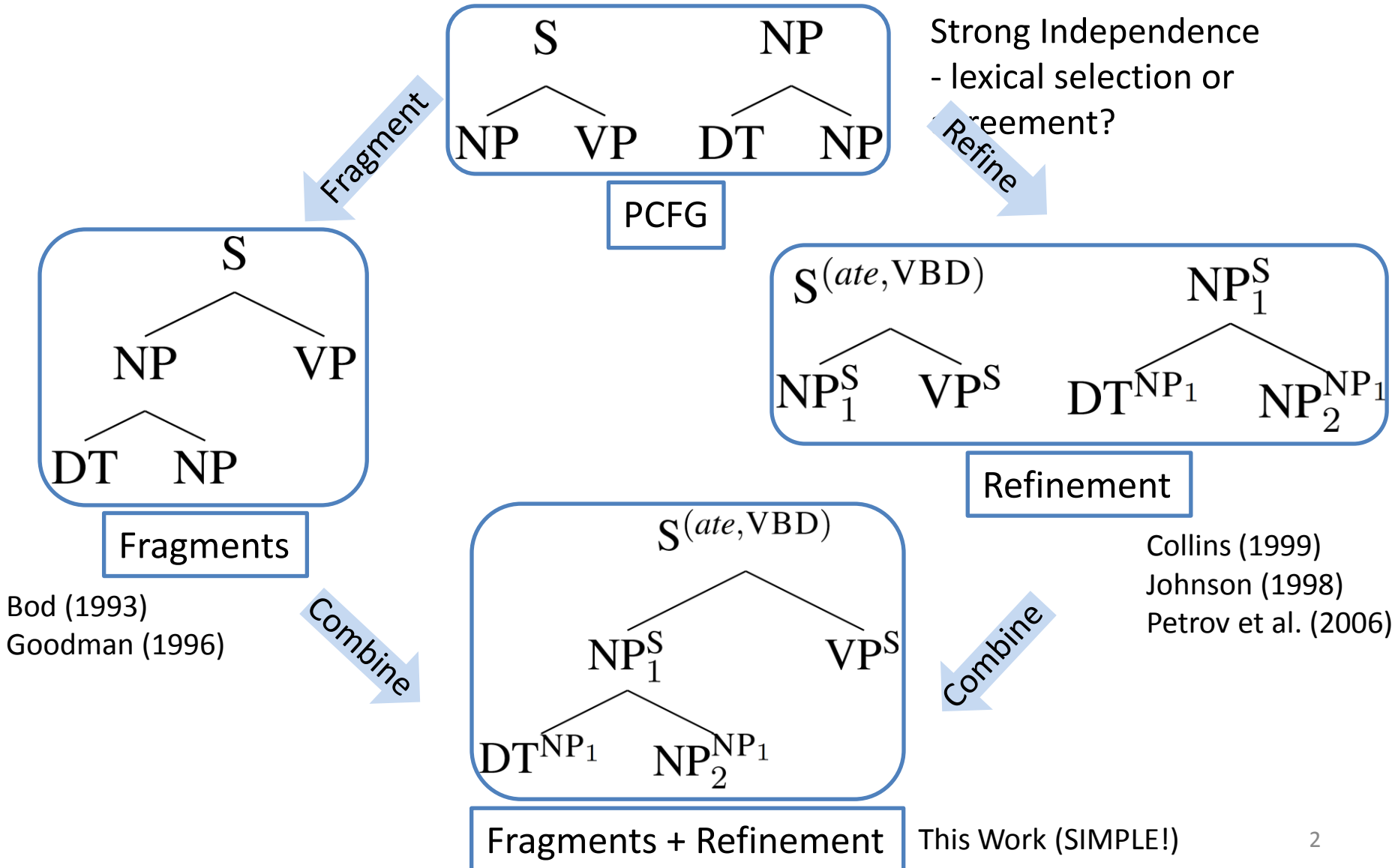


Simple, Accurate Parsing with an All-Fragments Grammar



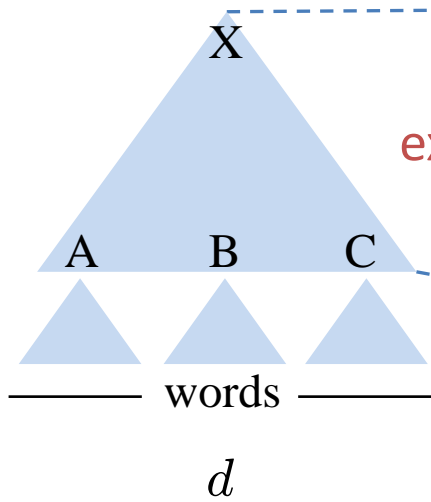
Mohit Bansal and Dan Klein
UC Berkeley

Independence Assumptions of PCFG



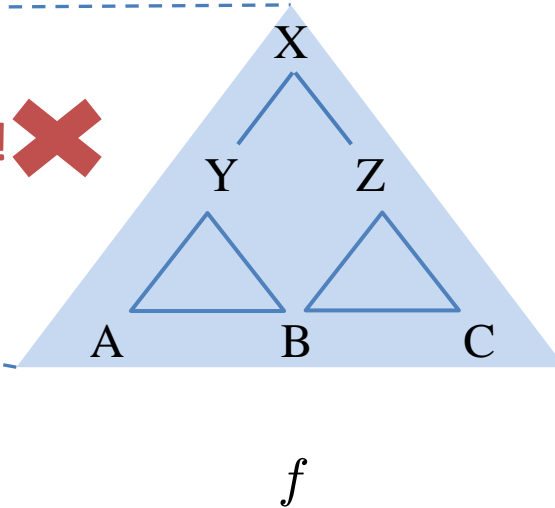
All-Fragments Grammar G

DERIVATIONS



exponential # of rules!! **✘**

FRAGMENTS

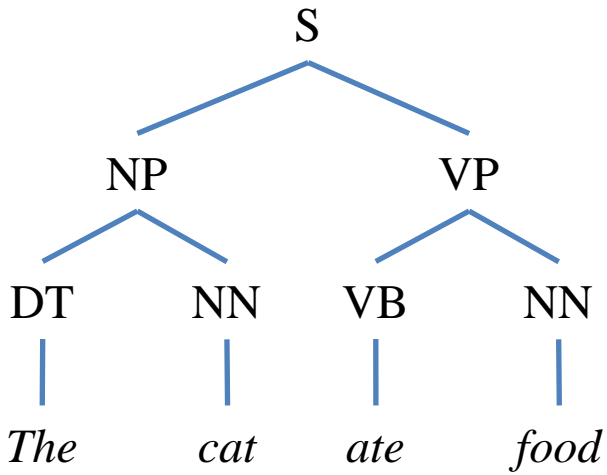


$$\omega(d) = \prod_{f \in d} \omega(f)$$

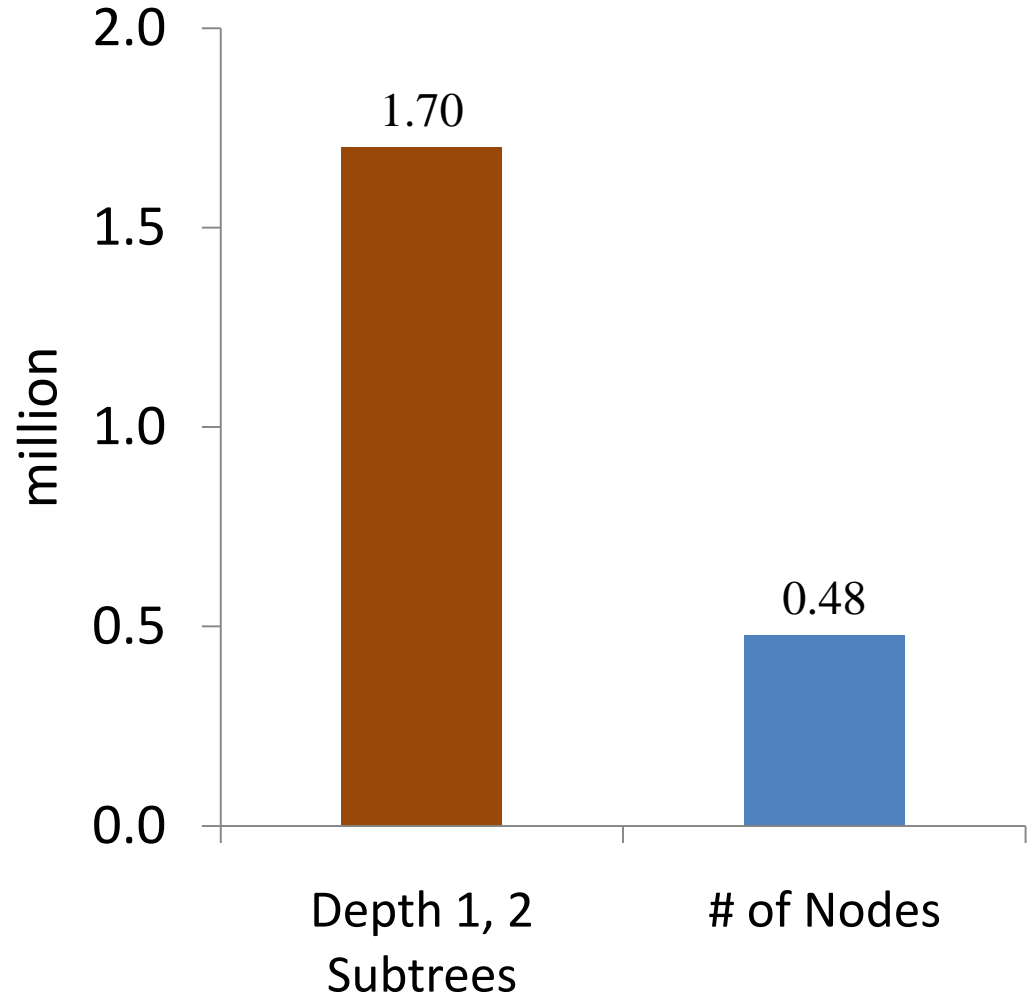
