Probabilistic context-free grammars, garden-pathing, and surprisal

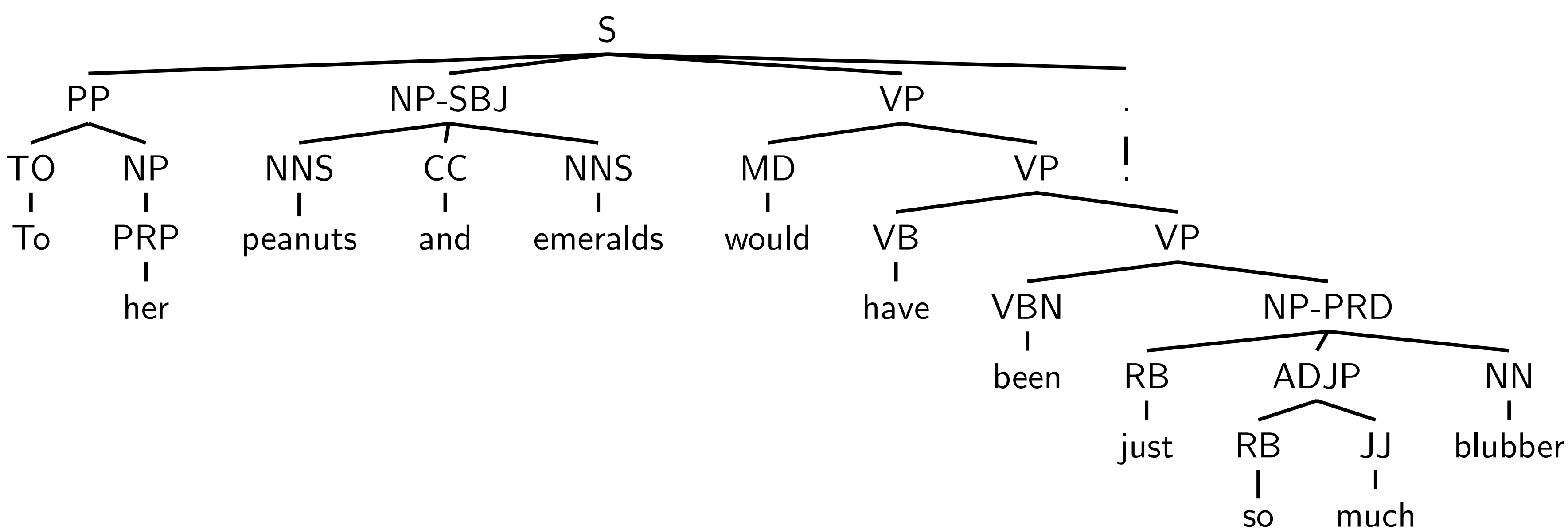
Roger Levy

9.19: Computational Psycholinguistics

Corpus annotation

- A CORPUS (pl. "corpora"): a collection of "naturalistic" text, or transcribed/recorded spoken or signed language
- It's useful to ANNOTATE the language's underlying structure
- An important SYNTACTICALLY ANNOTATED corpus: the **Penn Treebank** of English (Marcus et al., 1993)

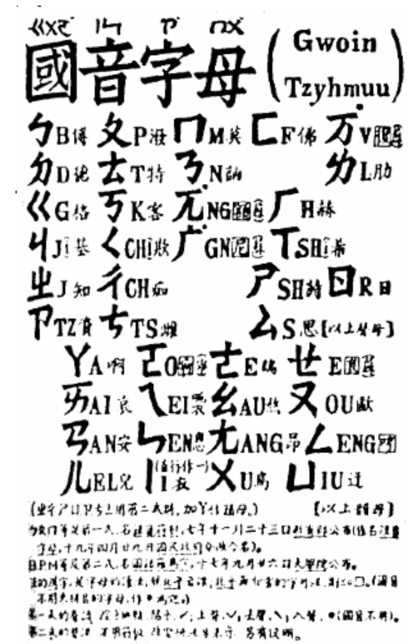
```
( (S (PP (TO To) (NP (PRP her))) (NP-SBJ (NNS peanuts) (CC and)
(NNS emeralds)) (VP (MD would) (VP (VB have) (VP (VBN been) (NP-
PRD (RB just) (ADJP (RB so) (JJ much)) (NN blubber))))) (. .)))
```



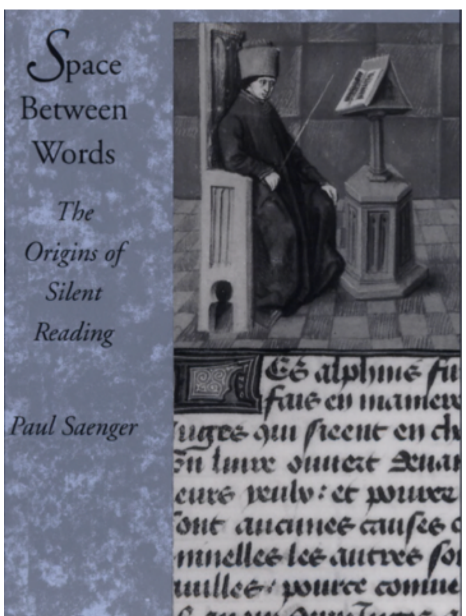
Naturally occurring linguistic annotation

Arabic short vowels and consonant lengths

bopomofo phonetic symbols (used in Taiwan for Mandarin)



Word boundary markers




```
Iwanttotellyouataleofalittlegirl
```

↓

```
I want to tell you a tale of a little girl
```

A brief & selective history of modern corpus annotation

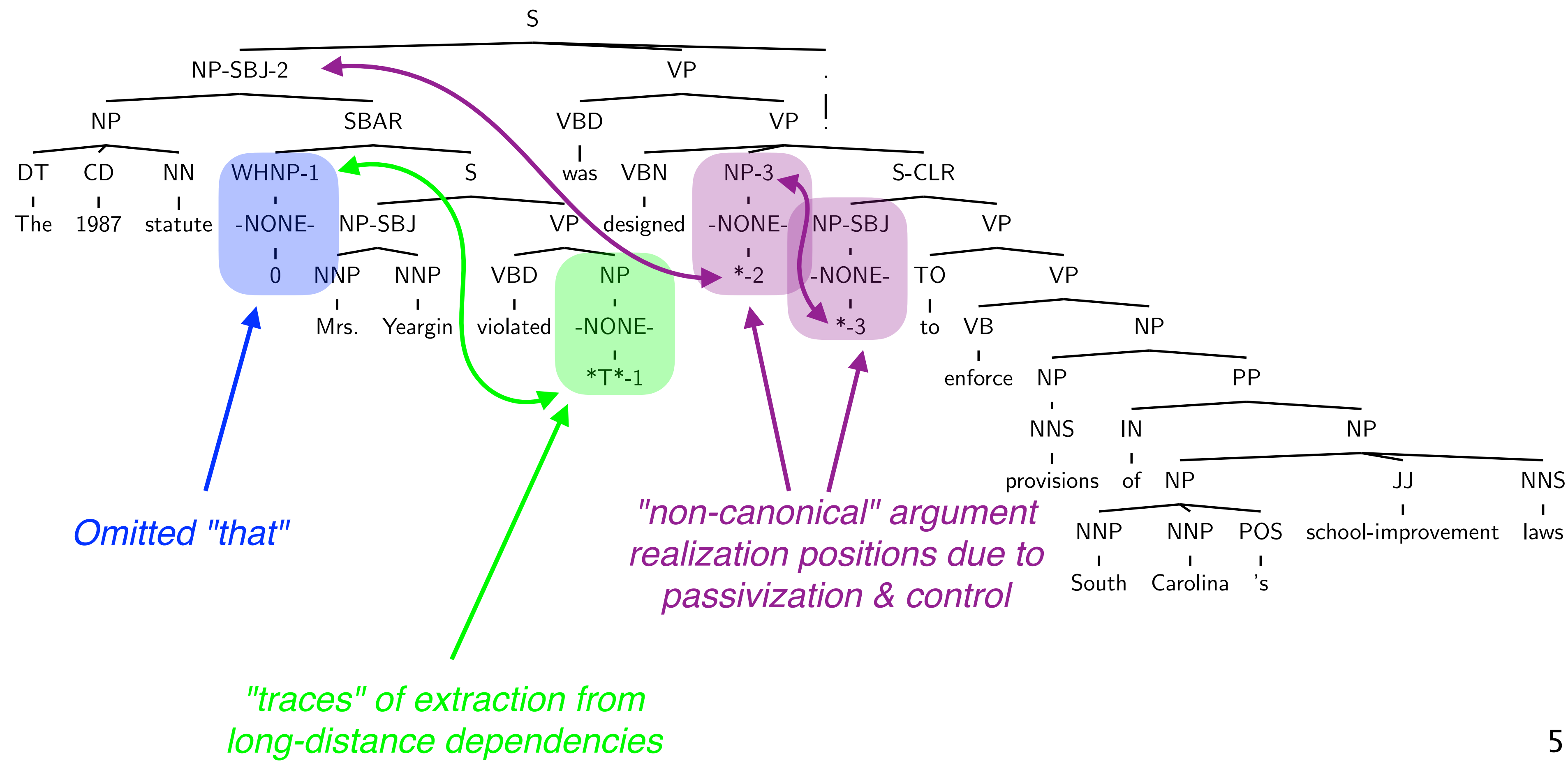
- In digital era, value of electronic readability of text & corpus annotations became quickly apparent
- 1960s: Brown Corpus of Standard American English (Kučera & Francis 1967)
  - Part-of-speech annotation added over next decade
- 1980s: large-scale language data, rise of statistical methods (Brown et al., 1990) led to many new projects
- First (morpho-)syntactic annotation project: Lancaster-Oslo-Bergen corpus of English (Garside et al., 1987)
- **Penn Treebank** project ~late '80s (Marcus et al., 1993)
  - Brown Corpus, 1989 Wall Street Journal, spoken Switchboard
- There are now treebanks in dozens of languages!

(Ide & Pustejovsky 2017, Chapter 1)

Penn Treebank conventions to know about

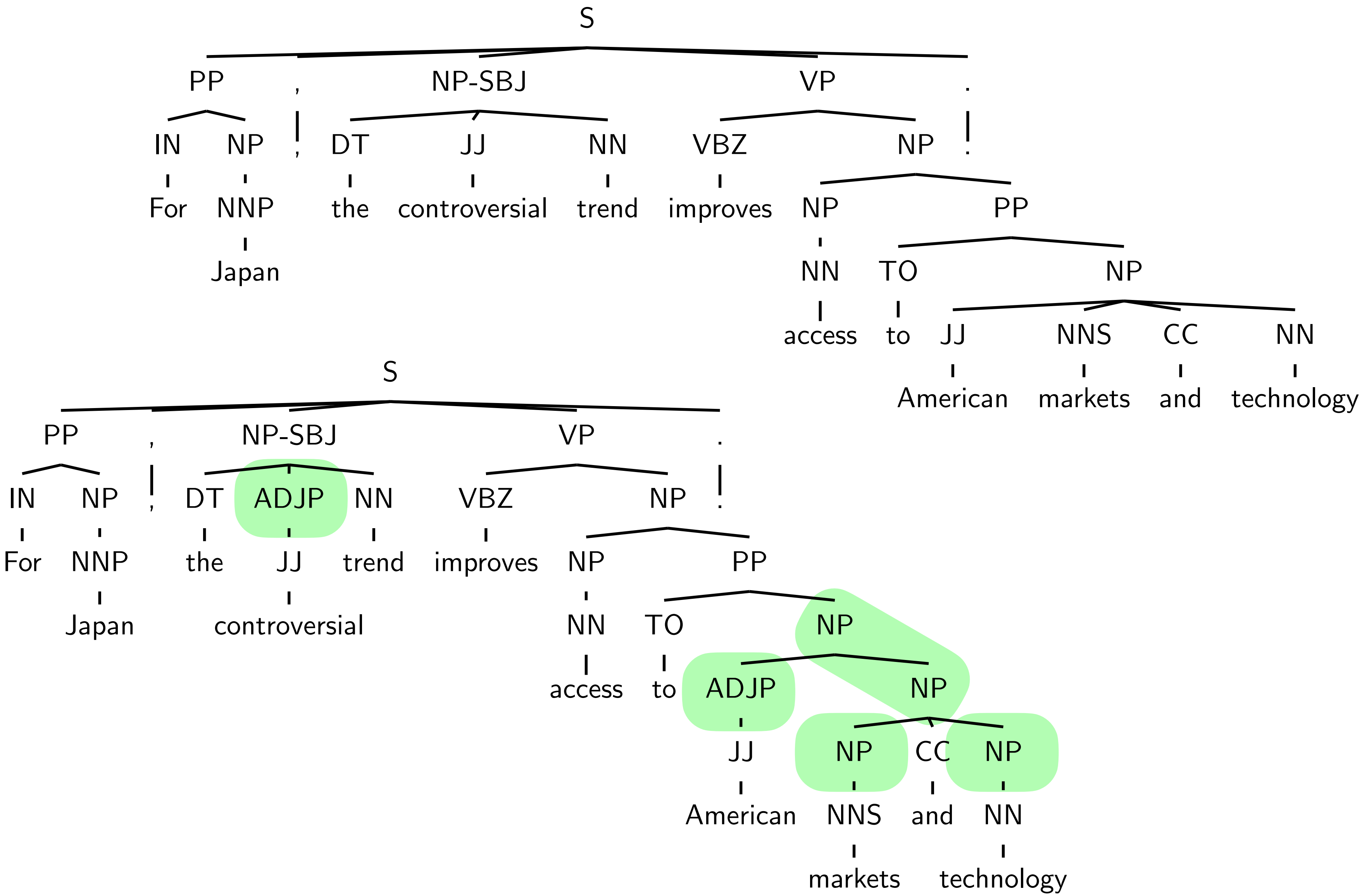
- Not all nodes of the tree dominate any words at all: there are empty categories!

The 1987 statute Mrs. Yeargin violated was designed to enforce provisions of South Carolina's school-improvement laws



Penn Treebank conventions to know about

- Annotations are often "flatter" than often (theoretically) ideal



Penn Treebank phrasal categories

1	ADJP	Adjective phrase
2	ADVP	Adverb phrase
3	NP	Noun phrase
4	PP	Prepositional phrase
5	S	Simple declarative clause
6	SBAR	Clause introduced by subordinating
7	SBARQ	Direct question introduced by wh-word or
8	SINV	Declarative sentence with subject-auxiliary
9	SQ	Subconstituent of SBARQ excluding wh-word
10	VP	Verb phrase
11	WHADVP	Wh-adverb phrase
12	WHNP	Wh-noun phrase
13	WHPP	Wh-prepositional phrase
14	X	Constituent of unknown or uncertain

There are some other phrasal categories to annotate spoken transcripts, in the Switchboard part of the Penn Treebank, too

(Marcus et al., 1993)

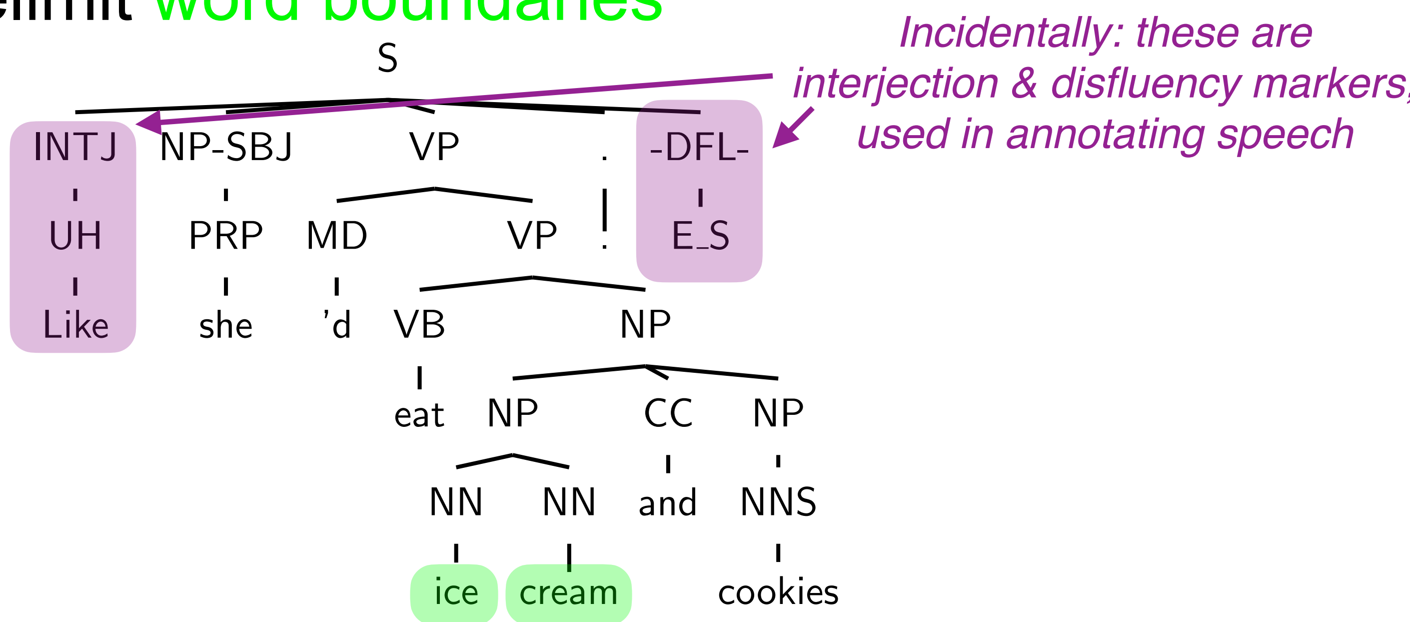
Penn Treebank tagset

1. CC Coordinating conjunction
2. CD Cardinal number
3. DT Determiner
4. EX Existential there
5. FW Foreign word
6. IN Preposition/subordinating conjunction
7. JJ Adjective
8. JJR Adjective, comparative
9. JJS Adjective, superlative
10. LS List item marker
11. MD Modal
12. NN Noun, singular or mass
13. NNS Noun, plural
14. NNP Proper noun, singular
15. NNPS Proper noun, plural
16. PDT Predeterminer
17. POS Possessive ending
18. PRP Personal pronoun
19. PP Possessive pronoun
20. RB Adverb
21. RBR Adverb, comparative
22. RBS Adverb, superlative
23. RP Particle
24. SYM Symbol (mathematical or scientific)
25. TO to
26. UH Interjection
27. VB Verb, base form
28. VBD Verb, past tense
29. VBG Verb, gerund/present participle
30. VBN Verb, past participle
31. VBP Verb, non-3rd ps. sing. present
32. VBZ Verb, 3rd ps. sing. present
33. WDT wh-determiner
34. WP wh-pronoun
35. WP Possessive wh-pronoun
36. WRB wh-adverb
37. # Pound sign
38. $ Dollar sign
39. . Sentence-final punctuation
40. , Comma
41. : Colon, semi-colon
42. (Left bracket character
43.) Right bracket character
44. " Straight double quote
45. ` Left open single quote
46. " Left open double quote
47. ' Right close single quote
48. " Right close double quote

(Marcus et al., 1993)

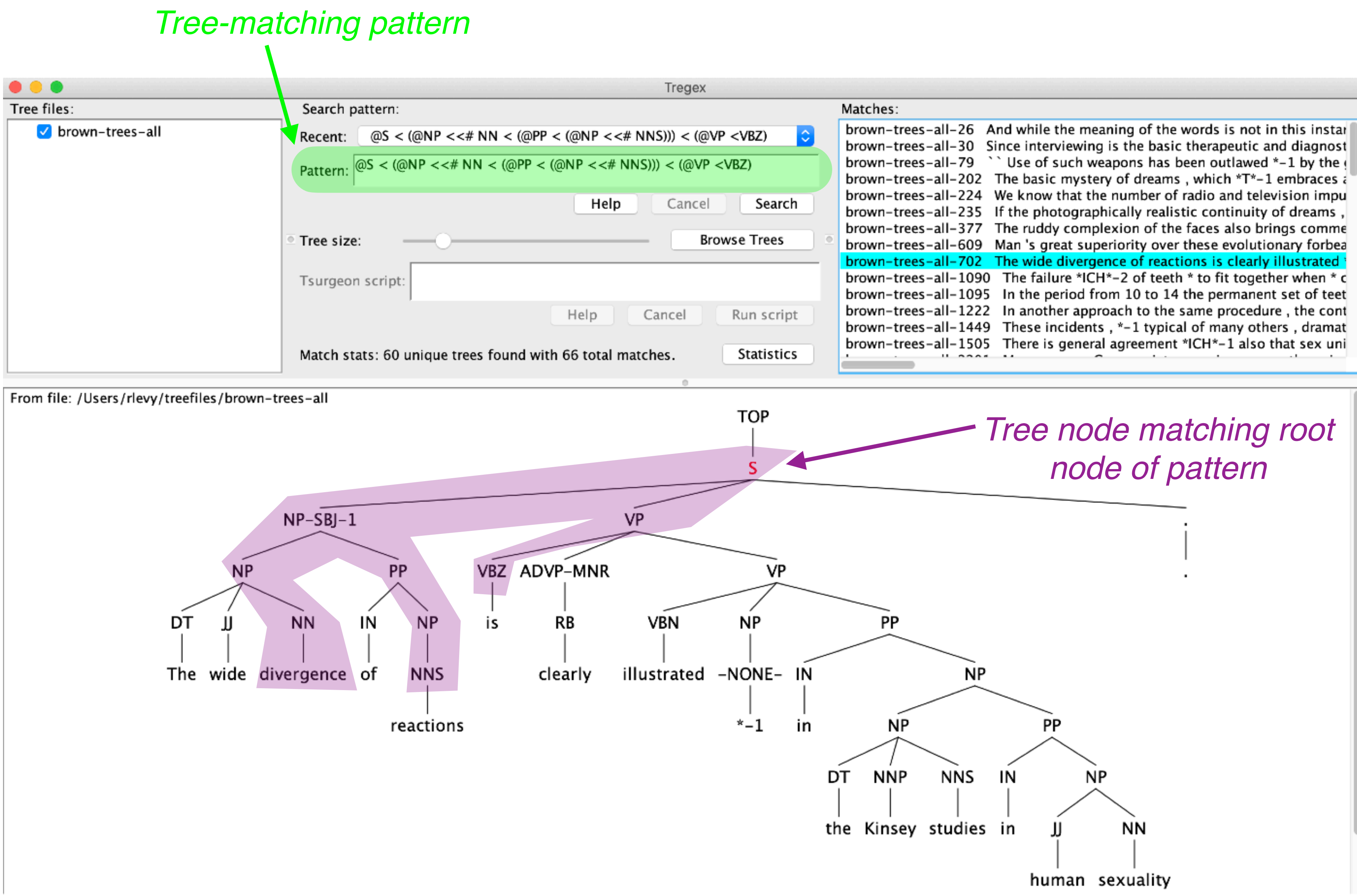
A few more Penn Treebank tidbits

- Spaces delimit word boundaries



- All tree leaves (words and empty categories) are dominated by their part-of-speech tag alone
- You can treat Treebank annotations (mostly) as derivations trees from a context-free grammar, BUT best to treat the annotations as *information about syntactic syntactic structure* that we want grammars that will accurately recover

