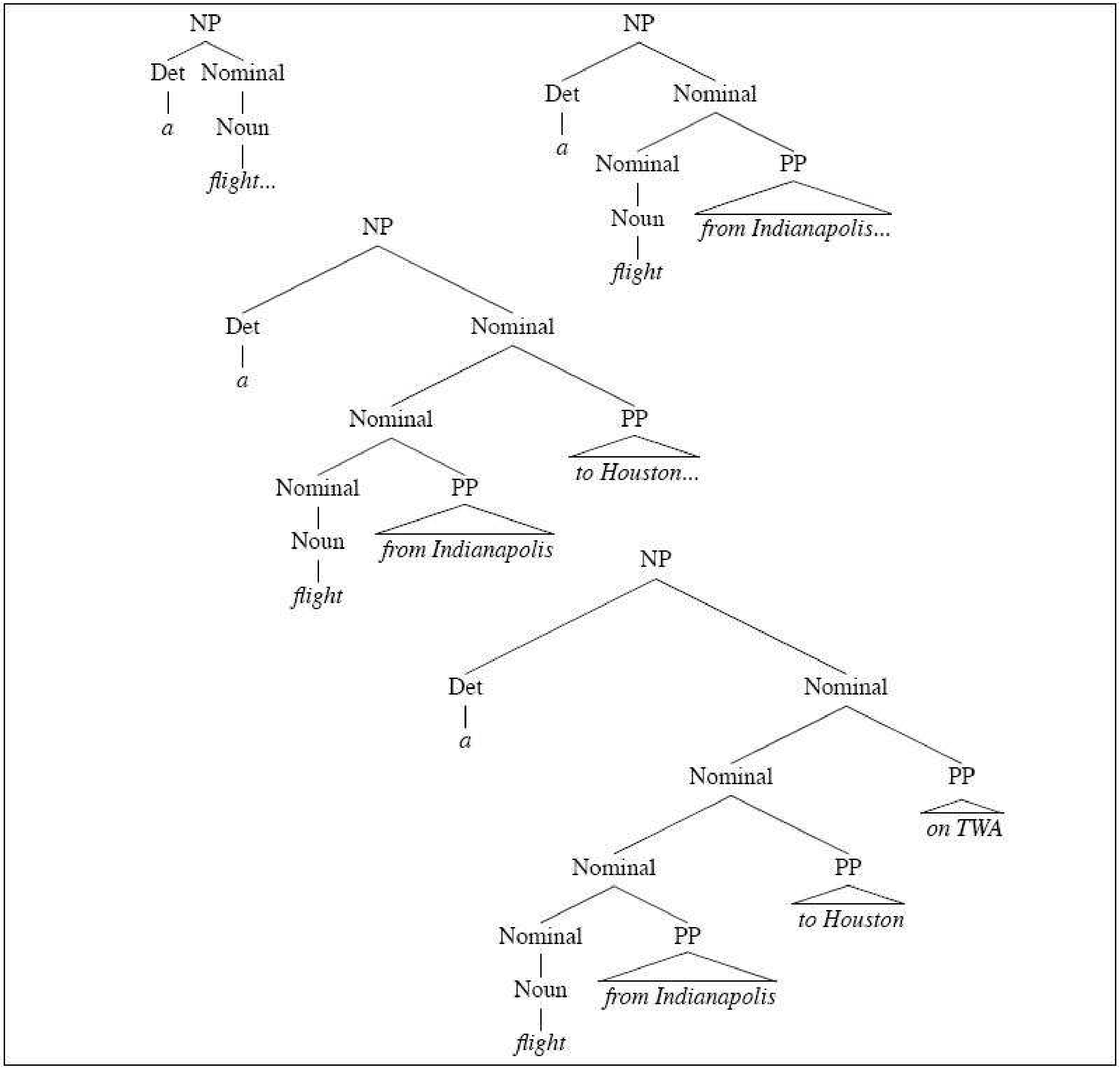
CS460/626 : Natural Language Processing/Speech, NLP and the Web

Lecture 19, 24: Algorithmics of Probabilistic Parsing; HMM-PCFG correspondence

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay

1st and 11th Oct, 2012

Shared Sub-Problems: Example



CKY Parsing: CNF

- CKY parsing requires that the grammar consist of ε-free, binary rules = Chomsky Normal Form
  - All rules of the form:
    - $A \rightarrow BC$ or $A \rightarrow a$
  - What does the tree look like?
- What if my CFG isn't in CNF?

CKY Algorithm

function CKY-PARSE(*words, grammar*) **returns** *table*

for $j \leftarrow$ **from** 1 **to** LENGTH(*words*) **do**

$table[j-1, j] \leftarrow \{A \mid A \rightarrow words[j] \in grammar\}$

for $i \leftarrow$ **from** $j-2$ **downto** 0 **do**

for $k \leftarrow i+1$ **to** $j-1$ **do**

$$table[i,j] \leftarrow table[i,j] \cup \{A \mid A \rightarrow BC \in grammar, B \in table[i,k], C \in table[k,j]\}$$

Illustrating CYK [Cocke, Younger, Kashmi] Algo

- S $\rightarrow$ NP VP 1.0
- NP $\rightarrow$ DT NN 0.5
- NP $\rightarrow$ NNS 0.3
- NP $\rightarrow$ NP PP 0.2
- PP $\rightarrow$ P NP 1.0
- VP $\rightarrow$ VP PP 0.6
- VP $\rightarrow$ VBD NP 0.4

- DT $\rightarrow$ the 1.0
- NN $\rightarrow$ gunman 0.5
- NN $\rightarrow$ building 0.5
- VBD $\rightarrow$ sprayed 1.0
- NNS $\rightarrow$ bullets 1.0

CYK: Start with (0,1)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7 *.*

To From	1	2	3	4	5	6	7
0	DT						
1	-------						
2	-------	---------					
3	-------	---------	--------				
4	--------	---------	--------	---------			
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK: Keep filling diagonals

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7 *.*

To From	1	2	3	4	5	6	7
0	DT						
1	-------	NN					
2	-------	---------					
3	-------	---------	--------				
4	--------	---------	--------	---------			
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK: Try getting higher level structures

0 *The* **1** *gunman* **2** *sprayed* **3** *the* **4** *building* **5** *with* **6** *bullets* **7** *.*

To From	1	2	3	4	5	6	7
0	DT	NP					
1	-------	NN					
2	-------	---------					
3	-------	---------	--------				
4	--------	---------	--------	---------			
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK: Diagonal continues

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7 *.*

To **From**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	DT	NP					
1	-------	NN					
2	-------	------- --	VBD				
3	-------	------- --	------- -				
4	------- -	------- --	------- -	-------- -			
5	------- -	------- --	------- -	-------- -	-------- -		
6	------- -	------- --	------- -	-------- -	-------- -	-------- -	

CYK (cont…)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7 *.*

To From	1	2	3	4	5	6	7
0	DT	NP	---------				
1	-------	NN	---------				
2	-------	---------	VBD				
3	-------	---------	--------				
4	--------	---------	--------	---------			
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK (cont…)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	--------				
1	-------	NN	--------				
2	-------	---------	VBD				
3	-------	---------	--------	DT			
4	--------	---------	--------	---------			
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK (cont...)

