

# Context-Free Parsing: Outside Probabilities, Earley's Algorithm, and Treebank Transformations

Introduction to Natural Language Processing  
Computer Science 585—Fall 2009  
University of Massachusetts Amherst

David Smith

# Announcements

- If you haven't gotten your HW #1, please let me know
- HW #2: think of part 2 as tagging when you don't know what the tags mean; please do email source code
- Jason's request: please hand in printout of non-source-code if you want comments
- I'd like to postpone the midterm till next Thursday

# Overview

- Complete inside-outside: outside computations
- Earley's algorithm: combining bottom-up and top-down parsing
- Better treebank grammars: annotating nonterminals

# Inside & Viterbi Algorithms

Let  $\beta_A(i, j) = p(\text{constit}(A, i, j))$

$$= p(w_{ij} \mid \text{nonterminal A from } i \text{ to } j)$$

NB: index *between* words;  
M&S index words

$$\beta_A(i, k) = \sum_{B, C, j} \beta_B(i, j) \cdot \beta_C(j, k) \cdot p(A \rightarrow B \ C)$$

Let  $\delta_A(i, j) = p_{\text{best}}(\text{constit}(A, i, j))$

$$\delta_A(i, k) = \max_{B, C, j} \delta_B(i, j) \cdot \delta_C(j, k) \cdot p(A \rightarrow B \ C)$$

