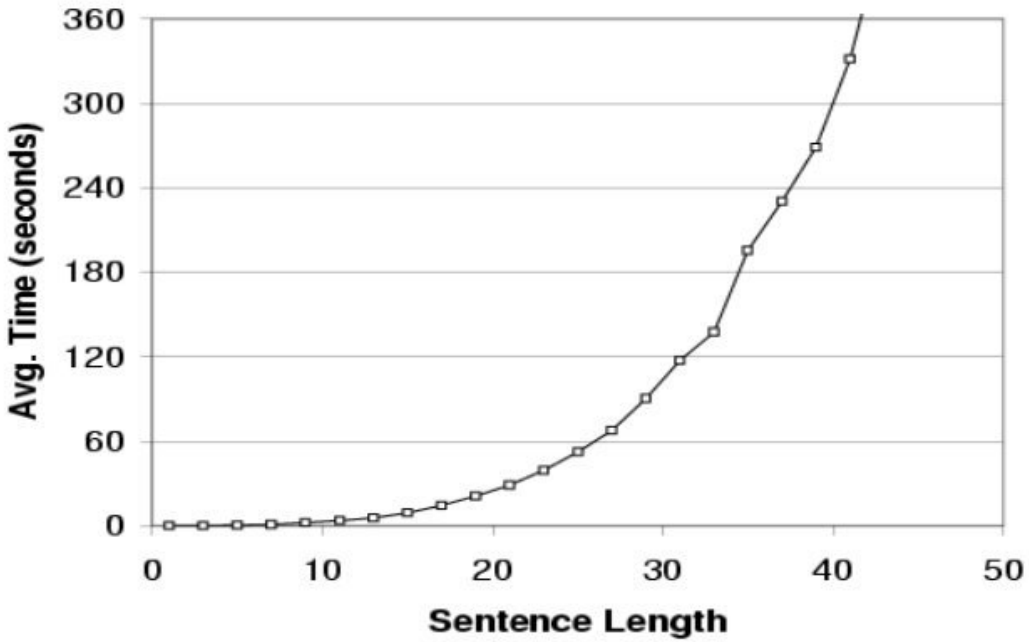
CS 4650/7650
Context Free Grammars and Parsing

Jacob Eisenstein

September 30, 2014

Complexity of CKY parsing



~ 20K Rules

(not an optimized parser!)

Observed exponent:

3.6

Ambiguity in parsing

Syntactic ambiguity is endemic to natural language:[1]

- Attachment ambiguity: we eat sushi with chopsticks, I shot an elephant in my pajamas.

[1] Examples borrowed from Dan Klein

Ambiguity in parsing

Syntactic ambiguity is endemic to natural language:[1]

- Attachment ambiguity: we eat sushi with chopsticks, I shot an elephant in my pajamas.
- Modifier scope: southern food store

[1] Examples borrowed from Dan Klein

Ambiguity in parsing

Syntactic ambiguity is endemic to natural language:[1]

- Attachment ambiguity: we eat sushi with chopsticks, I shot an elephant in my pajamas.
- Modifier scope: southern food store
- Particle versus preposition: The puppy tore up the staircase.

[1] Examples borrowed from Dan Klein

Ambiguity in parsing

Syntactic ambiguity is endemic to natural language:[1]

- Attachment ambiguity: we eat sushi with chopsticks, I shot an elephant in my pajamas.
- Modifier scope: southern food store
- Particle versus preposition: The puppy tore up the staircase.
- Complement structure: The tourists objected to the guide that they couldn't hear.

[1] Examples borrowed from Dan Klein

Ambiguity in parsing

Syntactic ambiguity is endemic to natural language:[1]

- Attachment ambiguity: we eat sushi with chopsticks, I shot an elephant in my pajamas.
- Modifier scope: southern food store
- Particle versus preposition: The puppy tore up the staircase.
- Complement structure: The tourists objected to the guide that they couldn't hear.
- Coordination scope: "I see," said the blind man, as he picked up the hammer and saw.

