Computational Psycholinguistics Lecture 3: Inference over infinite tree sets

Roger Levy

UC San Diego
Department of Linguistics

ESSLLI 2009
22 July 2009

Context-free Grammars

A context-free grammar (CFG) consists of a tuple (N, V, S, R) 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 $X \longrightarrow \alpha$ where $X \in N$ and α is a sequence of symbols drawn from $N \cup V$.

A CFG *derivation* is the recursive expansion of non-terminal symbols in a string by rules in R, starting with S, and a *derivation tree* T is the history of those rule applications.

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 $X \to \alpha$ 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 $X \in N$,

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

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

$1 \quad S \rightarrow NP\ VP$

$0.8 \quad NP \rightarrow Det\ N$

$0.2 \quad NP \rightarrow NP\ PP$

$1 \quad PP \rightarrow P\ NP$

$1 \quad VP \rightarrow V$

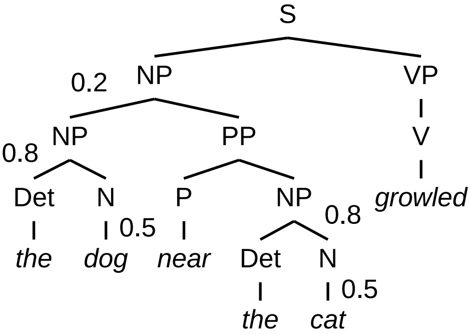
$1 \quad Det \rightarrow the$

$0.5 \quad N \rightarrow dog$

$0.5 \quad N \rightarrow cat$

$1 \quad P \rightarrow near$

$1 \quad V \rightarrow growled$



$$P(T) = 1 \times 0.2 \times 0.8 \times 1 \times 0.5 \times 0.8 \times 1 \times \times 0.8 \times 1 \times 0.5 \times 1 \times 1$$
$$= 0.032$$

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:

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

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

Inference over infinite tree sets

Consider the following noun-phrase grammar:

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

1 Det $\longrightarrow$ the
$\frac{2}{3}$ N $\longrightarrow$ dog
$\frac{1}{3}$ N $\longrightarrow$ cat
1 P $\longrightarrow$ 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:

$$\begin{array}{lllll} \frac{2}{3} & \mathrm{NP} & \longrightarrow & \mathrm{Det\ N} \\ \frac{1}{3} & \mathrm{NP} & \longrightarrow & \mathrm{NP\ PP} \\ 1 & \mathrm{PP} & \longrightarrow & \mathrm{P\ NP} \end{array} \qquad \begin{array}{llll} 1 & \mathrm{Det} & \longrightarrow & \mathrm{the} \\ \frac{2}{3} & \mathrm{N} & \longrightarrow & \mathrm{dog} \\ \frac{1}{3} & \mathrm{N} & \longrightarrow & \mathrm{cat} \\ 1 & \mathrm{P} & \longrightarrow & \mathrm{near} \end{array}$$

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:

$\frac{2}{3}$ 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

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

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

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

Inference over infinite tree sets (2)

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

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

Inference over infinite tree sets (2)

$$\begin{array}{lll} \frac{2}{3} & \text{NP} & \longrightarrow \text{Det N} \\ \frac{1}{3} & \text{NP} & \longrightarrow \text{NP PP} \\ 1 & \text{PP} & \longrightarrow \text{P NP} \end{array}$$

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

- 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)

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

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

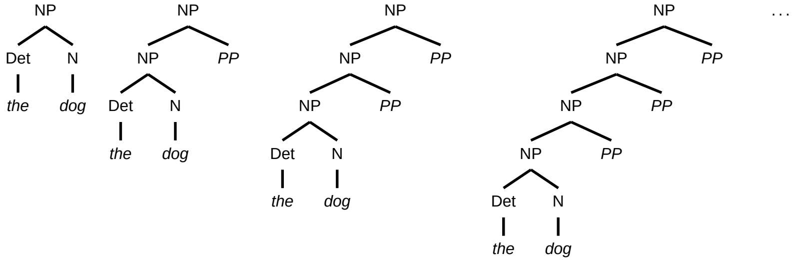
- We “should” just enumerate the trees that cover *the dog . . .* , and divide their total probability by that of *the . . .*
- . . . but there are infinitely many trees.

Inference over infinite tree sets (2)

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

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

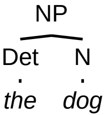
- We “should” just enumerate the trees that cover *the dog* . . . , and divide their total probability by that of *the* . . .
- . . . but there are infinitely many trees.



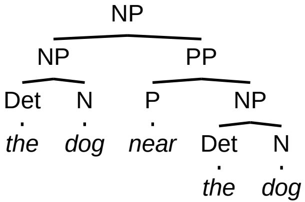
$\frac{2}{3}$ NP → Det N
$\frac{1}{3}$ NP → NP PP
1 PP → P NP

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

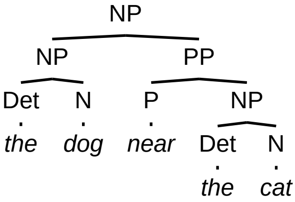
Shortcut 1: you can think of a *partial* tree as marginalizing over all completions of the partial tree.
It has a corresponding marginal probability in the PCFG.