Software for searching treebanks: `Tregex`

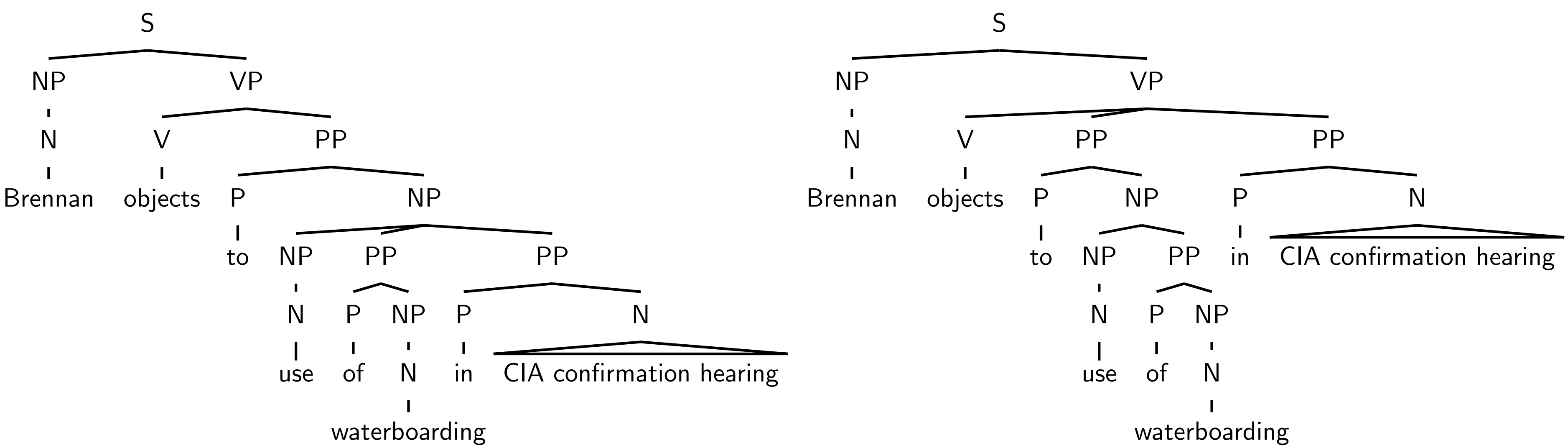


(Levy & Andrew, 2006)

Syntactic ambiguity

- Context-free grammars predict multiple derivations for many word strings
- This can capture many cases of AMBIGUITY in language

Brennan objects to use of waterboarding in CIA confirmation hearing

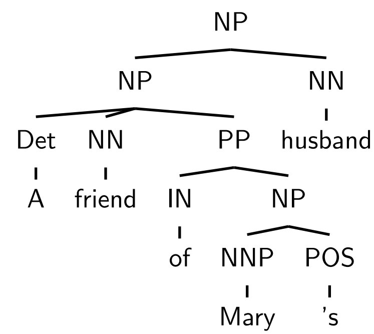
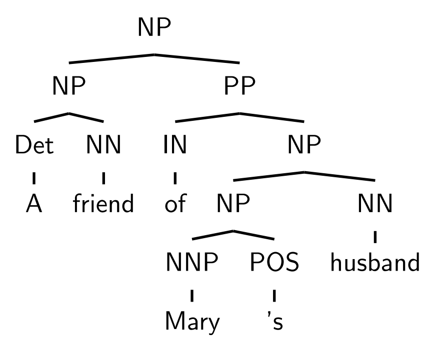


- But CFGs don't explain *where our interpretation preferences come from*

https://languagelog.ldc.upenn.edu/nll/?p=17431

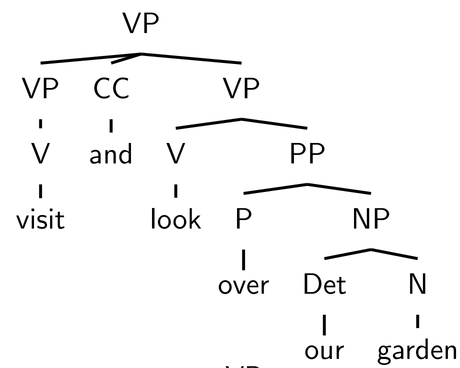
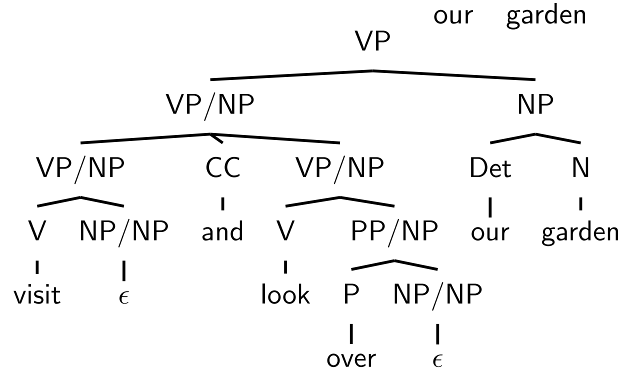
Example from in-class survey

A friend of Mary's husband wanted to visit and look over our garden last Tuesday.

Question	**Syntax**	**Proportion of choices**
Who wanted to visit and see our garden?		
The husband of one of Mary's friends		
Someone who is friends with Mary's husband		
Someone else		

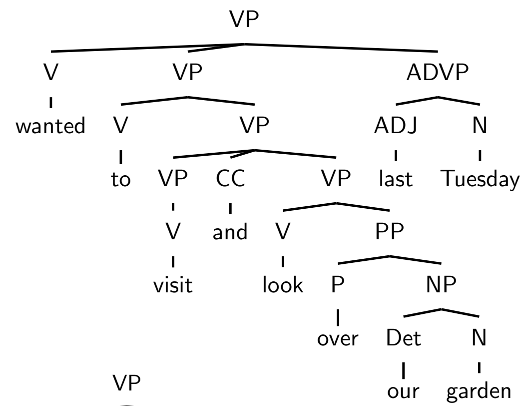
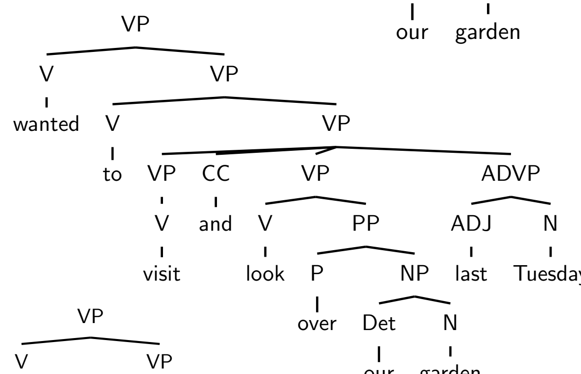
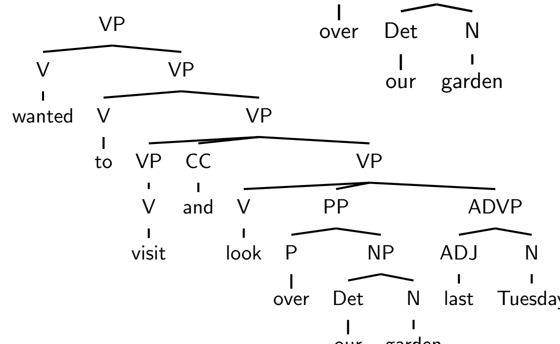
Example from in-class survey

A friend of Mary's husband wanted to visit and look over our garden last Tuesday.

Question	**Syntax**	**People choosing**
Who or what did this person want to visit?		
Us		
Our garden		
Someone or something else		

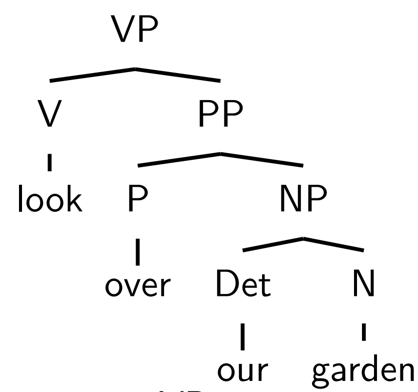
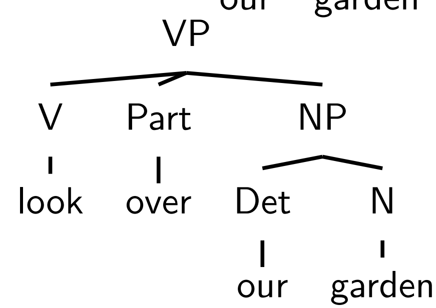
Example from in-class survey

A friend of Mary's husband wanted to visit and look over our garden last Tuesday.

Question	Syntax	People choosing
How does "Last Tuesday" relate to the rest of the sentence?		
This was the time that the person's desire (to visit and learn about our garden) arose		
This was the person's preferred time both to visit and to look over our garden		
This was the person's preferred time to look over our garden		

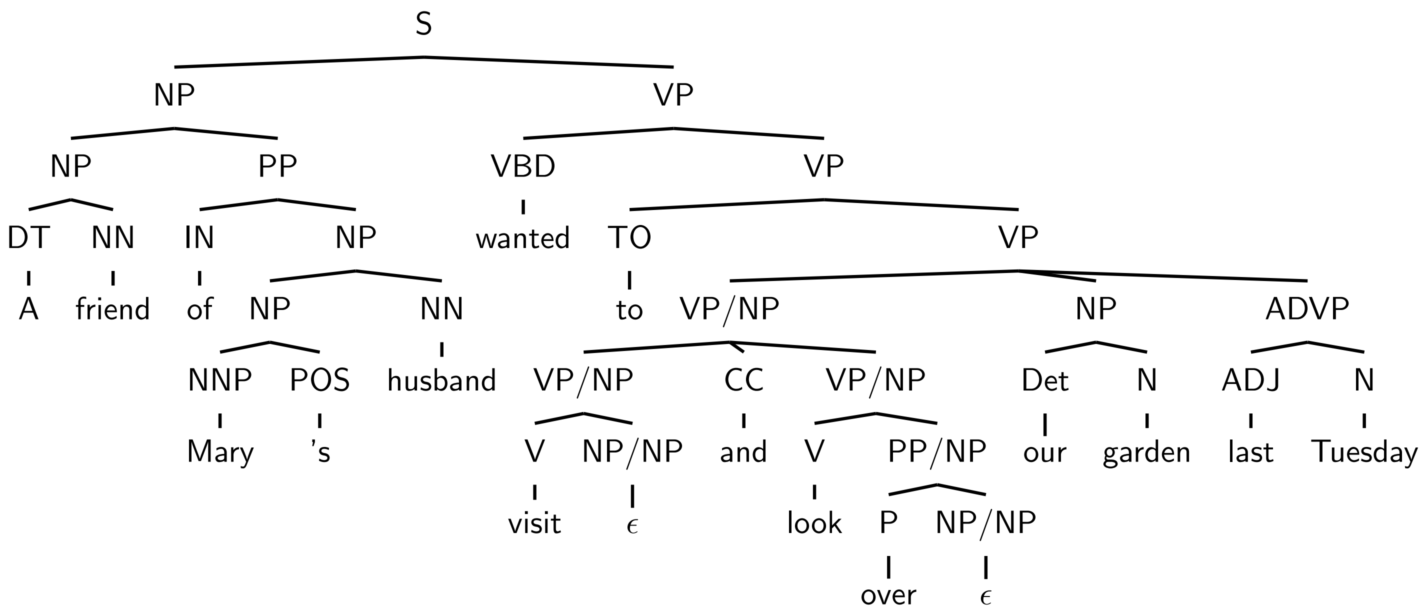
Example from in-class survey

A friend of Mary's husband wanted to visit and look over our garden last Tuesday.

Question	Syntax	People choosing
What is meant by "look over our garden"?		
From one side of the garden, look over to what's on the other side of the garden		
Look our garden over		
Something else		

Preferred analysis for our example

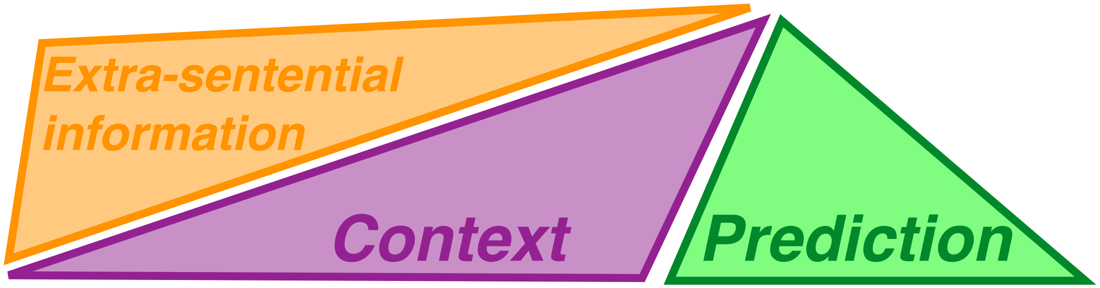
- There are 20 trees available from these 4 ambiguities*
- Yet 66% of respondents chose this analysis:



- 18% preferred an analysis differing in only 1 ambiguity
- 18% preferred analysis differing in 2 ambiguities
- **Theoretical challenge:** what determines the "preferred" analysis, and how do we find it?

recommended question: why 20, not 2×3×2×2=24?

Expectations in incremental comprehension



Previous Input **_Current Input_**

- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

  Terry ate a… nectarine/banana/sandwich

- Semantic & situational knowledge:

  The children went outside to…play

  The squirrel stored some nuts in the…~~statue~~

  tree

Rational analysis for syntactic processing

1. Specify precisely the goals of the cognitive system

Efficiently analyze ("process") incoming linguistic input, and identify intended meaning

2. Formalize model of the environment to which the cognitive system is adapted

Statistics of the linguistic environment; knowledge of interlocutors and their goals:
$P(\textit{\textbf{Structure}})$ ***and*** $P(\textit{\textbf{Input}} \mid \textit{\textbf{Structure}})$

3. Make minimal assumptions re: computational limitations

Fast, near-normative Bayesian inference: $P(\textit{\textbf{Structure}} \mid \textit{\textbf{Input}}) \propto P(\textit{\textbf{Input}} \mid \textit{\textbf{Structure}}) P(\textit{\textbf{Structure}})$

4. Derive predicted optimal behavior given 1–3

5. Compare predictions with empirical data

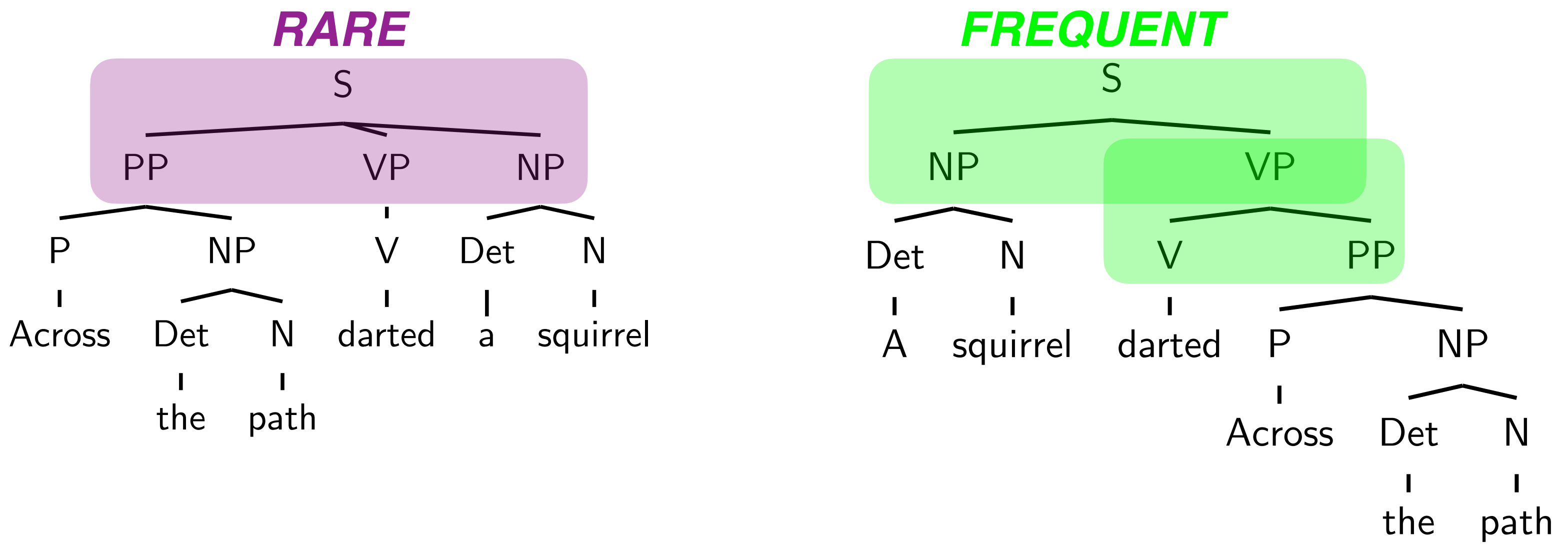
Use controlled, experimental case studies to investigate real-time human language understanding

6. If necessary, iterate 1–5

(Anderson, 1990)

Putting probabilities on structures

- Some syntactic structures are rarer than others



- We want a model that will probabilistically score *parts of a tree*
- One simple model for this is the PROBABILISTIC (or STOCHASTIC) CONTEXT-FREE GRAMMAR (**PCFG** or **SCFG**)

Probabilistic Context-Free Grammars

A *probabilistic* context-free grammar (PCFG) consists of a tuple (N, V, S, R, P) such that:

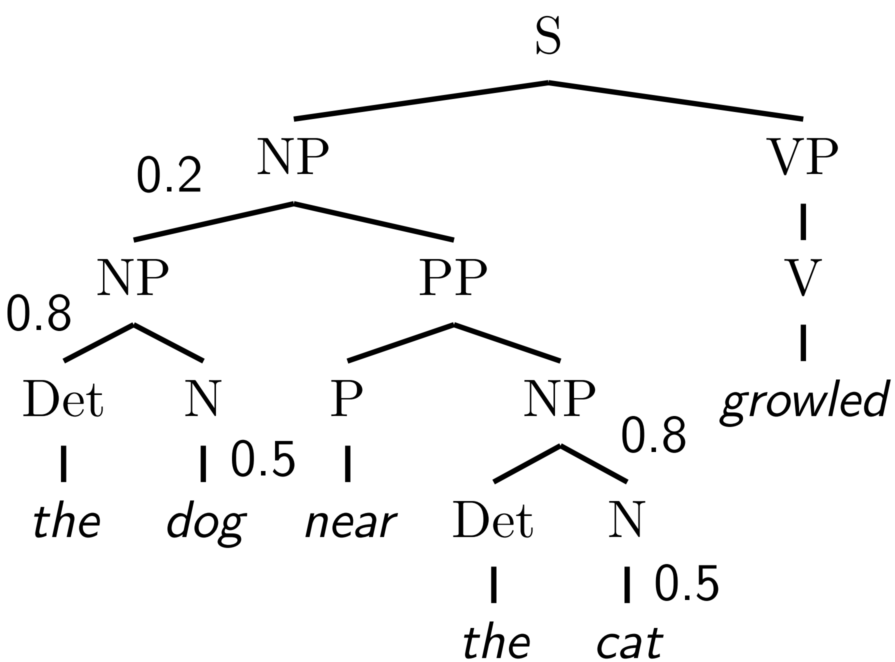
- N is a finite set of non-terminal symbols;
- V is a finite set of terminal symbols;
- S is the start symbol;
- R is a finite set of rules of the form $X \rightarrow \alpha$ where $X \in N$ and α is a sequence of symbols drawn from $N \cup V$;
- P is a mapping from R into probabilities, such that for each $X \in N$,

$$\sum_{[X \rightarrow \alpha] \in R} P(X \rightarrow \alpha) = 1$$

PCFG *derivations* and *derivation trees* are just like for CFGs. The probability $P(T)$ of a derivation tree is simply the product of the probabilities of each rule application.

Example PCFG

1	S	$\rightarrow$ NP VP	1	Det	$\rightarrow$ the
0.8	NP	$\rightarrow$ Det N	0.5	N	$\rightarrow$ dog
0.2	NP	$\rightarrow$ NP PP	0.5	N	$\rightarrow$ cat
1	PP	$\rightarrow$ P NP	1	P	$\rightarrow$ near
1	VP	$\rightarrow$ V	1	V	$\rightarrow$ growled



$$
\begin{aligned}
P(T) &= 1 \times 0.2 \times 0.8 \times 1 \times 0.5 \times 1 \times 1 \times 0.8 \times 1 \times 0.5 \times 1 \times 1 \\
&= 0.032
\end{aligned}
$$

PCFG review (2)

- We just learned how to calculate the *probability of a tree*
- The *probability of a string* $w_{1\ldots n}$ is the sum of the probabilities of all trees whose yield **is** $w_{1\ldots n}$
- The *probability of a string prefix* $w_{1\ldots i}$ is the sum of the probabilities of all trees whose yield **begins with** $w_{1\ldots i}$
- If we had the probabilities of two string prefixes $w_{1\ldots i-1}$ and $w_{1\ldots i}$, we could calculate the conditional probability $P(w_i|w_{1\ldots i-1})$ as their ratio:

$$P(w_i|w_{1\ldots i-1}) = \frac{P(w_{1\ldots i})}{P(w_{1\ldots i-1})}$$

Inference over infinite tree sets

Consider the following noun-phrase grammar:

		1	$\text{Det} \rightarrow \text{the}$
$\frac{2}{3}$	$\text{NP} \rightarrow \text{Det N}$	$\frac{2}{3}$	$\text{N} \rightarrow \text{dog}$
$\frac{1}{3}$	$\text{NP} \rightarrow \text{NP PP}$	$\frac{1}{3}$	$\text{N} \rightarrow \text{cat}$
1	$\text{PP} \rightarrow \text{P NP}$	1	$\text{P} \rightarrow \text{near}$

Inference over infinite tree sets

Consider the following noun-phrase grammar:

			1	Det	$\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N		$\frac{2}{3}$	N	$\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP		$\frac{1}{3}$	N	$\rightarrow$ cat
1	PP $\rightarrow$ P NP		1	P	$\rightarrow$ near

Question: given a sentence starting with

the...

what is the probability that the next word is *dog*?