0 *The* **1** *gunman* **2** *sprayed* **3** *the* **4** *building* **5** *with* **6** *bullets* **7** *.*

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------			
1	-------	NN	--------	---------			
2	-------	---------	VBD	---------			
3	-------	---------	--------	DT			
4	--------	---------	--------	---------	NN		
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK: starts filling the 5th column

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------			
1	-------	NN	--------	---------			
2	-------	---------	VBD	---------			
3	-------	---------	--------	DT	NP		
4	--------	---------	--------	---------	NN		
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK (cont...)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	---------	----------			
1	-------	NN	---------	----------			
2	-------	---------	VBD	----------	VP		
3	-------	---------	---------	DT	NP		
4	---------	---------	---------	----------	NN		
5	---------	---------	---------	----------	----------		
6	---------	---------	---------	----------	----------	----------	

CYK (cont...)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	------- -	-------- -			
1	-------	NN	------- -	-------- -	-------- -		
2	-------	------- --	VBD	-------- -	VP		
3	-------	------- --	------- -	DT	NP		
4	------- -	------- --	------- -	-------- -	NN		
5	------- -	------- --	------- -	-------- -	-------- -		
6	------- -	------- --	------- -	-------- -	-------- -	-------- -	

CYK: S found, but NO termination!

0 *The* **1** *gunman* **2** *sprayed* **3** *the* **4** *building* **5** *with* **6** *bullets* **7** *.*

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------	S		
1	-------	NN	--------	---------	---------		
2	-------	---------	VBD	---------	VP		
3	-------	---------	--------	DT	NP		
4	--------	---------	--------	---------	NN		
5	--------	---------	--------	---------	---------		
6	--------	---------	--------	---------	---------	---------	

CYK (cont...)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------	S		
1	-------	NN	--------	---------	---------		
2	-------	---------	VBD	---------	VP		
3	-------	---------	--------	DT	NP		
4	--------	---------	--------	---------	NN		
5	--------	---------	--------	---------	---------	P	
6	--------	---------	--------	---------	---------	---------	

CYK (cont…)

**0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	------- -	-------- -	S	-------- -	
1	-------	NN	------- -	-------- -	-------- -	-------- -	
2	-------	------- --	VBD	-------- -	VP	-------- -	
3	-------	------- --	------- -	DT	NP	-------- -	
4	------- -	------- --	------- -	-------- -	NN	-------- -	
5	------- -	------- --	------- -	-------- -	-------- -	P	
6	------- -	------- --	------- -	-------- -	-------- -	-------- -	

CYK: Control moves to last column

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	---------	---------	S	---------	
1	-------	NN	---------	---------	---------	---------	
2	-------	---------	VBD	---------	VP	---------	
3	-------	---------	---------	DT	NP	---------	
4	--------	---------	--------	---------	NN	---------	
5	--------	---------	--------	---------	---------	P	
6	--------	---------	--------	---------	---------	---------	NP NNS

CYK (cont…)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------	S	---------	
1	-------	NN	--------	---------	---------	---------	
2	-------	---------	VBD	---------	VP	---------	
3	-------	---------	--------	DT	NP	---------	
4	--------	---------	--------	---------	NN	---------	
5	--------	---------	--------	---------	---------	P	PP
6	--------	---------	--------	---------	---------	---------	NP NNS

CYK (cont…)

0 *The* **1** *gunman* **2** *sprayed* **3** *the* **4** *building* **5** *with* **6** *bullets* **7** *.*

To From	1	2	3	4	5	6	7
0	DT	NP	---------	---------	S	---------	
1	-------	NN	---------	---------	---------	---------	
2	-------	---------	VBD	---------	VP	---------	
3	-------	---------	---------	DT	NP	---------	NP
4	--------	---------	---------	---------	NN	---------	---------
5	--------	---------	---------	---------	---------	P	PP
6	--------	---------	---------	---------	---------	---------	NP NNS

CYK (cont...)

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------	S	---------	
1	-------	NN	--------	---------	---------	---------	
2	-------	---------	VBD	---------	VP	---------	VP
3	-------	---------	--------	DT	NP	---------	NP
4	--------	---------	--------	---------	NN	---------	---------
5	--------	---------	--------	---------	---------	P	PP
6	--------	---------	--------	---------	---------	---------	NP NNS

CYK: filling the last column

0 *The* 1 *gunman* 2 *sprayed* 3 *the* 4 *building* 5 *with* 6 *bullets* 7.

To From	1	2	3	4	5	6	7
0	DT	NP	--------	---------	S	---------	
1	-------	NN	--------	---------	---------	---------	---------
2	-------	---------	VBD	---------	VP	---------	VP
3	-------	---------	--------	DT	NP	---------	NP
4	--------	---------	--------	---------	NN	---------	---------
5	--------	---------	--------	---------	---------	P	PP
6	--------	---------	--------	---------	---------	---------	NP NNS

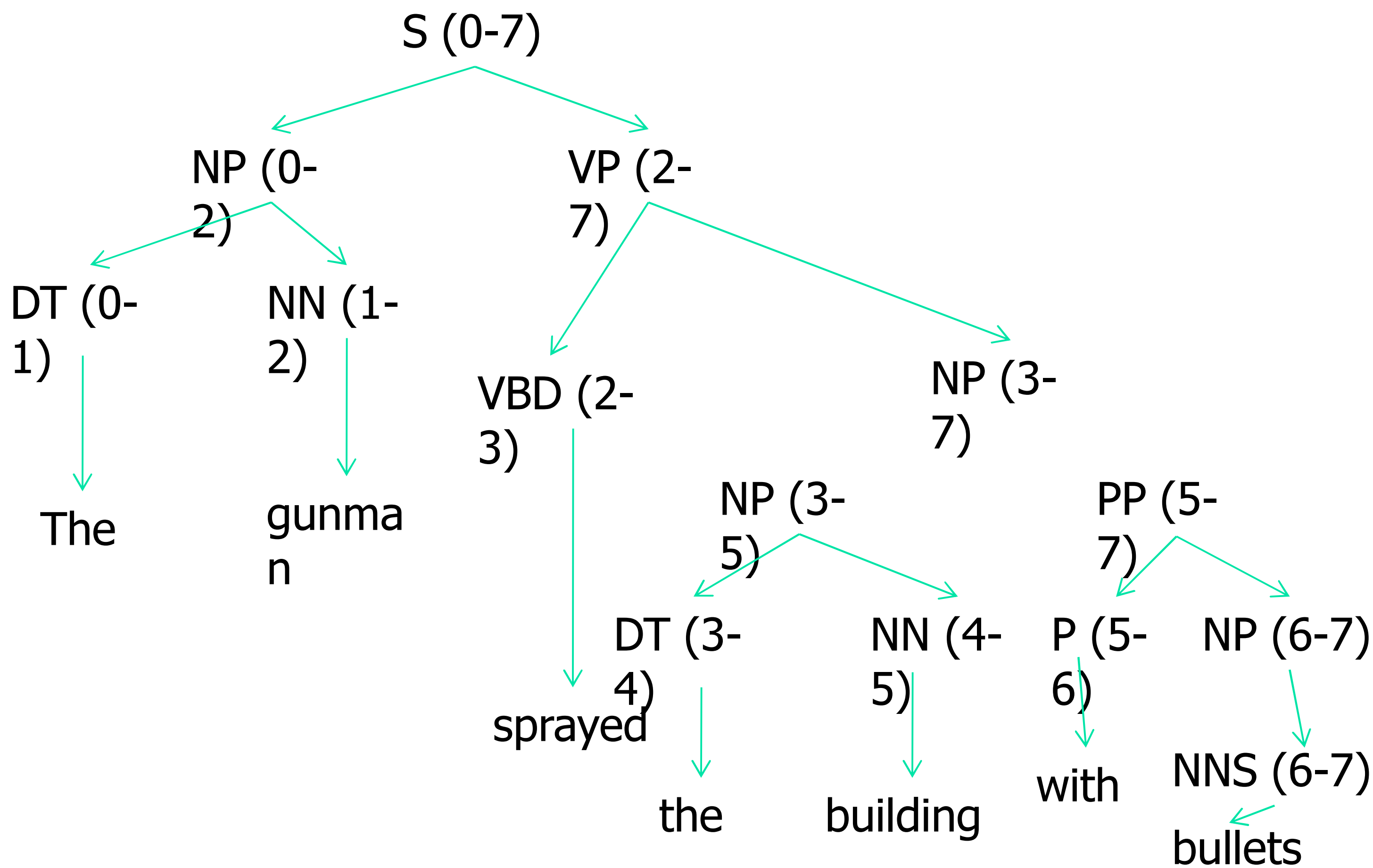
CYK: terminates with S in (0,7)