$\frac{4}{9}$



$\frac{8}{81}$

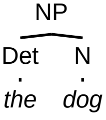


$\frac{4}{81}$

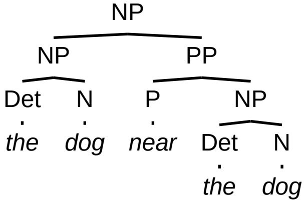
$\frac{2}{3}$ NP → Det N
$\frac{1}{3}$ NP → NP PP
1 PP → P NP

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

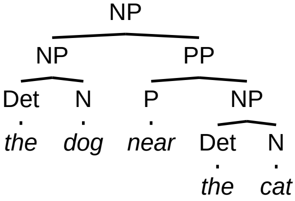
Shortcut 1: you can think of a *partial* tree as marginalizing over all completions of the partial tree.
It has a corresponding marginal probability in the PCFG.



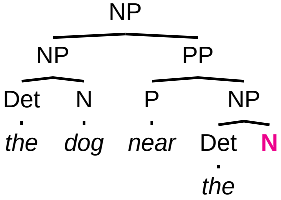
$\frac{4}{9}$



$\frac{8}{81}$



$\frac{4}{81}$

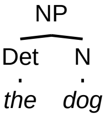


$\frac{12}{81}$

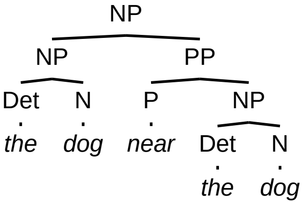
$\frac{2}{3}$ NP → Det N
$\frac{1}{3}$ NP → NP PP
1 PP → P NP

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

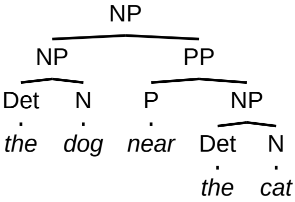
Shortcut 1: you can think of a *partial* tree as marginalizing over all completions of the partial tree.
It has a corresponding marginal probability in the PCFG.



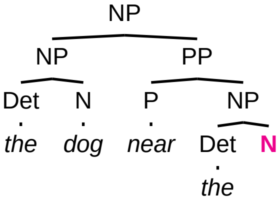
$\frac{4}{9}$



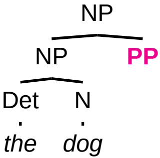
$\frac{8}{81}$



$\frac{4}{81}$

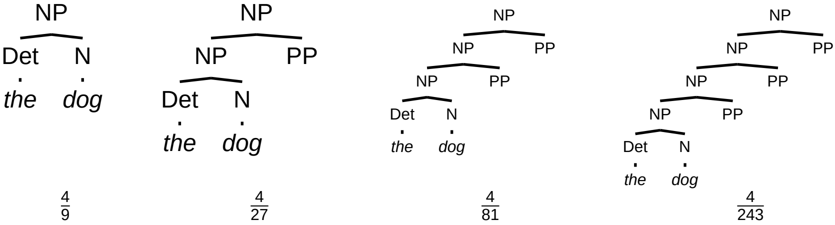


$\frac{12}{81}$

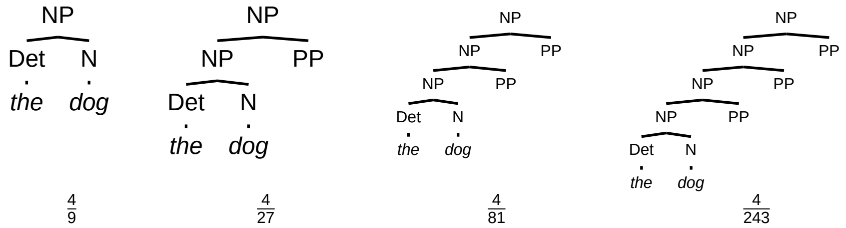


$\frac{4}{27}$

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



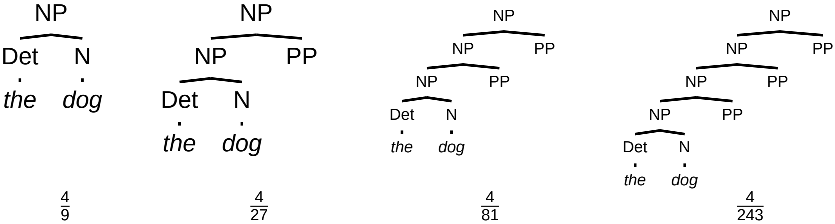
Problem 2: there are still an infinite number of incomplete trees covering a partial input.



BUT! These tree probabilities form a geometric series:

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

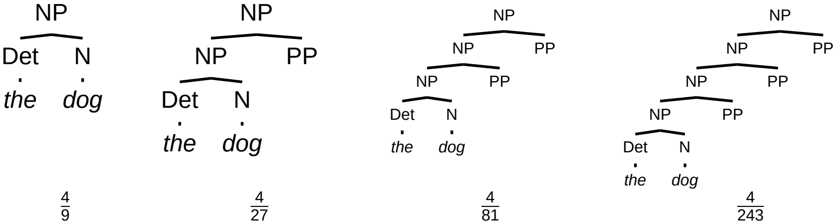
Problem 2: there are still an infinite number of incomplete trees covering a partial input.



BUT! These tree probabilities form a geometric series:

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

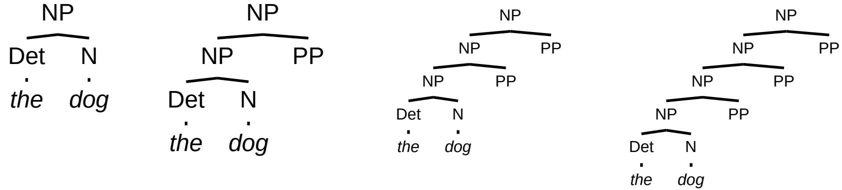
Problem 2: there are still an infinite number of incomplete trees covering a partial input.