Inference over infinite tree sets

Consider the following noun-phrase grammar:

$$\begin{array}{lllll} & & & 1 & \text{Det} \rightarrow \text{the} \\ \frac{2}{3} & \text{NP} \rightarrow \text{Det N} & & \frac{2}{3} & \text{N} \quad \rightarrow \text{dog} \\ \frac{1}{3} & \text{NP} \rightarrow \text{NP PP} & & \frac{1}{3} & \text{N} \quad \rightarrow \text{cat} \\ 1 & \text{PP} \rightarrow \text{P NP} & & 1 & \text{P} \quad \rightarrow \text{near} \end{array}$$

Question: given a sentence starting with

the...

what is the probability that the next word is *dog*?
Intuitively, the answers to this question should be

$$P(\mathrm{dog}|\mathrm{the}) = \frac{2}{3}$$

Inference over infinite tree sets

Consider the following noun-phrase grammar:

$$
\begin{array}{lll}
\frac{2}{3} & NP \rightarrow Det\ N \\
\frac{1}{3} & NP \rightarrow NP\ PP \\
1 & PP \rightarrow P\ NP
\end{array}
\qquad
\begin{array}{lll}
1 & Det & \rightarrow \text{the} \\
\frac{2}{3} & N & \rightarrow \text{dog} \\
\frac{1}{3} & N & \rightarrow \text{cat} \\
1 & P & \rightarrow \text{near}
\end{array}
$$

Question: given a sentence starting with

the...

what is the probability that the next word is *dog*?

Intuitively, the answers to this question should be

$$P(\text{dog}|\text{the}) = \frac{2}{3}$$

because the second word HAS to be either *dog* or *cat*.

Inference over infinite tree sets (2)

		1	Det $\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N	$\frac{2}{3}$	N $\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP	$\frac{1}{3}$	N $\rightarrow$ cat
1	PP $\rightarrow$ P NP	1	P $\rightarrow$ near

- We "should" just enumerate the trees that cover *the dog* . . . ,

Inference over infinite tree sets (2)

			1	Det	$\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N		$\frac{2}{3}$	N	$\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP		$\frac{1}{3}$	N	$\rightarrow$ cat
1	PP $\rightarrow$ P NP		1	P	$\rightarrow$ near

- We "should" just enumerate the trees that cover *the dog* . . . , and divide their total probability by that of *the* . . .

Inference over infinite tree sets (2)

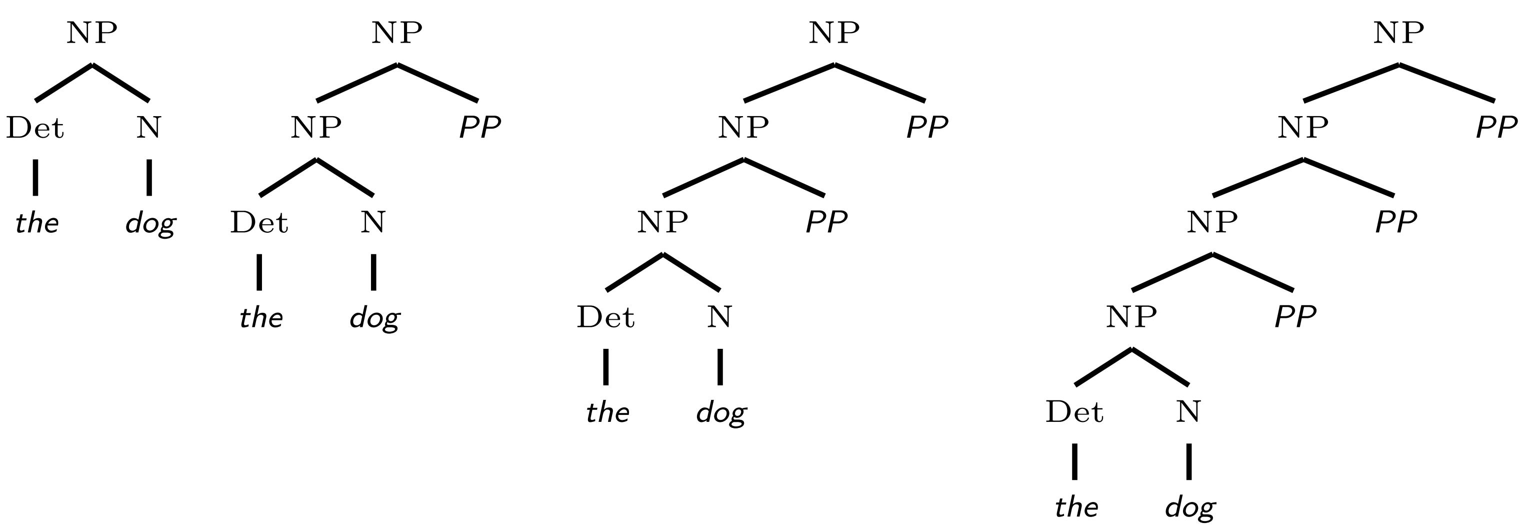
		1	Det $\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N	$\frac{2}{3}$	N $\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP	$\frac{1}{3}$	N $\rightarrow$ cat
1	PP $\rightarrow$ P NP	1	P $\rightarrow$ near

- We "should" just enumerate the trees that cover *the dog . . .* , and divide their total probability by that of *the . . .*
- . . . but there are infinitely many trees.

Inference over infinite tree sets (2)

$\frac{2}{3}$	NP $\rightarrow$ Det N		1	Det	$\rightarrow$ the
$\frac{1}{3}$	NP $\rightarrow$ NP PP		$\frac{2}{3}$	N	$\rightarrow$ dog
1	PP $\rightarrow$ P NP		$\frac{1}{3}$	N	$\rightarrow$ cat
			1	P	$\rightarrow$ near

- We "should" just enumerate the trees that cover *the dog* . . . , and divide their total probability by that of *the* . . .
- . . . but there are infinitely many trees.

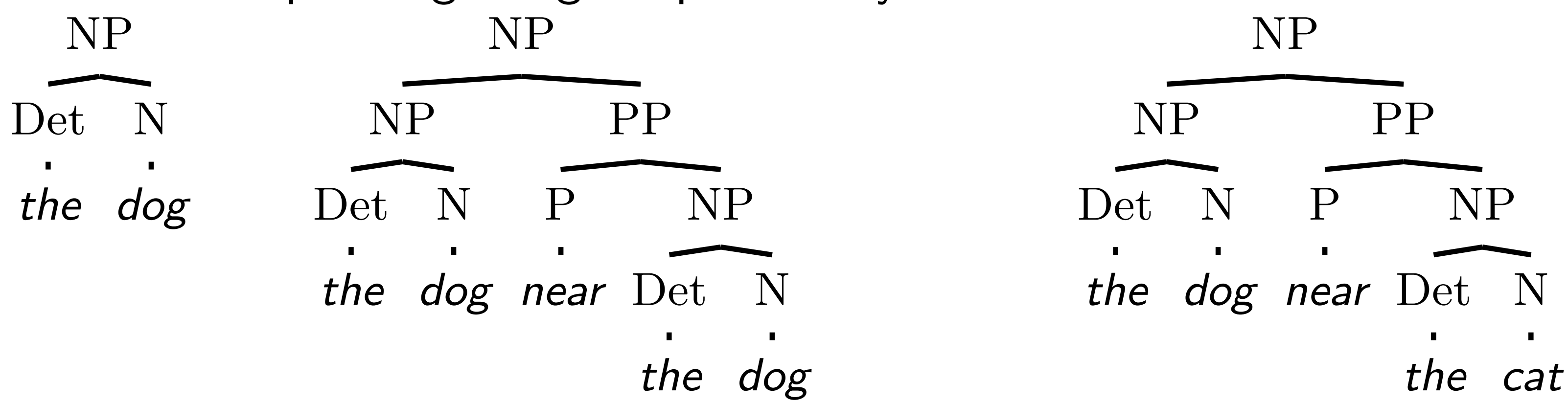


. . .

		1	Det $\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N	$\frac{2}{3}$	N $\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP	$\frac{1}{3}$	N $\rightarrow$ cat
1	PP $\rightarrow$ P NP	1	P $\rightarrow$ near

You can think of a *partial* tree as marginalizing over all completions of the partial tree.

It has a corresponding marginal probability in the PCFG.

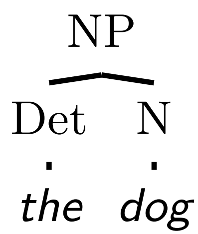


$\frac{4}{9}$ $\frac{8}{81}$ $\frac{4}{81}$

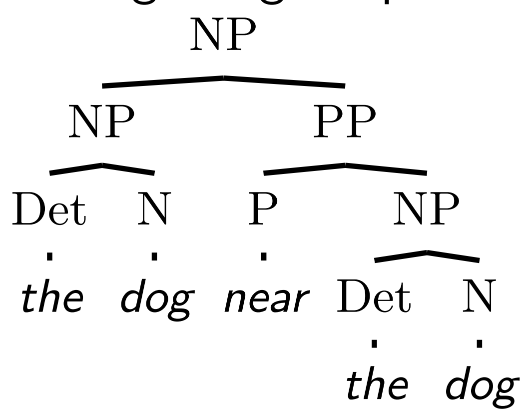
		1	Det $\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N	$\frac{2}{3}$	N $\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP	$\frac{1}{3}$	N $\rightarrow$ cat
1	PP $\rightarrow$ P NP	1	P $\rightarrow$ near

You can think of a *partial* tree as marginalizing over all completions of the partial tree.

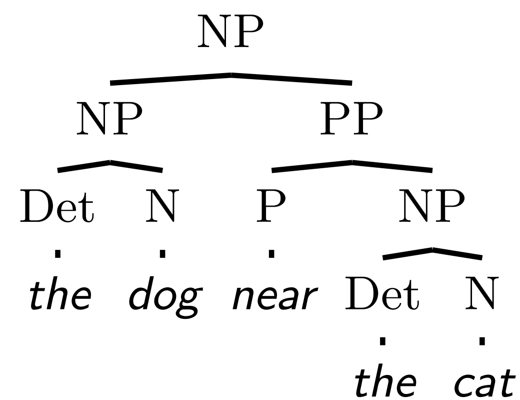
It has a corresponding marginal probability in the PCFG.



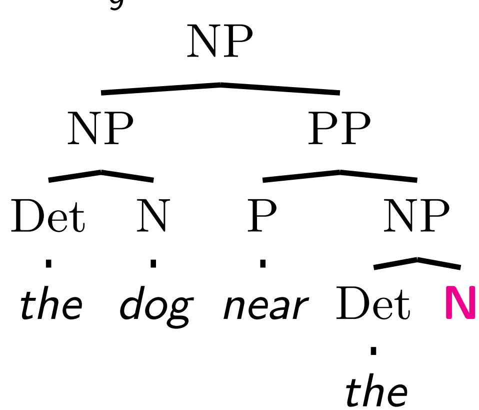
$\frac{4}{9}$



$\frac{8}{81}$



$\frac{4}{81}$

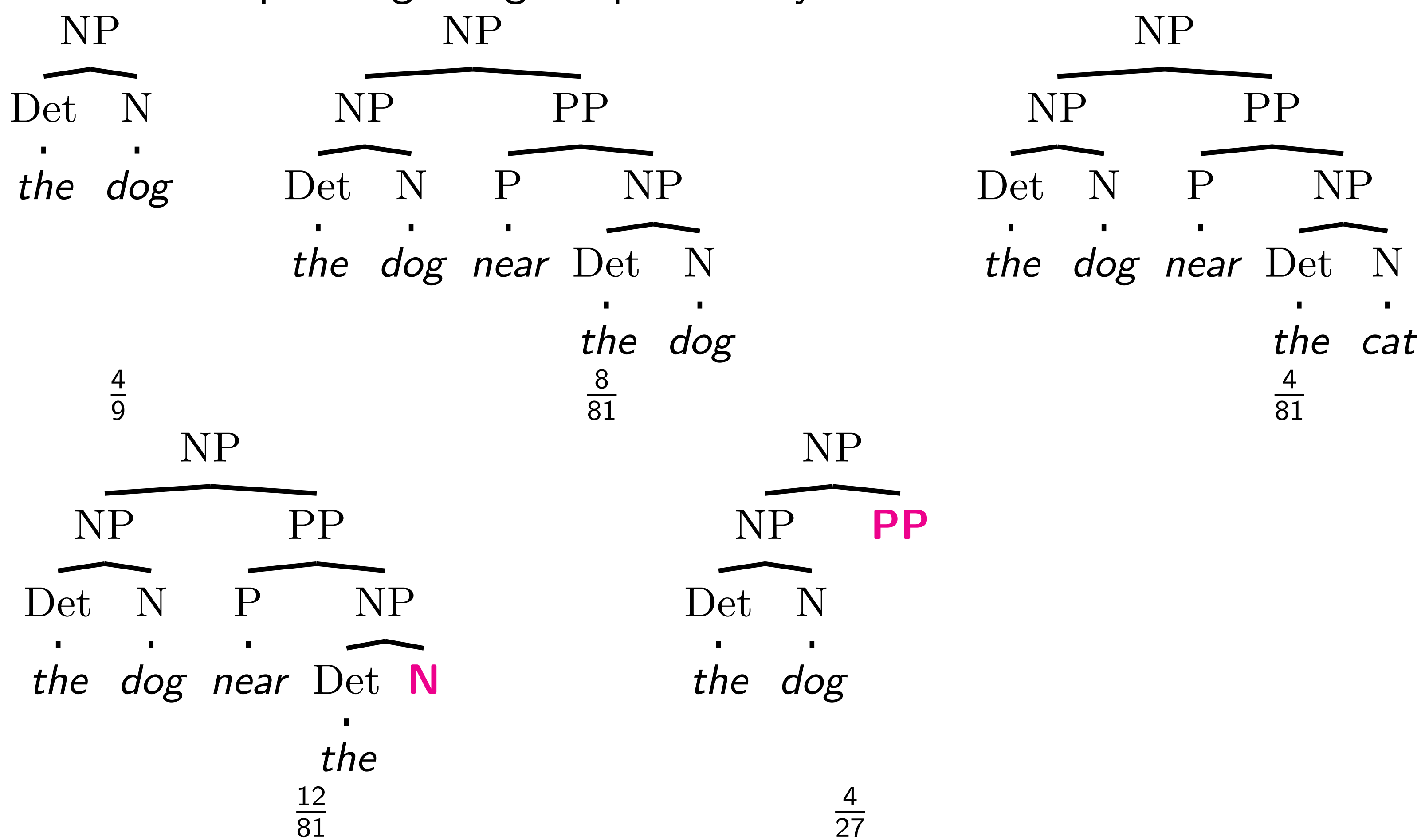


$\frac{12}{81}$

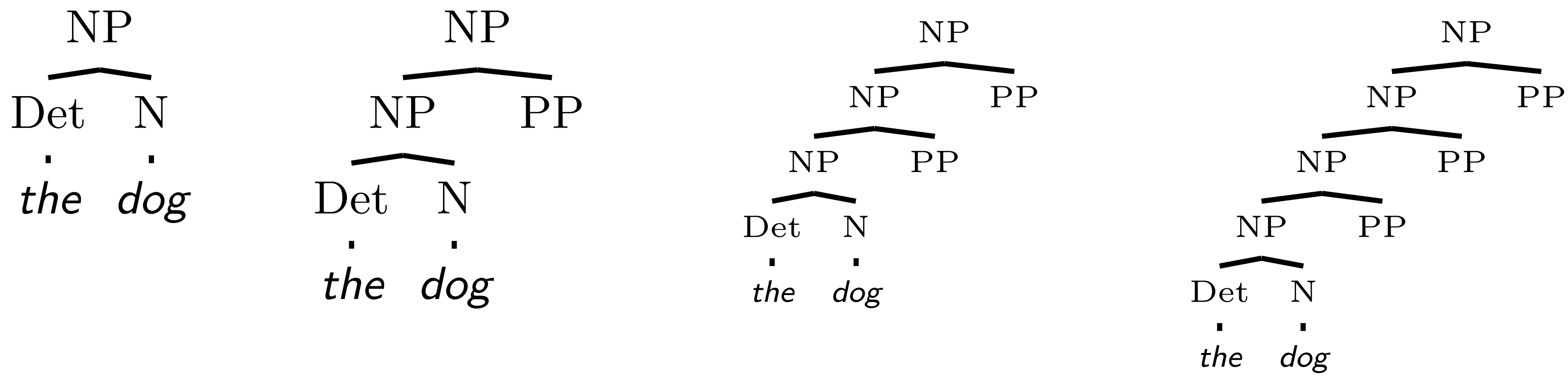
$\frac{2}{3}$	NP → Det N			1	Det → the
$\frac{1}{3}$	NP → NP PP			$\frac{2}{3}$	N → dog
1	PP → P NP			$\frac{1}{3}$	N → cat
				1	P → near

You can think of a *partial* tree as marginalizing over all completions of the partial tree.

It has a corresponding marginal probability in the PCFG.

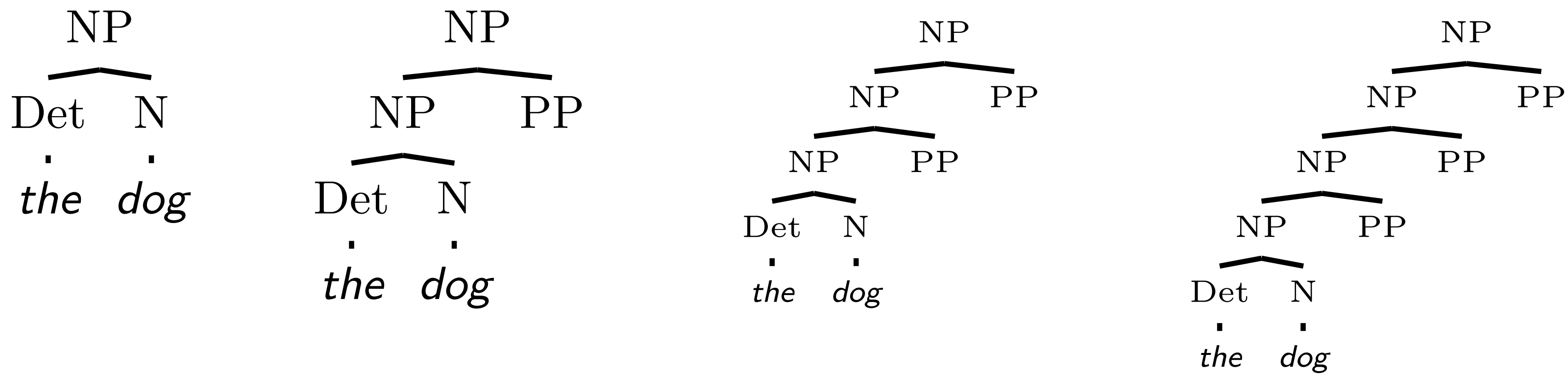


Problem 2: there are still an infinite number of incomplete trees covering a partial input.



$\frac{4}{9}$ $\frac{4}{27}$ $\frac{4}{81}$ $\frac{4}{243}$

Problem 2: there are still an infinite number of incomplete trees covering a partial input.

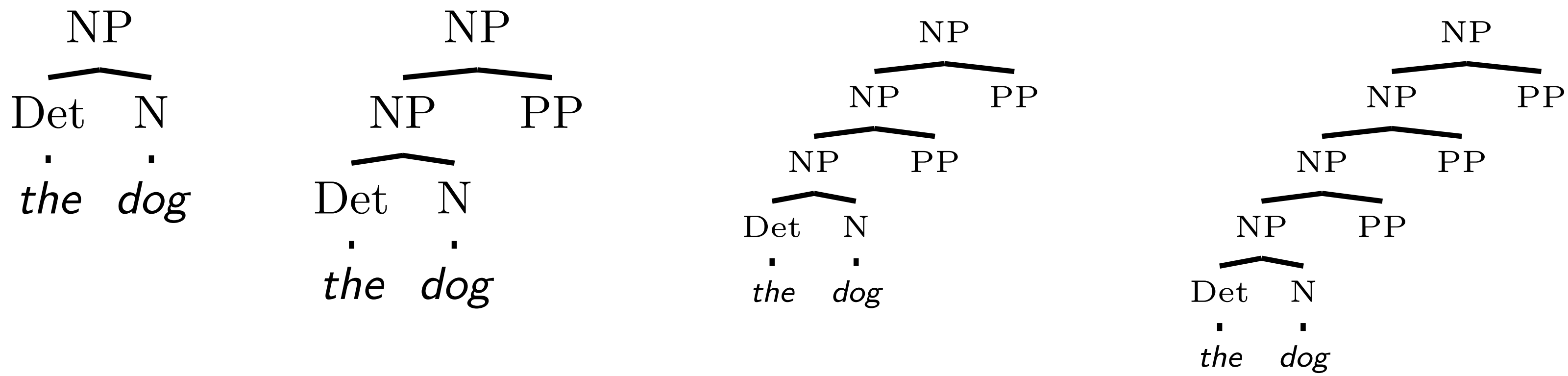


$\frac{4}{9}$ $\frac{4}{27}$ $\frac{4}{81}$ $\frac{4}{243}$

BUT! These tree probabilities form a geometric series:

$$P(the\ dog \ldots) = \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243} + \cdots$$

Problem 2: there are still an infinite number of incomplete trees covering a partial input.