0 *The* **1** *gunman* **2** *sprayed* **3** *the* **4** *building* **5** *with* **6** *bullets* **7**.

To From	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	DT	NP	------- -	-------- -	S	-------- -	S
1	-------	NN	------- -	-------- -	-------- -	-------- -	-------- -
2	-------	------- --	VBD	-------- -	VP	-------- -	VP
3	-------	------- --	------- -	DT	NP	-------- -	NP
4	------- -	------- --	------- -	-------- -	NN	-------- -	-------- -
5	------- -	------- --	------- -	-------- -	-------- -	P	PP
6	------- -	------- --	------- -	-------- -	-------- -	-------- -	NP NNS

CYK: Extracting the Parse Tree

- The parse tree is obtained by keeping back pointers.



Probabilistic parse tree construction

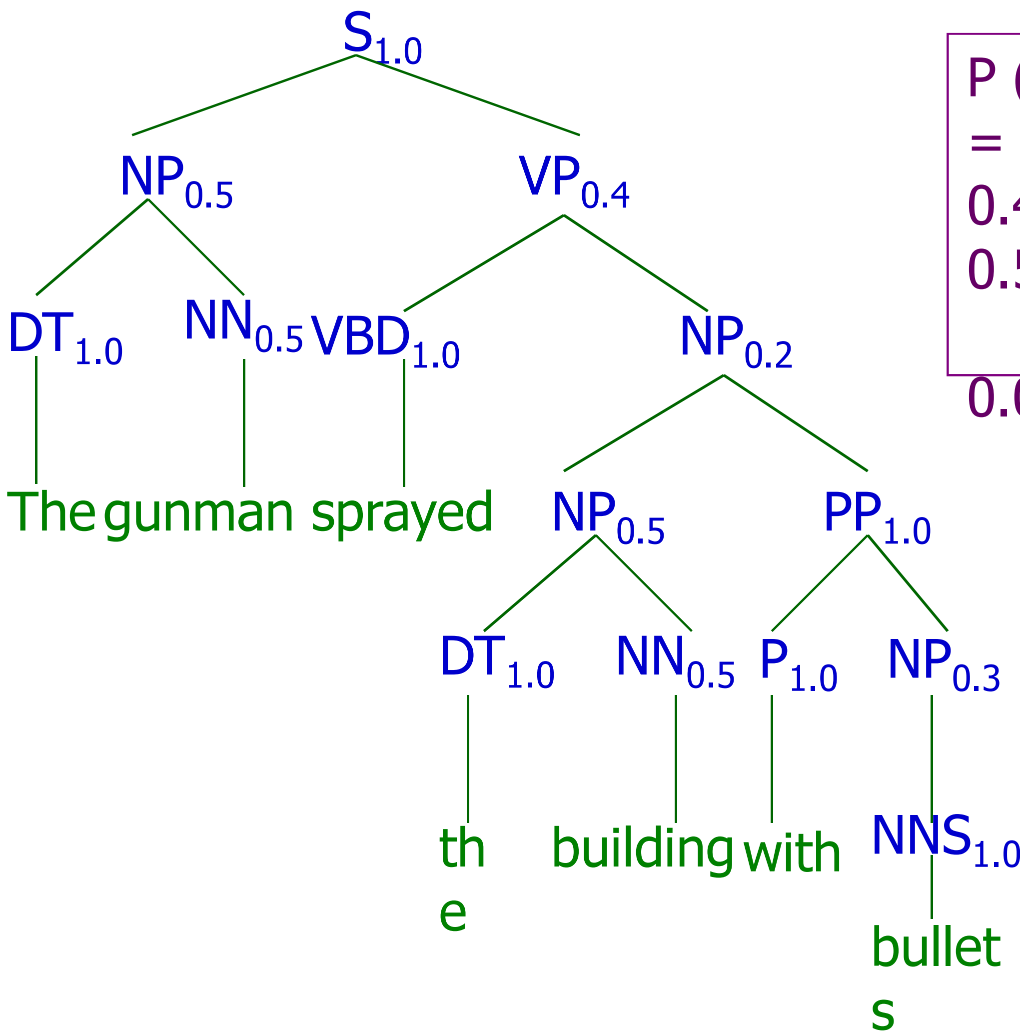
Probabilistic Context Free Grammars

- $S \rightarrow NP\ VP$ 1.0
- $NP \rightarrow DT\ NN$ 0.5
- $NP \rightarrow NNS$ 0.3
- $NP \rightarrow NP\ PP$ 0.2
- $PP \rightarrow P\ NP$ 1.0
- $VP \rightarrow VP\ PP$ 0.6
- $VP \rightarrow VBD\ NP$ 0.4

- $DT \rightarrow$ the 1.0
- $NN \rightarrow$ gunman 0.5
- $NN \rightarrow$ building 0.5
- $VBD \rightarrow$ sprayed 1.0
- $NNS \rightarrow$ bullets 1.0