BUT! These tree probabilities form a geometric series:

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

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:

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

…which matches the original rule probability

$\frac{2}{3}\ N \longrightarrow \mathit{dog}$

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

$$A \to B\ \alpha \qquad\qquad B \to A\ \beta$$

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:

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

and solve for its closure $R_L = (I - P_L)^{-1}$. *(Stolcke, 1995)*

Generalizing the geometric series

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

- The closure of our left-corner matrix is

$$R_L = \begin{array}{c} \\ \text{ROOT} \\ \text{NP} \\ \text{PP} \\ \text{Det} \\ \text{N} \\ \text{P} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{ROOT} & \text{NP} & \text{PP} & \text{Det} & \text{N} & \text{P} \end{array} \\ \left(\begin{array}{cccccc} 1 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array}$$

Generalizing the geometric series

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

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

- The closure of our left-corner matrix is

$$R_L = \begin{array}{c c} & \begin{array}{cccccc} \text{ROOT} & \text{NP} & \text{PP} & \text{Det} & \text{N} & \text{P} \end{array} \\ \begin{array}{c} \text{ROOT} \\ \text{NP} \\ \text{PP} \\ \text{Det} \\ \text{N} \\ \text{P} \end{array} & \left(\begin{array}{cccccc} 1 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array}$$

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

Generalizing the geometric series

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

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

- The closure of our left-corner matrix is

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

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

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

- The closure of our left-corner matrix is

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

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

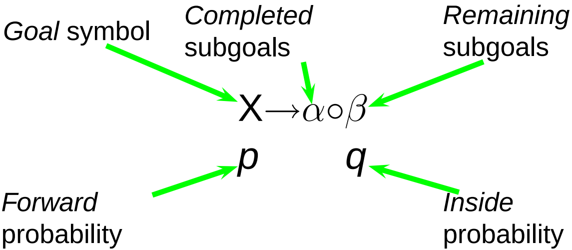
We can use the Earley algorithm (Earley, 1970) in a probabilistic incarnation (Stolcke, 1995) to deal with these infinite tree sets.

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

$$\underset{p}{X} \rightarrow \alpha \circ \underset{q}{\beta}$$

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**:

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

Efficient incremental parsing: the probabilistic Earley algorithm

Inference rules for probabilistic Earley:

- **Prediction**:

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

- **Completion**:

$$\frac{\begin{array}{ll} X\rightarrow\beta\circ Y\gamma & \\ p & q \end{array} \qquad a : R(Y \overset{*}{\Rightarrow}_U Z) \qquad \begin{array}{ll} Z\rightarrow\alpha\circ & \\ b & c \end{array}}{\begin{array}{ll} X\rightarrow\beta Y\circ\gamma & \\ acp & acq \end{array}}$$

Efficient incremental parsing: probabilistic Earley

the dog near the

Efficient incremental parsing: probabilistic Earley

ROOT→○NP

1 1

the dog near the

Efficient incremental parsing: probabilistic Earley

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

$1 \qquad 1$

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

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

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

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

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

$1 \qquad 1$

the dog near the

Efficient incremental parsing: probabilistic Earley

Det→∘the

1 1

NP→∘Det N

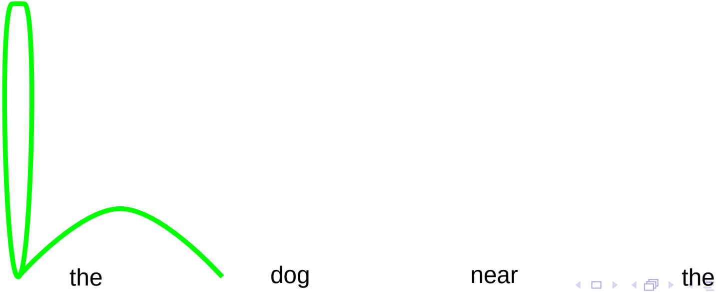
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

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

Det→∘the

1 1

NP→∘Det N

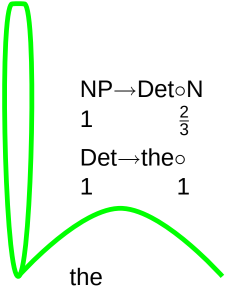
$\frac{2}{3} \times \frac{3}{2}$ $\frac{2}{3}$

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

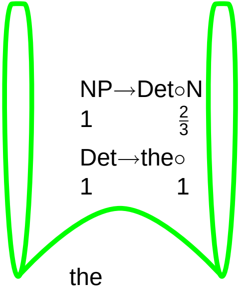
Efficient incremental parsing: probabilistic Earley