$$t_{max} = \operatorname{argmax}_t \sum_{d \in t} \omega(d)$$



Fragment (data-oriented) Approach



of fragments = exponential
in length of sentence!!





G^I - Implicit Representation of G

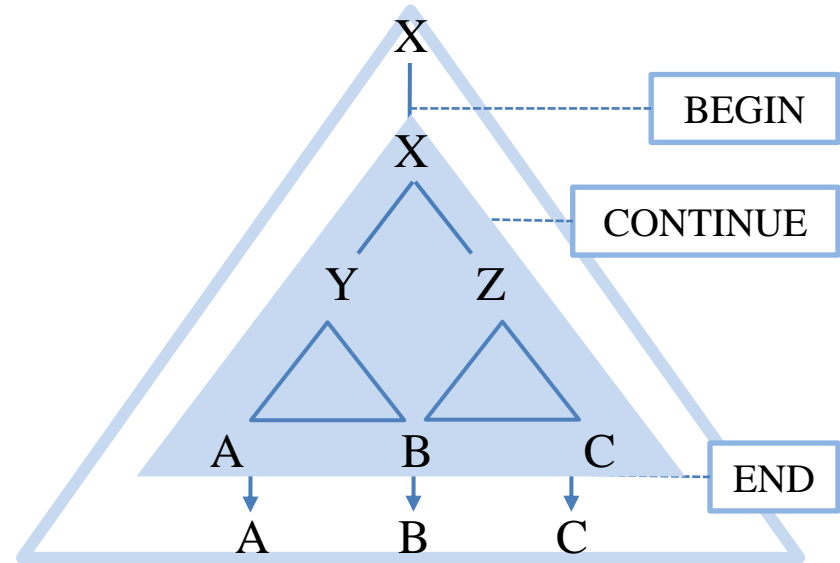
SYMBOLS:

- ▶ Base: X
- ▶ Indexed: X_i

RULES:

- ▶ Continue: $X_i \rightarrow Y_j Z_k$
- ▶ End: $X_i \rightarrow X$
- ▶ Begin: $X \rightarrow X_i$

FRAGMENTS



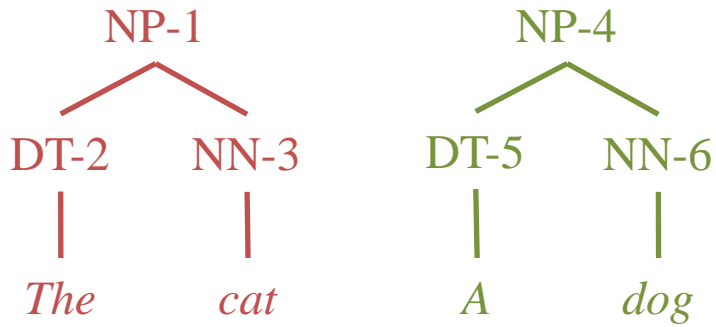
Goodman (1996)