Another Parse t_2

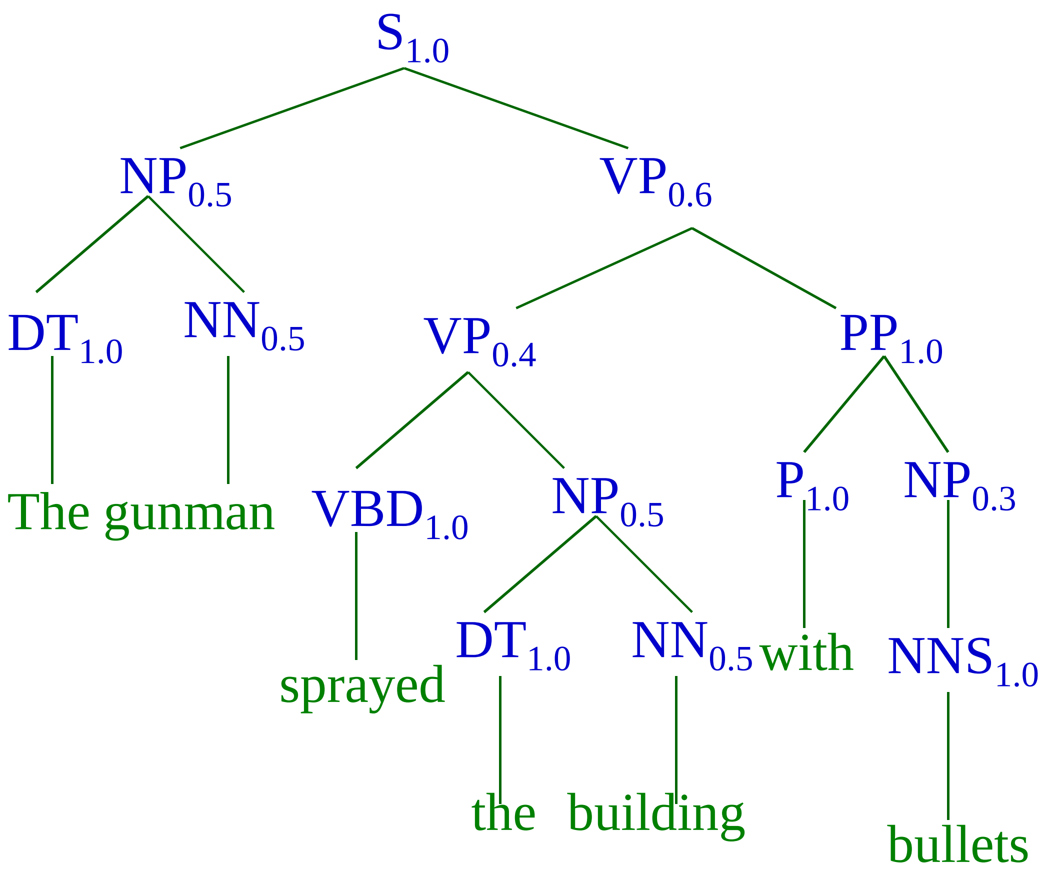
- The gunman sprayed the building with bullets.



P (t_2)
= 1.0 * 0.5 * 1.0 * 0.5 * 0.4 * 1.0 * 0.2 * 0.5 * 1.0 * 0.5 * 1.0 * 1.0 * 0.3 * 1.0
= 0.0015

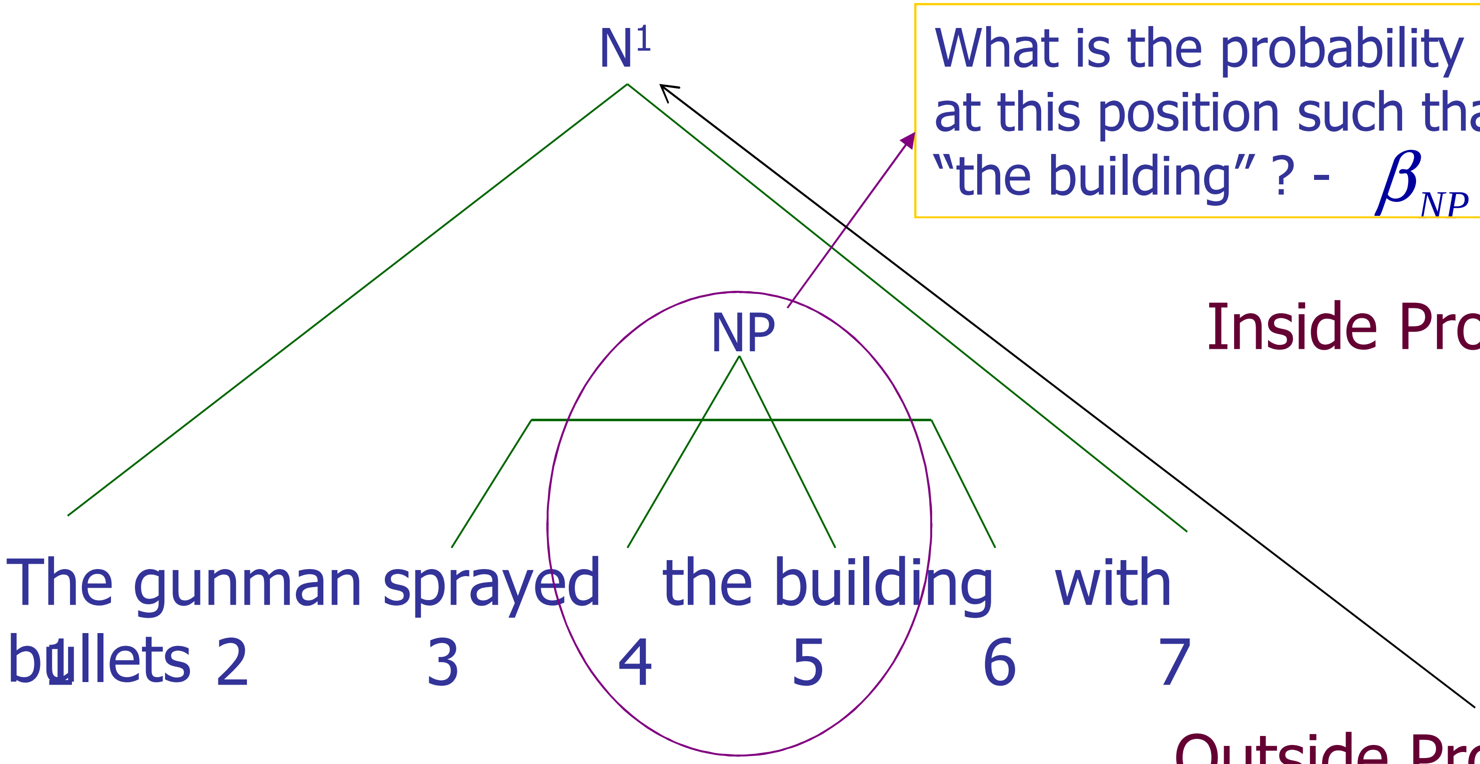
Example Parse t_1`

- The gunman sprayed the building with bullets.