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

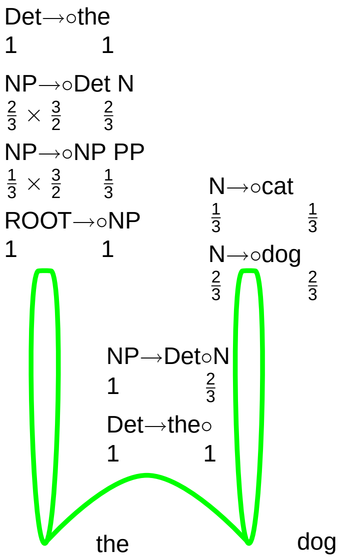


NP→Det∘N
1 $\frac{2}{3}$

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

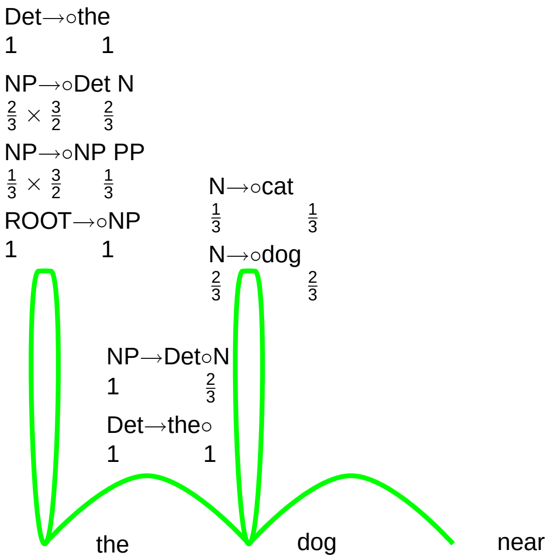
1 $\frac{2}{3}$

Det→the∘

1 1

the dog near the

Efficient incremental parsing: probabilistic Earley



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 $\frac{2}{3}$

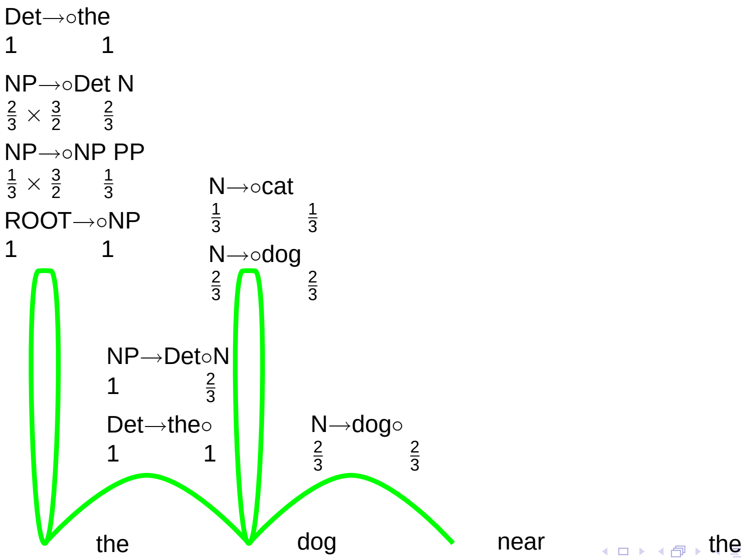
Det→the○
1 1

N→○cat
$\frac{1}{3}$ $\frac{1}{3}$

N→○dog
$\frac{2}{3}$ $\frac{2}{3}$

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 $\frac{2}{3}$

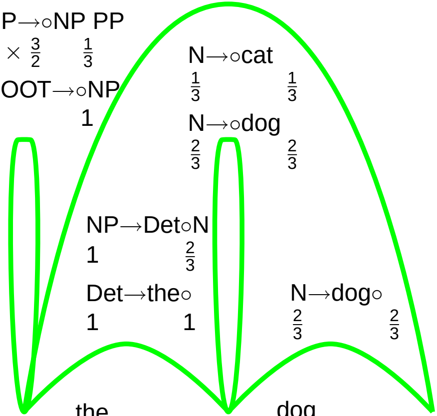
Det→the○

1 1

N→dog○

$\frac{2}{3}$ $\frac{2}{3}$

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
$\frac{2}{3}$ $\frac{2}{3}$