$$\beta_S(0, n) = ? \qquad \qquad \delta_S(0, n) = ?$$

# Inside & Outside

$\text{constit}(A, i, j)$

$p(\text{words } 0-i, \text{words } j-n, \text{constit})$

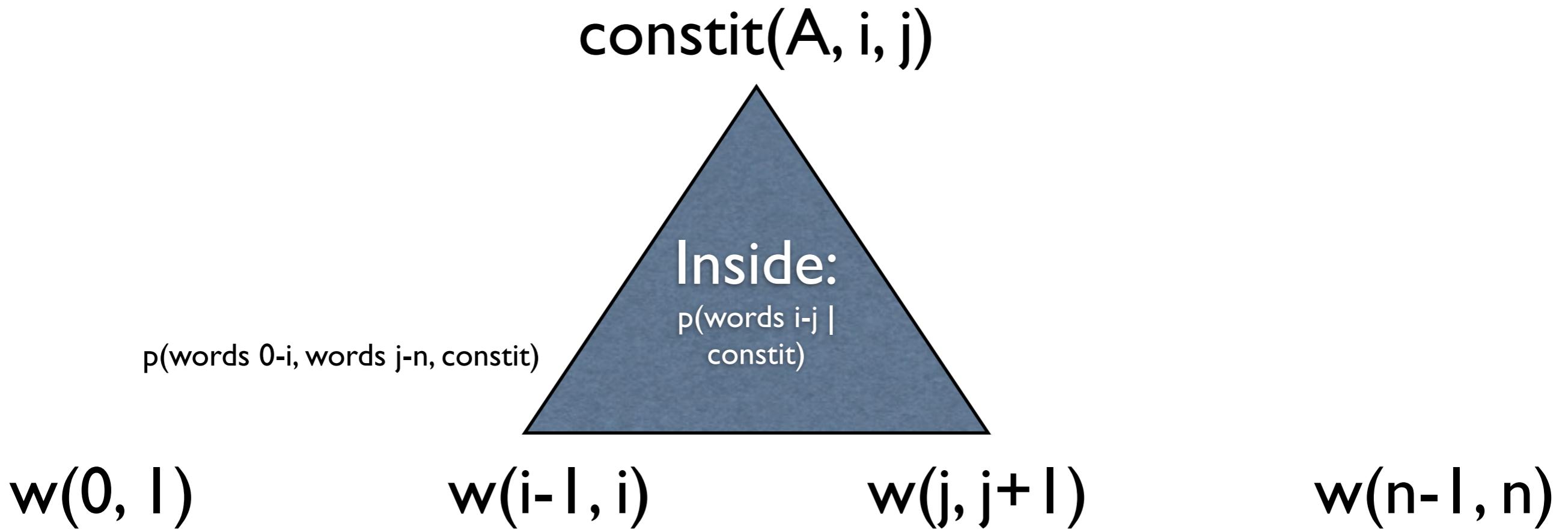
$w(0, l)$

$w(i-l, i)$

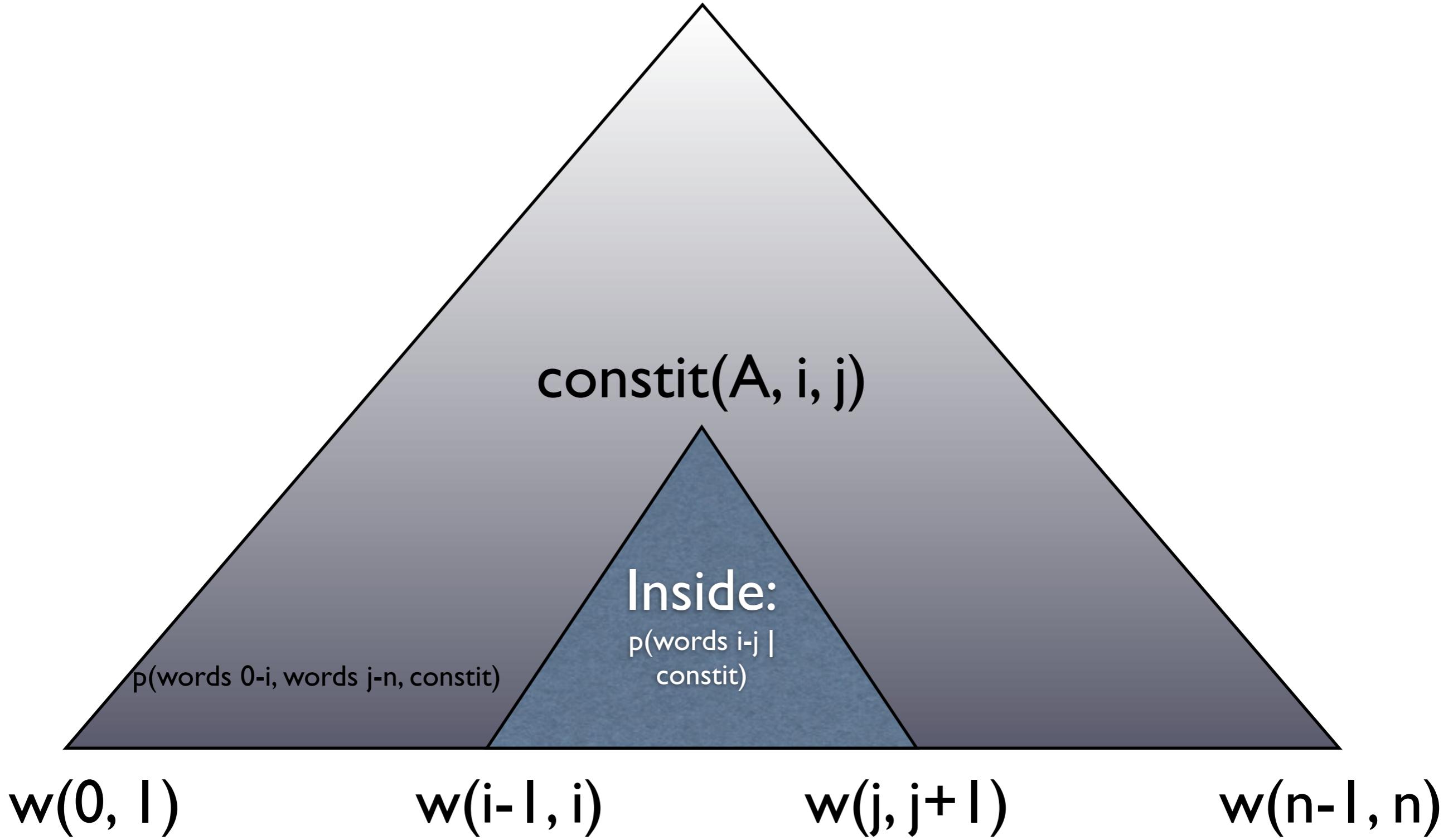
$w(j, j+l)$

$w(n-l, n)$

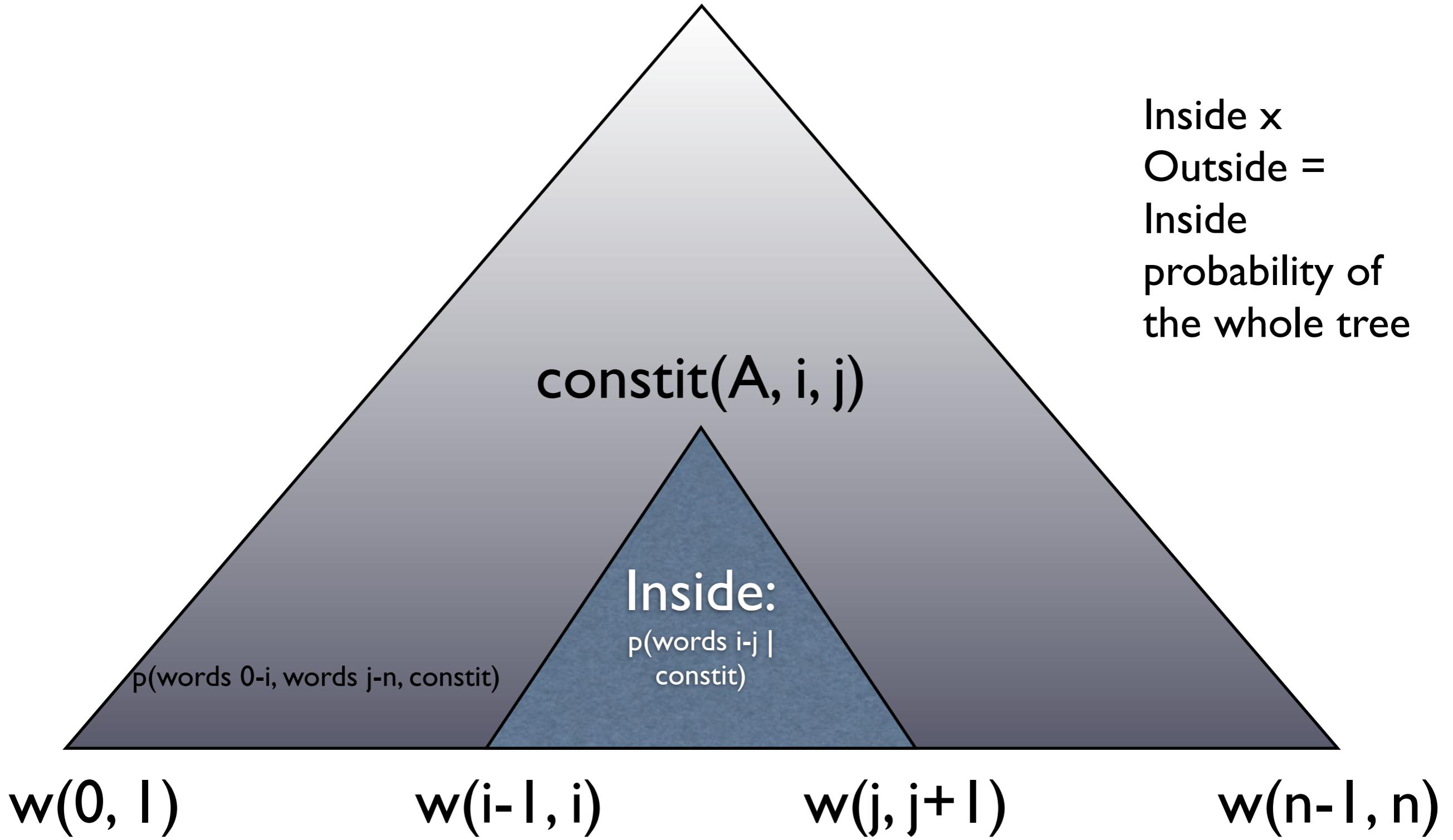
# Inside & Outside



# Inside & Outside



# Inside & Outside



# Outside Algorithm

$$\alpha_A(i, j) = p(w_{0,i}, A_{i,j}, w_{j,n})$$

Uses inside  
probs.

$$\begin{aligned}\alpha_A(i, j) &= \sum_{B,C,k=j}^n \alpha_B(i, k) \cdot \beta_C(j, k) \cdot p(B \rightarrow A \mid C) \\ &\quad + \sum_{B,C,k=0}^i \alpha_B(k, j) \cdot \beta_C(k, i) \cdot p(B \rightarrow C \mid A)\end{aligned}$$

$$\alpha_S(0, n) = ?$$

Some  
resemblance  
to derivative  
product rule

$$\alpha_{PP}(0, n) = ?$$

# Top-Down/Bottom-Up

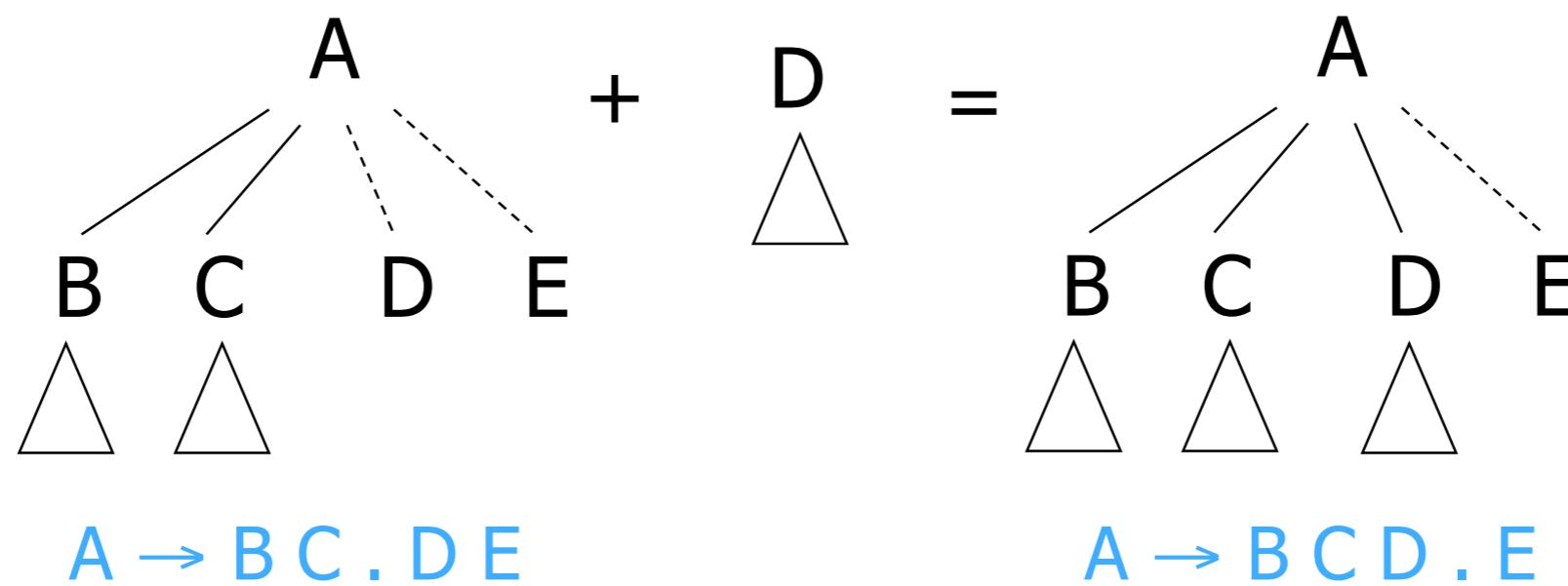
- Top-down parsers
  - Can get caught in infinite loops
  - Take exponential time backtracking
- CKY
  - Needs Chomsky normal form
  - Builds all possible constituents

