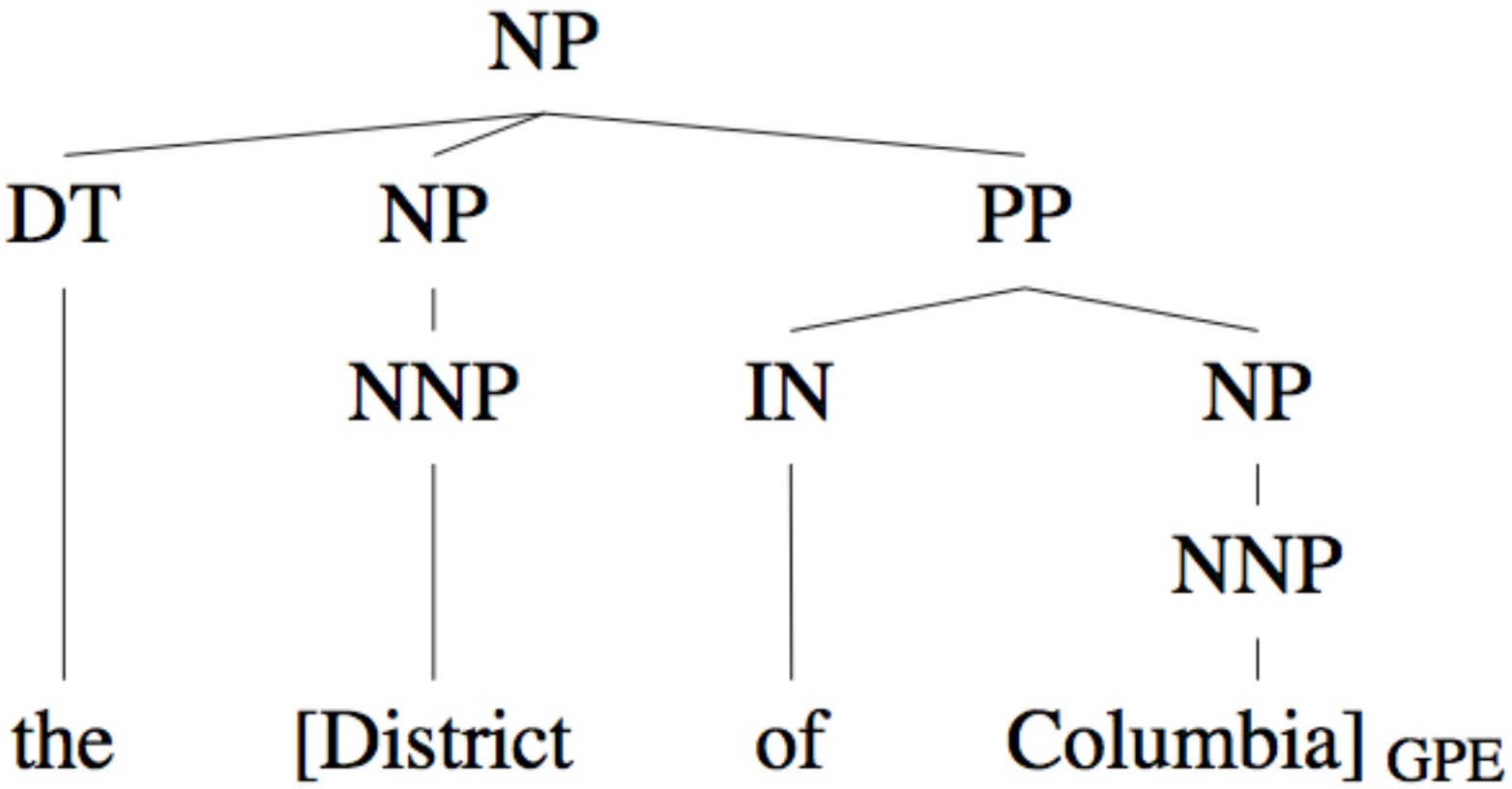


$$f_{parse}(PP_{3,5} \rightarrow IN_{3,4} \;\; NN_{4,5})$$