of rules = |treebank B| ✓

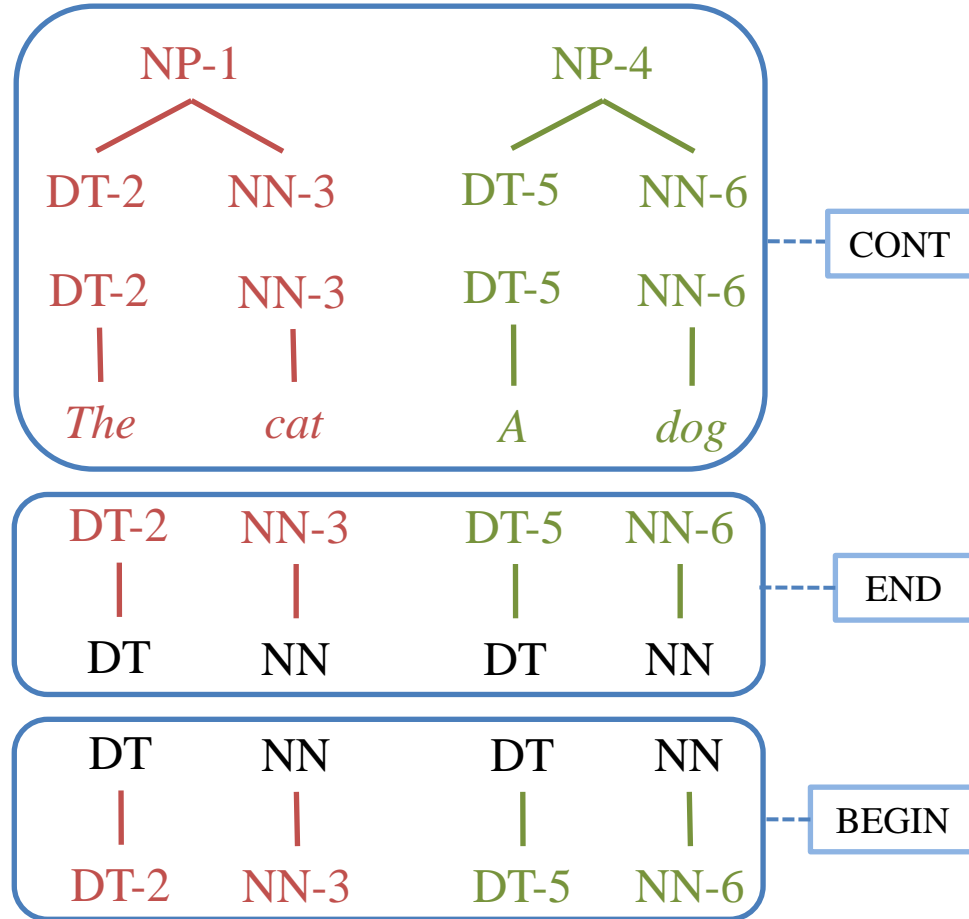


Example

Training Data

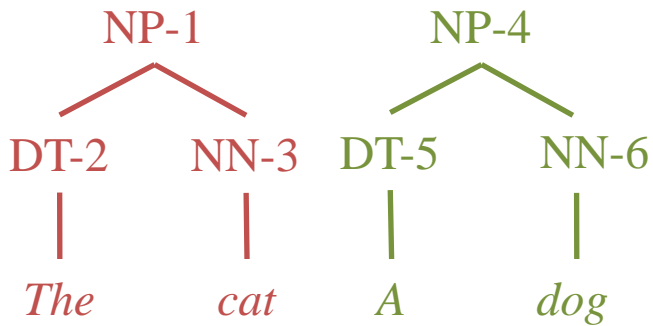


Rules



Parsing a Novel Test Sentence

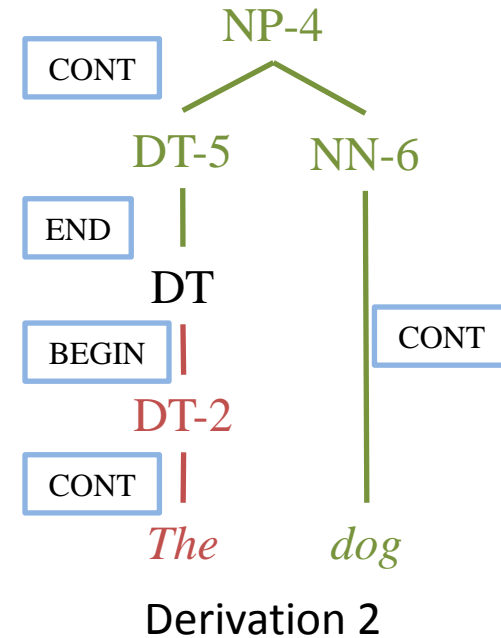
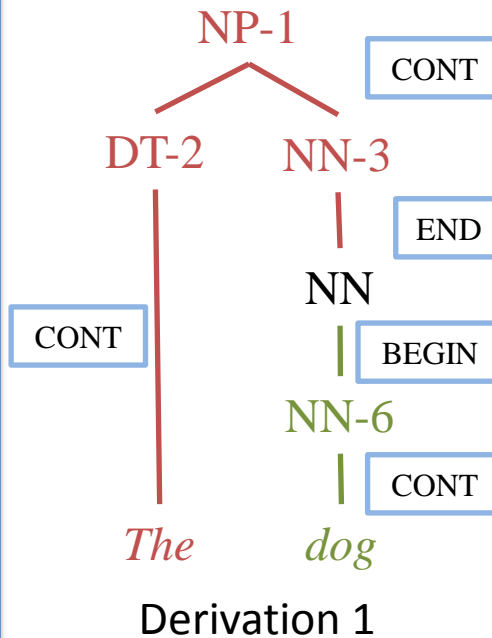
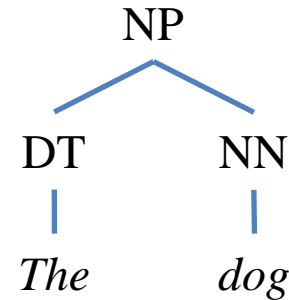
Training Data



Novel Test Sentence

The dog

Test Parse





Equivalence of G and G^I

- Each derivation d in G reproducible in G^I
- Multiple derivations in G^I correspond to same d in G , differing in indices