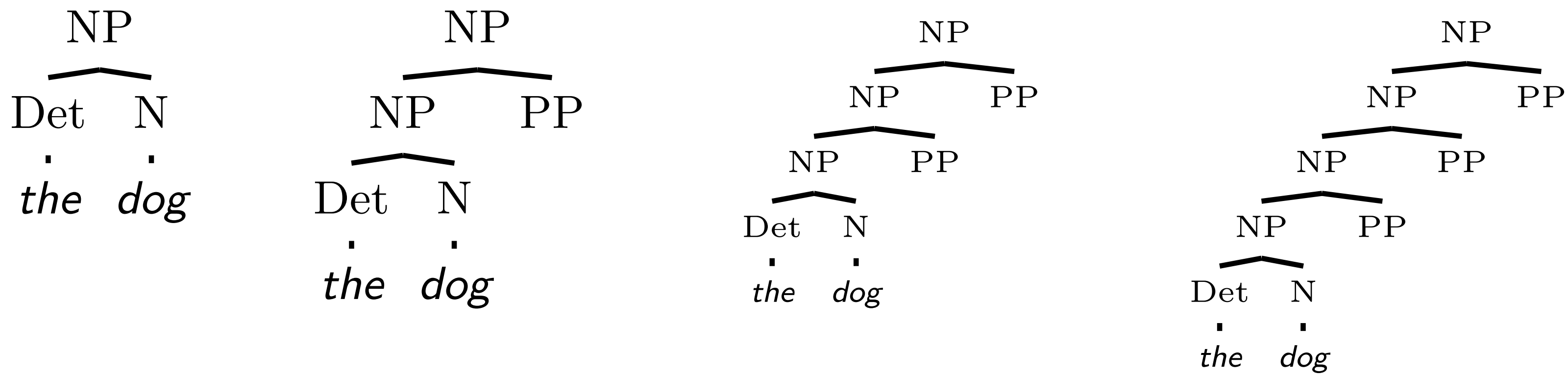


$\frac{4}{9}$ $\frac{4}{27}$ $\frac{4}{81}$ $\frac{4}{243}$

BUT! These tree probabilities form a geometric series:

$$\begin{aligned} P(\textit{the dog}\ldots) &= \frac{4}{9}+\frac{4}{27}+\frac{4}{81}+\frac{4}{243}+\cdots \\ &= \frac{4}{9}\sum_{i=0}^{\infty}\left(\frac{1}{3}\right)^{i} \end{aligned}$$

Problem 2: there are still an infinite number of incomplete trees covering a partial input.

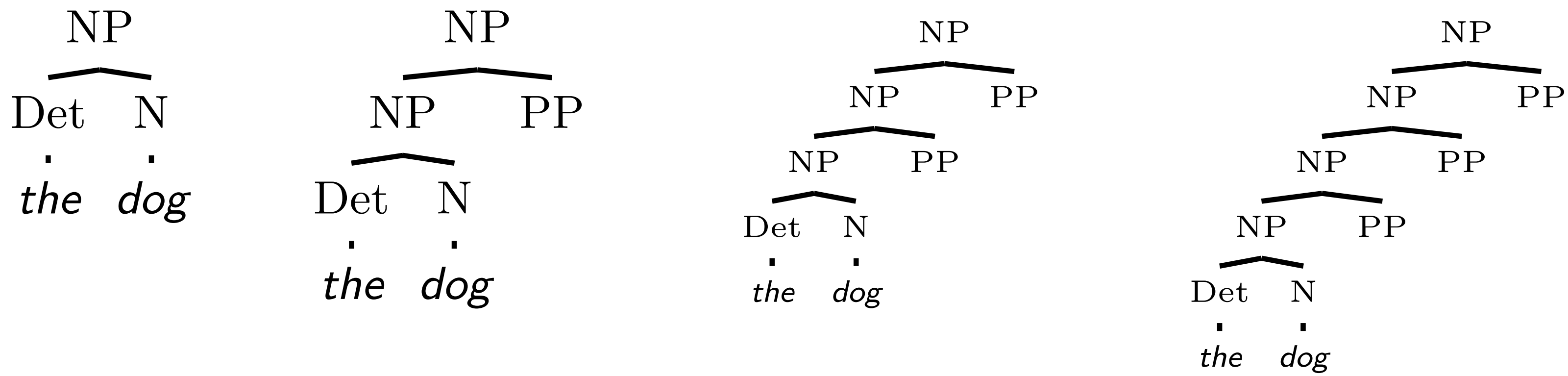


$\frac{4}{9}$ $\frac{4}{27}$ $\frac{4}{81}$ $\frac{4}{243}$

BUT! These tree probabilities form a geometric series:

$$
\begin{aligned}
P(\textit{the dog} \ldots) &= \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243} + \cdots \\
&= \frac{4}{9} \sum_{i=0}^{\infty} \left(\frac{1}{3}\right)^i \\
&= \frac{2}{3}
\end{aligned}
$$

Problem 2: there are still an infinite number of incomplete trees covering a partial input.



$\frac{4}{9}$ $\frac{4}{27}$ $\frac{4}{81}$ $\frac{4}{243}$

BUT! These tree probabilities form a geometric series:

$$\begin{aligned} P(\mathit{the\ dog}\ldots) &= \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243} + \cdots \\ &= \frac{4}{9} \sum_{i=0}^{\infty} \left(\frac{1}{3}\right)^i \\ &= \frac{2}{3} \end{aligned}$$

...which matches the original rule probability

$\frac{2}{3}$ $N \rightarrow dog$

Generalizing the geometric series induced by rule recursion

In general, these infinite tree sets arise due to *left recursion* in a probabilistic grammar

$$A \rightarrow B\ \alpha \qquad\qquad B \rightarrow A\ \beta$$

(Stolcke, 1995)

Generalizing the geometric series induced by rule recursion

In general, these infinite tree sets arise due to *left recursion* in a probabilistic grammar

$$A \rightarrow B\ \alpha \qquad\qquad B \rightarrow A\ \beta$$

We can formulate a stochastic *left-corner matrix* of transitions between categories:

$$P_L = \begin{array}{c|cccc} & A & B & \dots & K \\ \hline A & 0.3 & 0.7 & \cdots & 0 \\ B & 0.1 & 0.1 & \cdots & 0.2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & 0.2 & 0.1 & \cdots & 0.2 \end{array}$$

(Stolcke, 1995)

Generalizing the geometric series induced by rule recursion

In general, these infinite tree sets arise due to *left recursion* in a probabilistic grammar

$$A \rightarrow B\ \alpha \qquad\qquad B \rightarrow A\ \beta$$

We can formulate a stochastic *left-corner matrix* of transitions between categories:

$$P_L = \begin{array}{c|ccccc} & A & B & \dots & K \\ \hline A & 0.3 & 0.7 & \cdots & 0 \\ B & 0.1 & 0.1 & \cdots & 0.2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & 0.2 & 0.1 & \cdots & 0.2 \end{array}$$

and solve for its closure $R_L = (I - P_L)^{-1}$.

(Stolcke, 1995)

Generalizing the geometric series

1	ROOT	$\rightarrow$	NP	1	Det	$\rightarrow$	the
$\frac{2}{3}$	NP	$\rightarrow$	Det N	$\frac{2}{3}$	N	$\rightarrow$	dog
$\frac{1}{3}$	NP	$\rightarrow$	NP PP	$\frac{1}{3}$	N	$\rightarrow$	cat
1	PP	$\rightarrow$	P NP	1	P	$\rightarrow$	near

- The closure of our left-corner matrix is

$$
R_L = \begin{array}{c c}
 & \begin{array}{cccccc} \mathrm{ROOT} & \mathrm{NP} & \mathrm{PP} & \mathrm{Det} & \mathrm{N} & \mathrm{P} \end{array} \\
\begin{array}{c} \mathrm{ROOT} \\ \mathrm{NP} \\ \mathrm{PP} \\ \mathrm{Det} \\ \mathrm{N} \\ \mathrm{P} \end{array} &
\begin{pmatrix}
1 & \frac{3}{2} & 0 & 1 & 0 & 0 \\
0 & \frac{3}{2} & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
$$

Generalizing the geometric series

1	ROOT	$\rightarrow$ NP		1	Det	$\rightarrow$ the
$\frac{2}{3}$	NP	$\rightarrow$ Det N		$\frac{2}{3}$	N	$\rightarrow$ dog
$\frac{1}{3}$	NP	$\rightarrow$ NP PP		$\frac{1}{3}$	N	$\rightarrow$ cat
1	PP	$\rightarrow$ P NP		1	P	$\rightarrow$ near

- The closure of our left-corner matrix is

$$R_L = \begin{array}{c} \\ \mathrm{ROOT} \\ \mathrm{NP} \\ \mathrm{PP} \\ \mathrm{Det} \\ \mathrm{N} \\ \mathrm{P} \end{array} \begin{array}{c} \begin{array}{cccccc} \mathrm{ROOT} & \mathrm{NP} & \mathrm{PP} & \mathrm{Det} & \mathrm{N} & \mathrm{P} \end{array} \\ \begin{pmatrix} 1 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

- Refer to an entry (X, Y) in this matrix as $R(X \stackrel{*}{\Rightarrow}_L Y)$

Generalizing the geometric series

1	ROOT $\rightarrow$ NP		1	Det $\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N		$\frac{2}{3}$	N $\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP		$\frac{1}{3}$	N $\rightarrow$ cat
1	PP $\rightarrow$ P NP		1	P $\rightarrow$ near

- The closure of our left-corner matrix is

$$R_L = \begin{array}{c|c} & \begin{array}{cccccc} \mathrm{ROOT} & \mathrm{NP} & \mathrm{PP} & \mathrm{Det} & \mathrm{N} & \mathrm{P} \end{array} \\ \begin{array}{c} \mathrm{ROOT} \\ \mathrm{NP} \\ \mathrm{PP} \\ \mathrm{Det} \\ \mathrm{N} \\ \mathrm{P} \end{array} & \begin{pmatrix} 1 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

- Refer to an entry (X, Y) in this matrix as $R(X \stackrel{*}{\Rightarrow}_L Y)$
- Note that the $\frac{3}{2}$ "bonus" accrued for left-recursion of NPs appears in the (ROOT,NP) and (NP,NP) cells of the matrix

Generalizing the geometric series

1	ROOT	$\rightarrow$	NP	1	Det	$\rightarrow$	the
$\frac{2}{3}$	NP	$\rightarrow$	Det N	$\frac{2}{3}$	N	$\rightarrow$	dog
$\frac{1}{3}$	NP	$\rightarrow$	NP PP	$\frac{1}{3}$	N	$\rightarrow$	cat
1	PP	$\rightarrow$	P NP	1	P	$\rightarrow$	near

- The closure of our left-corner matrix is

$$R_L = \begin{array}{c} \\ \mathrm{ROOT} \\ \mathrm{NP} \\ \mathrm{PP} \\ \mathrm{Det} \\ \mathrm{N} \\ \mathrm{P} \end{array} \begin{array}{c} \begin{array}{cccccc} \mathrm{ROOT} & \mathrm{NP} & \mathrm{PP} & \mathrm{Det} & \mathrm{N} & \mathrm{P} \end{array} \\ \begin{pmatrix} 1 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

- Refer to an entry (X, Y) in this matrix as $R(X \stackrel{*}{\Rightarrow}_L Y)$
- Note that the $\frac{3}{2}$ "bonus" accrued for left-recursion of NPs appears in the (ROOT,NP) and (NP,NP) cells of the matrix
- We need to do the same with unary chains, constructing a unary-closure matrix R_U.

Efficient incremental parsing: the probabilistic Earley algorithm

We can use the Earley algorithm (Earley, 1970) in a probabilistic incarnation (Stolcke, 1995) to deal with these infinite tree sets.

The (slightly oversimplified) probabilistic Earley algorithm has two fundamental types of operations:

- **Prediction**: if Y is a possible goal, and Y can lead to Z through a left corner, choose a rule $Z \rightarrow \alpha$ and set up α as a new sequence of possible goals.
- **Completion**: if Y is a possible goal, Y can lead to Z through unary rewrites, and we encounter a completed Z, absorb it and move on to the next sub-goal in the sequence.

Efficient incremental parsing: the probabilistic Earley algorithm

- Parsing consists of constructing a *chart* of *states* (items)
- A state has the following structure:

Goal symbol

Completed subgoals

Remaining subgoals

$$\mathrm{X} \rightarrow \alpha \circ \beta$$

$$p \qquad q$$

Forward probability

Inside probability

- The *forward* probability is the total probability of getting from the root at the start of the sentence through to this state
- The *inside* probability is the "bottom-up" probability of the state

Efficient incremental parsing: the probabilistic Earley algorithm

Inference rules for probabilistic Earley:

- **Prediction**:

$$\frac{\begin{array}{ll} X \rightarrow \beta \circ Y\gamma \\ p \qquad\qquad q \end{array} \quad a : R(Y \stackrel{*}{\Rightarrow}_L Z) \quad b : Z \rightarrow \alpha}{\begin{array}{l} Z \rightarrow \circ\alpha \\ abp \qquad b \end{array}}$$

Efficient incremental parsing: the probabilistic Earley algorithm

Inference rules for probabilistic Earley:

- **Prediction**:

$$\frac{\underset{p}{X} \rightarrow \beta \circ \underset{q}{Y} \gamma \qquad a : R(Y \stackrel{*}{\Rightarrow}_L Z) \qquad b : Z \rightarrow \alpha}{\underset{abp}{Z} \rightarrow \circ \underset{b}{\alpha}}$$

- **Completion**:

$$\frac{\underset{p}{X} \rightarrow \beta \circ \underset{q}{Y} \gamma \qquad a : R(Y \stackrel{*}{\Rightarrow}_U Z) \qquad \underset{b}{Z} \rightarrow \underset{c}{\alpha \circ}}{\underset{acp}{X} \rightarrow \beta Y \circ \underset{acq}{\gamma}}$$

Efficient incremental parsing: probabilistic Earley

the dog near the

Efficient incremental parsing: probabilistic Earley

ROOT$\rightarrow\circ$NP
1 1

the dog near the

Efficient incremental parsing: probabilistic Earley

Det→∘the
1 1

NP→∘Det N
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

NP→∘NP PP
$\frac{1}{3} \times \frac{3}{2}$ $\frac{1}{3}$

ROOT→∘NP
1 1

the dog near the

Efficient incremental parsing: probabilistic Earley

$\text{Det} \rightarrow \circ\,\text{the}$

$1 \qquad\qquad 1$

$\text{NP} \rightarrow \circ\,\text{Det N}$

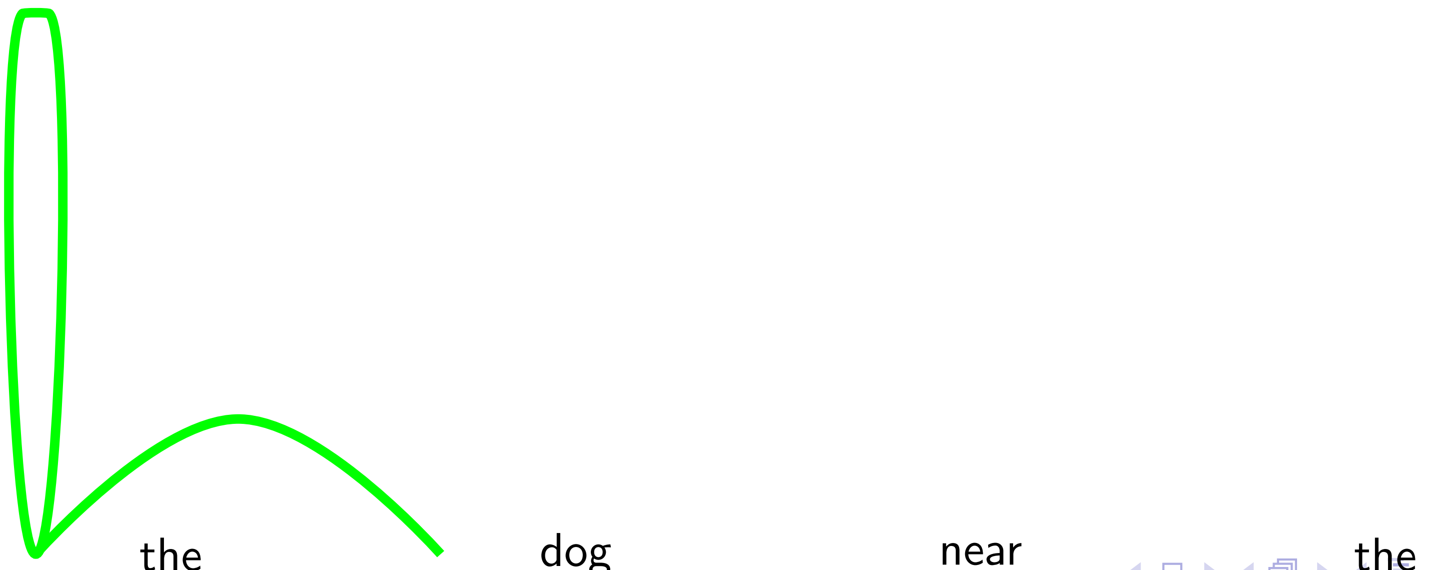
$\frac{2}{3} \times \frac{3}{2} \qquad \frac{2}{3}$