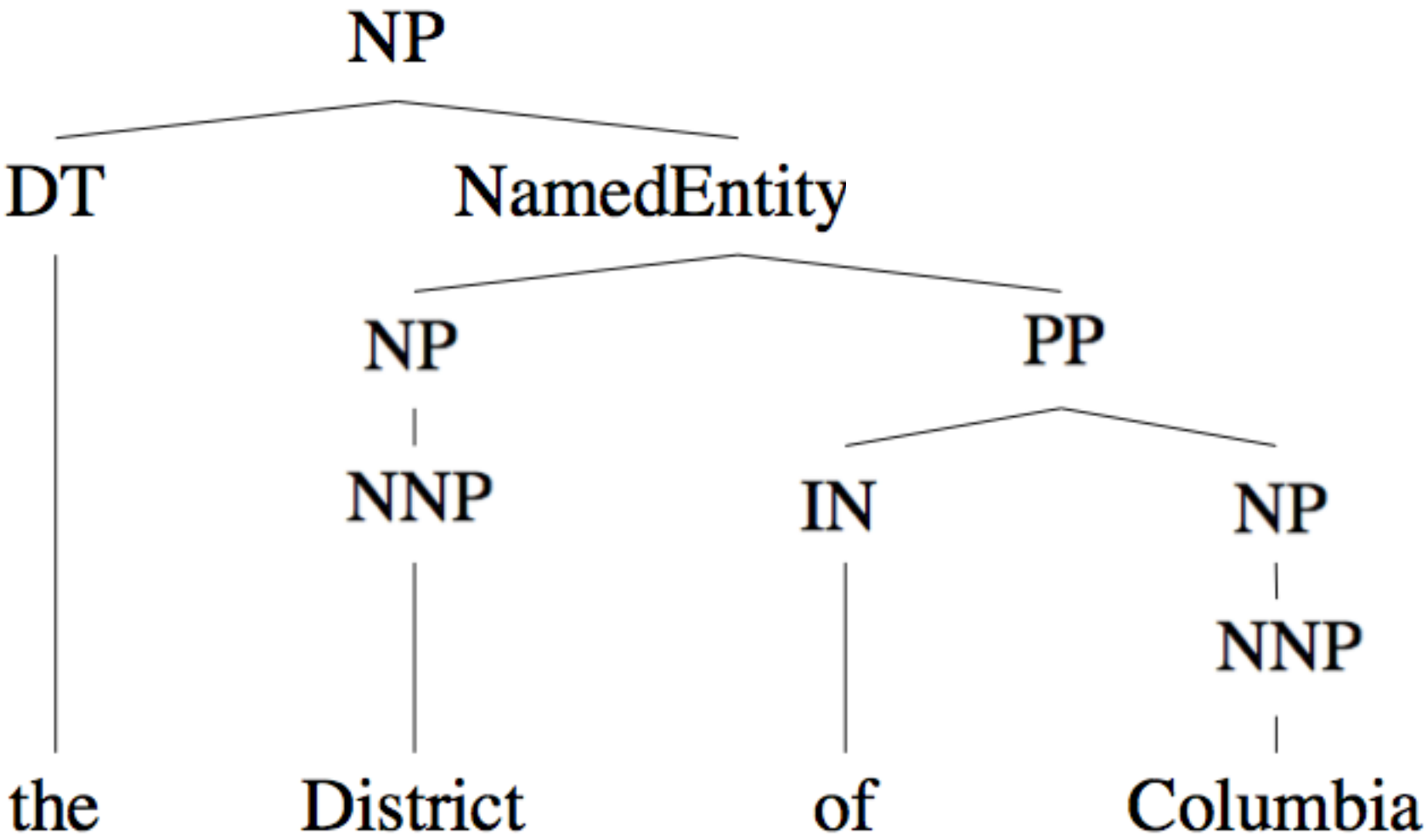
$$f_{NER}(O_{3,5} \rightarrow O_{3,4} \;\; MISC_{4,5})$$