$P(t_1)$ = 1.0 * 0.5 * 1.0 * 0.5 * 0.6 * 0.4 * 1.0 * 0.5 * 1.0 * 0.5 * 1.0 * 1.0 * 0.3 * 1.0 = 0.00225

Interesting Probabilities



What is the probability of having a NP at this position such that it will derive "the building" ? - $\beta_{NP}(4,5)$

Inside Probabilities

Outside Probabilities

What is the probability of starting from N1 and deriving "The gunman sprayed", a NP and "with bullets" ? - $\alpha_{NP}(4,5)$

Interesting Probabilities

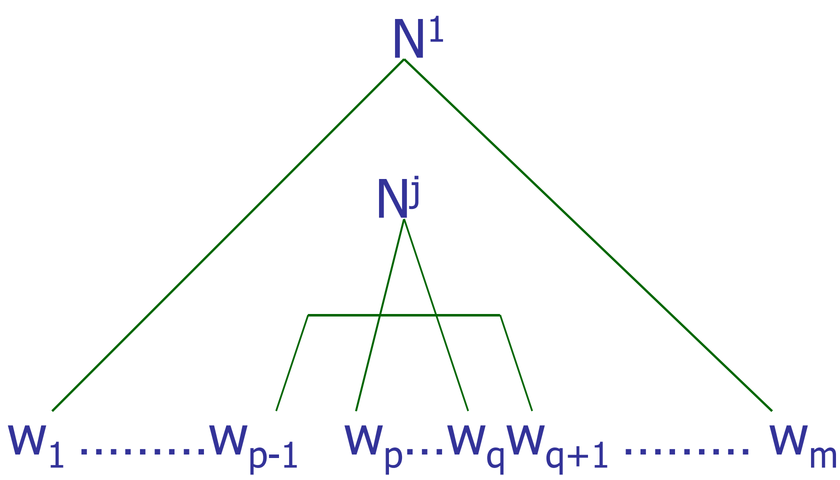
- Random variables to be considered
  - The non-terminal being expanded.
    E.g., NP
  - The word-span covered by the non-terminal.
    *E.g., (*4,5) refers to words "the building"

- While calculating probabilities, consider:
  - The rule to be used for expansion :
    E.g., NP $\rightarrow$ DT NN
  - The probabilities associated with the RHS non-terminals : *E.g.,* DT subtree's inside/outside probabilities & NN subtree's inside/outside probabilities

Outside Probability

- $\alpha_j(p,q)$: The probability of beginning with N^1 & generating the non-terminal N^j_{pq} and all words outside $w_p..w_q$

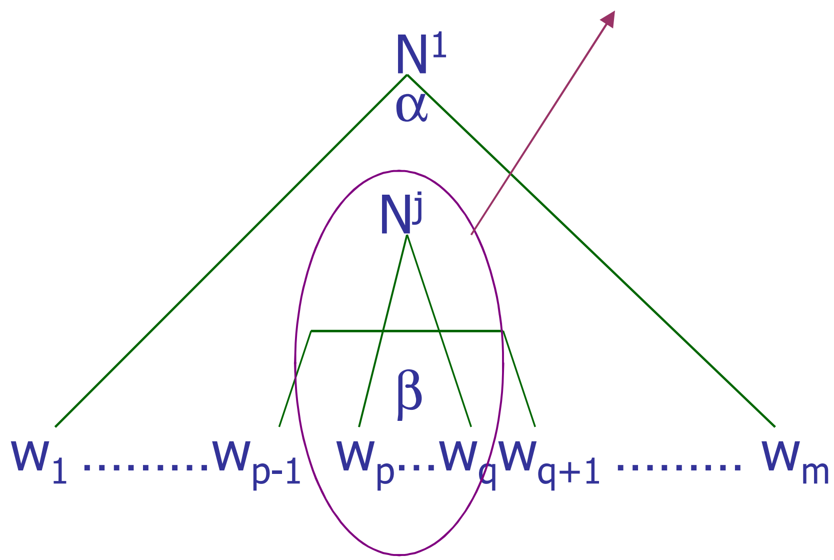
$$\alpha_j(p,q) = P(w_{1(p-1)}, N^j_{pq}, w_{(q+1)m} \mid G)$$



Inside Probabilities

- $\beta_j(p,q)$: The probability of generating the words $w_p..w_q$ starting with the non-terminal N^j_{pq}.

$$\beta_j(p,q) = P(w_{pq} \mid N^j_{pq}, G)$$

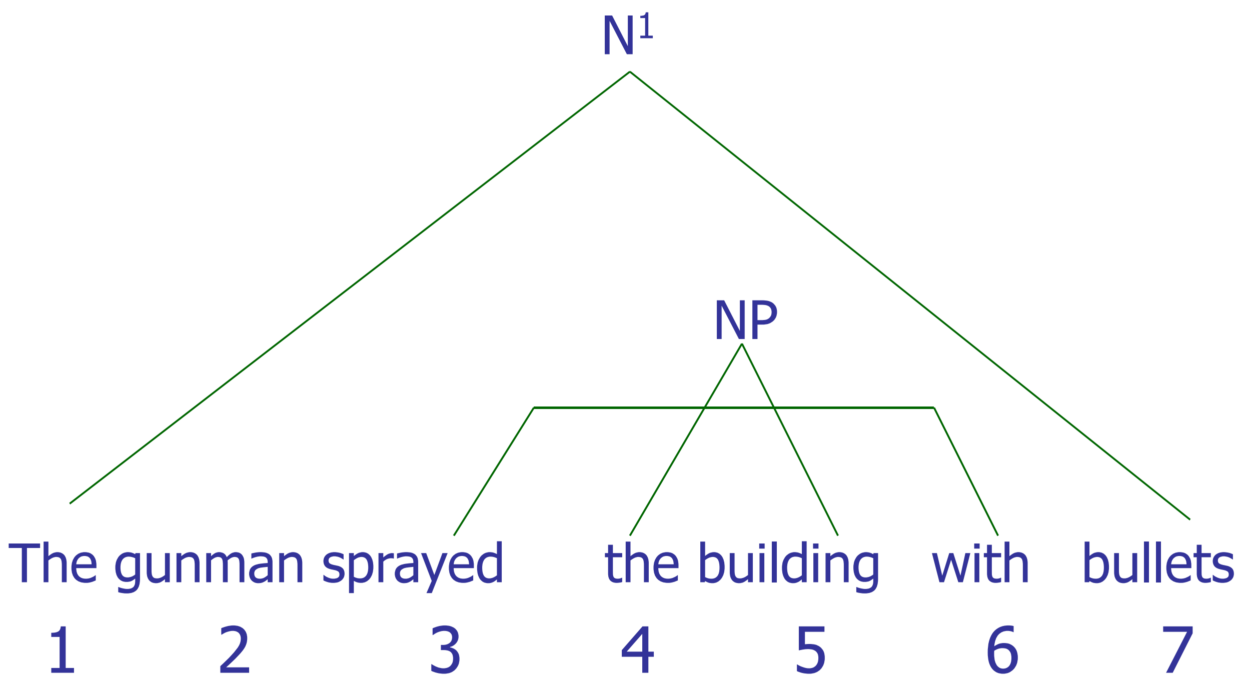


Outside & Inside Probabilities: example

$$\alpha_{NP}(4,5) \text{ for "the building"}$$

$$= P(\text{The gunman sprayed}, NP_{4,5}, \text{with bullets} \mid G)$$

$$\beta_{NP}(4,5) \text{ for "the building"} = P(\text{the building} \mid NP_{4,5}, G)$$