$\text{NP} \rightarrow \circ\,\text{NP PP}$

$\frac{1}{3} \times \frac{3}{2} \qquad \frac{1}{3}$

$\text{ROOT} \rightarrow \circ\,\text{NP}$

$1 \qquad\qquad 1$



Efficient incremental parsing: probabilistic Earley

Det$\rightarrow\circ$the
1 1

NP$\rightarrow\circ$Det N
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

NP$\rightarrow\circ$NP PP
$\frac{1}{3} \times \frac{3}{2}$ $\frac{1}{3}$

ROOT$\rightarrow\circ$NP
1 1

NP$\rightarrow$Det$\circ$N
1 $\frac{2}{3}$

Det$\rightarrow$the$\circ$
1 1

the dog near the

Efficient incremental parsing: probabilistic Earley

Det→∘the
1 1

NP→∘Det N
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

NP→∘NP PP
$\frac{1}{3} \times \frac{3}{2}$ $\frac{1}{3}$

ROOT→∘NP
1 1

NP→Det∘N
1 $\frac{2}{3}$

Det→the∘
1 1

the dog near the

Efficient incremental parsing: probabilistic Earley

Det→∘the
1 1

NP→∘Det N
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

NP→∘NP PP
$\frac{1}{3} \times \frac{3}{2}$ $\frac{1}{3}$

ROOT→∘NP
1 1

N→∘cat
$\frac{1}{3}$ $\frac{1}{3}$

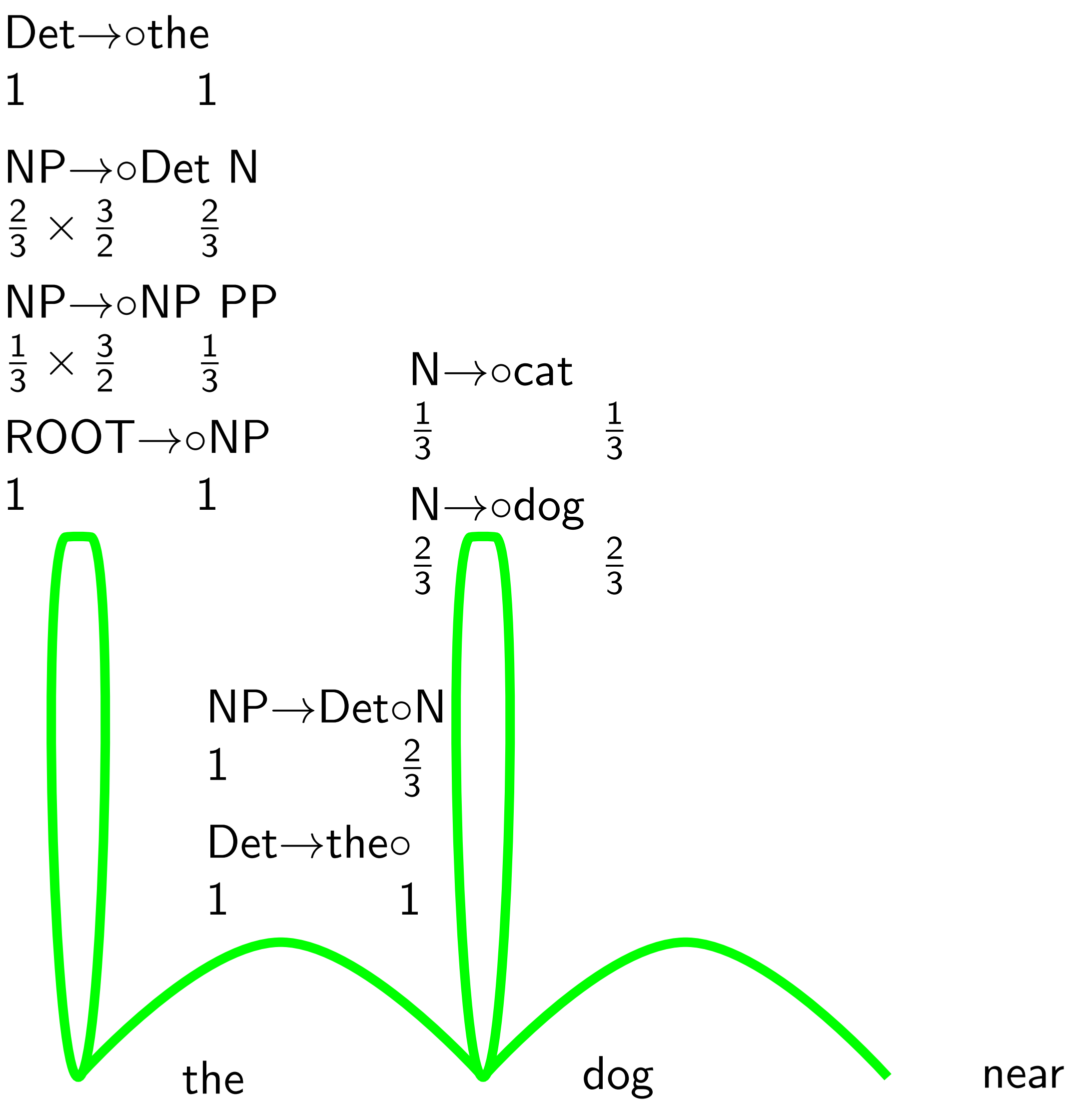
N→∘dog
$\frac{2}{3}$ $\frac{2}{3}$

NP→Det∘N
1 $\frac{2}{3}$

Det→the∘
1 1

the dog near the

Efficient incremental parsing: probabilistic Earley



Efficient incremental parsing: probabilistic Earley

Det→∘the
1 1

NP→∘Det N
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

NP→∘NP PP
$\frac{1}{3} \times \frac{3}{2}$ $\frac{1}{3}$

ROOT→∘NP
1 1

N→∘cat
$\frac{1}{3}$ $\frac{1}{3}$

N→∘dog
$\frac{2}{3}$ $\frac{2}{3}$

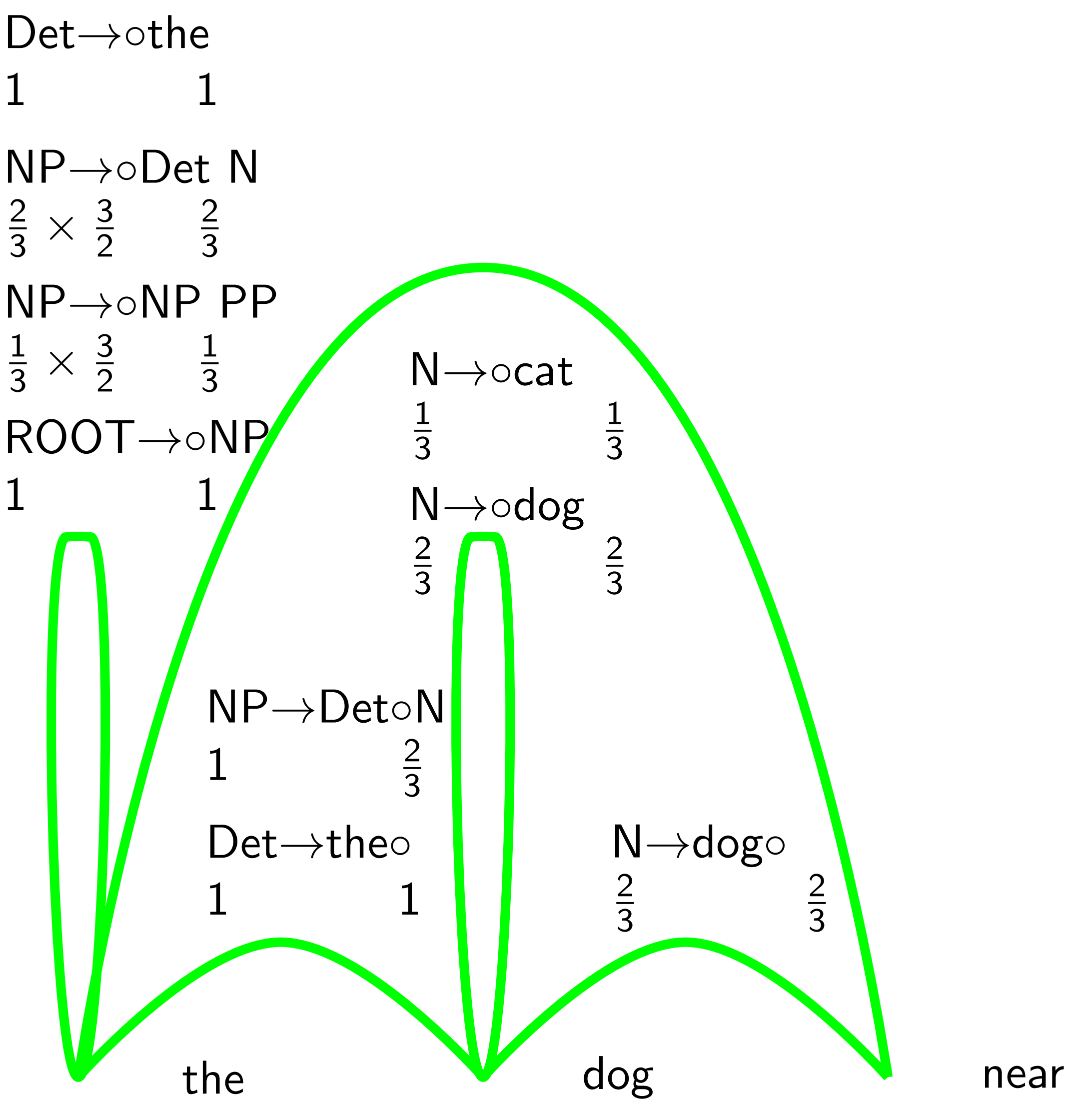
NP→Det∘N
1 $\frac{2}{3}$

Det→the∘
1 1

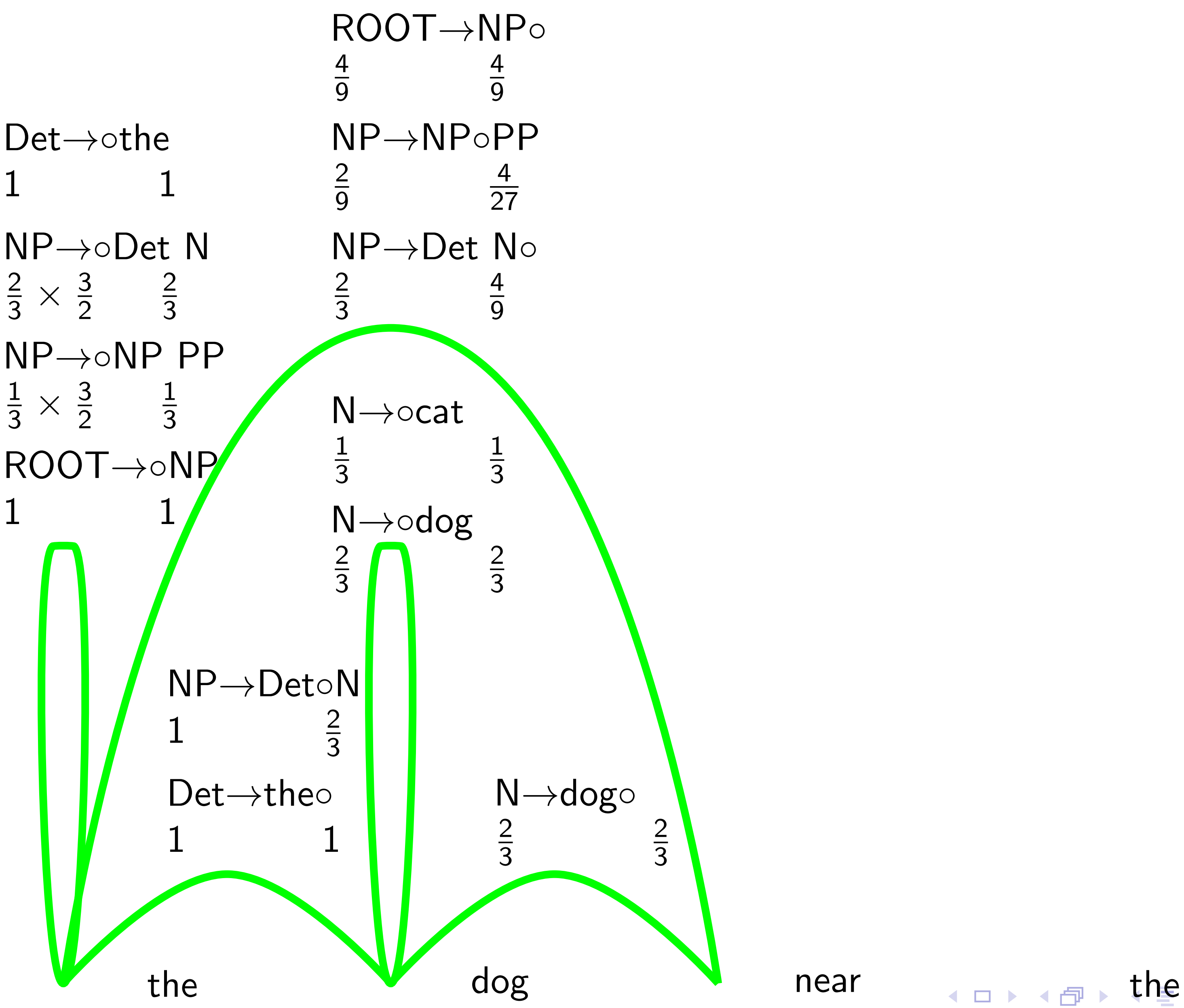
N→dog∘
$\frac{2}{3}$ $\frac{2}{3}$

the dog near the

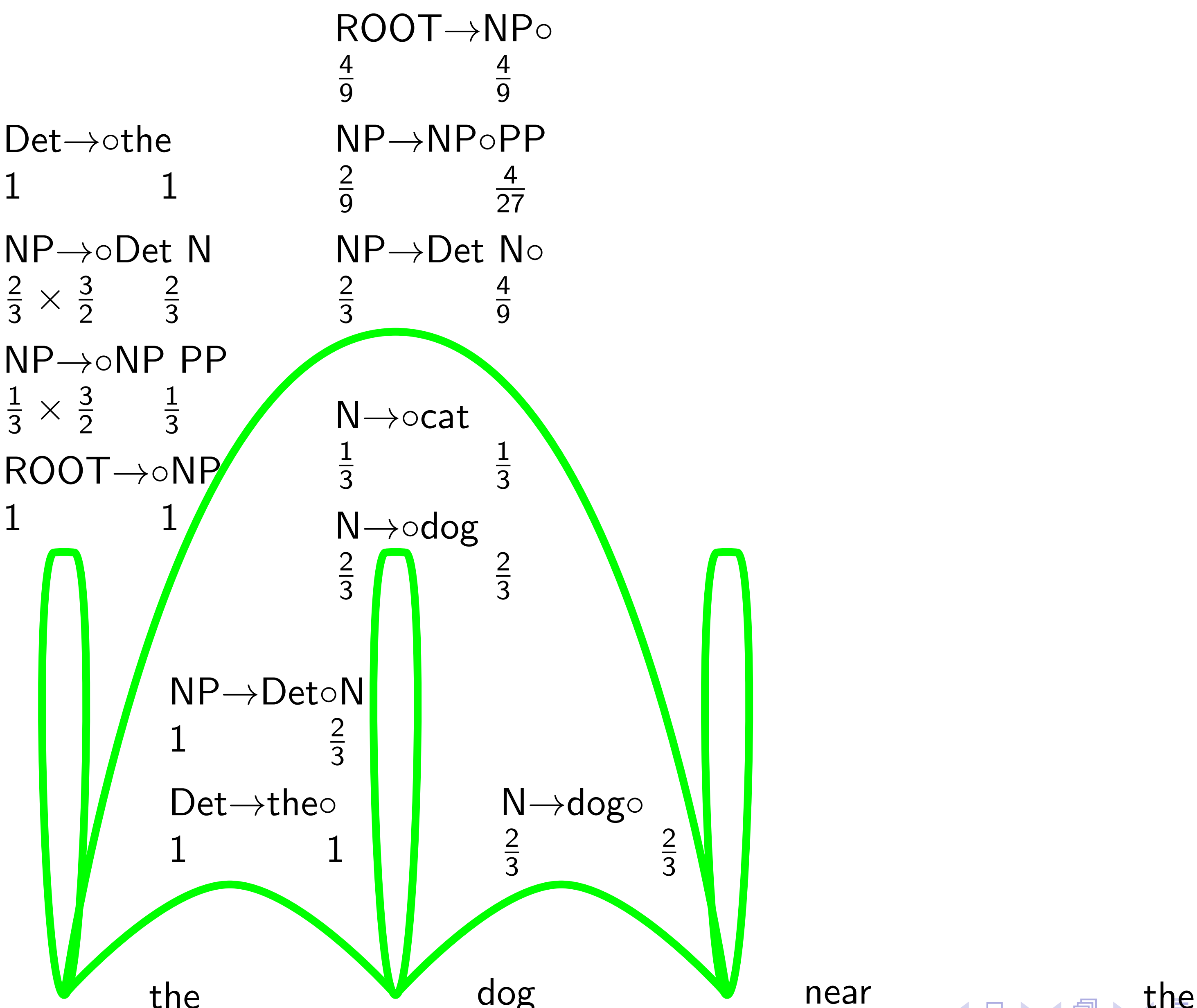
Efficient incremental parsing: probabilistic Earley



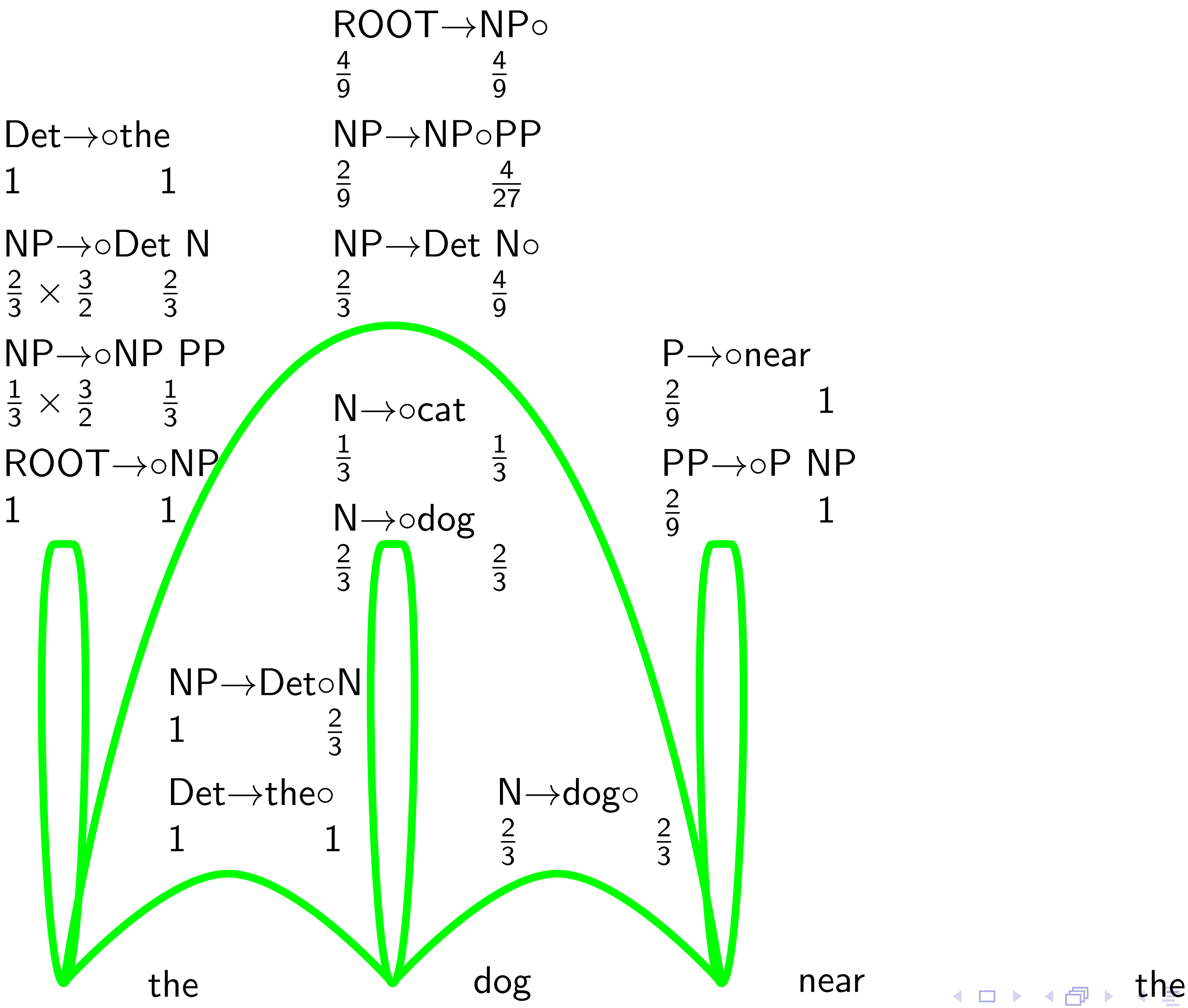
Efficient incremental parsing: probabilistic Earley



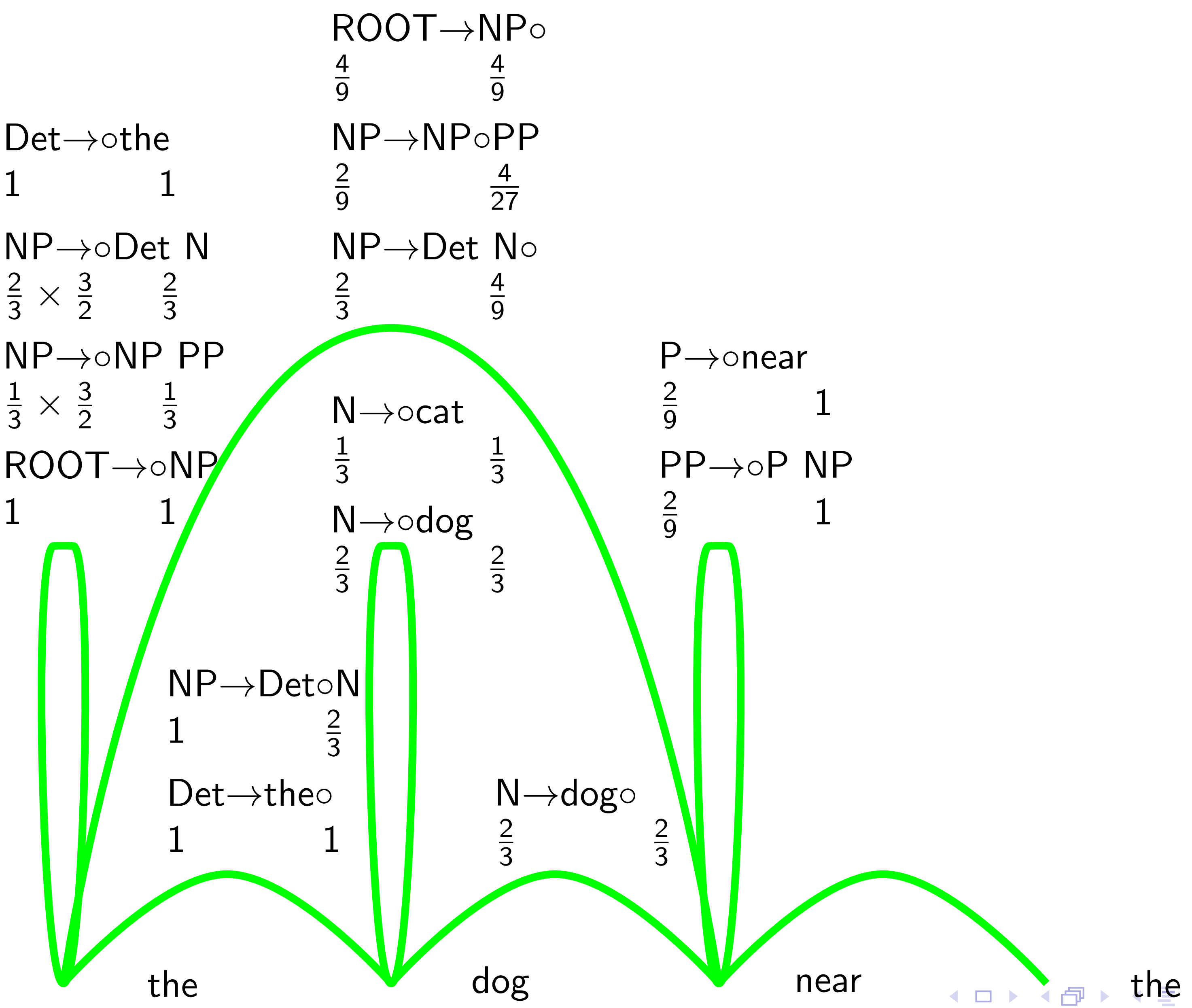
Efficient incremental parsing: probabilistic Earley



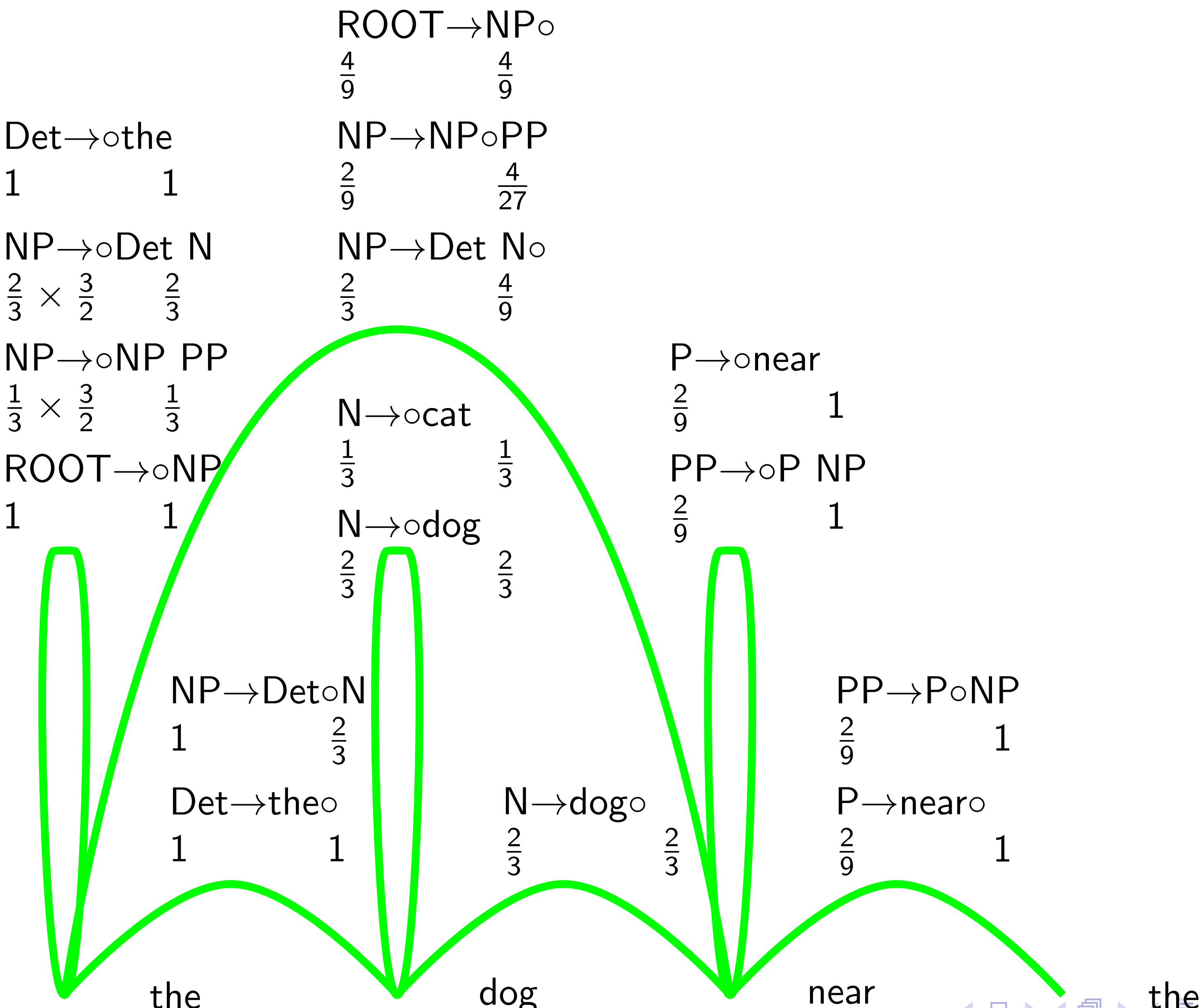
Efficient incremental parsing: probabilistic Earley