## Earley Parser (1970)

- Nice combination of
    - dynamic programming
    - incremental interpretation
    - avoids infinite loops
    - no restrictions on the form of the context-free grammar.
- $A \rightarrow B C$  the D of** causes no problems
- $O(n^3)$  worst case, but faster for many grammars
  - Uses left context and optionally right context to constrain search.

## Overview of the Algorithm

- Finds constituents and partial constituents in input
  - $A \rightarrow B C . D E$  is partial: only the first half of the  $A$



## Overview of the Algorithm

- Proceeds incrementally left-to-right
  - Before it reads word 5, it has already built all hypotheses that are consistent with first 4 words
  - Reads word 5 & attaches it to immediately preceding hypotheses. Might yield new constituents that are then attached to hypotheses immediately preceding *them* ...
  - E.g., attaching D to  $A \rightarrow B C . D E$  gives  $A \rightarrow B C D . E$
  - Attaching E to that gives  $A \rightarrow B C D E$ .
  - Now we have a complete A that we can attach to hypotheses immediately preceding the A, etc.

## The Parse Table

- Columns 0 through n corresponding to the gaps between words
- Entries in column 5 look like  $(3, NP \rightarrow NP . PP)$   
(but we'll omit the  $\rightarrow$  etc. to save space)
  - Built while processing word 5
  - Means that the input substring from 3 to 5 matches the initial NP portion of a  $NP \rightarrow NP PP$  rule
  - Dot shows how much we've matched as of column 5
  - Perfectly fine to have entries like  $(3, VP \rightarrow is\ it\ .\ true\ that\ S)$

## The Parse Table

- Entries in column 5 look like  $(3, NP \rightarrow NP . PP)$
- What will it mean that we have this entry?
  - *Unknown right context*: Doesn't mean we'll necessarily be able to find a VP starting at column 5 to complete the S.
  - *Known left context*: Does mean that some dotted rule back in column 3 is looking for an S that starts at 3.
    - So if we actually do find a VP starting at column 5, allowing us to complete the S, then we'll be able to attach the S to something.
    - And when that something is complete, it too will have a customer to *its* left ...
    - In short, a top-down (i.e., goal-directed) parser: it chooses to start building a constituent not because of the input but because that's what the left context needs. In **the spoon**, won't build **spoon** as a verb because there's no way to use a verb there.
    - So any hypothesis in column 5 *could* get used in the correct parse, if words 1-5 are continued in just the right way by words 6-n.

## Earley's Algorithm, recognizer version

- Add  $\text{ROOT} \rightarrow . S$  to column 0.
- For each  $j$  from 0 to  $n$ :
  - For each dotted rule in column  $j$ ,  
(including those we add as we go!)  
look at what's after the dot:
    - If it's a word  $w$ , SCAN:
      - If  $w$  matches the input word between  $j$  and  $j+1$ , advance the dot  
and add the resulting rule to column  $j+1$
    - If it's a non-terminal  $X$ , PREDICT:
      - Add all rules for  $X$  to the bottom of column  $j$ , wth the dot at the  
start: e.g.  $X \rightarrow . Y Z$
    - If there's nothing after the dot, ATTACH:
      - We've finished some constituent,  $A$ , that started in column  $I < j$ . So  
for each rule in column  $j$  that has  $A$  after the dot: Advance the dot  
and add the result to the bottom of column  $j$ .
- Output “yes” just if last column has  $\text{ROOT} \rightarrow S .$
- **NOTE: Don't add an entry to a column if it's already there!**

## Summary of the Algorithm

- Process all hypotheses one at a time in order.  
(**Current hypothesis** is shown in blue.)
- This may add **new hypotheses** to the end of the to-do list, or try to add **old hypotheses** again.
- Process a hypothesis according to what follows the dot:
  - If a word, **scan** input and see if it matches
  - If a nonterminal, **predict** ways to match it
    - (we'll predict blindly, but could reduce # of predictions by *looking ahead* k symbols in the input and only making predictions that are compatible with this limited *right context*)
    - If nothing, then we have a complete constituent, so **attach** it to all its customers

## A Grammar

$S \rightarrow NP\ VP$	$NP \rightarrow Papa$
$NP \rightarrow Det\ N$	$N \rightarrow caviar$
$NP \rightarrow NP\ PP$	$N \rightarrow spoon$
$VP \rightarrow V\ NP$	$V \rightarrow ate$
$VP \rightarrow VP\ PP$	$P \rightarrow with$
$PP \rightarrow P\ NP$	$Det \rightarrow the$
	$Det \rightarrow a$

## An Input Sentence

*Papa ate the caviar with a spoon.*

0
0 ROOT . S

initialize

*Remember this stands for (0, ROOT → . S)*

0  
0 ROOT . S  
0 S . NP VP

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

**predict** the kind of S we are looking for

*Remember this stands for (0, S → . NP VP)*

0
0 ROOT . S
0 S . NP VP
0 NP . Det N
0 NP . NP PP
0 NP . Papa

**predict** the kind of NP we are looking for  
*(actually we'll look for 3 kinds: any of the 3 will do)*

0
0 ROOT . S
0 S . NP VP
0 NP . Det N
0 NP . NP PP
0 NP . Papa
0 Det . the
0 Det . a

**predict the kind of Det we are looking for (*2 kinds*)**

0
0 ROOT . S
0 S . NP VP
0 NP . Det N
0 NP . NP PP
0 NP . Papa
0 Det . the
0 Det . a

**predict** the kind of NP we're looking for  
*but we were already looking for these so  
don't add duplicate goals! Note that this happened  
when we were processing a left-recursive rule.*

0	Papa	1
0 ROOT . S	0 NP Papa .	
0 S . NP VP		
0 NP . Det N		
0 NP . NP PP		
0 NP . Papa		scan: the desired word is in the input!
0 Det . the		
0 Det . a		

0	Papa	1
0 ROOT . S	0 NP Papa .	
0 S . NP VP		
0 NP . Det N		
0 NP . NP PP		
0 NP . Papa		
0 Det . the	scan: failure	
0 Det . a		

0	Papa	1
0 ROOT . S	0 NP Papa .	
0 S . NP VP		
0 NP . Det N		
0 NP . NP PP		
0 NP . Papa		
0 Det . the		
0 Det . a	scan: failure	