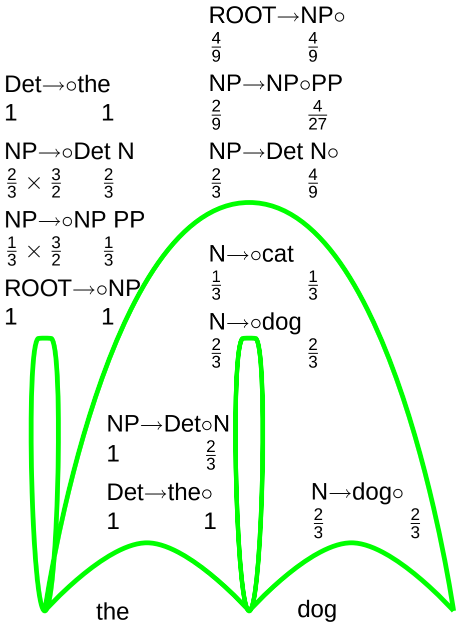
NP→Det○N
1 $\frac{2}{3}$

Det→the○
1 1

N→dog○
$\frac{2}{3}$ $\frac{2}{3}$

the dog near the

Efficient incremental parsing: probabilistic Earley



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

ROOT→NP∘

$\frac{4}{9}$ $\frac{4}{9}$

NP→NP∘PP

$\frac{2}{9}$ $\frac{4}{27}$

NP→Det N∘

$\frac{2}{3}$ $\frac{4}{9}$

N→∘cat

$\frac{1}{3}$ $\frac{1}{3}$

N→∘dog

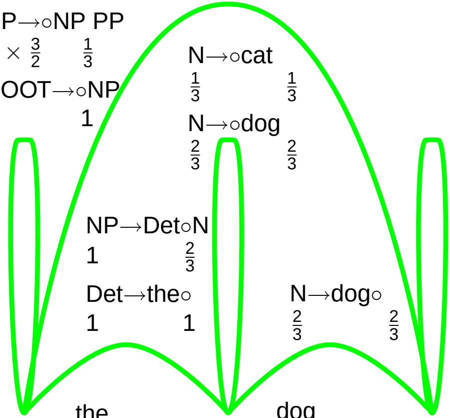
$\frac{2}{3}$ $\frac{2}{3}$

N→dog∘

$\frac{2}{3}$ $\frac{2}{3}$

the dog near the

Efficient incremental parsing: probabilistic Earley



Det→∘the

$1 \qquad 1$

NP→∘Det N

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

NP→∘NP PP

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

ROOT→∘NP

$1 \qquad 1$

NP→Det∘N

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

Det→the∘

$1 \qquad 1$

ROOT→NP∘

$\frac{4}{9} \qquad \frac{4}{9}$

NP→NP∘PP

$\frac{2}{9} \qquad \frac{4}{27}$

NP→Det N∘

$\frac{2}{3} \qquad \frac{4}{9}$

N→∘cat

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

N→∘dog

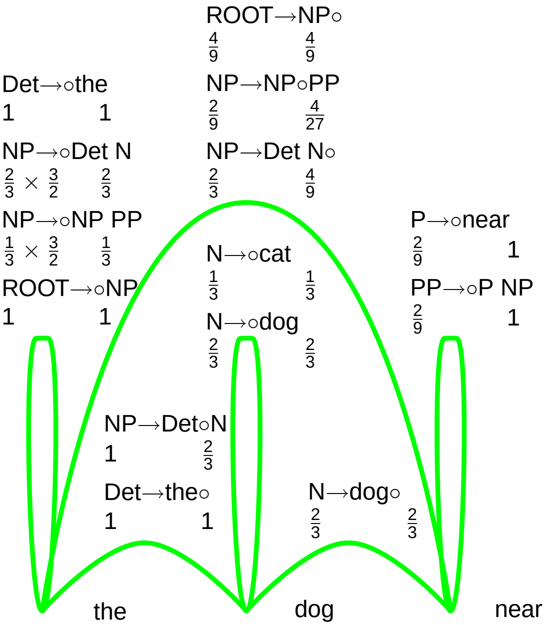
$\frac{2}{3} \qquad \frac{2}{3}$

N→dog∘

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

the dog near the

Efficient incremental parsing: probabilistic Earley



$\mathrm{Det}\rightarrow\circ\mathrm{the}$

$1 \qquad 1$

$\mathrm{NP}\rightarrow\circ\mathrm{Det}\ \mathrm{N}$

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

$\mathrm{NP}\rightarrow\circ\mathrm{NP}\ \mathrm{PP}$

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

$\mathrm{ROOT}\rightarrow\circ\mathrm{NP}$

$1 \qquad 1$

$\mathrm{NP}\rightarrow\mathrm{Det}\circ\mathrm{N}$

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

$\mathrm{Det}\rightarrow\mathrm{the}\circ$

$1 \qquad 1$

$\mathrm{ROOT}\rightarrow\mathrm{NP}\circ$

$\frac{4}{9} \qquad \frac{4}{9}$

$\mathrm{NP}\rightarrow\mathrm{NP}\circ\mathrm{PP}$

$\frac{2}{9} \qquad \frac{4}{27}$

$\mathrm{NP}\rightarrow\mathrm{Det}\ \mathrm{N}\circ$

$\frac{2}{3} \qquad \frac{4}{9}$

$\mathrm{N}\rightarrow\circ\mathrm{cat}$

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

$\mathrm{N}\rightarrow\circ\mathrm{dog}$

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

$\mathrm{N}\rightarrow\mathrm{dog}\circ$

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

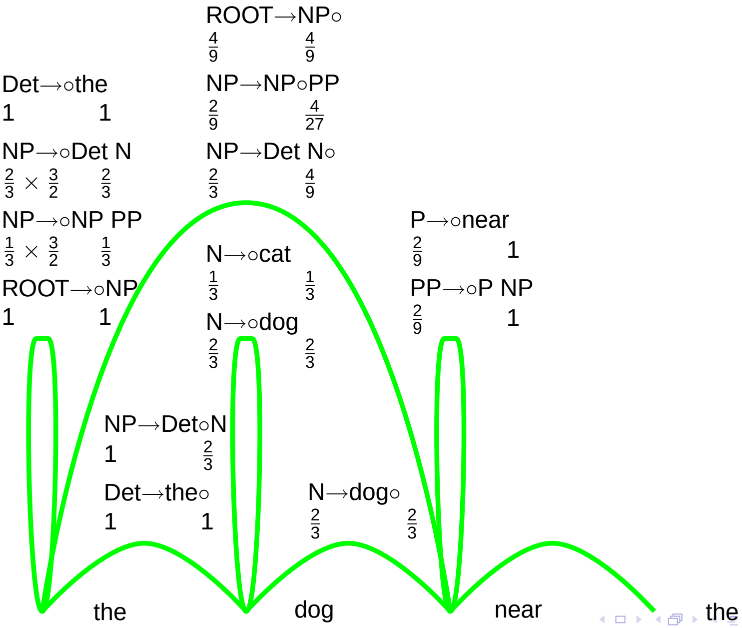
$\mathrm{P}\rightarrow\circ\mathrm{near}$