Calculating Inside probabilities $\beta_j(p,q)$

Base case:

$$\beta_j(k,k) = P(w_k \mid N^j_{kk}, G) = P(N^j_{kk} \rightarrow w_k \mid G)$$

- Base case is used for rules which derive the words or terminals directly

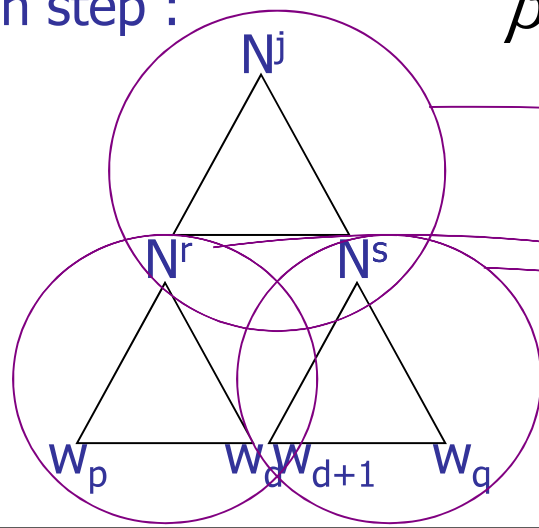
E.g., Suppose N^j = NN is being considered & NN $\rightarrow$ building is one of the rules with probability 0.5

$$\beta_{NN}(5,5) = P(building \mid NN_{5,5}, G)$$

$$= P(NN_{5,5} \rightarrow building \mid G) = 0.5$$

Induction Step: Assuming Grammar in Chomsky Normal Form

Induction step :



$$\beta_j(p,q) = P(w_{pq} \mid N^j_{pq}, G)$$

$$= \sum_{r,s} \sum_{d=p}^{q-1} P(N^j \rightarrow N^r N^s) * \beta_r(p,d) * \beta_s(d+1,q)$$

- Consider different splits of the words - indicated by *d*

  E.g., the huge building

  Split here for d=2 d=3

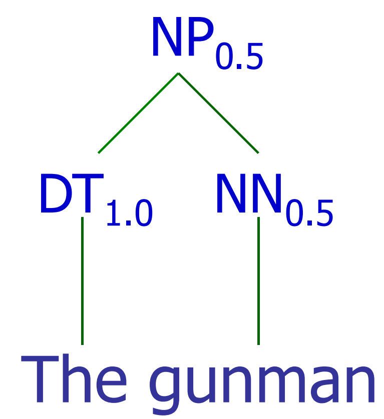
- Consider different non-terminals to be used in the rule:

  NP $\rightarrow$ DT NN, NP $\rightarrow$ DT NNS are available options

  Consider summation over all these.

The Bottom-Up Approach

- The idea of induction
- Consider "the gunman"



- Base cases : Apply unary rules

  DT $\rightarrow$ the Prob = 1.0

  NN $\rightarrow$ gunman Prob = 0.5

- Induction : Prob that a NP covers these 2 words

  = P (NP $\rightarrow$ DT NN) * P (DT deriving the word "the") * P (NN deriving the word "gunman")

  = 0.5 * 1.0 * 0.5 = 0.25

Parse Triangle (probablity of output observation, the sentence)

- A parse triangle is constructed for calculating $\beta_j(p,q)$
- Probability of a sentence using $\beta_j(p,q)$:

$$P(w_{1m} \mid G) = P(N^1 \rightarrow w_{1m} \mid G) = P(w_{1m} \mid N^1_{1m}, G) = \beta_1(1, m)$$

Parse Triangle

to / from	The (1)	gunman (2)	sprayed (3)	the (4)	building (5)	with (6)	bullets (7)
0	$\beta_{DT} = 1.0$						
1		$\beta_{NN} = 0.5$					
2			$\beta_{VBD} = 1.0$				
3				$\beta_{DT} = 1.0$			
4					$\beta_{NN} = 0.5$		
5						$\beta_{P} = 1.0$	
6							$\beta_{NNS} = 1.0$

- Fill diagonals with $\beta_j(k,k)$

Parse Triangle

	The (1)	gunman (2)	sprayed (3)	the (4)	building (5)	with (6)	bullets (7)
1	$\beta_{DT} = 1.0$	$\beta_{NP} = 0.25$					
2		$\beta_{NN} = 0.5$					
3			$\beta_{VBD} = 1.0$				
4				$\beta_{DT} = 1.0$			
5					$\beta_{NN} = 0.5$		
6						$\beta_{P} = 1.0$	
7							$\beta_{NNS} = 1.0$

- Calculate using induction formula

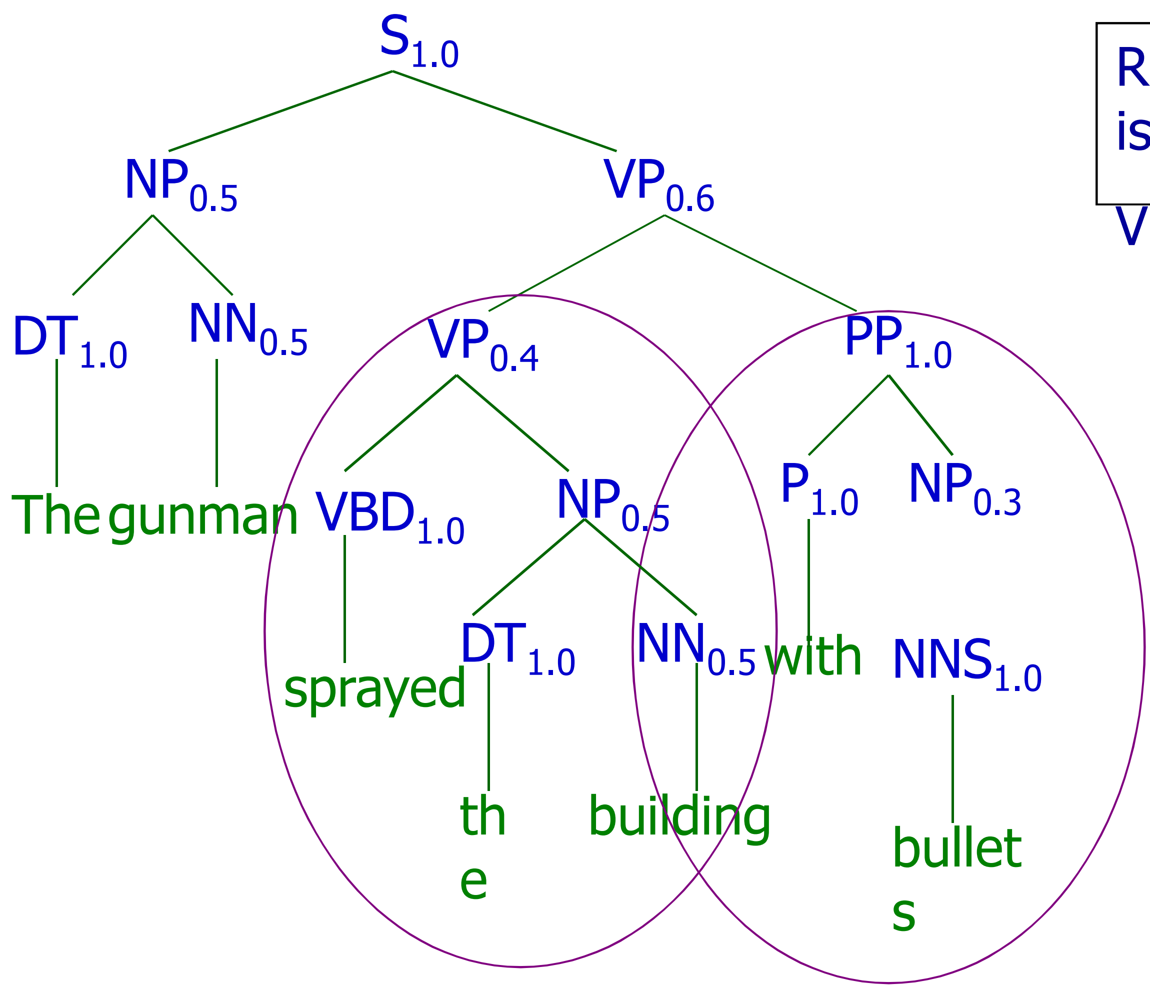
$$\beta_{NP}(1,2) = P(\text{the gunman} \mid NP_{1,2}, G)$$