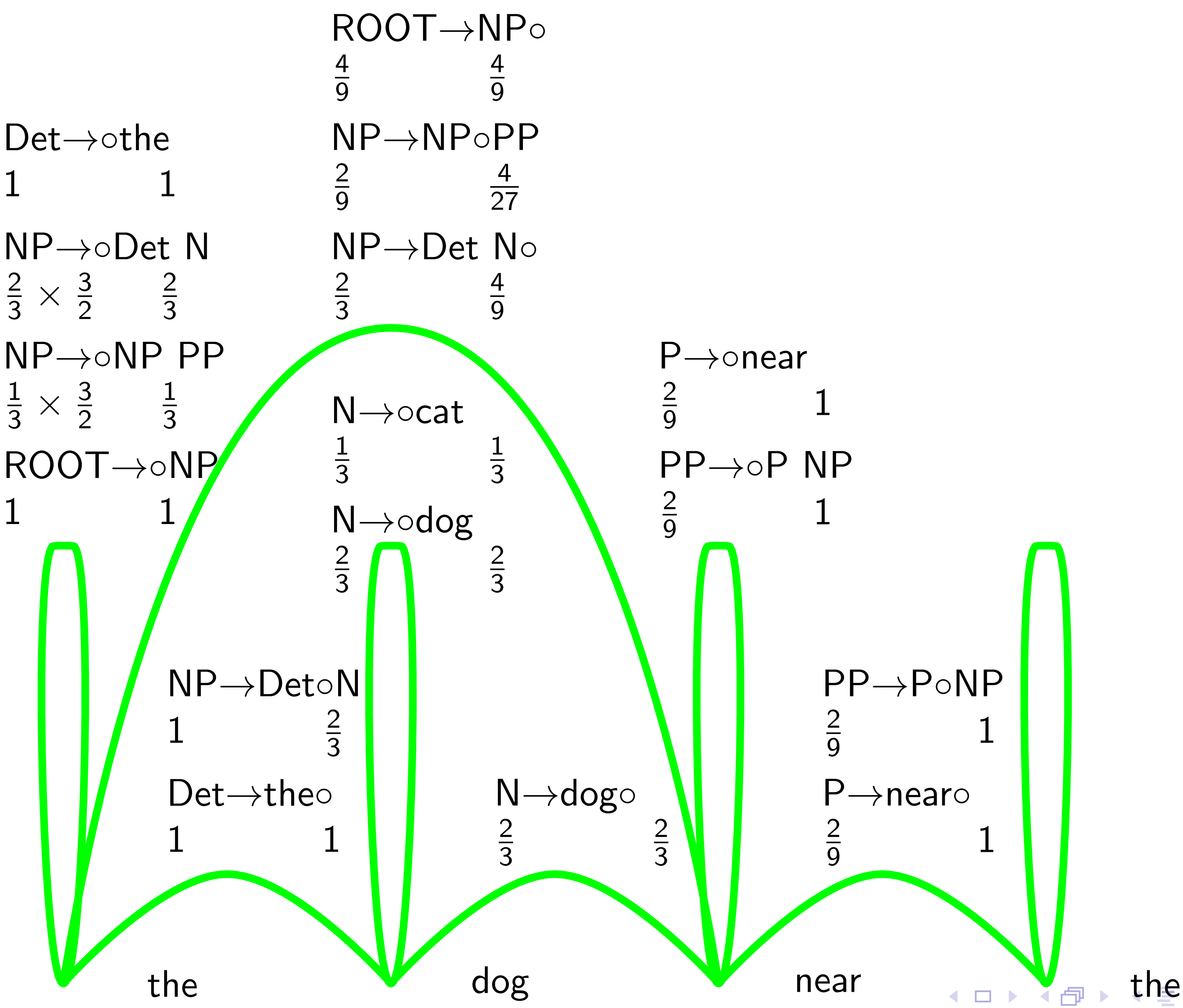
Efficient incremental parsing: probabilistic Earley



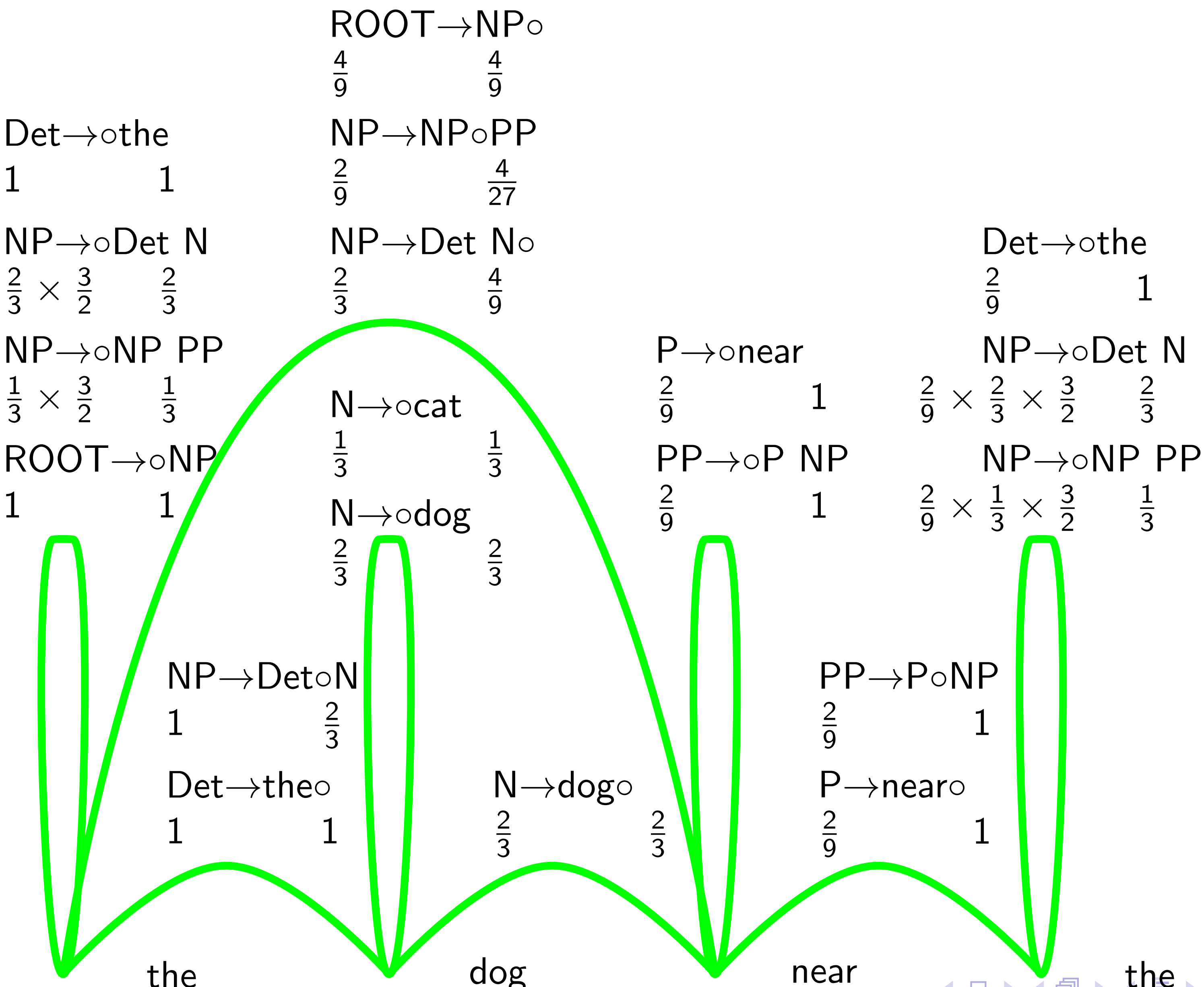
Efficient incremental parsing: probabilistic Earley



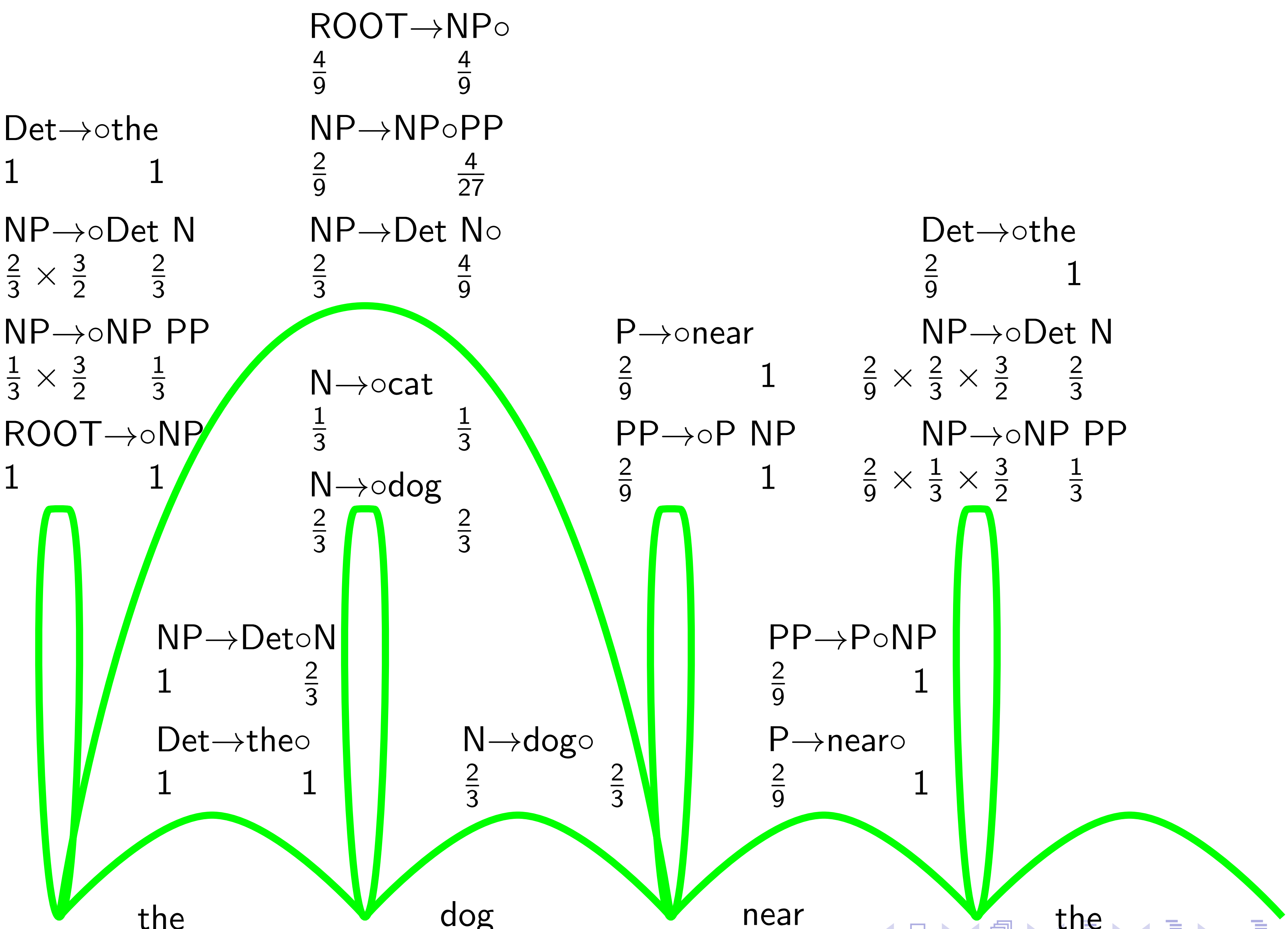
Efficient incremental parsing: probabilistic Earley



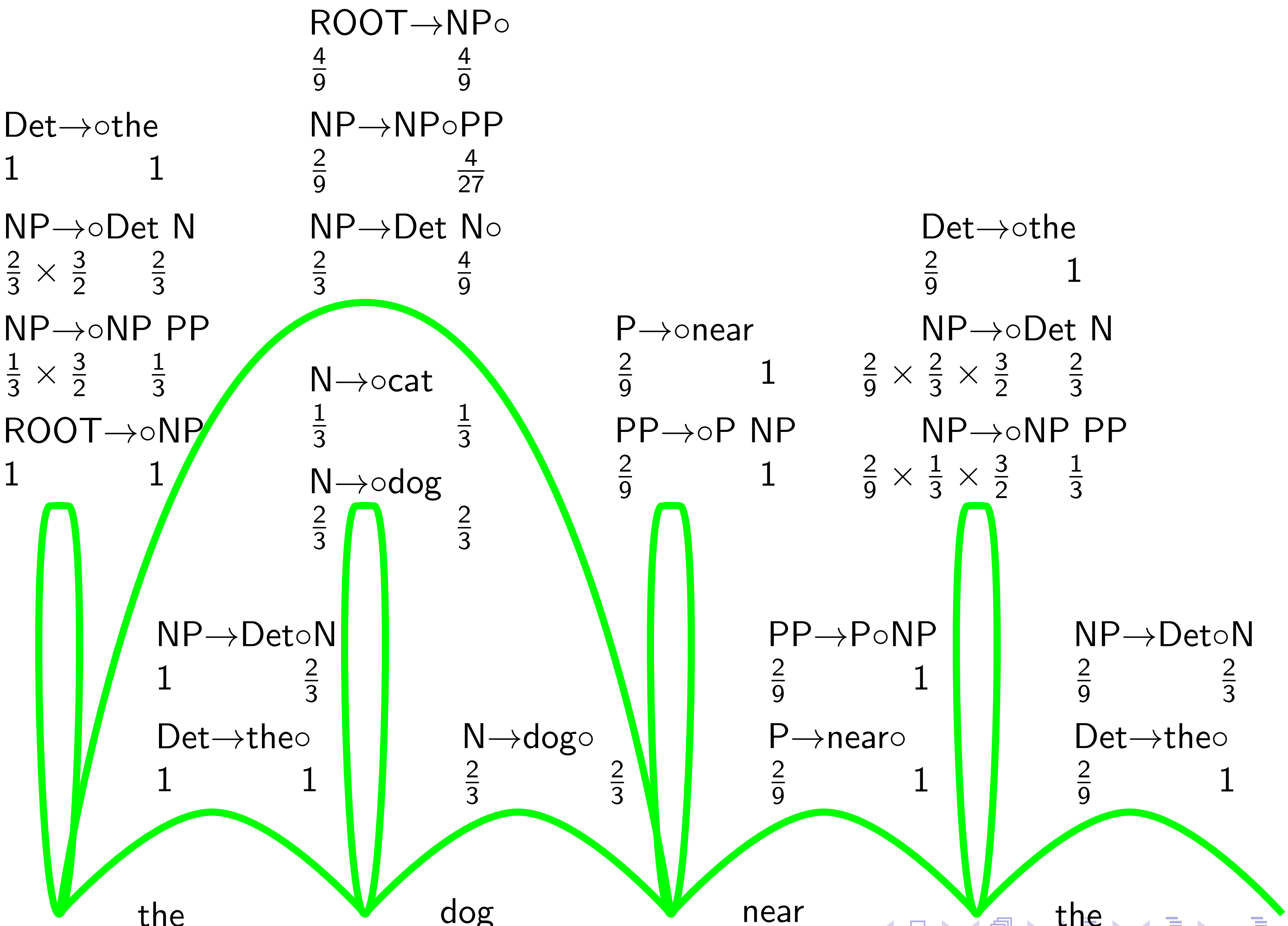
Efficient incremental parsing: probabilistic Earley



Efficient incremental parsing: probabilistic Earley



Efficient incremental parsing: probabilistic Earley



Prefix probabilities from probabilistic Earley

- If you have just processed word w_i, then the prefix probability of $w_{1...i}$ can be obtained by summing all forward probabilities of items that have the form $X \to \alpha w_i \circ \beta$

Prefix probabilities from probabilistic Earley

- If you have just processed word w_i, then the prefix probability of $w_{1\ldots i}$ can be obtained by summing all forward probabilities of items that have the form $X \to \alpha w_i \circ \beta$
- In our example, we see:

$$
\begin{aligned}
&P(\text{the}) = 1 \\
&P(\text{the dog}) = \tfrac{2}{3} \\
&P(\text{the dog near}) = \tfrac{2}{9} \\
&P(\text{the dog near the}) = \tfrac{2}{9}
\end{aligned}
$$

Prefix probabilities from probabilistic Earley

- If you have just processed word w_i, then the prefix probability of $w_{1\ldots i}$ can be obtained by summing all forward probabilities of items that have the form $X \to \alpha w_i \circ \beta$
- In our example, we see:

$$
\begin{aligned}
&P(\text{the}) = 1 \\
&P(\text{the dog}) = \tfrac{2}{3} \\
&P(\text{the dog near}) = \tfrac{2}{9} \\
&P(\text{the dog near the}) = \tfrac{2}{9}
\end{aligned}
$$

- Taking the ratios of these prefix probabilities can give us conditional word probabilities

Probabilistic Earley as an "eager" algorithm

- From the *inside probabilities* of the states on the chart, the posterior distribution on (incremental) trees can be directly calculated
- This posterior distribution is *precisely* the correct result of the application of Bayes' rule:

$$P(T_{\text{incremental}} | w_{1\ldots i}) = \frac{P(w_{1\ldots i}, T_{\text{incremental}})}{P(w_{1\ldots i})}$$

- Hence, probabilistic Earley is also performing rational disambiguation
- Hale (2001) called this the "eager" property of an incremental parsing algorithm.

Probabilistic Earley algorithm: key ideas

- We want to use probabilistic grammars for both disambiguation and calculating probability distributions over upcoming events
- Infinitely many trees can be constructed in polynomial time ( ) and space ( )
- The *prefix probability* of the string is calculated in the process
- By taking the log-ratio of two prefix probabilities, the surprisal of a word in its context can be calculated

Probabilistic Earley algorithm: key ideas

- We want to use probabilistic grammars for both disambiguation and calculating probability distributions over upcoming events
- Infinitely many trees can be constructed in polynomial time ($O(n^3)$) and space ( )
- The *prefix probability* of the string is calculated in the process
- By taking the log-ratio of two prefix probabilities, the surprisal of a word in its context can be calculated

Probabilistic Earley algorithm: key ideas

- We want to use probabilistic grammars for both disambiguation and calculating probability distributions over upcoming events
- Infinitely many trees can be constructed in polynomial time ($O(n^3)$) and space ($O(n^2)$)
- The *prefix probability* of the string is calculated in the process
- By taking the log-ratio of two prefix probabilities, the surprisal of a word in its context can be calculated

Probabilistic ambiguity resolution

- Let's consider another case of ambiguity:

The complex houses married students and their families.

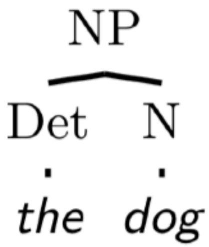
The prime number few.

- **In-class exercise:** develop a PCFG in which which the "garden-path" analysis is strongly disfavored

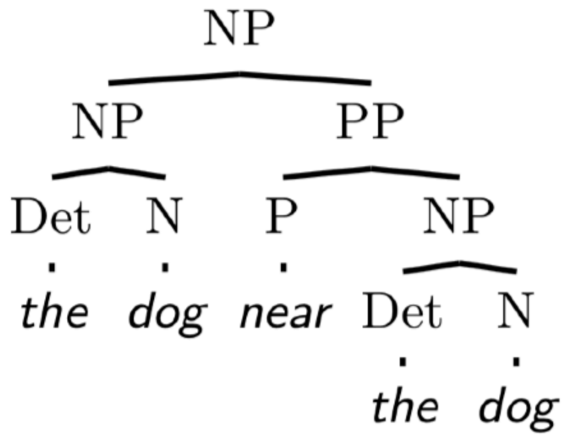
		1	Det $\rightarrow$ the
$\frac{2}{3}$	NP $\rightarrow$ Det N	$\frac{2}{3}$	N $\rightarrow$ dog
$\frac{1}{3}$	NP $\rightarrow$ NP PP	$\frac{1}{3}$	N $\rightarrow$ cat
1	PP $\rightarrow$ P NP	1	P $\rightarrow$ near

Incrementality: you can think of a *partial* tree as marginalizing over all completions of the partial tree.

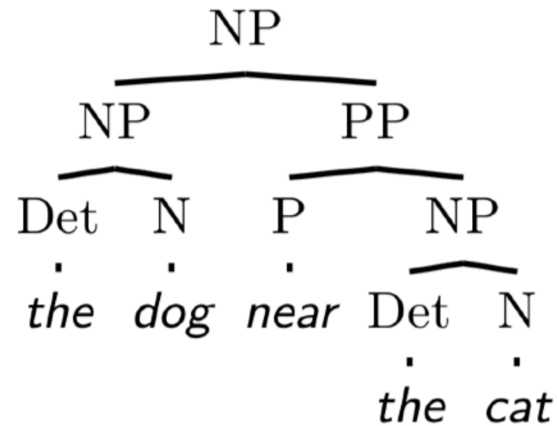
It has a corresponding marginal probability in the PCFG.