0	Papa	1
0 ROOT . S	0 NP Papa .	
0 S . NP VP	0 S NP . VP	
0 NP . Det N	0 NP NP . PP	
0 NP . NP PP		
0 NP . Papa		
0 Det . the		
0 Det . a		

**attach** the newly created NP  
 (which starts at 0) to its **customers**  
 (incomplete constituents that *end* at 0  
 and have NP after the dot)

0	Papa	1
0 ROOT . S	0 NP Papa .	
0 S . NP VP	0 S NP . VP	
0 NP . Det N	0 NP NP . PP	
0 NP . NP PP	1 VP . V NP	
0 NP . Papa	1 VP . VP PP	
0 Det . the		
0 Det . a		

**predict**

0	Papa	1
0 ROOT . S	0 NP Papa .	
0 S . NP VP	0 S NP . VP	
0 NP . Det N	0 NP NP . PP	
0 NP . NP PP	1 VP . V NP	
0 NP . Papa	1 VP . VP PP	
0 Det . the	1 PP . P NP	
0 Det . a		

**predict**

0	Papa 1
0 ROOT . S	0 NP Papa .
0 S . NP VP	0 S NP . VP
0 NP . Det N	0 NP NP . PP
0 NP . NP PP	1 VP . V NP
0 NP . Papa	1 VP . VP PP
0 Det . the	1 PP . P NP
0 Det . a	1 V . ate

**predict**

0	Papa 1
0 ROOT . S	0 NP Papa .
0 S . NP VP	0 S NP . VP
0 NP . Det N	0 NP NP . PP
0 NP . NP PP	1 VP . V NP
0 NP . Papa	1 VP . VP PP
0 Det . the	1 PP . P NP
0 Det . a	1 V . ate

**predict**

0	Papa 1
0 ROOT . S	0 NP Papa .
0 S . NP VP	0 S NP . VP
0 NP . Det N	0 NP NP . PP
0 NP . NP PP	1 VP . V NP
0 NP . Papa	1 VP . VP PP
0 Det . the	1 PP . P NP
0 Det . a	1 V . ate
	1 P . with

**predict**

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP			
0 NP . Det N	0 NP NP . PP			
0 NP . NP PP	1 VP . V NP			
0 NP . Papa	1 VP . VP PP			
0 Det . the	1 PP . P NP			
0 Det . a	1 V . ate	scan: success!		
	1 P . with			

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP			
0 NP . Det N	0 NP NP . PP			
0 NP . NP PP	1 VP . V NP			
0 NP . Papa	1 VP . VP PP			
0 Det . the	1 PP . P NP			
0 Det . a	1 V . ate			
	1 P . with	scan: failure		

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		
0 NP . Det N	0 NP NP . PP			
0 NP . NP PP	1 VP . V NP			
0 NP . Papa	1 VP . VP PP			
0 Det . the	1 PP . P NP			
0 Det . a	1 V . ate			
	1 P . with			

attach

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa		
0 Det . the	1 PP . P NP			
0 Det . a	1 V . ate			
	1 P . with			

**predict**

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the		
0 Det . a	1 V . ate	2 Det . a		
	1 P . with			

**predict** (*these next few steps should look familiar*)

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the		
0 Det . a	1 V . ate	2 Det . a		
	1 P . with			

**predict**

0	Papa	1	ate	2
0 ROOT . S	0 NP Papa .	1 V ate .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the		
0 Det . a	1 V . ate	2 Det . a		
	1 P . with			

**scan** (*this time we fail since  
Papa is not the next word*)

0	Papa	1	ate	2	the	3
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		
0 S . NP VP	0 S NP . VP	1 VP V . NP				
0 NP . Det N	0 NP NP . PP	2 NP . Det N				
0 NP . NP PP	1 VP . V NP	2 NP . NP PP				
0 NP . Papa	1 VP . VP PP	2 NP . Papa				
0 Det . the	1 PP . P NP	2 Det . the		scan: success!		
0 Det . a	1 V . ate	2 Det . a				
	1 P . with					

0	Papa	1	ate	2	the	3
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		
0 S . NP VP	0 S NP . VP	1 VP V . NP				
0 NP . Det N	0 NP NP . PP	2 NP . Det N				
0 NP . NP PP	1 VP . V NP	2 NP . NP PP				
0 NP . Papa	1 VP . VP PP	2 NP . Papa				
0 Det . the	1 PP . P NP	2 Det . the				
0 Det . a	1 V . ate	2 Det . a				
	1 P . with					

0	Papa	1	ate	2	the	3
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		
0 NP . Det N	0 NP NP . PP	2 NP . Det N				
0 NP . NP PP	1 VP . V NP	2 NP . NP PP				
0 NP . Papa	1 VP . VP PP	2 NP . Papa				
0 Det . the	1 PP . P NP	2 Det . the				
0 Det . a	1 V . ate	2 Det . a				
	1 P . with					

0	Papa	1	ate	2	the	3
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				
0 Det . the	1 PP . P NP	2 Det . the				
0 Det . a	1 V . ate	2 Det . a				
	1 P . with					

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N				
0 NP . Det N	0 NP NP . PP	2 NP . Det N	3 N . caviar					
0 NP . NP PP	1 VP . V NP	2 NP . NP PP	3 N . spoon					
0 NP . Papa	1 VP . VP PP	2 NP . Papa						
0 Det . the	1 PP . P NP	2 Det . the						
0 Det . a	1 V . ate	2 Det . a						
	1 P . with							



0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		2 NP Det N .		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar				
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon				
0 NP . Papa	1 VP . VP PP	2 NP . Papa						
0 Det . the	1 PP . P NP	2 Det . the						
0 Det . a	1 V . ate	2 Det . a						
	1 P . with							

attach

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP		1 VP V . NP		2 NP Det . N		2 NP Det N .	
0 NP . Det N	0 NP NP . PP		2 NP . Det N		3 N . caviar		1 VP V NP .	
0 NP . NP PP	1 VP . V NP		2 NP . NP PP		3 N . spoon		2 NP NP . PP	
0 NP . Papa	1 VP . VP PP		2 NP . Papa					
0 Det . the	1 PP . P NP		2 Det . the					
0 Det . a	1 V . ate		2 Det . a					
	1 P . with							

**attach**  
*(again!)*

# attach *(again!)*

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		
	1 P . with							

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		2 NP Det N .		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		
	1 P . with					0 ROOT S .		

**attach**  
*(again!)*

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		2 NP Det N .		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		
	1 P . with					0 ROOT S .		