$\frac{2}{9} \qquad 1$

$\mathrm{PP}\rightarrow\circ\mathrm{P}\ \mathrm{NP}$

$\frac{2}{9} \qquad 1$

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

ROOT→NP∘
$\frac{4}{9}$ $\frac{4}{9}$

NP→NP∘PP
$\frac{2}{9}$ $\frac{4}{27}$

NP→Det N∘
$\frac{2}{3}$ $\frac{4}{9}$

N→∘cat
$\frac{1}{3}$ $\frac{1}{3}$

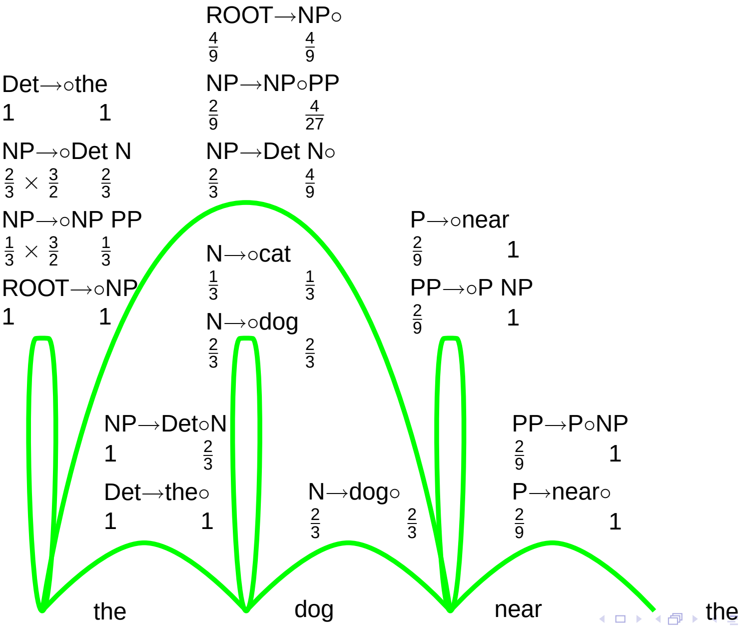
N→∘dog
$\frac{2}{3}$ $\frac{2}{3}$

N→dog∘
$\frac{2}{3}$ $\frac{2}{3}$

P→∘near
$\frac{2}{9}$ 1

PP→∘P NP
$\frac{2}{9}$ 1

Efficient incremental parsing: probabilistic Earley



$\mathrm{ROOT} \longrightarrow \mathrm{NP} \circ$

$\frac{4}{9} \qquad \frac{4}{9}$

$\mathrm{NP} \longrightarrow \mathrm{NP} \circ \mathrm{PP}$

$\frac{2}{9} \qquad \frac{4}{27}$

$\mathrm{NP} \longrightarrow \mathrm{Det}\ \mathrm{N} \circ$

$\frac{2}{3} \qquad \frac{4}{9}$

$\mathrm{Det} \longrightarrow \circ \mathrm{the}$

$1 \qquad 1$

$\mathrm{NP} \longrightarrow \circ \mathrm{Det}\ \mathrm{N}$

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

$\mathrm{NP} \longrightarrow \circ \mathrm{NP}\ \mathrm{PP}$

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

$\mathrm{ROOT} \longrightarrow \circ \mathrm{NP}$

$1 \qquad 1$

$\mathrm{N} \longrightarrow \circ \mathrm{cat}$

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

$\mathrm{N} \longrightarrow \circ \mathrm{dog}$

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

$\mathrm{P} \longrightarrow \circ \mathrm{near}$

$\frac{2}{9} \qquad 1$

$\mathrm{PP} \longrightarrow \circ \mathrm{P}\ \mathrm{NP}$

$\frac{2}{9} \qquad 1$

$\mathrm{NP} \longrightarrow \mathrm{Det} \circ \mathrm{N}$

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

$\mathrm{Det} \longrightarrow \mathrm{the} \circ$

$1 \qquad 1$

$\mathrm{N} \longrightarrow \mathrm{dog} \circ$

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

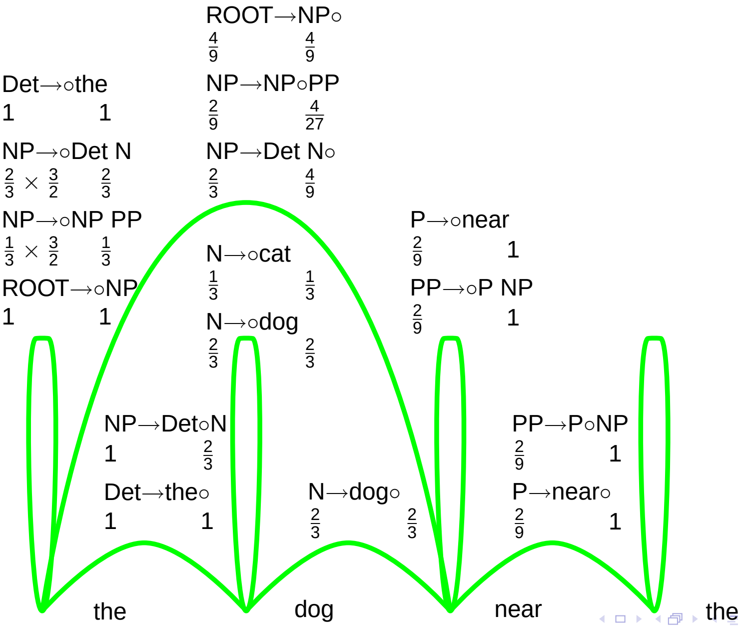
$\mathrm{PP} \longrightarrow \mathrm{P} \circ \mathrm{NP}$