[1] Examples borrowed from Dan Klein

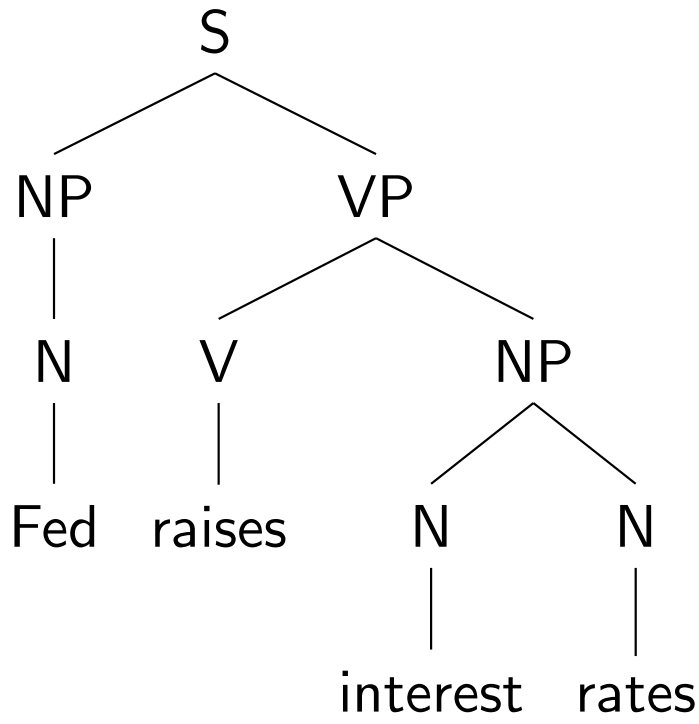
Ambiguity in parsing

Syntactic ambiguity is endemic to natural language:[1]

- Attachment ambiguity: we eat sushi with chopsticks, I shot an elephant in my pajamas.
- Modifier scope: southern food store
- Particle versus preposition: The puppy tore up the staircase.
- Complement structure: The tourists objected to the guide that they couldn't hear.
- Coordination scope: "I see," said the blind man, as he picked up the hammer and saw.
- Multiple gap constructions: The chicken is ready to eat

[1] Examples borrowed from Dan Klein

Another example



- A minimal grammar permits 36 parses!
- Broad-coverage grammars permit millions of parses of moderate-size sentences.

Attachment ambiguity

Probability of attachment sites

- [imposed [a ban [on asbestos]]]
- [imposed [a ban][on asbestos]]

Attachment ambiguity

Probability of attachment sites

- [imposed [a ban [on asbestos]]]
- [imposed [a ban][on asbestos]]

Include head of embedded NP

- ...[it [would end [its venture [with Maserati]]]]
- ...[it [would end [its venture][with Maserati]]]

Attachment ambiguity

Probability of attachment sites

- [imposed [a ban [on asbestos]]]
- [imposed [a ban][on asbestos]]

Include head of embedded NP

- ...[it [would end [its venture [with Maserati]]]]
- ...[it [would end [its venture][with Maserati]]]