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		
	1 P . with					0 ROOT S .		
						4 P . with		

0	Papa	1	ate	2	the	3	caviar	4
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		2 NP Det N .		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		
	1 P . with					0 ROOT S .		
						4 P . with		

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .					
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .					
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP					
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .					
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP					
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP					
	1 P . with				0 ROOT S .					
					4 P . with					

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		4 PPP . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .				
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP				
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .				
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP				
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP				
	1 P . with					0 ROOT S .				
						4 P . with				

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		2 NP Det N .		4 PPP . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		5 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		5 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		5 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP				
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP				
	1 P . with					0 ROOT S .				
						4 P . with				

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		4 PP P . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		5 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		5 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		5 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		5 Det . the		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		5 Det . a		
	1 P . with					0 ROOT S .				
						4 P . with				

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		4 PP P . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		5 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		5 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		5 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		5 Det . the		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		5 Det . a		
	1 P . with					0 ROOT S .				
						4 P . with				

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		4 PP P . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		5 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		5 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		5 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		5 Det . the		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		5 Det . a		
	1 P . with					0 ROOT S .				
						4 P . with				

0	Papa	1	ate	2	the	3	caviar	4	with	5
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .		3 N caviar .		4 P with .		
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N		3 NP Det N .		4 PP P . NP		
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar		1 VP V NP .		5 NP . Det N		
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon		2 NP NP . PP		5 NP . NP PP		
0 NP . Papa	1 VP . VP PP	2 NP . Papa				0 S NP VP .		5 NP . Papa		
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP		5 Det . the		
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP		5 Det . a		
	1 P . with					0 ROOT S .				
						4 P . with				









e	2	the	3	caviar	4	with	5	a	6	spoon	7
	1 V ate .	2 Det the .	3 N caviar .	4 P with .	5 Det a .	6 N spoon .					
	1 VP V . NP	2 NP Det . N	2 NP Det N .	4 PP P . NP	5 NP Det . N						
P	2 NP . Det N	3 N . caviar	1 VP V NP .	5 NP . Det N	6 N . caviar						
	2 NP . NP PP	3 N . spoon	2 NP NP . PP	5 NP . NP PP	6 N . spoon						
D	2 NP . Papa		0 S NP VP .	5 NP . Papa							
	2 Det . the		1 VP VP . PP	5 Det . the							
	2 Det . a		4 PP . P NP	5 Det . a							
			0 ROOT S .								
			4 P . with								

e	2	the	3	caviar	4	with	5	a	6	spoon	7
	1 V ate .	2 Det the .	3 N caviar .	4 P with .	5 Det a .	6 N spoon .					
	1 VP V . NP	2 NP Det . N	2 NP Det N .	4 PP P . NP	5 NP Det . N	5 NP Det N .					
P	2 NP . Det N	3 N . caviar	1 VP V NP .	5 NP . Det N	6 N . caviar						
	2 NP . NP PP	3 N . spoon	2 NP NP . PP	5 NP . NP PP	6 N . spoon						
D	2 NP . Papa		0 S NP VP .	5 NP . Papa							
	2 Det . the		1 VP VP . PP	5 Det . the							
	2 Det . a		4 PP . P NP	5 Det . a							
			0 ROOT S .								
			4 P . with								

e	2	the	3	caviar	4	with	5	a	6	spoon	7
	1 V ate .	2 Det the .	3 N caviar .	4 P with .	5 Det a .	6 N spoon .					
	1 VP V . NP	2 NP Det . N	2 NP Det N .	4 PPP . NP	5 NP Det . N	5 NP Det N .					
P	2 NP . Det N	3 N . caviar	1 VP V NP .	5 NP . Det N	6 N . caviar	4 PPP NP .					
	2 NP . NP PP	3 N . spoon	2 NP NP . PP	5 NP . NP PP	6 N . spoon	5 NP NP . PP					
D	2 NP . Papa		0 S NP VP .	5 NP . Papa							
	2 Det . the		1 VP VP . PP	5 Det . the							
	2 Det . a		4 PP . P NP	5 Det . a							
			0 ROOT S .								
			4 P . with								

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			...	6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP			1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP				
	1 P . with					0 ROOT S .				
						4 P . with				

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			...	6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .					
					4 P . with					

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			...	6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			...	6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP			1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP			7 PP . P NP	
	1 P . with					0 ROOT S .			1 VP V NP .	
						4 P . with			2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP			1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP			7 PP . P NP	
	1 P . with					0 ROOT S .			1 VP V NP .	
						4 P . with			2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	
									0 ROOT S .	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	
									0 ROOT S .	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			...	6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	
									0 ROOT S .	

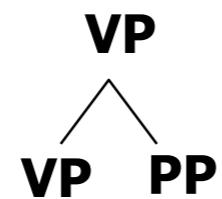
0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	
									0 ROOT S .	

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .				... 6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the			1 VP VP . PP				1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a			4 PP . P NP				7 PP . P NP	
	1 P . with				0 ROOT S .				1 VP V NP .	
					4 P . with				2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	
									0 ROOT S .	

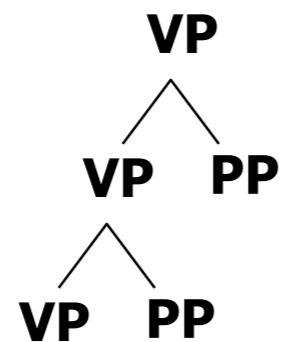
## **Left Recursion Kills Pure Top-Down Parsing ...**

**VP**

## Left Recursion Kills Pure Top-Down Parsing ...

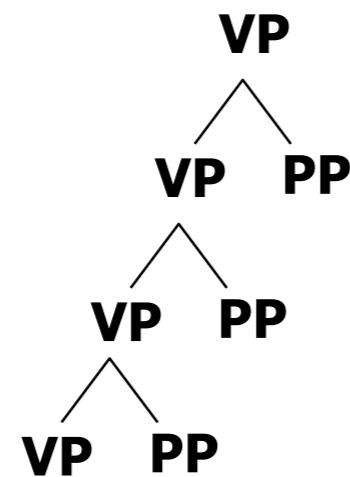


## Left Recursion Kills Pure Top-Down Parsing ...



Andrew McCallum, UMass Amherst

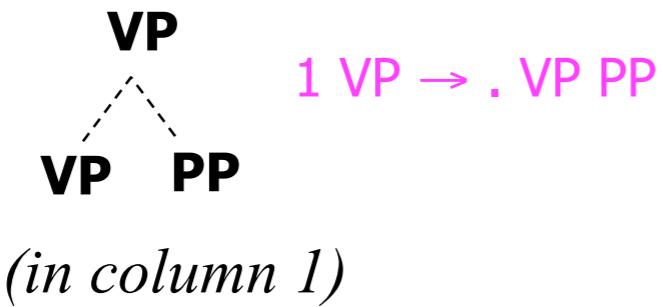
## Left Recursion Kills Pure Top-Down Parsing ...



makes new hypotheses  
ad infinitum before we've  
seen the PPs at all

hypotheses try to predict  
in advance how many  
PP's will arrive in input

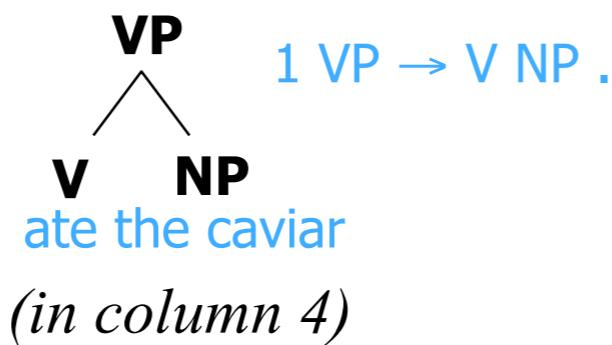
# ... but Earley's Alg is Okay!