$\frac{2}{9} \qquad 1$

$\mathrm{P} \longrightarrow \mathrm{near} \circ$

$\frac{2}{9} \qquad 1$

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

ROOT→NP∘
$\frac{4}{9}$ $\frac{4}{9}$

NP→NP∘PP
$\frac{2}{9}$ $\frac{4}{27}$

NP→Det N∘
$\frac{2}{3}$ $\frac{4}{9}$

N→∘cat
$\frac{1}{3}$ $\frac{1}{3}$

N→∘dog
$\frac{2}{3}$ $\frac{2}{3}$

N→dog∘
$\frac{2}{3}$ $\frac{2}{3}$

P→∘near
$\frac{2}{9}$ 1

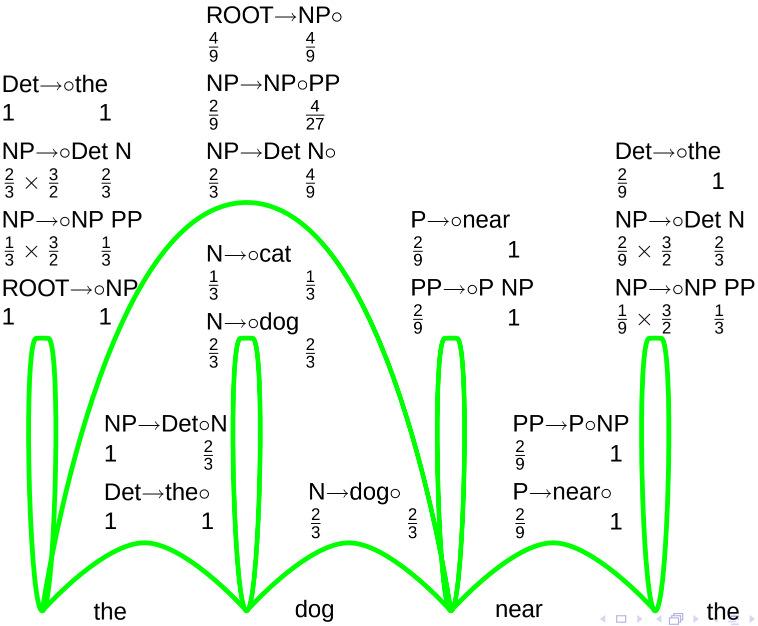
PP→∘P NP
$\frac{2}{9}$ 1

PP→P∘NP
$\frac{2}{9}$ 1

P→near∘
$\frac{2}{9}$ 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

ROOT→NP∘
$\frac{4}{9}$ $\frac{4}{9}$

NP→NP∘PP
$\frac{2}{9}$ $\frac{4}{27}$

NP→Det N∘
$\frac{2}{3}$ $\frac{4}{9}$

N→∘cat
$\frac{1}{3}$ $\frac{1}{3}$

N→∘dog
$\frac{2}{3}$ $\frac{2}{3}$

N→dog∘
$\frac{2}{3}$ $\frac{2}{3}$

P→∘near
$\frac{2}{9}$ 1

PP→∘P NP
$\frac{2}{9}$ 1

PP→P∘NP
$\frac{2}{9}$ 1

P→near∘
$\frac{2}{9}$ 1

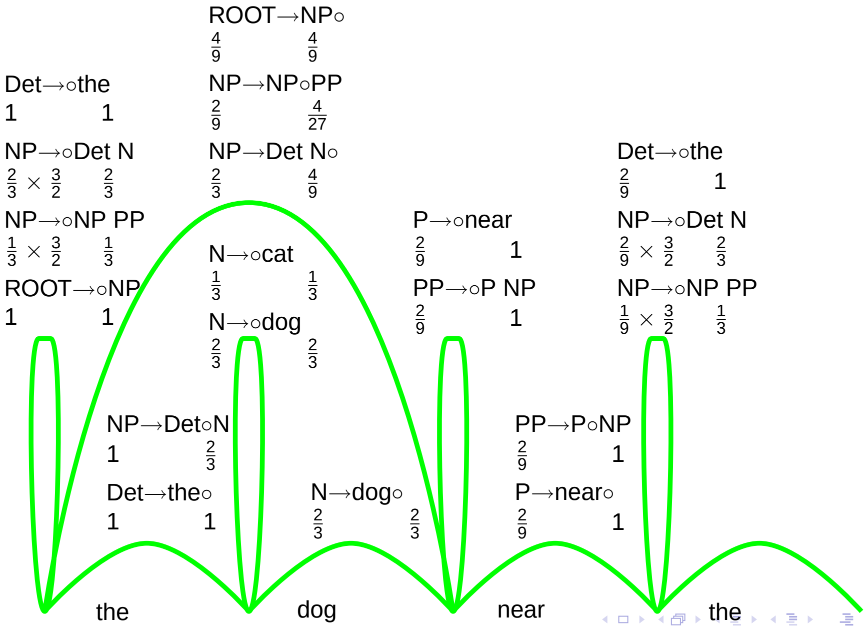
Det→∘the
$\frac{2}{9}$ 1

NP→∘Det N
$\frac{2}{9} \times \frac{3}{2}$ $\frac{2}{3}$

NP→∘NP PP
$\frac{1}{9} \times \frac{3}{2}$ $\frac{1}{3}$

the dog near the

Efficient incremental parsing: probabilistic Earley



Det→∘the

$1 \quad 1$

NP→∘Det N

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

NP→∘NP PP

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

ROOT→∘NP

$1 \quad 1$

NP→Det∘N

$1 \quad \frac{2}{3}$

Det→the∘

$1 \quad 1$

ROOT→NP∘

$\frac{4}{9} \quad \frac{4}{9}$

NP→NP∘PP

$\frac{2}{9} \quad \frac{4}{27}$

NP→Det N∘

$\frac{2}{3} \quad \frac{4}{9}$

N→∘cat

$\frac{1}{3} \quad \frac{1}{3}$

N→∘dog

$\frac{2}{3} \quad \frac{2}{3}$

N→dog∘

$\frac{2}{3} \quad \frac{2}{3}$

P→∘near

$\frac{2}{9} \quad 1$

PP→∘P NP

$\frac{2}{9} \quad 1$

PP→P∘NP

$\frac{2}{9} \quad 1$

P→near∘

$\frac{2}{9} \quad 1$

Det→∘the

$\frac{2}{9} \quad 1$

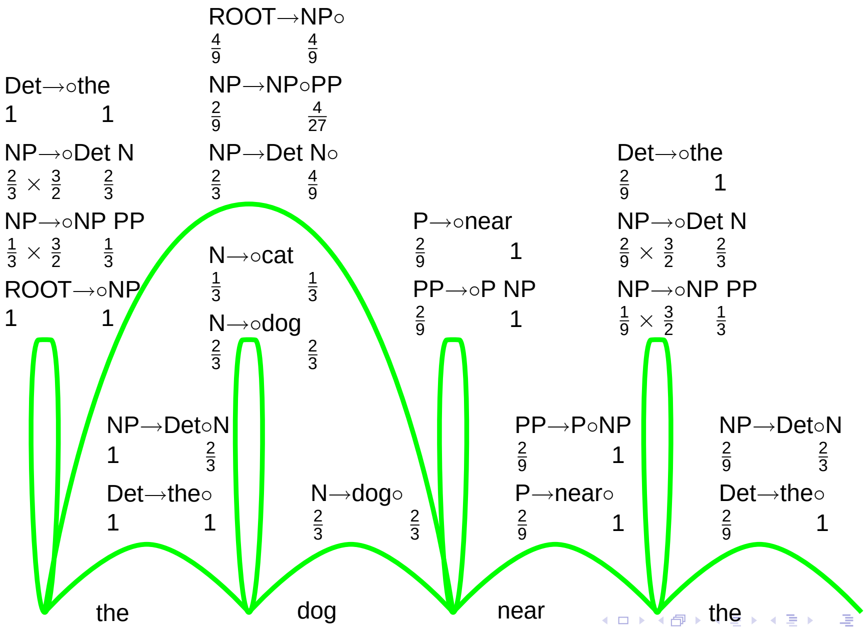
NP→∘Det N

$\frac{2}{9} \times \frac{3}{2} \quad \frac{2}{3}$

NP→∘NP PP

$\frac{1}{9} \times \frac{3}{2} \quad \frac{1}{3}$

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

ROOT→NP∘
$\frac{4}{9}$ $\frac{4}{9}$

NP→NP∘PP
$\frac{2}{9}$ $\frac{4}{27}$