Weights for Implicit Grammar

RULES

WEIGHTS

▶ CONT: $X_i \rightarrow Y_j Z_k$

$\omega_{\text{BODY}} (\omega_{\text{LEX}})$

▶ END: $X_i \rightarrow X$

ω_{SWITCH}

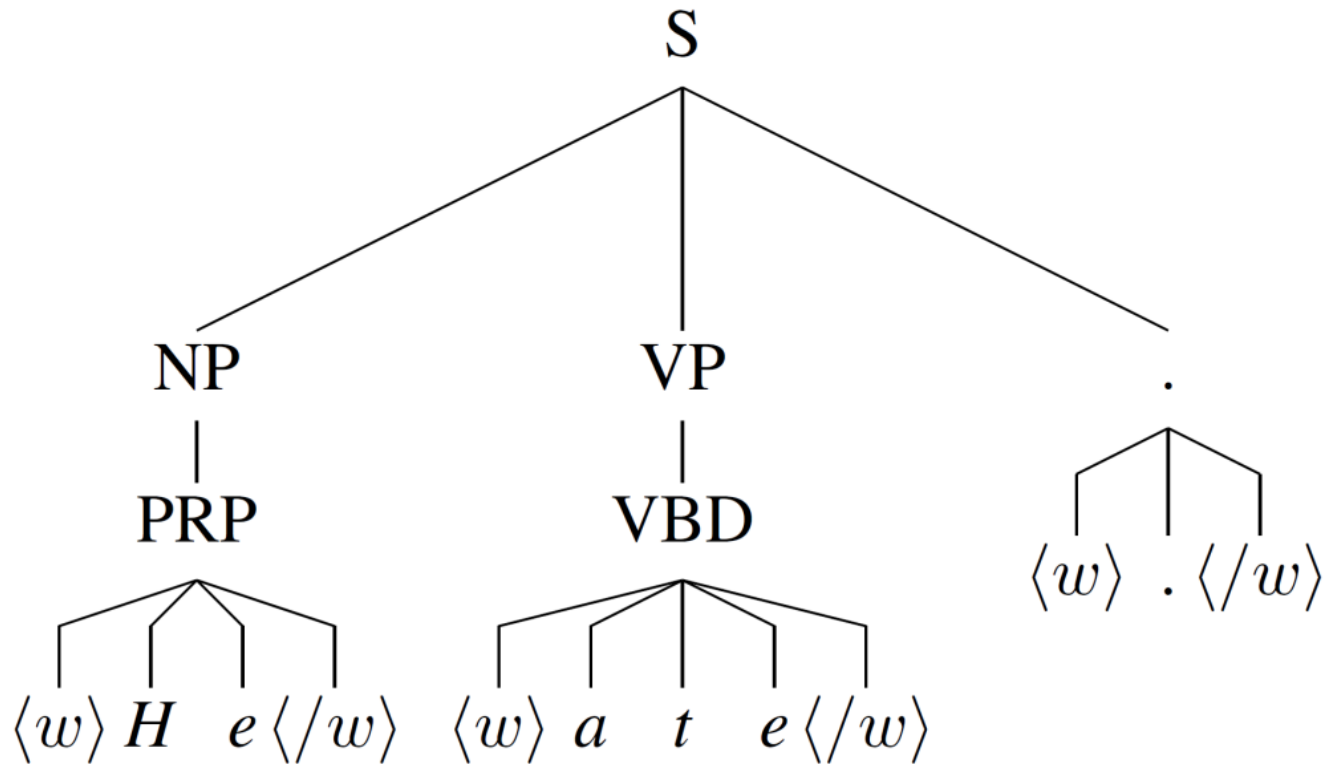
▶ BEGIN: $X \rightarrow X_i$

$\frac{1}{\# \text{ frags rooted at } X}$

JUST 3 PARAMETERS !