## ... but Earley's Alg is Okay!

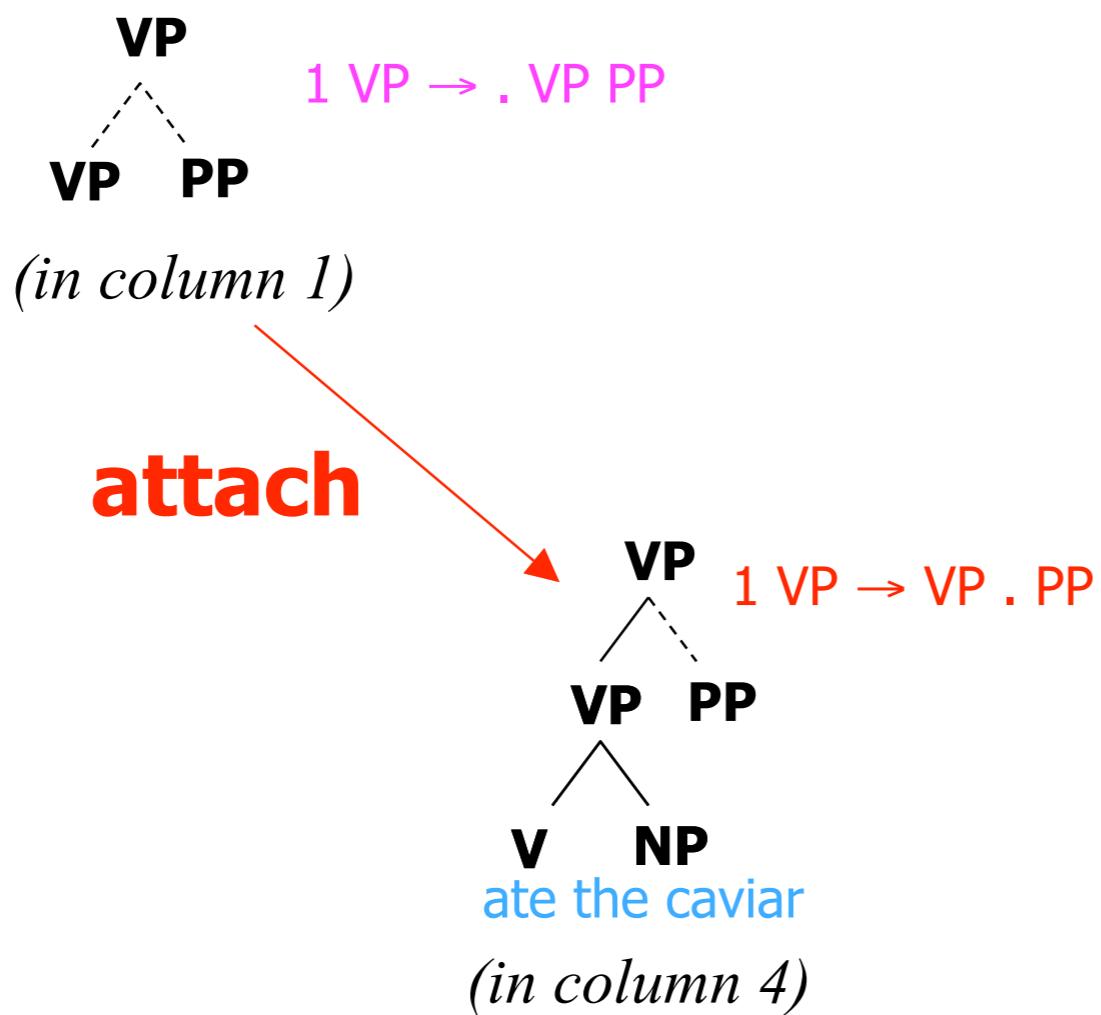


*(in column 1)*



*(in column 4)*

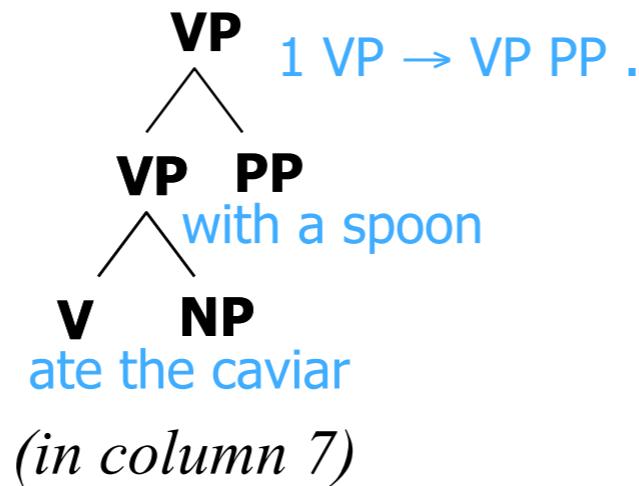
## ... but Earley's Alg is Okay!



## ... but Earley's Alg is Okay!

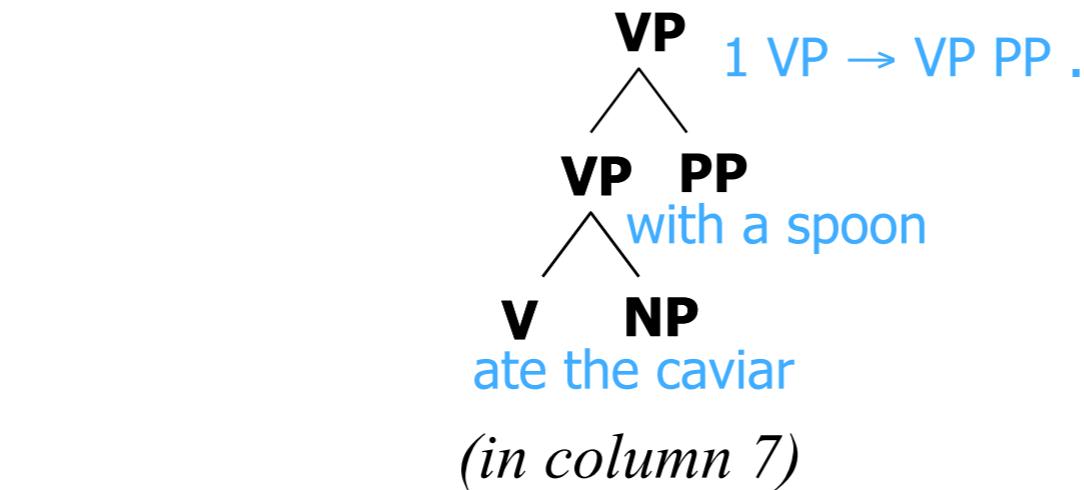
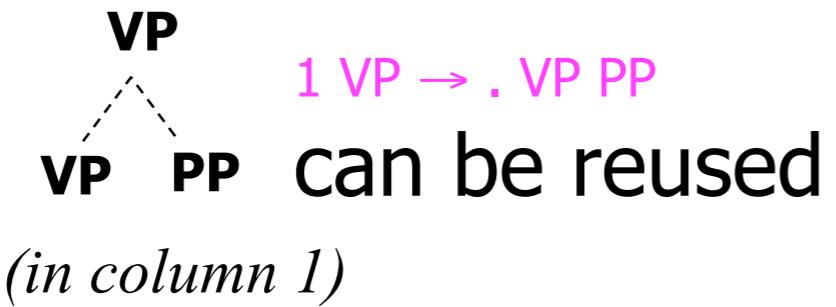