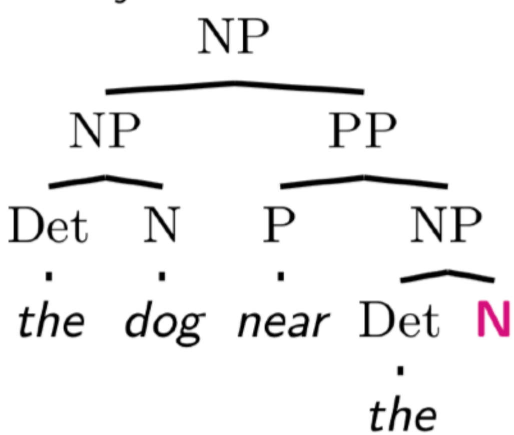
$\frac{4}{9}$



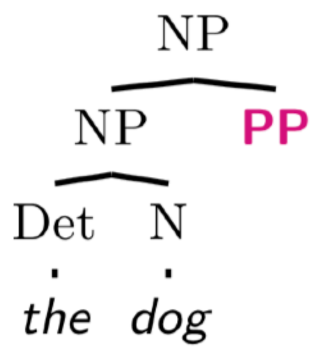
$\frac{8}{81}$



$\frac{4}{81}$

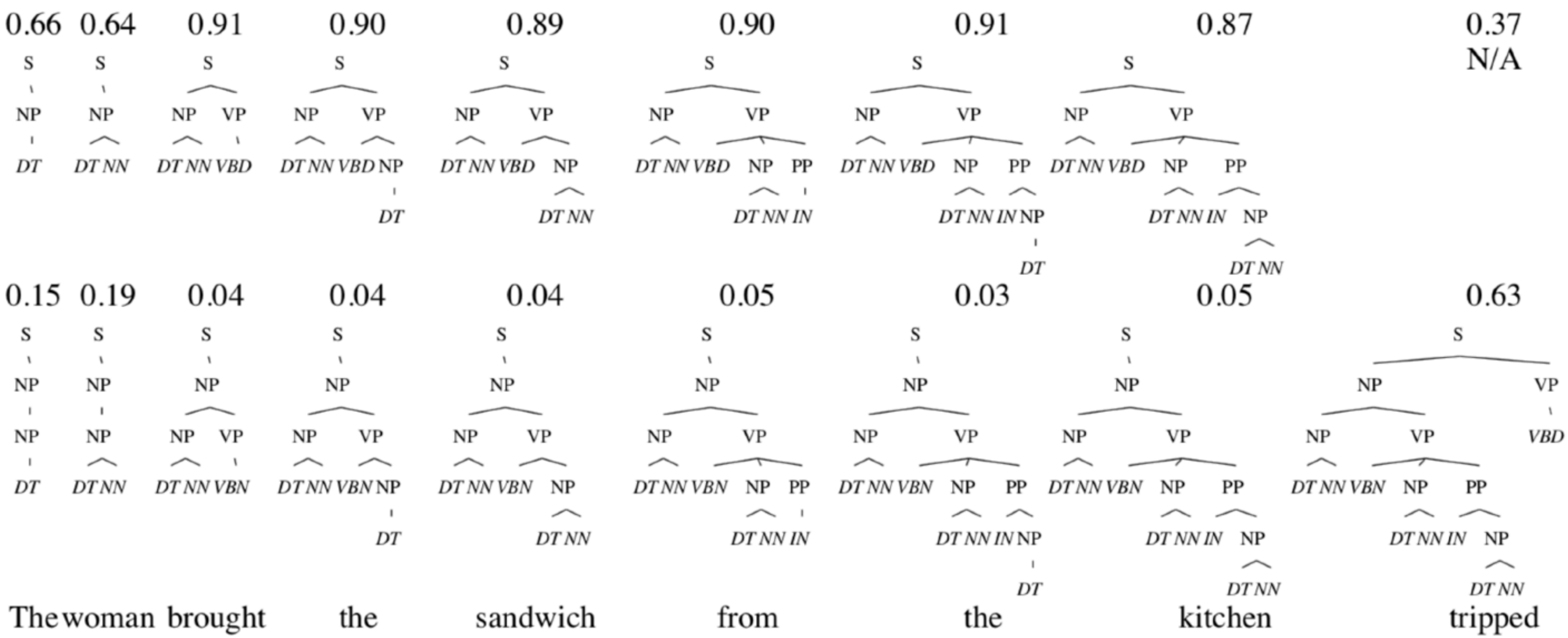


$\frac{12}{81}$



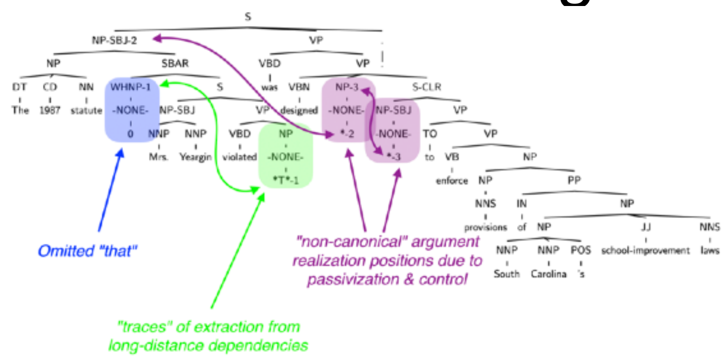
$\frac{4}{27}$

Our more complex examples

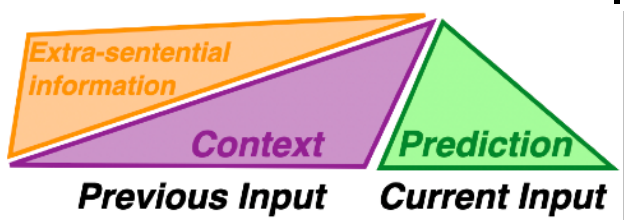


Ingredients for modeling human syntactic processing

- Estimate of statistics of the linguistic environment



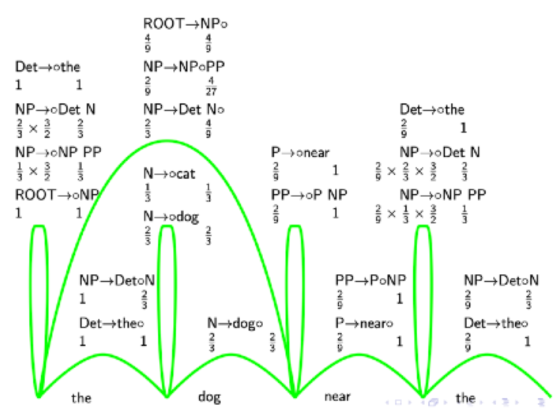
- Focus on predictive, incremental processing



- An incremental probabilistic (Earley) parsing model

$\frac{2}{3}$	NP	$\rightarrow$	Det	N
$\frac{1}{3}$	NP	$\rightarrow$	NP	PP
1	PP	$\rightarrow$	P	NP

1	Det	$\rightarrow$	the
$\frac{2}{3}$	N	$\rightarrow$	dog
$\frac{1}{3}$	N	$\rightarrow$	cat
1	P	$\rightarrow$	near



Human real-time syntactic processing

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$
$$\left[\approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

- Captures the *expectation* intuition: the more we expect an event, the easier it is to process
  - Brains are prediction engines!

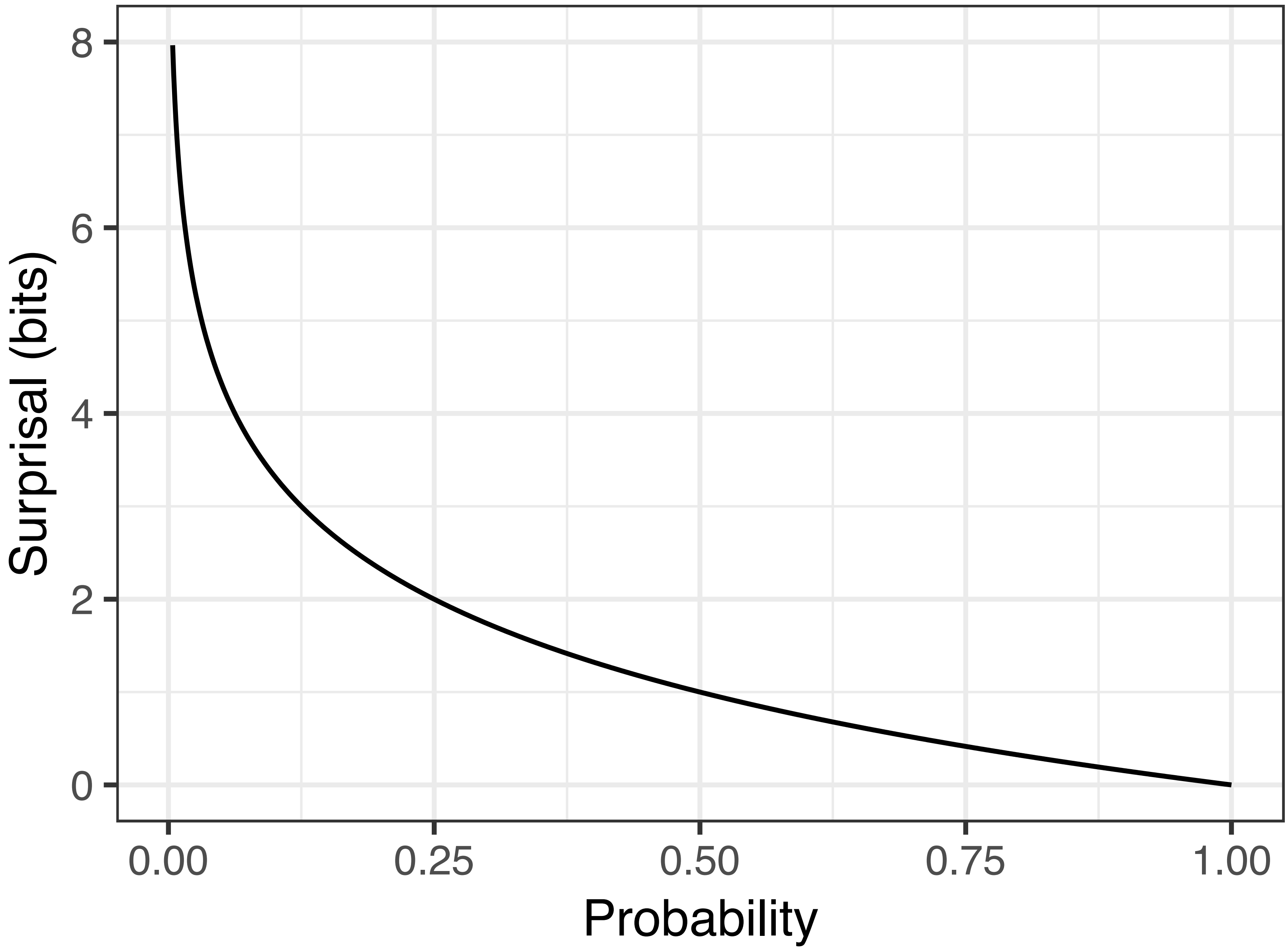
    my brother came inside to… chat? wash? get warm?

    the children went outside to… play

  - Predictable words are read faster (Ehrlich & Rayner, 1981) and have distinctive EEG responses (Kutas & Hillyard 1980)
- Combine with probabilistic grammars to give *grammatical expectations*

(*Hale, 2001, NAACL; Levy, 2008, Cognition*)

The surprisal graph



Garden-pathing and surprisal

- Here's a *local syntactic ambiguity*

When the dog scratched the vet and his new assistant removed the muzzle.

difficulty here
(68ms/char)

- Compare with:

When the dog scratched, the vet and his new assistant removed the muzzle.

When the dog scratched its owner the vet and his new assistant removed the muzzle.

easier
(50ms/char)

(Frazier & Rayner, 1982)

A small PCFG for this sentence type

S	$\rightarrow$	SBAR S	0.3
S	$\rightarrow$	NP VP	0.7
SBAR	$\rightarrow$	COMPL S	0.3
SBAR	$\rightarrow$	COMPL S COMMA	0.7
COMPL	$\rightarrow$	When	1
NP	$\rightarrow$	Det N	0.6
NP	$\rightarrow$	Det Adj N	0.2
NP	$\rightarrow$	NP Conj NP	0.2

Conj	$\rightarrow$	and	1
Det	$\rightarrow$	the	0.8
Det	$\rightarrow$	its	0.1
Det	$\rightarrow$	his	0.1
N	$\rightarrow$	dog	0.2
N	$\rightarrow$	vet	0.2
N	$\rightarrow$	assistant	0.2
N	$\rightarrow$	muzzle	0.2
N	$\rightarrow$	owner	0.2

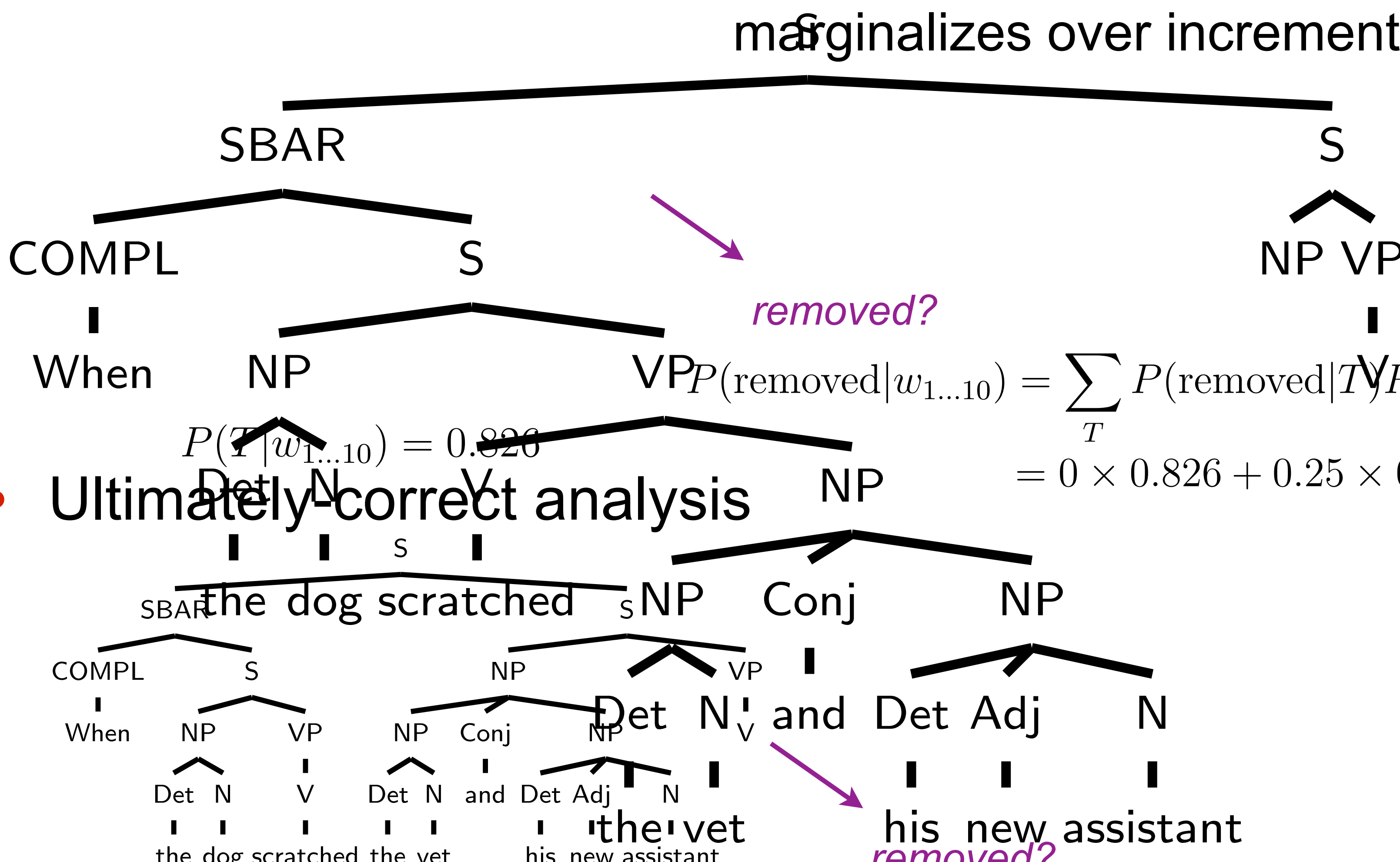
Adj	$\rightarrow$	new	1
VP	$\rightarrow$	V NP	0.5
VP	$\rightarrow$	V	0.5
V	$\rightarrow$	scratched	0.25
V	$\rightarrow$	removed	0.25
V	$\rightarrow$	arrived	0.5
COMMA	$\rightarrow$	,	1

(analysis in Levy, 2013)

Two incremental trees

- "Garden-path" analysis:

Disambiguating word probability marginalizes over incremental trees:



$P(T|w_{1\ldots10}) = 0.826$

removed?

$$P(\text{removed}|w_{1\ldots10}) = \sum_T P(\text{removed}|T)P(T|w_{1\ldots10})$$
$$= 0 \times 0.826 + 0.25 \times 0.174$$

- Ultimately-correct analysis

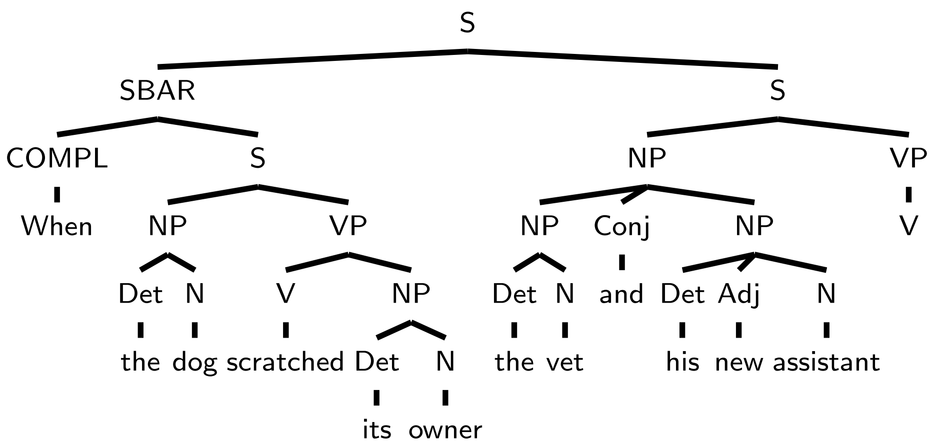
removed?

$P(T|w_{1\ldots10}) = 0.174$

(analysis in Levy, 2013)

Preceding context can disambiguate

- *"its owner"* takes up the object slot of *scratched*



Condition	Surprisal at Resolution
NP absent	4.2
NP present	2

Sensitivity to verb argument structure

- A superficially similar example:

When the dog arrived the vet and his new assistant removed the muzzle.

But harder here!

Easier here

(c.f. When the dog scratched the vet and his new assistant removed the muzzle.)

(Staub, 2007)

Modeling argument-structure sensitivity

S	→	SBAR S	0.3	Conj	→	and	1	Adj	→	new	1
S	→	NP VP	0.7	Det	→	the	0.8	VP	→	V NP	0.5
SBAR	→	COMPL S	0.3	Det	→	its	0.1	VP	→	V	0.5
SBAR	→	COMPL S COMMA	0.7	Det	→	his	0.1	V	→	scratched	0.25
COMPL	→	When	1	N	→	dog	0.2	V	→	removed	0.25
NP	→	Det N	0.6	N	→	vet	0.2	V	→	arrived	0.5
NP	→	Det Adj N	0.2	N	→	assistant	0.2	COMMA	→	,	1
NP	→	NP Conj NP	0.2	N	→	muzzle	0.2				
				N	→	owner	0.2				

- The "context-free" assumption doesn't preclude relaxing probabilistic locality:

VP	→	V NP	0.5
VP	→	V	0.5
V	→	scratched	0.25
V	→	removed	0.25
V	→	arrived	0.5

Replaced by ⇒

VP	→	Vtrans NP	0.45
VP	→	Vtrans	0.05
VP	→	Vintrans	0.45
VP	→	Vintrans NP	0.05
Vtrans	→	scratched	0.5
Vtrans	→	removed	0.5
Vintrans	→	arrived	1

(Johnson, 1998; Klein & Manning, 2003)

Result

When the dog arrived the vet and his new assistant removed the muzzle.

ambiguity onset

ambiguity resolution

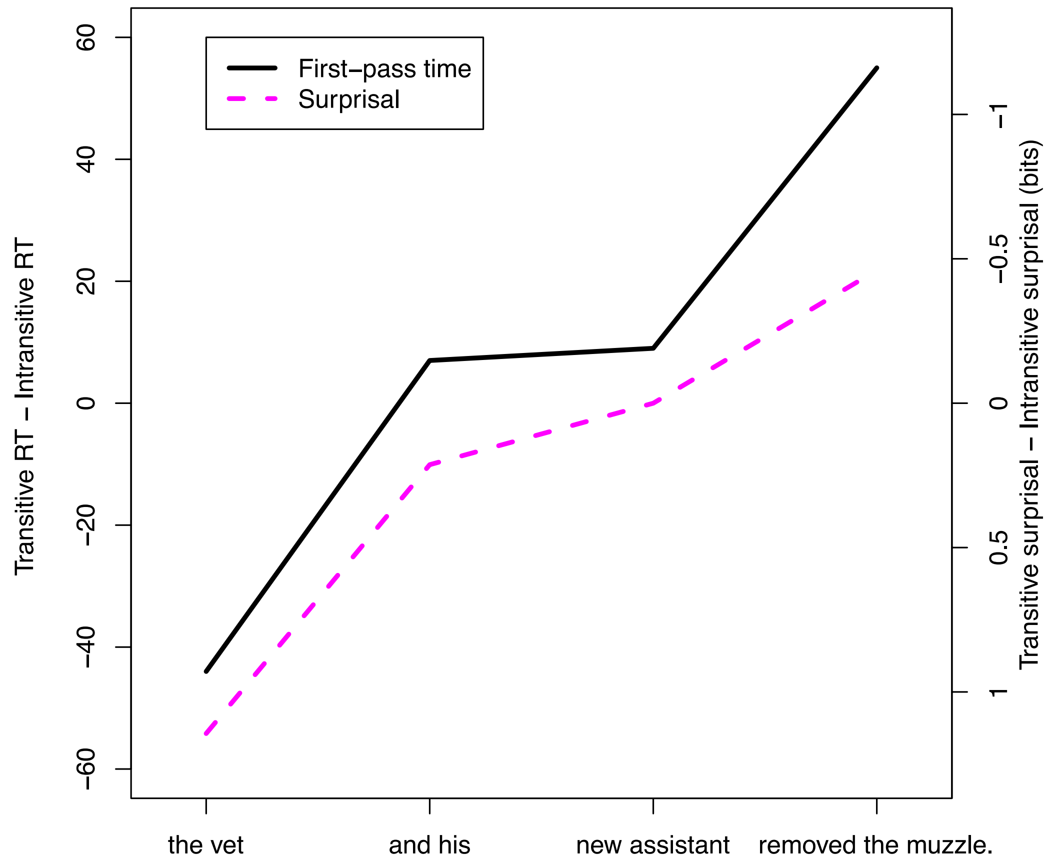
When the dog scratched the vet and his new assistant removed the muzzle.