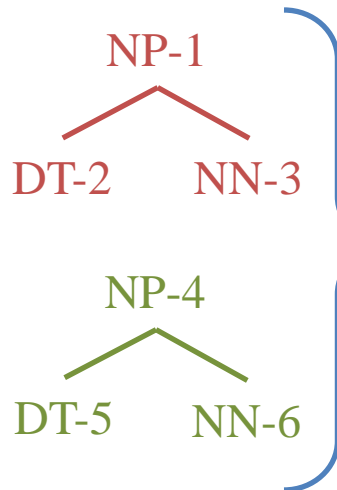
Character-level Parsing



words split into characters

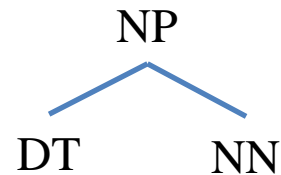
Coarse-to-Fine Inference

“Fine” Grammar



DT-2
|
The

“Coarse” Grammar



DT
|
The

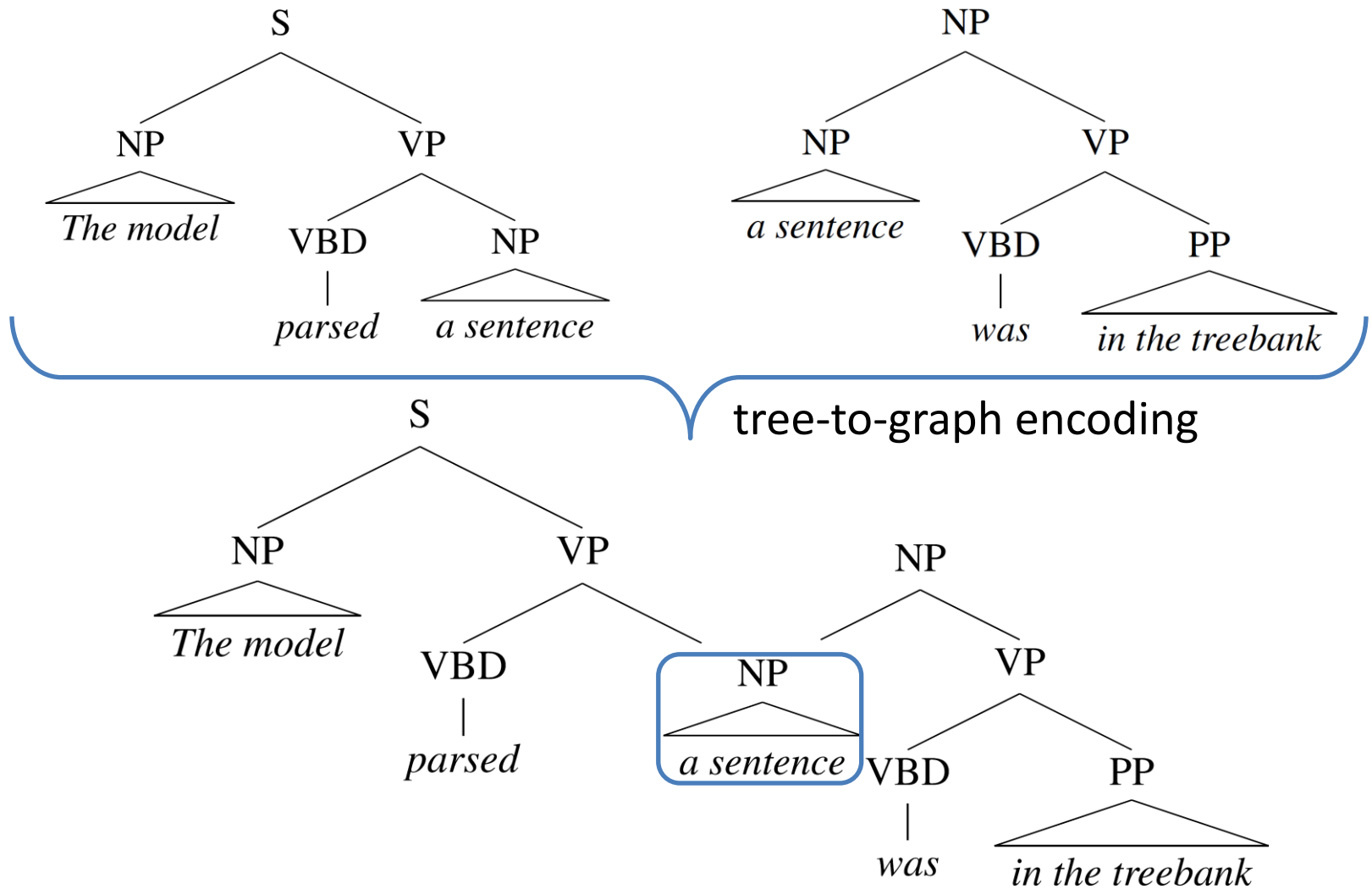
AVERAGE
OVER INDICES

For same accuracy,
– 40x speed up
– 10x memory reduction

PCFG



Packed Graph Encoding



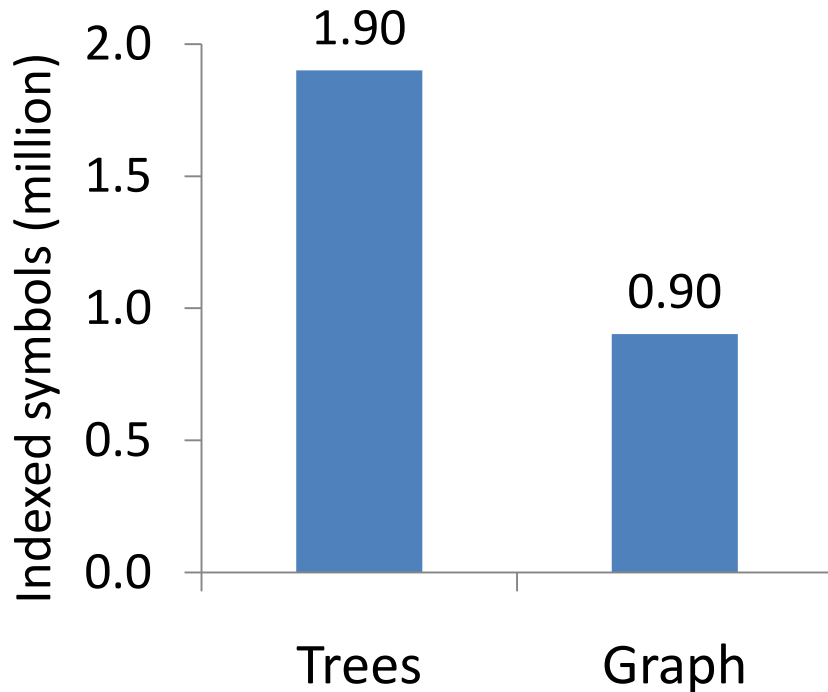


Savings from Packed Graph Encoding

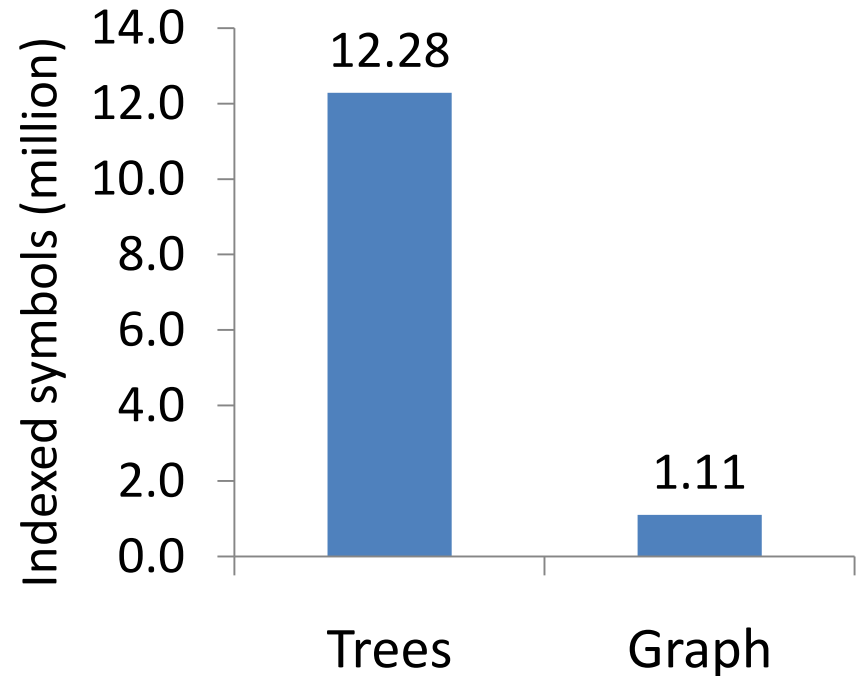
- 1.4x speed up
- memory-usage < 4GB

- 20x speed up
- memory-usage < 8GB

Word-level Parsing

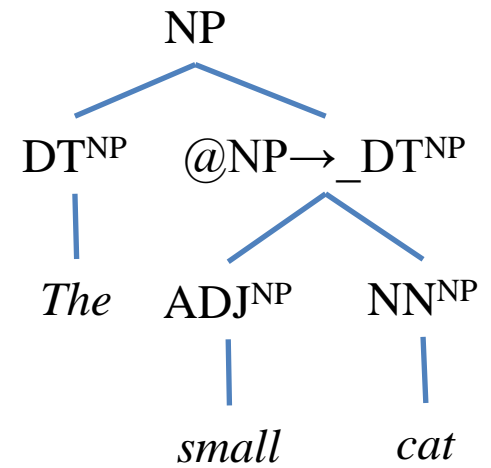
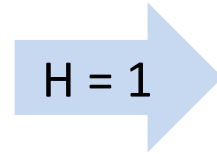
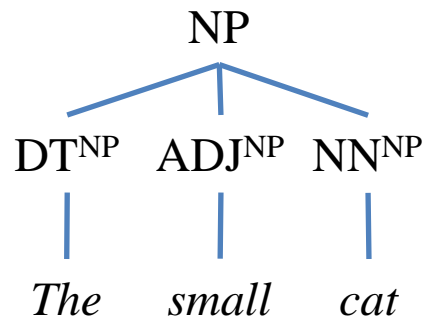
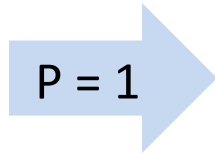
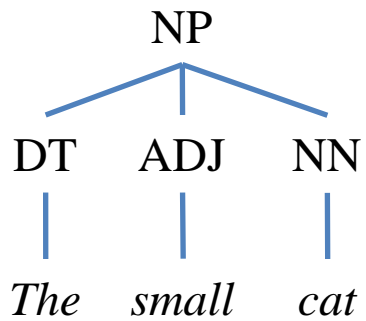