$$= P(NP \rightarrow DT\ NN) * \beta_{DT}(1,1) * \beta_{NN}(2,2)$$

$$= 0.5 * 1.0 * 0.5 = 0.25$$

Example Parse t_1

- The gunman sprayed the building with bullets.

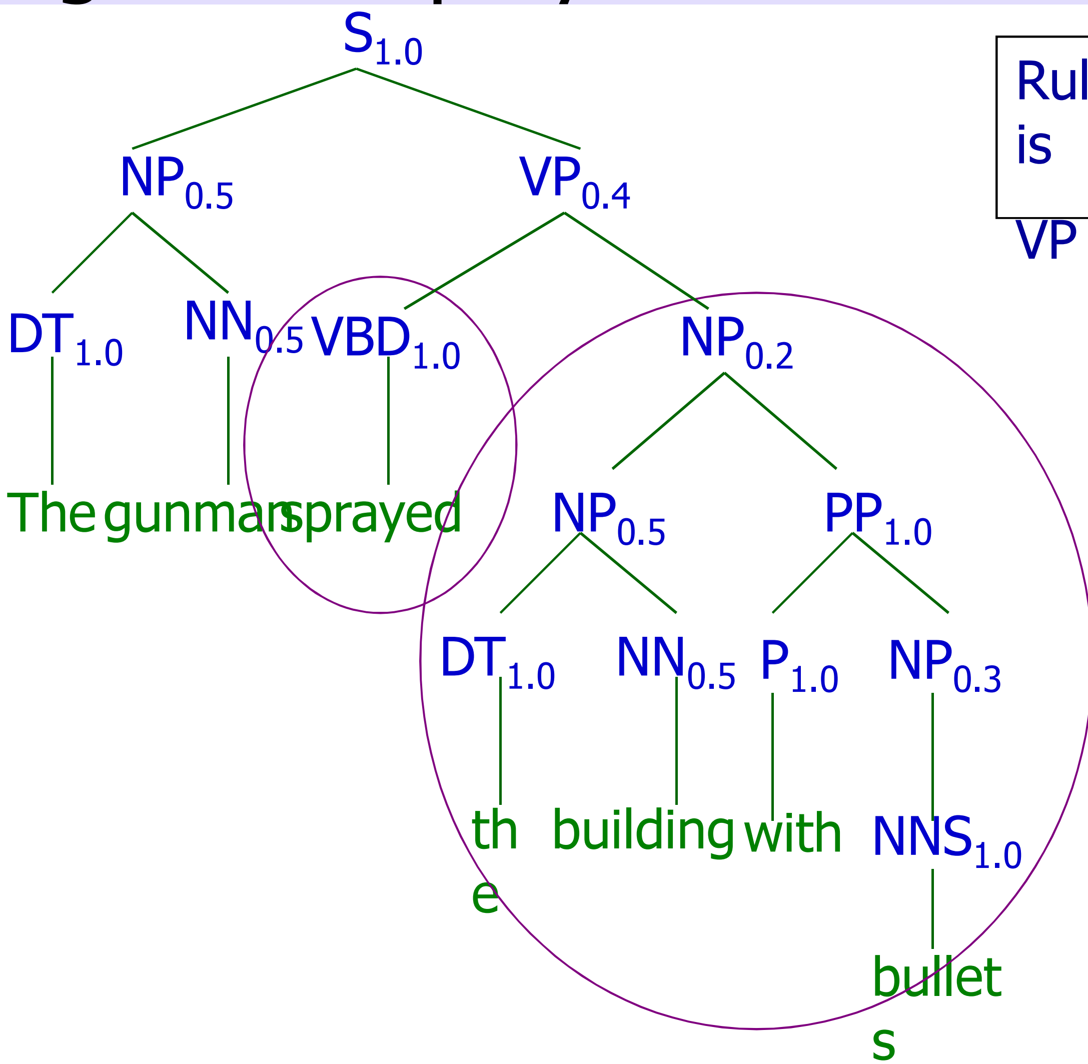


Rule used here is

$VP \rightarrow VP\ PP$

Another Parse t_2

- The gunman sprayed the building with bullets.



Rule used here is

$VP \rightarrow VBD\ NP$

Parse Triangle

	The (1)	gunman (2)	sprayed (3)	the (4)	building (5)	with (6)	bullets (7)
1	$\beta_{DT} = 1.0$	$\beta_{NP} = 0.25$					$\beta_S = 0.0465$
2		$\beta_{NN} = 0.5$					
3			$\beta_{VBD} = 1.0$		β_{VP}=0.1		$\beta_{VP} = 0.186$
4				$\beta_{DT} = 1.0$	$\beta_{NP} = 0.25$		$\beta_{NP} = 0.015$
5					$\beta_{NN} = 0.5$		
6						$\beta_P = 1.0$	$\beta_{PP} = 0.3$
7							$\beta_{NNS} = 1.0$

$$\beta_{VP}(3,7) = P(\text{sprayed the building with bullets} \mid VP_{3,7}, G)$$

$$= P(VP \to VP\ PP) * \beta_{VP}(3,5) * \beta_{PP}(6,7)$$

$$+\ P(VP \to VBD\ NP) * \beta_{VBD}(3,3) * \beta_{NP}(4,7)$$

$$= 0.6 * 1.0 * 0.3 + 0.4 * 1.0 * 0.015 = 0.186$$

Different Parses

- Consider
  - Different splitting points :

    E.g., 5th and 3rd position
  - Using different rules for VP expansion :

    E.g., VP $\rightarrow$ VP PP, VP $\rightarrow$ VBD NP
- Different parses for the VP "sprayed the building with bullets" can be constructed this way.