Transitivity-distinguishing PCFG

Condition	Ambiguity onset	Resolution
Intransitive (arrived)	2.11	3.20
Transitive (scratched)	0.44	8.04

Move to broad coverage

- Instead of the pedagogical grammar, a “broad-coverage” grammar from the parsed Brown corpus (11,984 rules)
- Relative-frequency estimation of rule probabilities (“vanilla” PCFG)
- (We’ll discuss these estimation techniques next class)



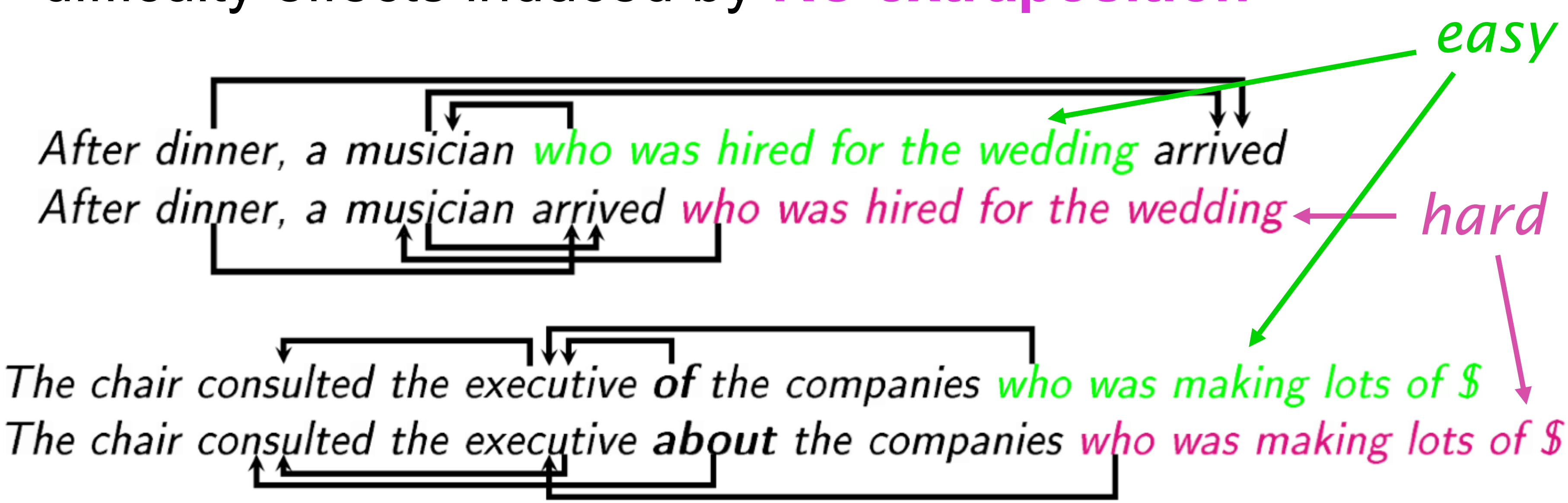
Syntactic complexity--non-probabilistic

- On the *resource limitation* view, memory demands are a "processing bottleneck"
- Gibson 1998, 2000 (DLT): multiple and/or more distant dependencies are harder to process

	Processing
the reporter who attacked the senator	***Easy***
the reporter who the senator attacked	***Hard***

Rethinking locality: RC extraposition

- Equipped with a theory of probabilistic expectations, let's revisit more "memory"-oriented results
- Example: Levy et al. (2012) found found consistent difficulty effects induced by **RC extraposition**

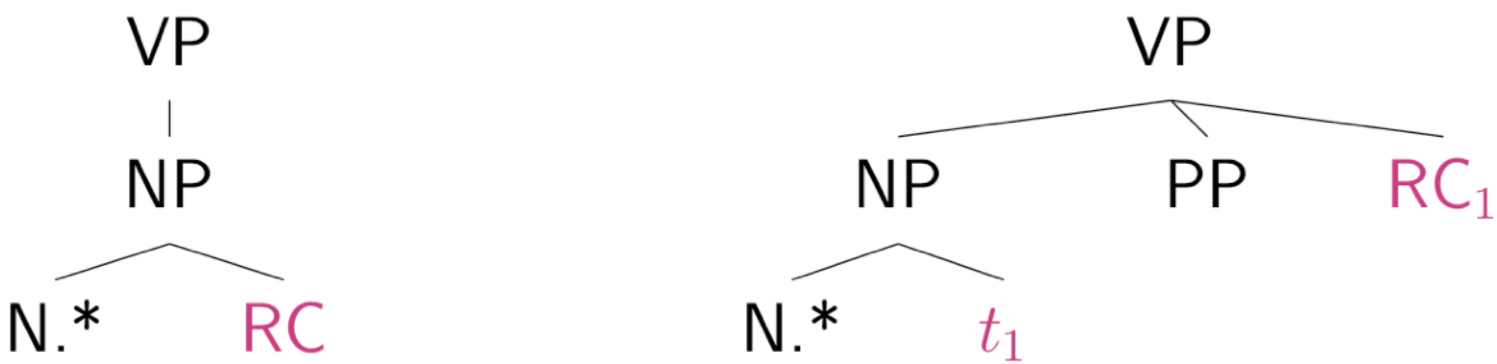


- Is this evidence for a special type of locality: a *phrasal adjacency constraint* (or a constraint against crossing dependencies)?

(Levy, Fedorenko, Breen, & Gibson, 2012)

Probability & extraposition

- But…
- …RC extraposition is relatively rare in English



In situ: $P_{VP}(RC|NP)=0.06$ Extraposed: $P_{VP}(RC|NP,PP)=0.003$

(estimated from the parsed Brown corpus)

- Alternative hypothesis: ***processing extraposed RCs is hard because they're unexpected***

Testing the role of expectations

- If extraposed RCs are hard because they're unexpected…
- …then making them more expected should make them easier
- Work by Wasow, Jaeger, and colleagues (Wasow et al., 2005, Levy & Jaeger 2007) has found that premodifier type can affect expectation for (in-situ) RCs

***a** barber…*	low RC expectation
***the** barber…*	higher RC expectation
***the only** barber…*	very high RC expectation

- If premodifier-induced expectations are carried over past the continuous NP domain, we may be able to manipulate extraposed RC expectations the same way*

The chair consulted the executives about the companies. . . RC less expected

The chair consulted only those executives about the companies. . . RC more expected

Experimental design

- We crossed *RC expectation* (low/high) with *RC extraposition* (extraposed/unextraposed)
- Example sentence: *The chairman consulted…*

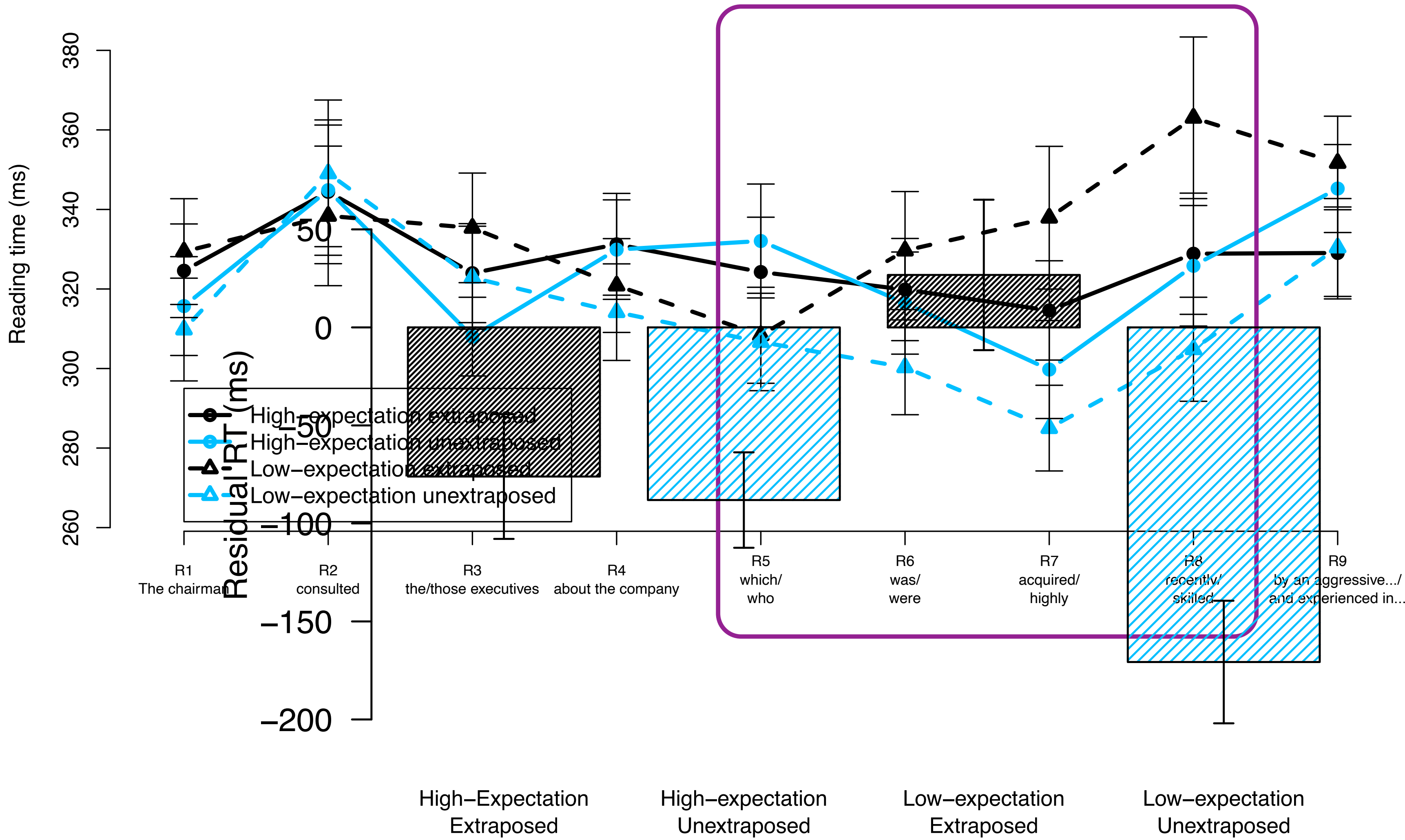
Expect	Extr	
low	–	*. . . the executives about the company which was making. . .*
low	+	*. . . the executives about the company who were making. . .*
high	–	*. . . only those executives about the company which was making. . .*
high	+	*. . . only those executives about the company who were making. . .*

- Our prediction is an ***interactive effect***: high RC expectation ("only those") will facilitate RC reading, but ***only*** in the extraposed condition
- We tested this in a self-paced reading study

Levy, Fedorenko, Breen, & Gibson (2012)

Online processing results

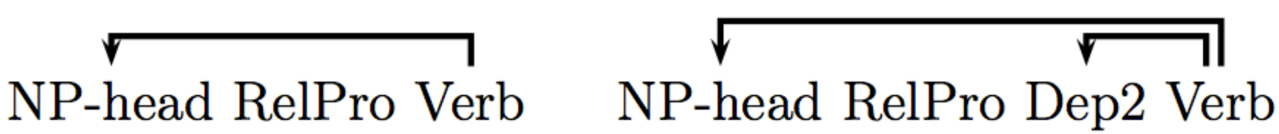
- The difficulty pattern emerges within the RC's first 4 words:



Levy, Fedorenko, Breen, & Gibson (2012)

Expectations versus memory

- Suppose you know that some event class X has to happen in the future, but you don't know:
  1. When X is going to occur
  2. Which member of X it's going to be
- The things W you see before X can give you hints about (1) and (2)
  - If expectations facilitate processing, then seeing W should generally speed processing of X
- But you also have to *keep W in memory* and retrieve it at X
  - This could slow processing at X



(a) (b)

What happens in German final-verb processing?

- Variation in pre-verbal dependency structure also found in verb-final clauses such as in German

```
Die Einsicht, dass der     Freund
The insight,  that the.NOM friend

dem     Kunden das     Auto aus Plastik
the.DAT client the.ACC car  of  plastic

verkaufte, erheiterte die Anderen.
sold,      amused     the others.
```

What happens in German final-verb processing?

...daß der Freund DEM Kunden das Auto verkaufte
...that the friend the client the car sold

'...that the friend sold the client a car...'

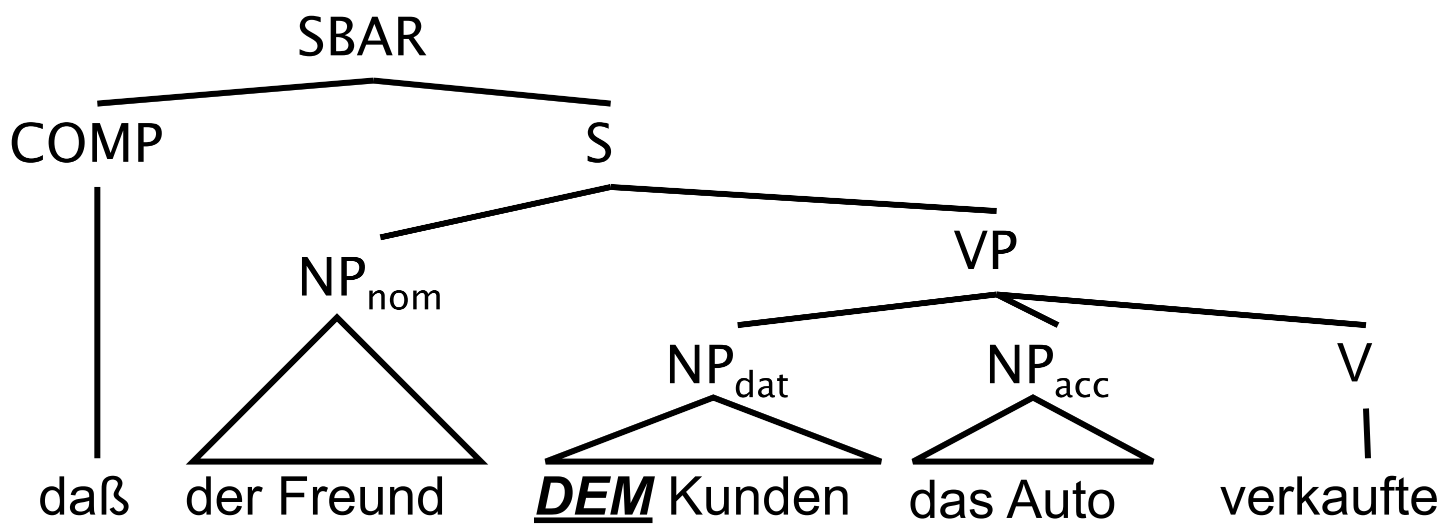
...daß der Freund DES Kunden das Auto verkaufte
...that the friend the client the car sold

'...that the friend of the client sold a car...'

Locality: final verb read faster in ***DES*** condition

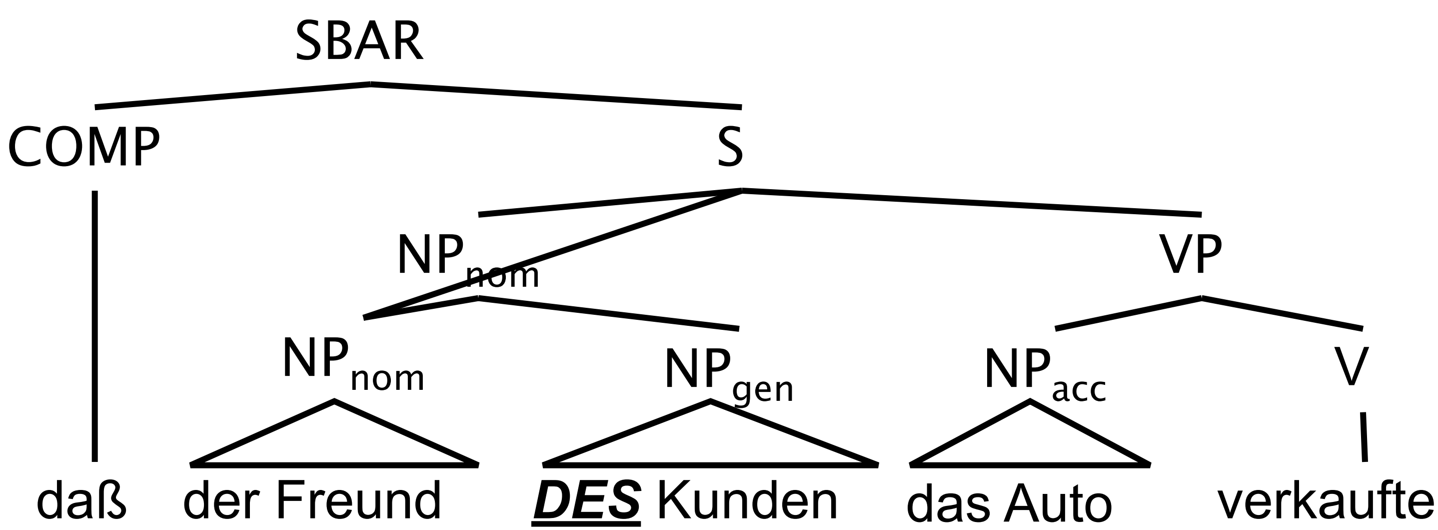
Observed: final verb read faster in ***DEM*** condition

(Konieczny & Döring 2003)



Next:

~~NP$_{nom}$~~
~~NP$_{acc}$~~
~~NP$_{dat}$~~
PP
ADVP
Verb



Next:

~~NP$_{nom}$~~
~~NP$_{acc}$~~
NP$_{dat}$
PP
ADVP
Verb

Model results

	Reading time (ms)	$P(w_i)$: word probability	Locality-based predictions
dem Kunden (dative)	555	8.38×10^{-8}	slower
des Kunden (genitive)	793	6.35×10^{-8}	faster

~30% greater expectation in dative condition

once again, wrong monotonicity

References

Bever, T. G. (1970). The cognitive basis for linguistic structures. *Cognition and the development of language*, *279*(362), 1-61.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210.

Hale, J. (2001, June). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8). Association for Computational Linguistics.

Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics*, *24*(4), 613-632.

Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics*.

Konieczny, L., & Döring, P. (2003, July). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In *Proceedings of iccs/ascs* (pp. 13-17).

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Levy, R. (2013). Memory and surprisal in human sentence comprehension. *Sentence processing*, *78*.

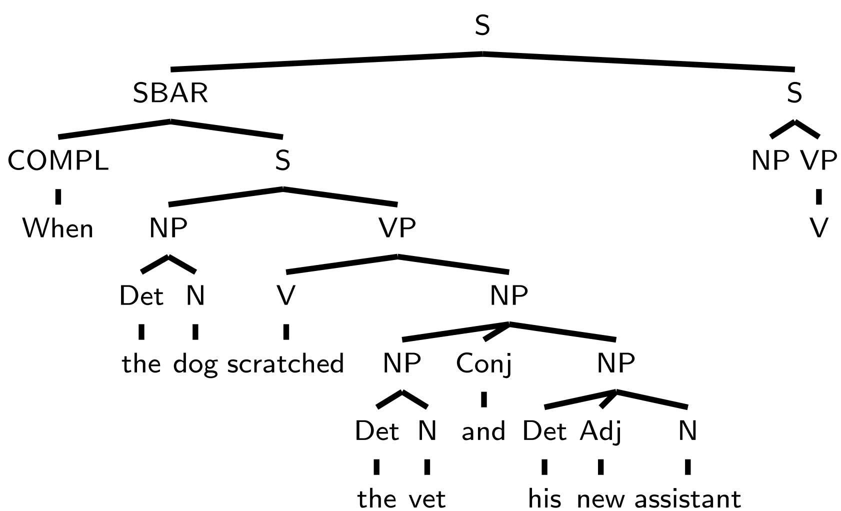
Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, *122*(1), 12-36.

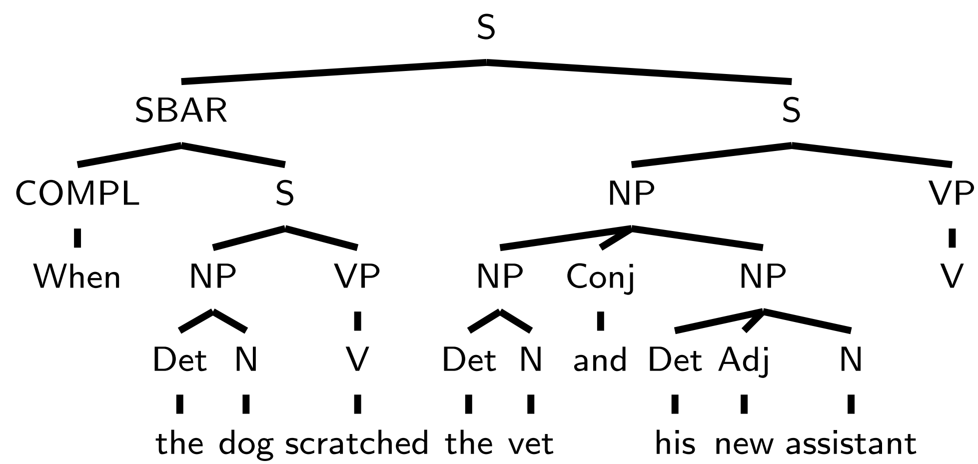
Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Back-pocket slides beyond here

S	$\rightarrow$	SBAR S	0.3	Conj	$\rightarrow$	and	1	Adj	$\rightarrow$	new	1
S	$\rightarrow$	NP VP	0.7	Det	$\rightarrow$	the	0.8	VP	$\rightarrow$	V NP	0.5
SBAR	$\rightarrow$	COMPL S	0.3	Det	$\rightarrow$	its	0.1	VP	$\rightarrow$	V	0.5
SBAR	$\rightarrow$	COMPL S COMMA	0.7	Det	$\rightarrow$	his	0.1	V	$\rightarrow$	scratched	0.25
COMPL	$\rightarrow$	When	1	N	$\rightarrow$	dog	0.2	V	$\rightarrow$	removed	0.25
NP	$\rightarrow$	Det N	0.6	N	$\rightarrow$	vet	0.2	V	$\rightarrow$	arrived	0.5
NP	$\rightarrow$	Det Adj N	0.2	N	$\rightarrow$	assistant	0.2	COMMA	$\rightarrow$	,	1
NP	$\rightarrow$	NP Conj NP	0.2	N	$\rightarrow$	muzzle	0.2				
				N	$\rightarrow$	owner	0.2				





References

- Saenger spaces between words
- Kucera & Francis 1967
- Brown et al. 1990
- Garside et al. 1987
- Marcus et al. 1993