Character-level Parsing



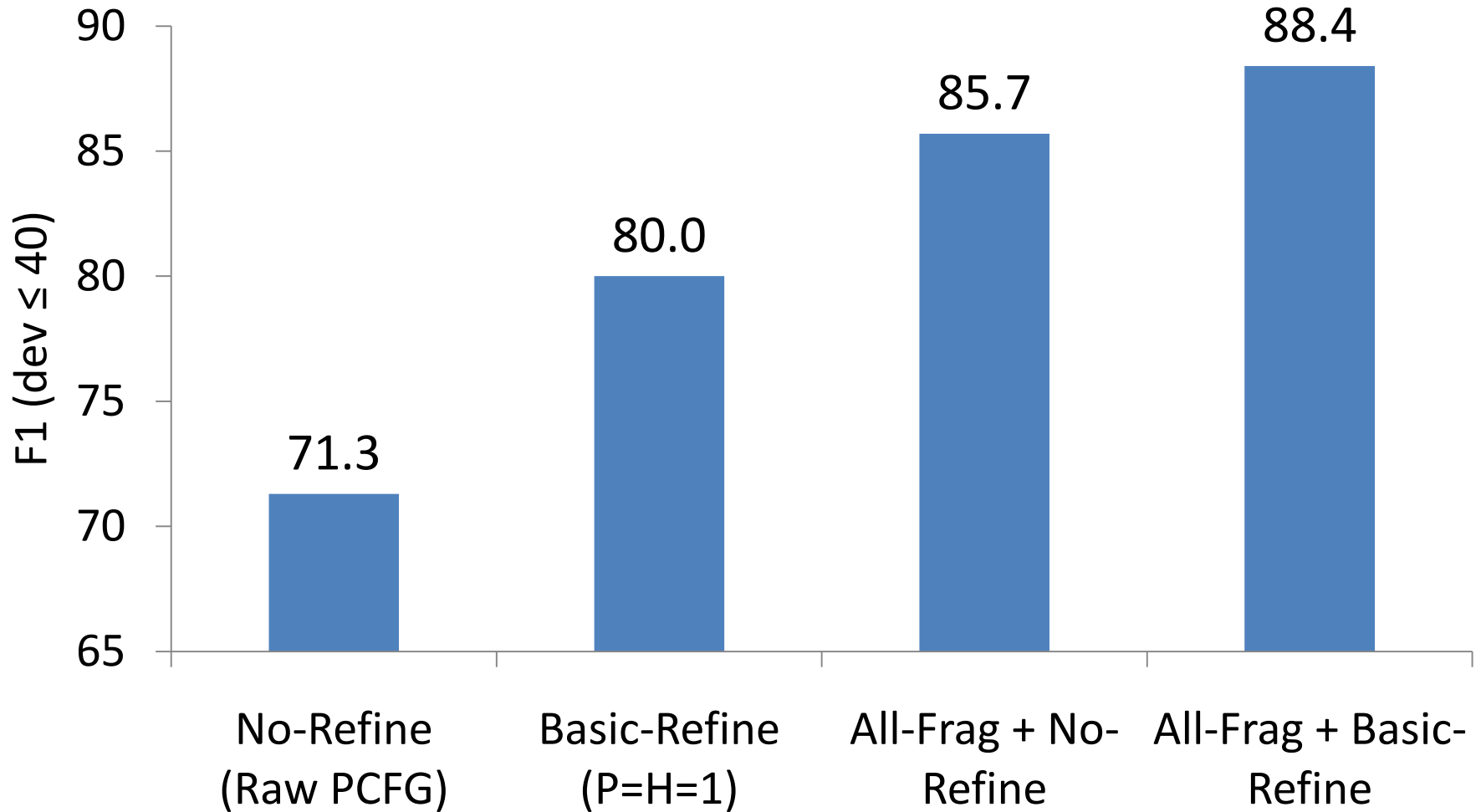


Basic Refinement





Fragments Complement Refinements





Parsing Accuracy

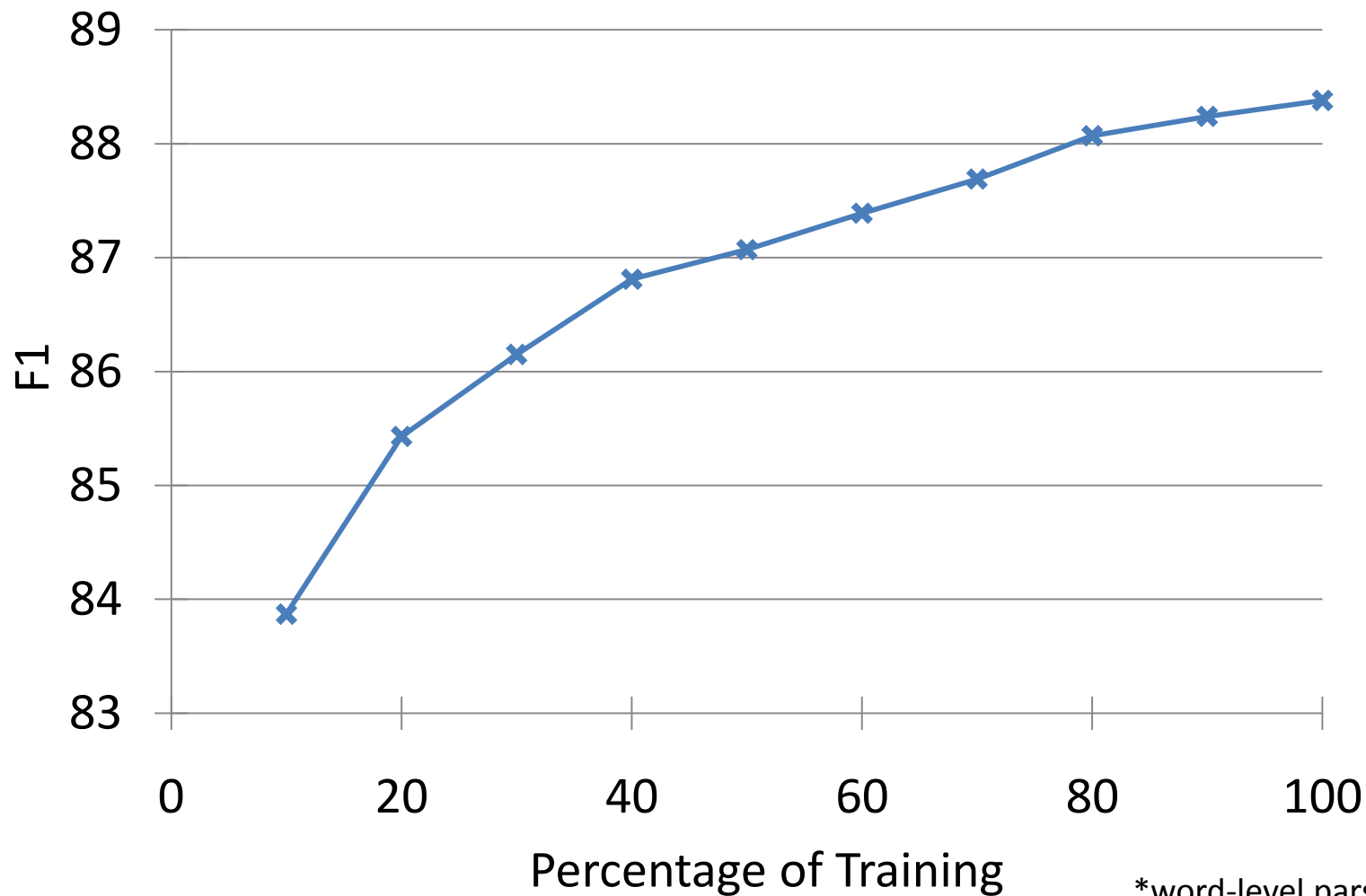
Word-level Parsing F1			
Decoding Objective	dev (≤ 40)	test (≤ 40)	test (all)
Max-Constituent ¹	88.4	88.5	87.6