$$f_{parse/NER}(PP_{3,5} \rightarrow IN_{3,4} \;\; NN - MISC_{4,5})$$

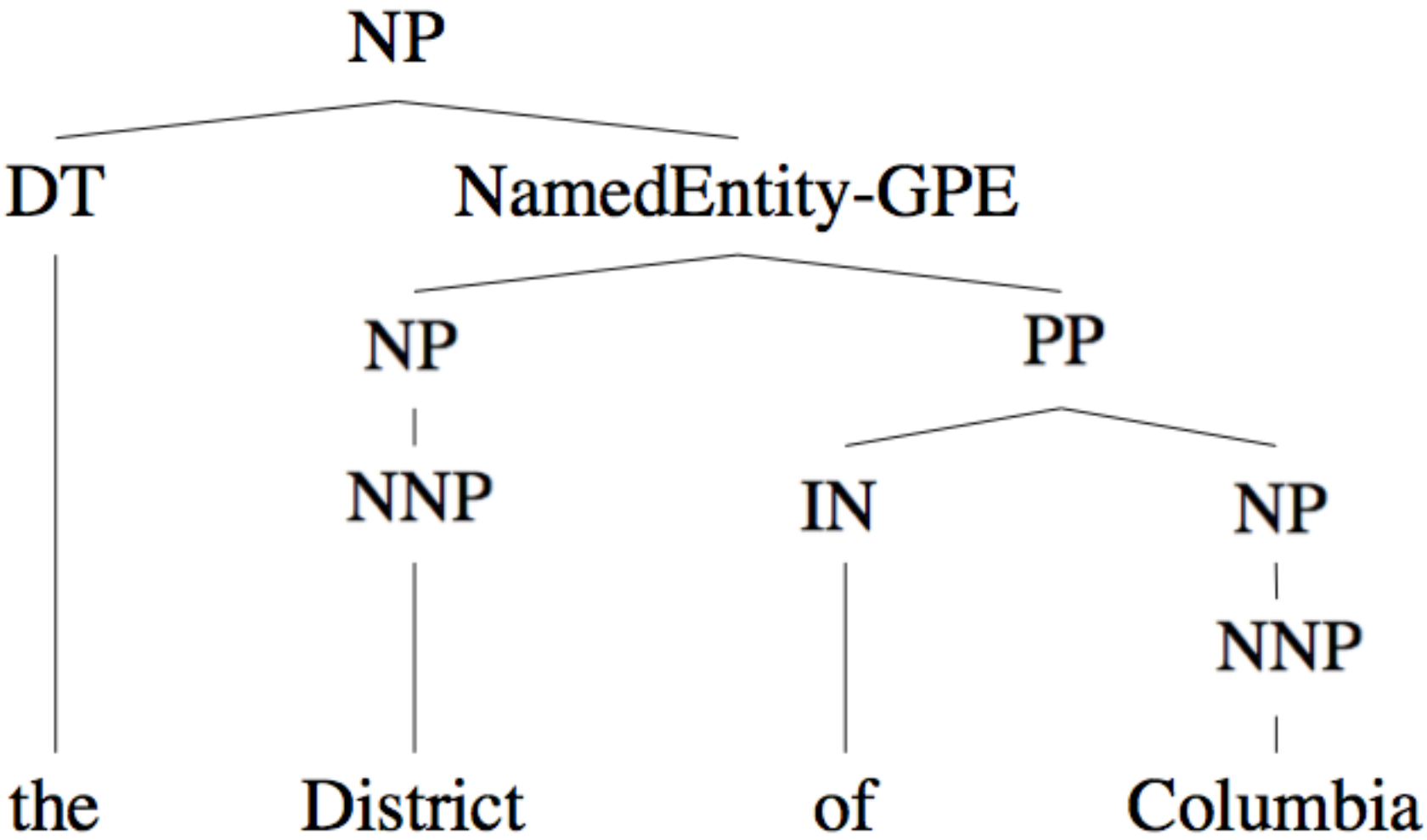
Joint Representation



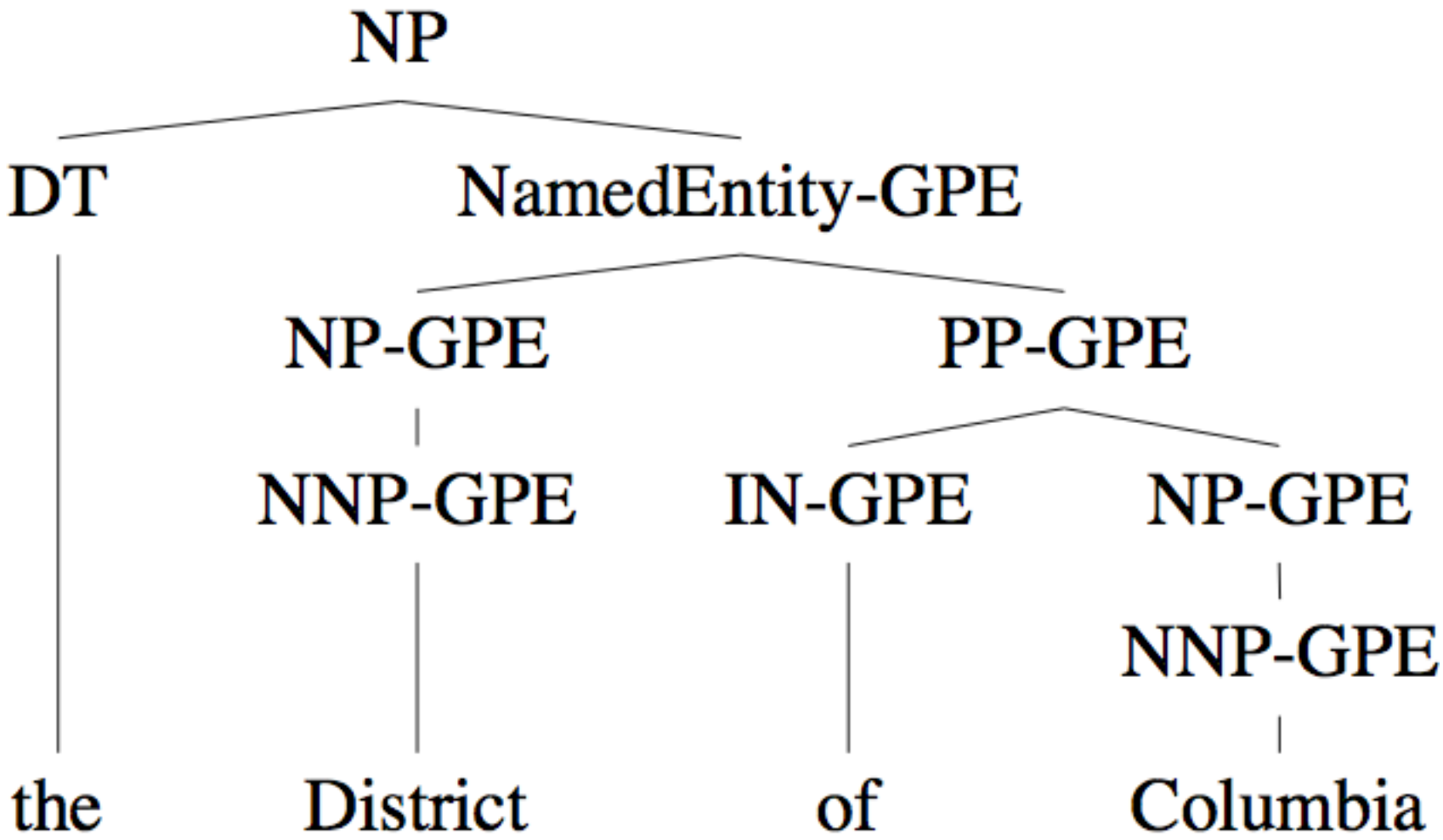
Joint Representation



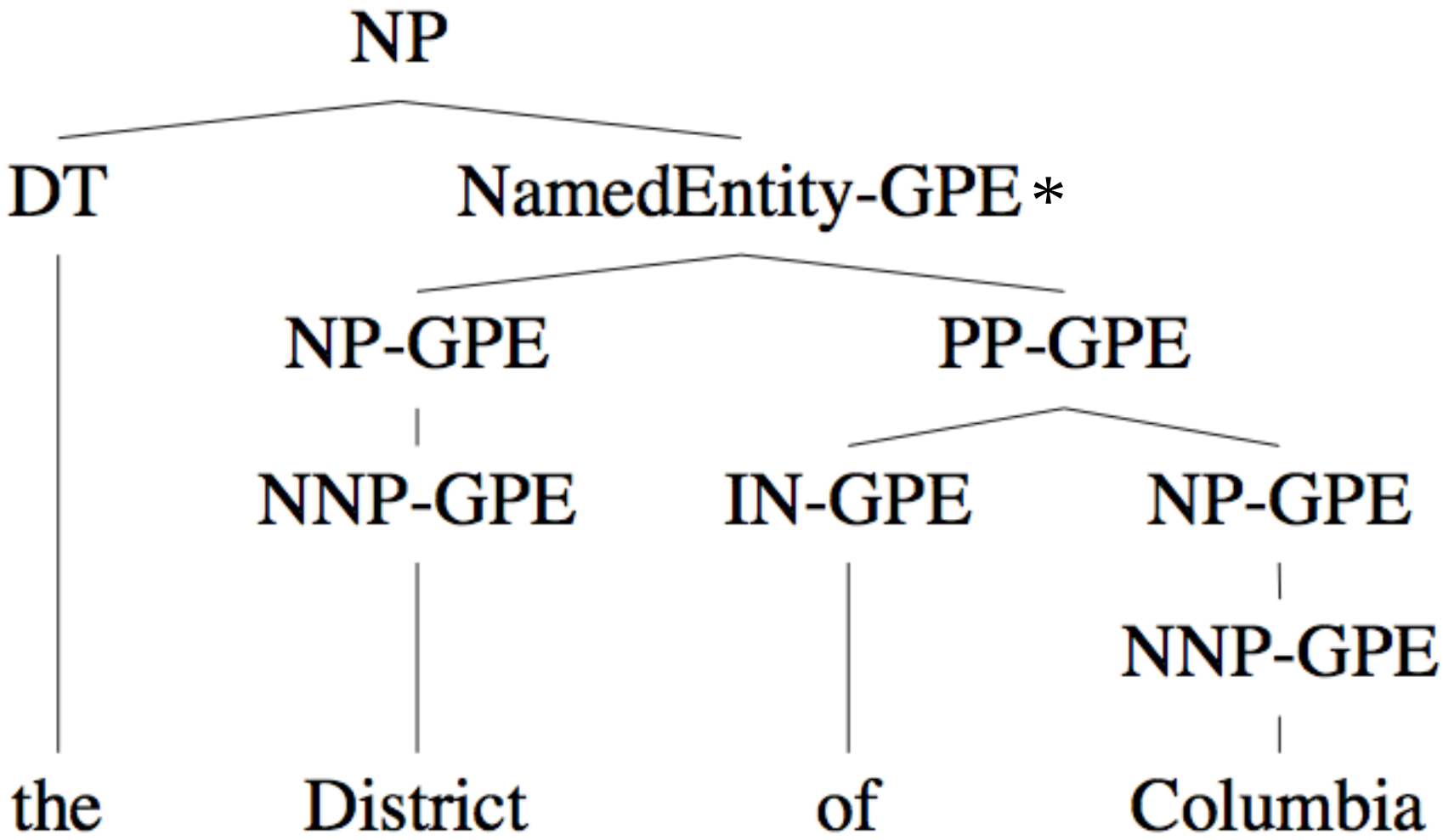
Joint Representation



Joint Representation



Joint Representation



Grammar Smoothing

- Add rules, stripped of all entities
- If a rule has been seen with one type of entity, duplicate it with all kinds of entities
- If a rule has an NP for a child, allow that NP to be any kind of entity

- We tried adding other ones, but they mostly slowed it down without improving performance

Features

- *Parse features* – same as baseline parser