Resolve multiple ambiguities simultaneously

- Cats scratch people with claws with knives

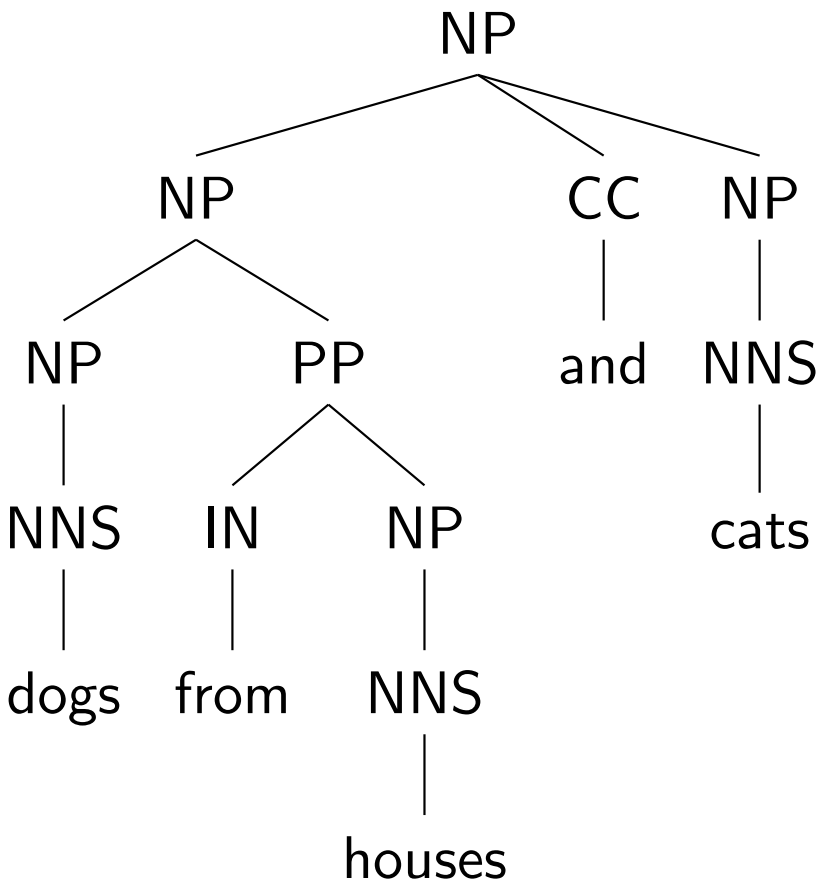
PCFGs

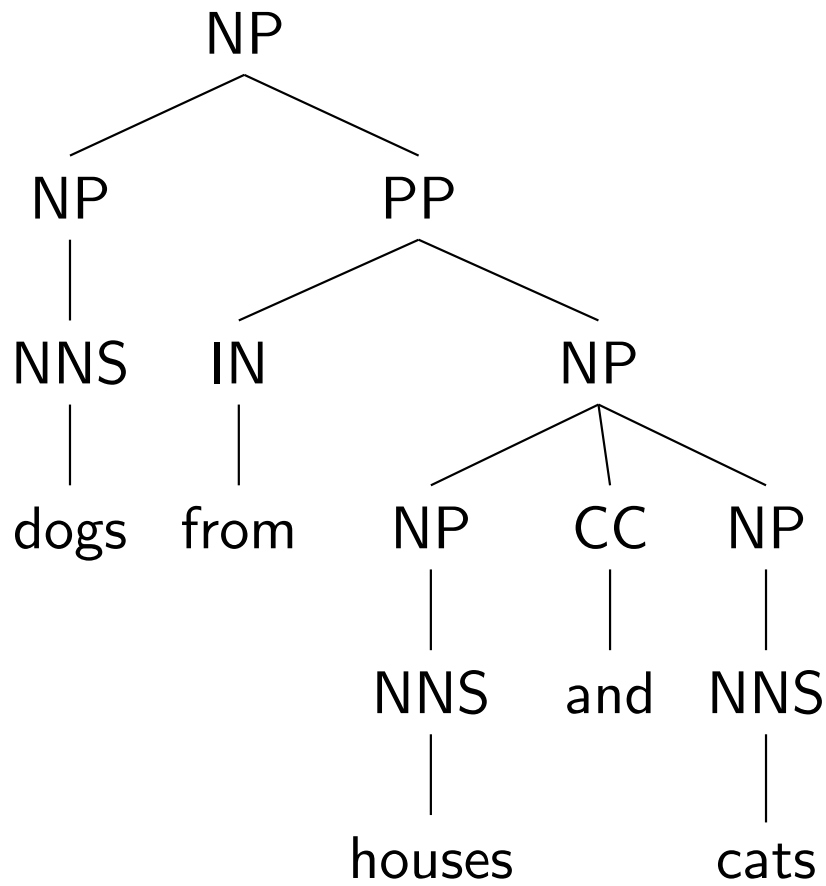
S	$\rightarrow$ NP VP	0.9
S	$\rightarrow$ S CC S	0.1
NP	$\rightarrow$ N	0.2
NP	$\rightarrow$ DT N	0.3
NP	$\rightarrow$ N NP	0.2
NP	$\rightarrow$ JJ NP	0.2
NP	$\rightarrow$ NP PP	0.1
VP	$\rightarrow$ V	0.4
VP	$\rightarrow$ V NP	0.3
VP	$\rightarrow$ V NP NP	0.1
VP	$\rightarrow$ VP PP	0.2
PP	$\rightarrow$ P NP	1.0