Character-level Parsing F1			
Decoding Objective	dev (≤ 40)	test (≤ 40)	test (all)
Max-Constituent ¹	88.2	88.0	87.1

¹ Goodman (1996)



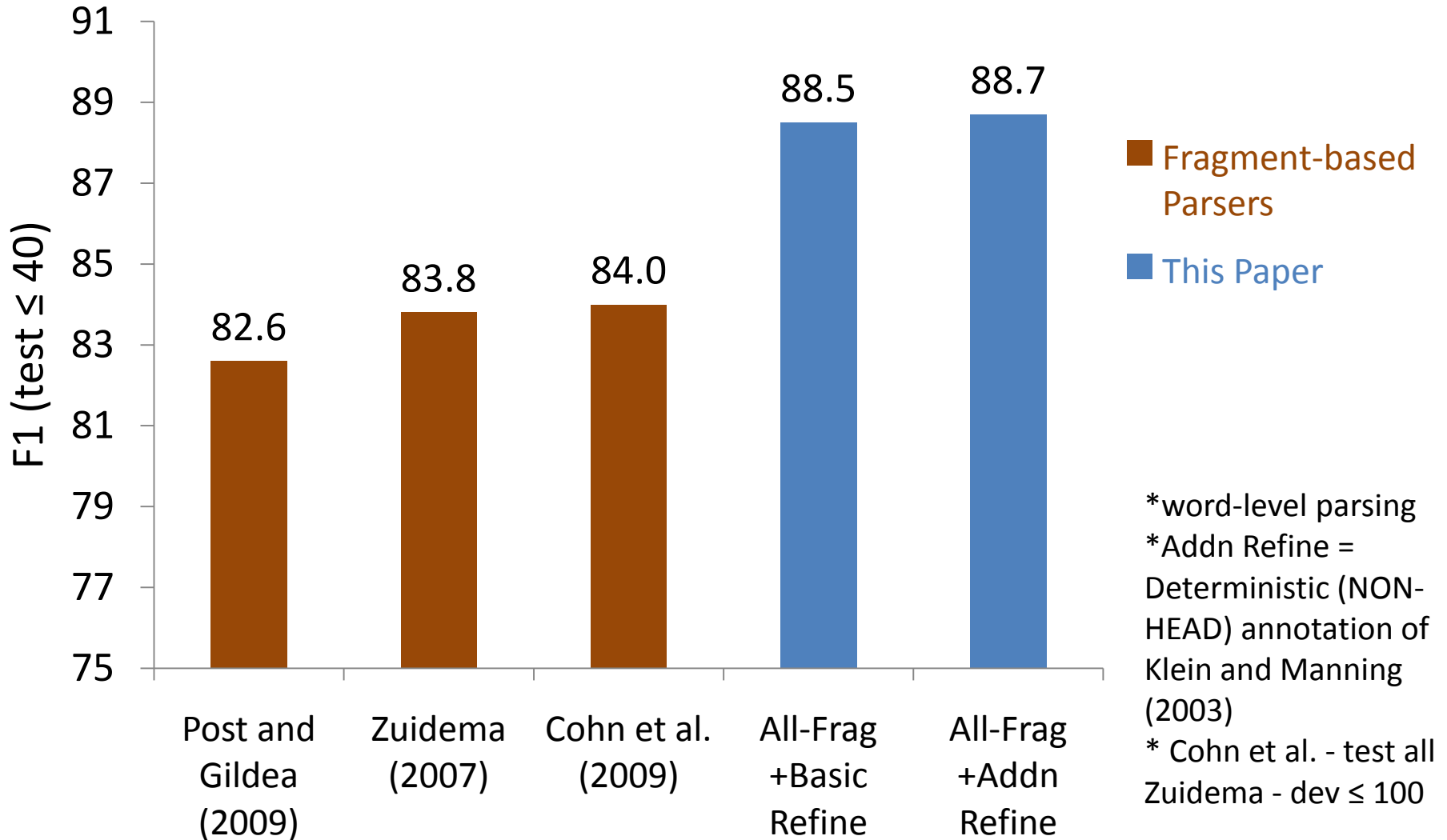
Full-scale Parsing



*word-level parsing results
on dev-set (≤ 40)