- *NER features* – same as baseline CRF named entity recognizer
- *Joint features* -

Joint parse/NER Results (F-score)

	ABC	CNN	MNB	NBC	PRI	VOA
Just NER	80.23	83.18	75.50	77.97	85.03	91.65
Joint Model NER	80.30	87.37	81.21	81.03	86.05	91.94
Just Parse	70.6	77.61	64.66	62.1	74.71	77.41
Joint Model Parse	70.71	78.79	67.26	62.79	76.49	79.38
# Train Sent	1099	5220	592	555	1645	1485
# Test Sent	295	1393	162	146	456	410
Parse Train Time	25 m	16.5 h	12 m	10 m	2.4 h	2.3 h
Joint Train Time	44 m	31.7 h	19 m	17 m	4.2 h	4.4 h

Talk Outline

1. Motivation
2. Discriminative Parsing
3. Adding Named Entity Recognition
4. Adding Semantic Role Labeling (work in progress)
5. Conclusions / Future Work

Conclusions

- Jointly learning multiple levels of annotation can help all levels (at least with parsing and NER)
- Picking the correct model structure is key
  - Too simple and you can't capture important aspects of the structure
  - Too complex and it becomes too expensive to do inference
  - Can be beneficial to restrict the model
- Separate and Joint features
  - Use all of your familiar single-task features
  - Add multi-task features to capture interactions between the different types of annotation

Future Work

- Finish getting SRL to work
- Multiple predicates at once
- Adding in the other levels of annotation in OntoNotes
- Require moving from exact inference to approximate inference
- Adding long distance dependencies for coreference resolution