NP→Det N∘
$\frac{2}{3}$ $\frac{4}{9}$

N→∘cat
$\frac{1}{3}$ $\frac{1}{3}$

N→∘dog
$\frac{2}{3}$ $\frac{2}{3}$

N→dog∘
$\frac{2}{3}$ $\frac{2}{3}$

P→∘near
$\frac{2}{9}$ 1

PP→∘P NP
$\frac{2}{9}$ 1

PP→P∘NP
$\frac{2}{9}$ 1

P→near∘
$\frac{2}{9}$ 1

Det→∘the
$\frac{2}{9}$ 1

NP→∘Det N
$\frac{2}{9} \times \frac{3}{2}$ $\frac{2}{3}$

NP→∘NP PP
$\frac{1}{9} \times \frac{3}{2}$ $\frac{1}{3}$

NP→Det∘N
$\frac{2}{9}$ $\frac{2}{3}$

Det→the∘
$\frac{2}{9}$ 1

the dog near the

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 $X \rightarrow \alpha \circ w_i \beta$

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 $X \rightarrow \alpha \circ w_i \beta$
- In our example, we see:

$$
\begin{aligned}
&P(\text{the}) = 1\\
&P(\text{the dog}) = \tfrac{2}{3}\\
&P(\text{the dog near}) = \tfrac{4}{9}\\
&P(\text{the dog near the}) = \tfrac{4}{9}
\end{aligned}
$$

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 $X \rightarrow \alpha \circ w_i \beta$
- In our example, we see:

$$\begin{aligned} &P(\text{the}) = 1 \\ &P(\text{the dog}) = \tfrac{2}{3} \\ &P(\text{the dog near}) = \tfrac{4}{9} \\ &P(\text{the dog near the}) = \tfrac{4}{9} \end{aligned}$$

- 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
- 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

Other introductions

- You can read about the (non-probabilistic) Earley algorithm in (Jurafsky and Martin, 2000, Chapter 13)
- Prefix probabilities can also be calculated with an extension of the CKY algorithm due to Jelinek and Lafferty (1991)

Now. . .

Applications of the idea of surprisal to online comprehension

References I

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the ACM*, 13(2):94–102.

Hale, J. (2001). 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*, pages 159–166.

Jelinek, F. and Lafferty, J. D. (1991). Computation of the probability of initial substring generation by stochastic context free grammars. *Computational Linguistics*, 17(3):315–323.

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall.

Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21(2):165–201.