*(in column 1)*

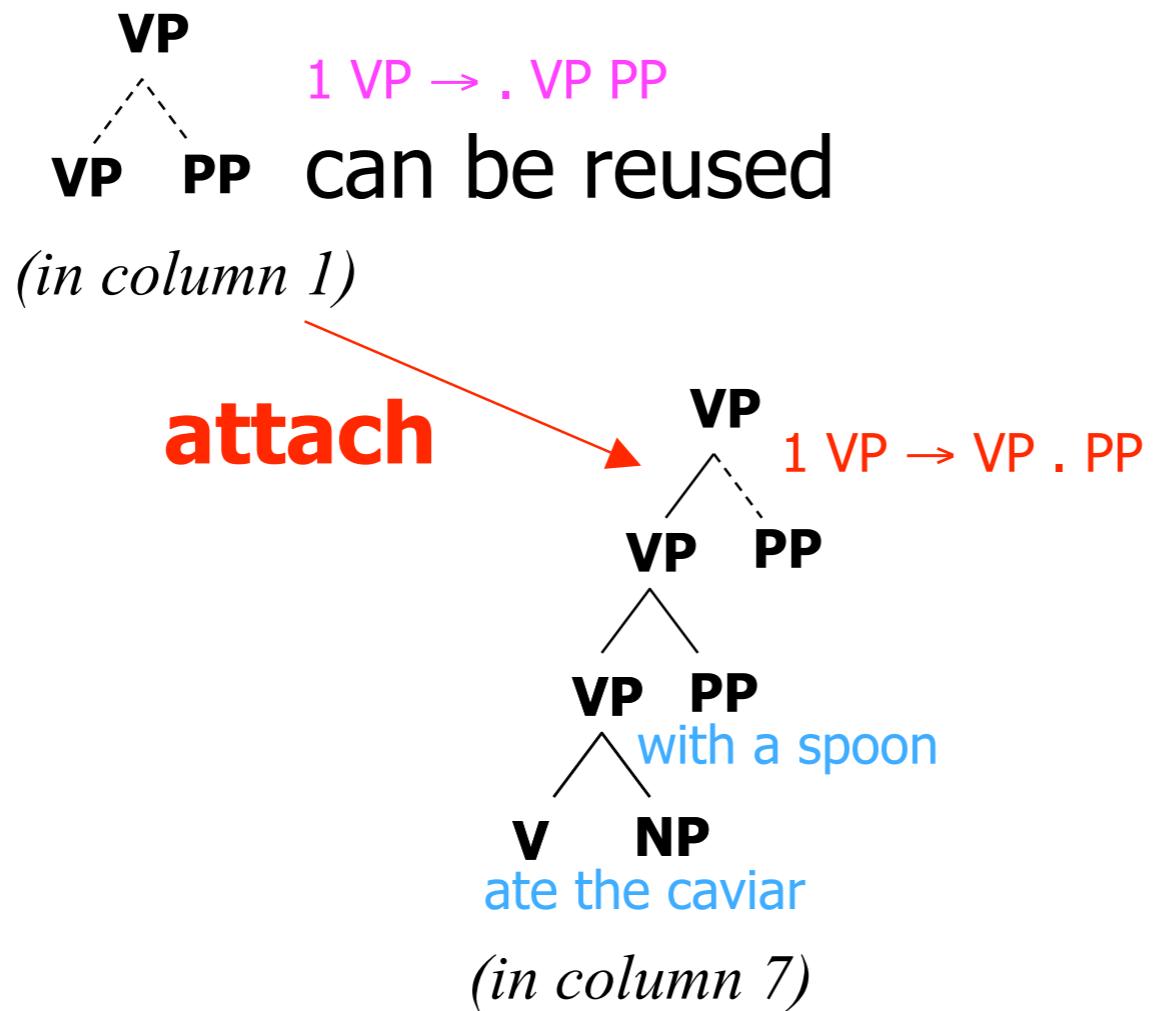


*(in column 7)*

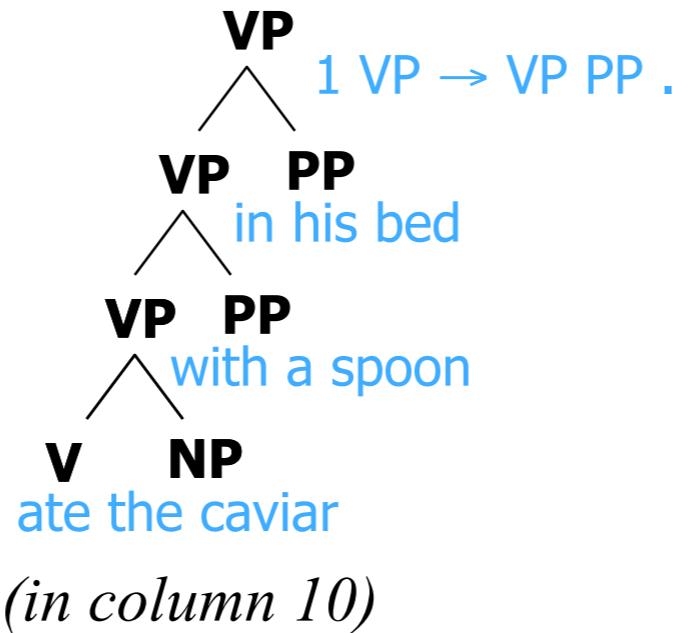
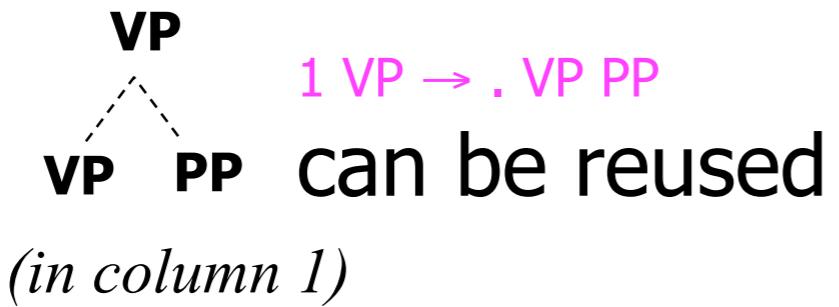
## ... but Earley's Alg is Okay!



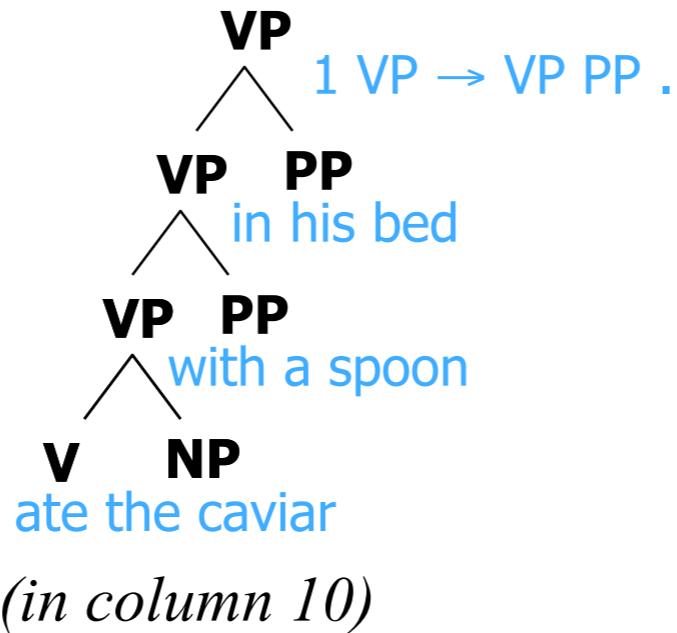
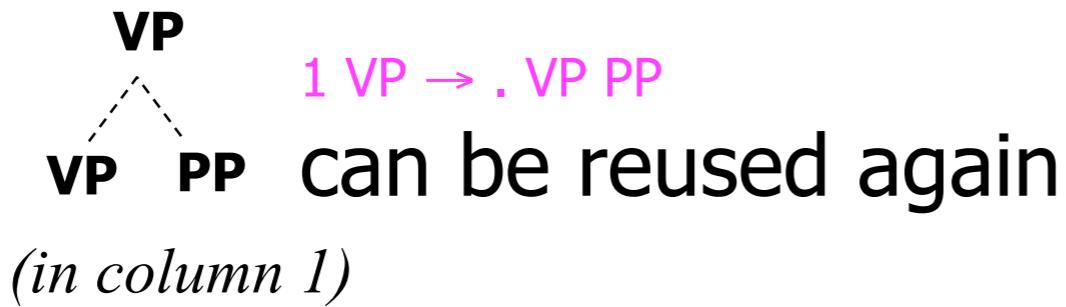
## ... but Earley's Alg is Okay!