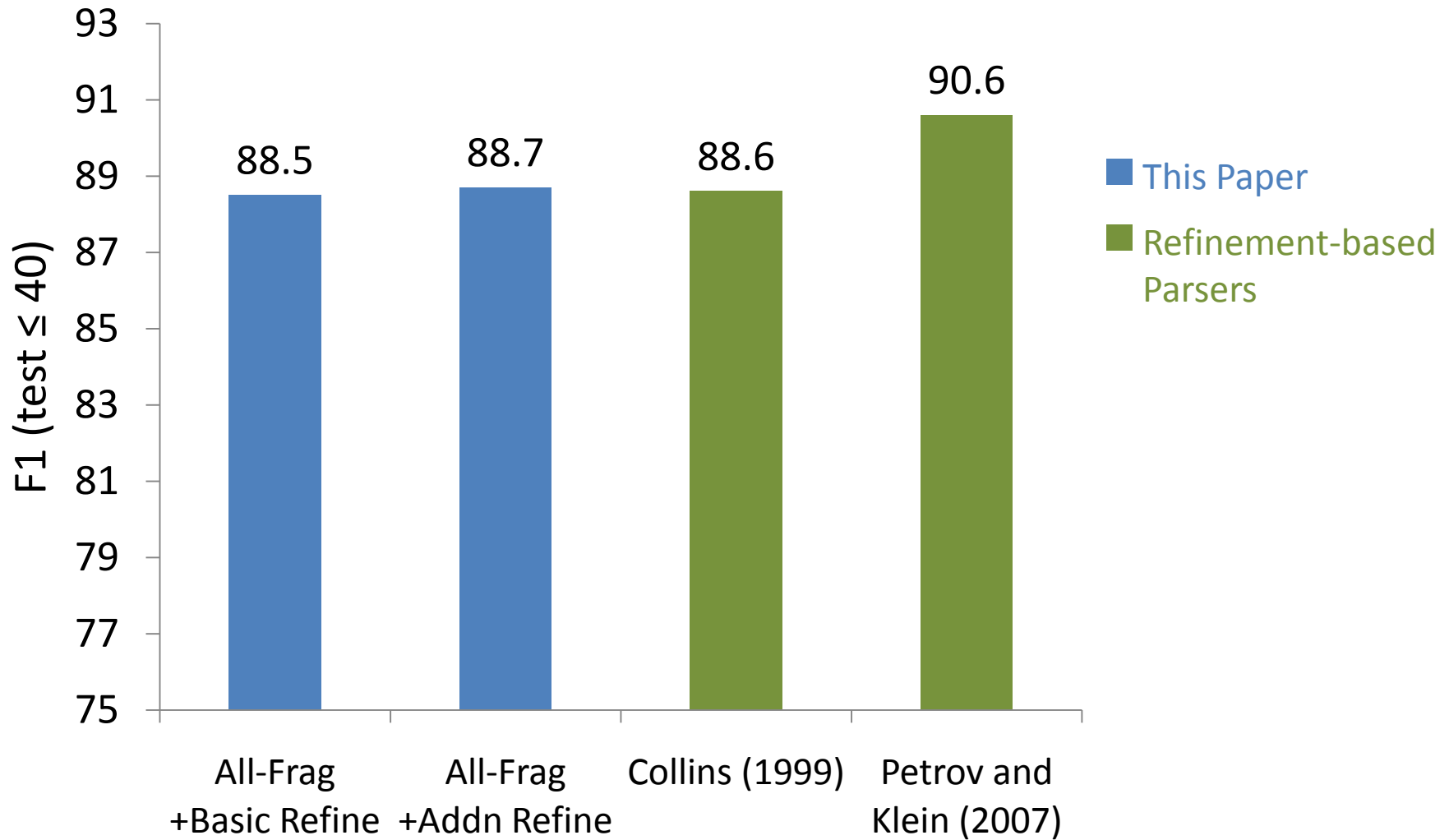


Final WSJ Results



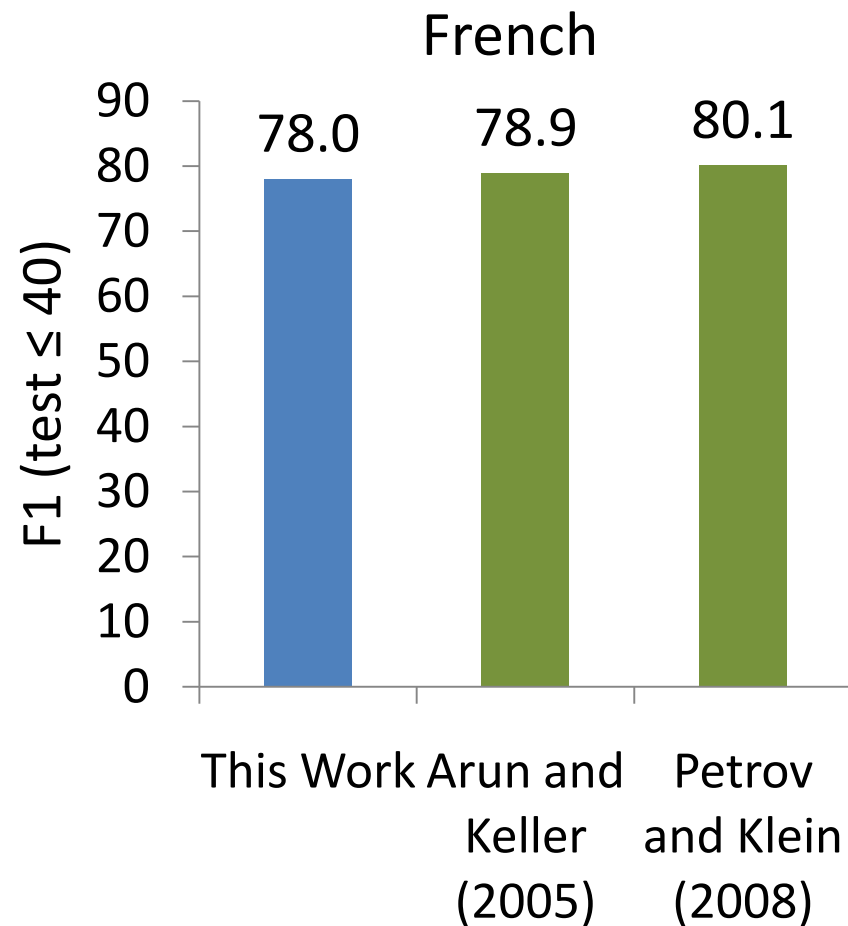
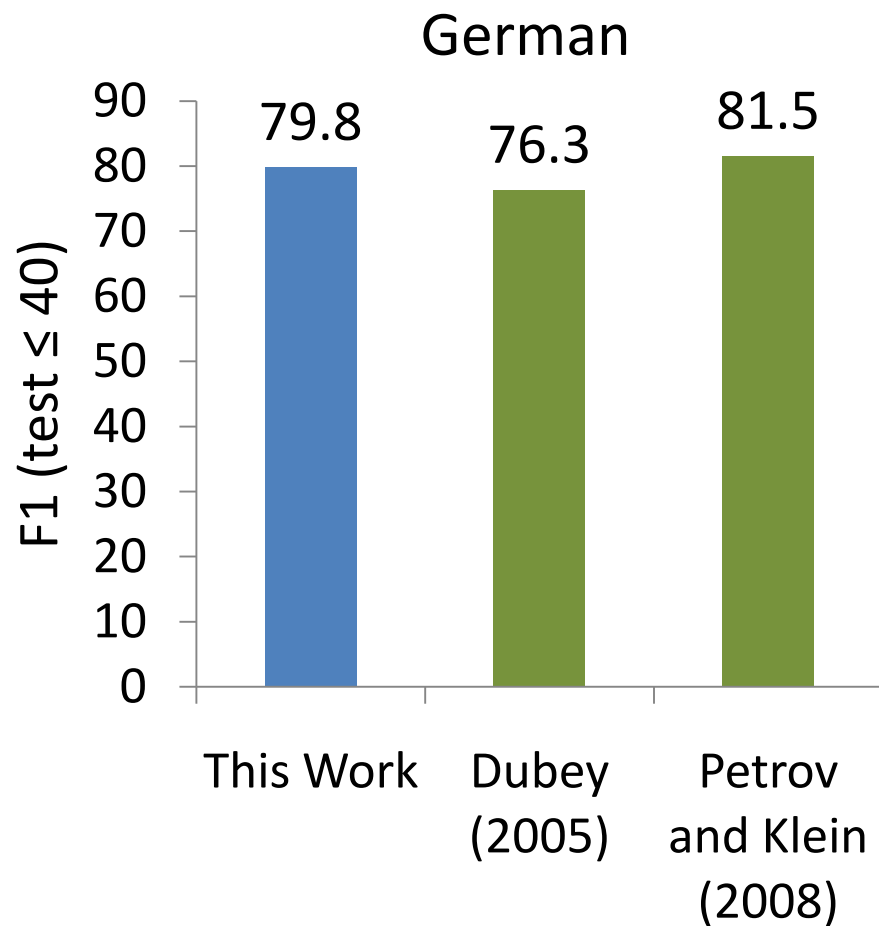


Final WSJ Results





Other Language Results





- Practical, full-scale parsing with an all fragments grammar
 - Indexed grammar boils down to only 2-3 hyperparameters
 - Practical with natural coarse-to-fine projections and graph encodings
- Fragments complement refinements
 - Simple refinement + fragments F1 \approx Collins 99
 - Accurate without an explicit lexicon
 - Zero training

Parsing Model	F1 (test \leq 40)	F1 (test all)
Collins (1999)	88.6	88.2
Our Model	88.7	88.1

Thank you!

Questions?