HMMs and PCFGs

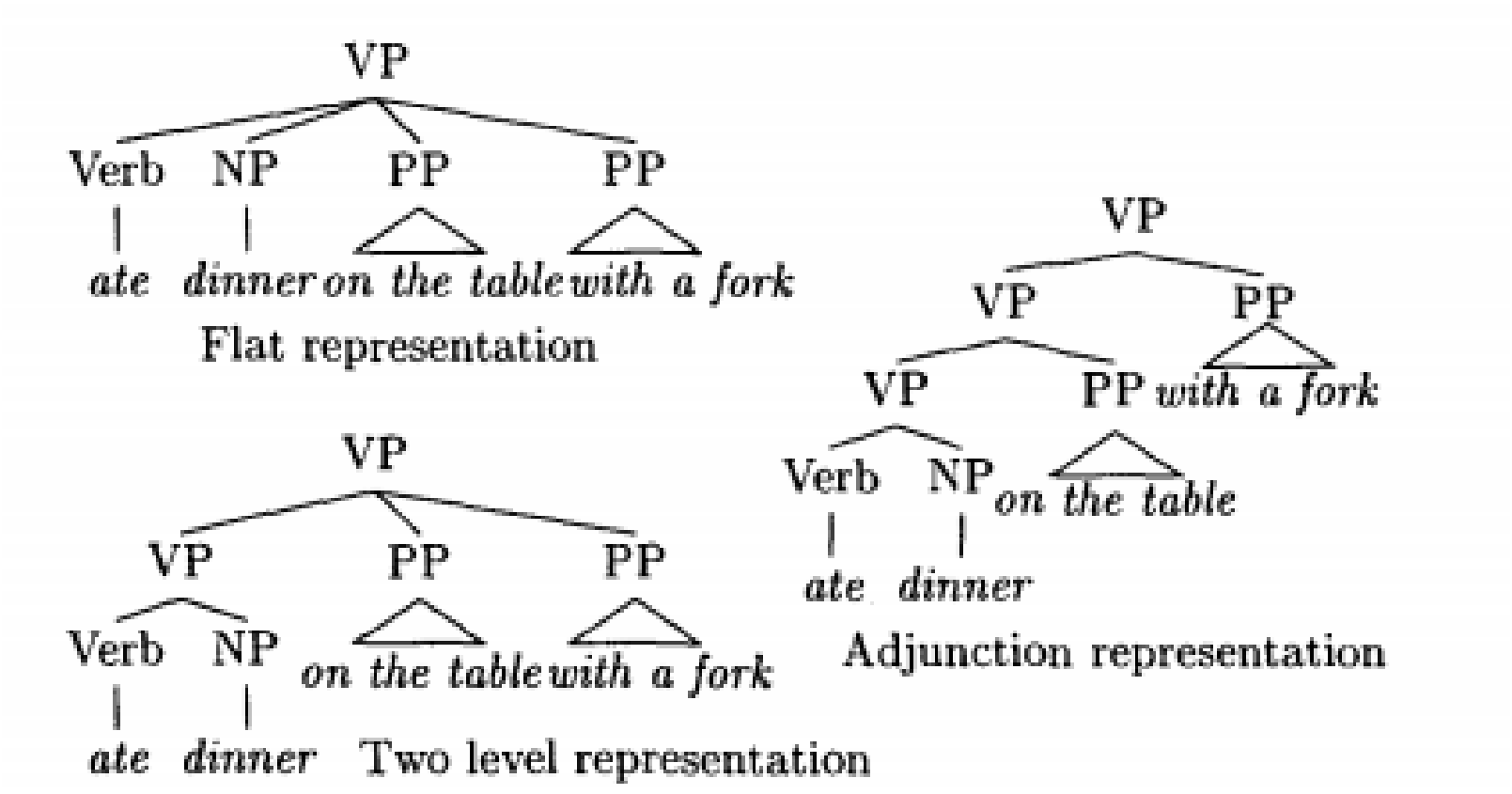
	Sequences	Trees
model	HMM	PCFG
decoding	Viterbi algorithm	CKY
decoding complexity	$\mathcal{O}(M^2\|K\|)$	$\mathcal{O}(M^3\|R\|)$
likelihood	forward algorithm	inside algorithm
marginals	forward-backward	inside-outside