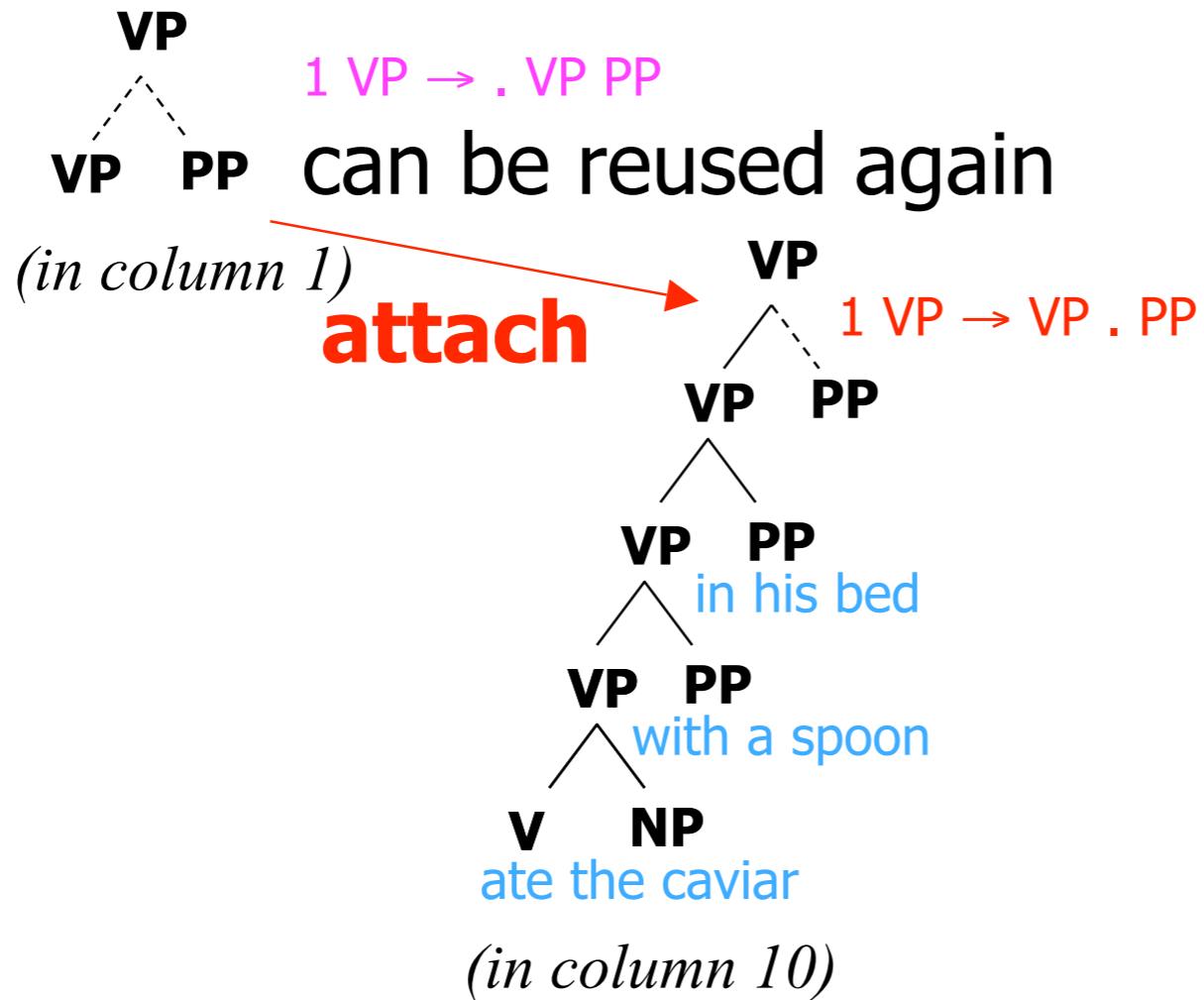
## ... but Earley's Alg is Okay!



## ... but Earley's Alg is Okay!



## ... but Earley's Alg is Okay!



0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .		1 V ate .		2 Det the .		3 N caviar .			... 6 N spoon .
0 S . NP VP	0 S NP . VP		1 VP V . NP		2 NP Det . N		3 N . caviar	1 VP V NP .		5 NP Det N .
0 NP . Det N	0 NP NP . PP		2 NP . Det N		3 N . caviar		4 PP P NP .			5 NP NP . PP
0 NP . NP PP	1 VP . V NP		2 NP . NP PP		3 N . spoon		5 NP NP PP .			2 NP NP PP .
0 NP . Papa	1 VP . VP PP		2 NP . Papa				6 N spoon .	0 S NP VP .		1 VP VP PP .
0 Det . the	1 PP . P NP		2 Det . the				7 PP . P NP	1 VP VP . PP		2 NP NP . PP
0 Det . a	1 V . ate		2 Det . a				8 P . with	0 S NP VP .		1 VP V NP .
	1 P . with							1 VP VP . PP		2 NP NP . PP
								7 P . with		0 S NP VP .
								0 ROOT S .		1 VP V NP .
										2 NP NP . PP
										0 S NP VP .
										1 VP VP . PP
										7 P . with
										0 ROOT S .

completed a VP in col 4  
 col 1 lets us use it in a VP PP structure

0	Papa	1	ate	2	the	3	caviar	4	with a spoon	7
0 ROOT . S	0 NP Papa .	1 V ate .		2 Det the .	3 N caviar .			...	6 N spoon .	
0 S . NP VP	0 S NP . VP	1 VP V . NP		2 NP Det . N	2 NP Det N .				5 NP Det N .	
0 NP . Det N	0 NP NP . PP	2 NP . Det N		3 N . caviar	1 VP V NP .				4 PP P NP .	
0 NP . NP PP	1 VP . V NP	2 NP . NP PP		3 N . spoon	2 NP NP . PP				5 NP NP . PP	
0 NP . Papa	1 VP . VP PP	2 NP . Papa			0 S NP VP .				2 NP NP PP .	
0 Det . the	1 PP . P NP	2 Det . the				1 VP VP . PP			1 VP VP PP .	
0 Det . a	1 V . ate	2 Det . a				4 PP . P NP			7 PP . P NP	
	1 P . with					0 ROOT S .			1 VP V NP .	
						4 P . with			2 NP NP . PP	
									0 S NP VP .	
									1 VP VP . PP	
									7 P . with	
									0 ROOT S .	

completed that VP = VP PP in col 7  
 col 1 would let us use *it* in a VP PP structure  
 can reuse col 1 as often as we need

# Beyond Recognition

- So far, we've described an Earley *recognizer*
- Note what we did when we tried to create entries that already existed
- What should we do when combining items?
- How to derive outside algorithm?

# Better Treebank Grammars

- Last time: CNF conversion, Markovization
- Clustering and splitting nonterminals
- Grandparent annotation
- Lexicalization