Problems with PCFGs

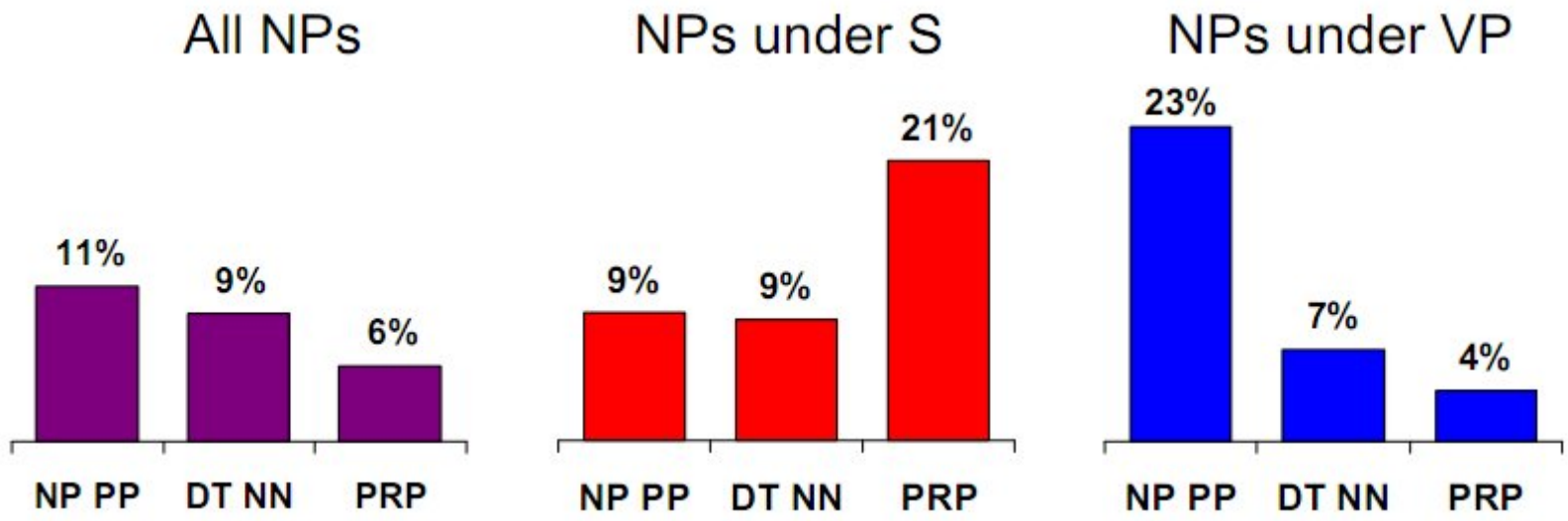




Subsumption



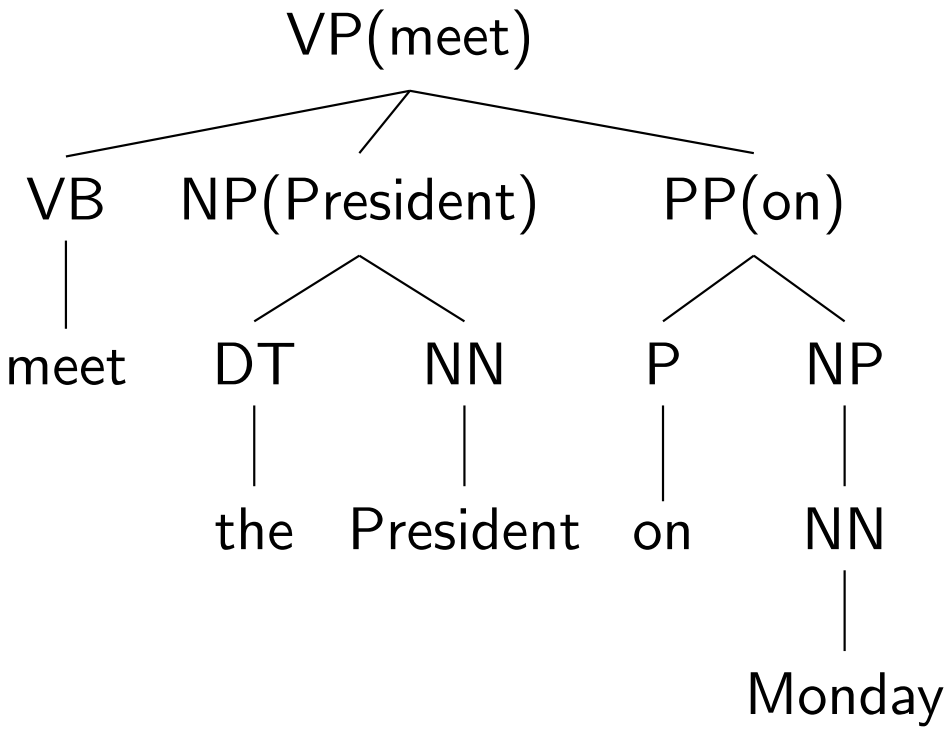
Parent annotation



Lexicalization

Non-terminal	Direction	Priority
S	right	VP SBAR ADJP UCP NP
VP	left	VBD VBN MD VBZ TO VB VP VBG
NP	right	N* EX $ CD QP PRP ...
PP	left	IN TO FW

Lexicalization



Lexicalization