The Viterbi-like Algorithm for PCFGs

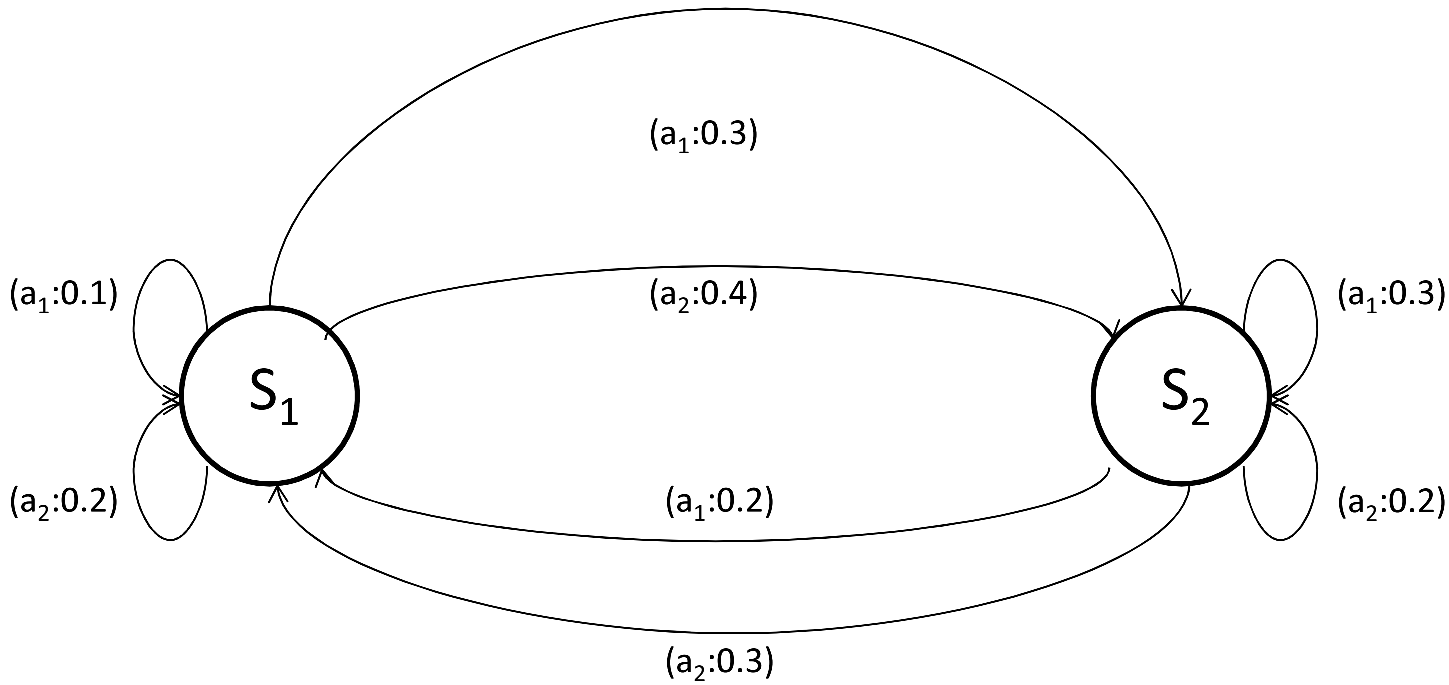
$$\delta_i(p,q) = \text{highest inside probability parse of } N^{i}_{pq}$$

- Very similar to calculation of inside probabilities $\beta_i(p,q)$
- Instead of summing over all ways of constructing the parse for w_{pq}
  - Choose only the best way (the maximum probability one!)

Three basic problems (contd.)

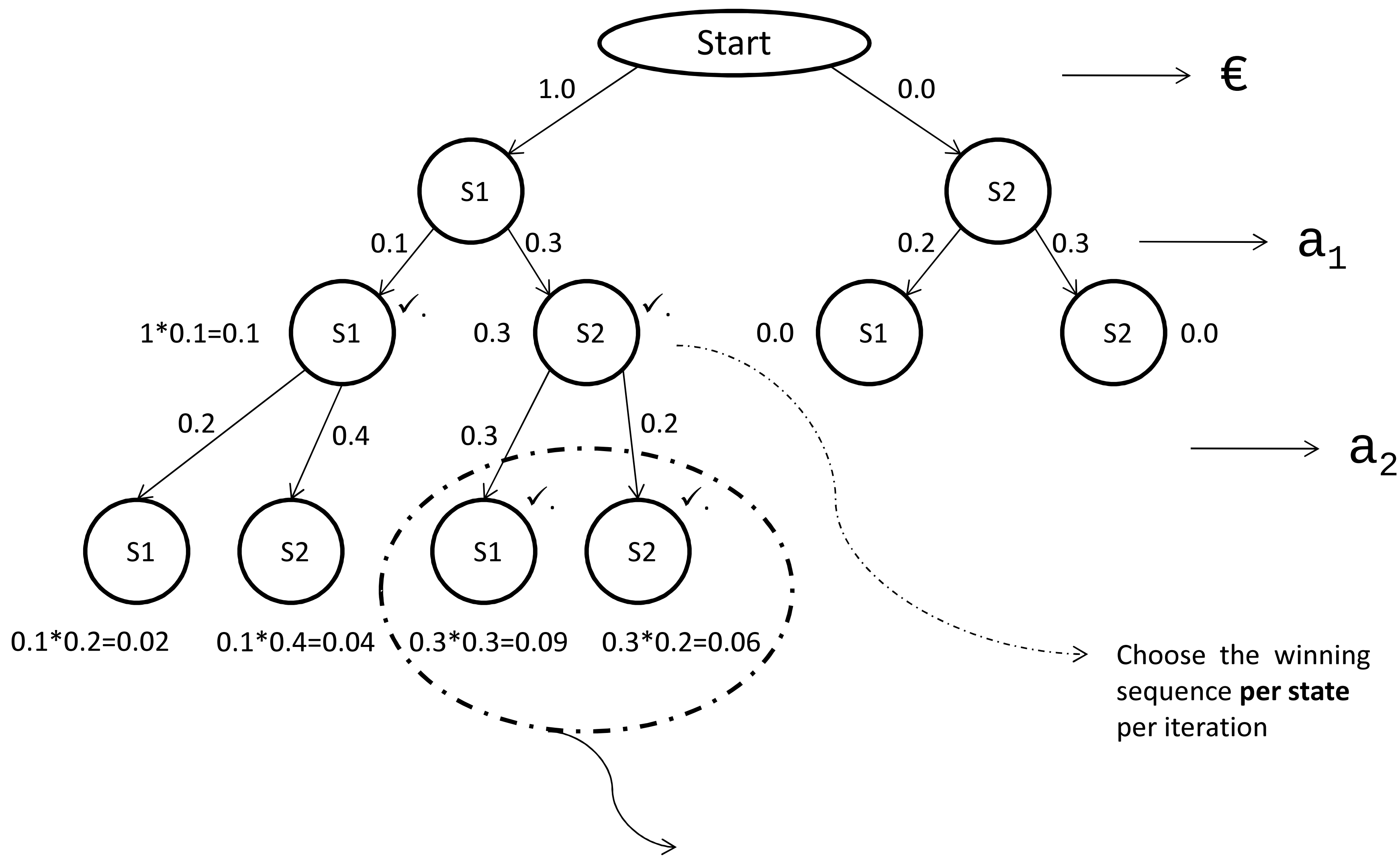
- Problem 1: Likelihood of a sequence
  - Forward Procedure
  - Backward Procedure
- Problem 2: Best state sequence
  - Viterbi Algorithm
- Problem 3: Re-estimation
  - Baum-Welch (Forward-Backward Algorithm)

Probabilistic FSM

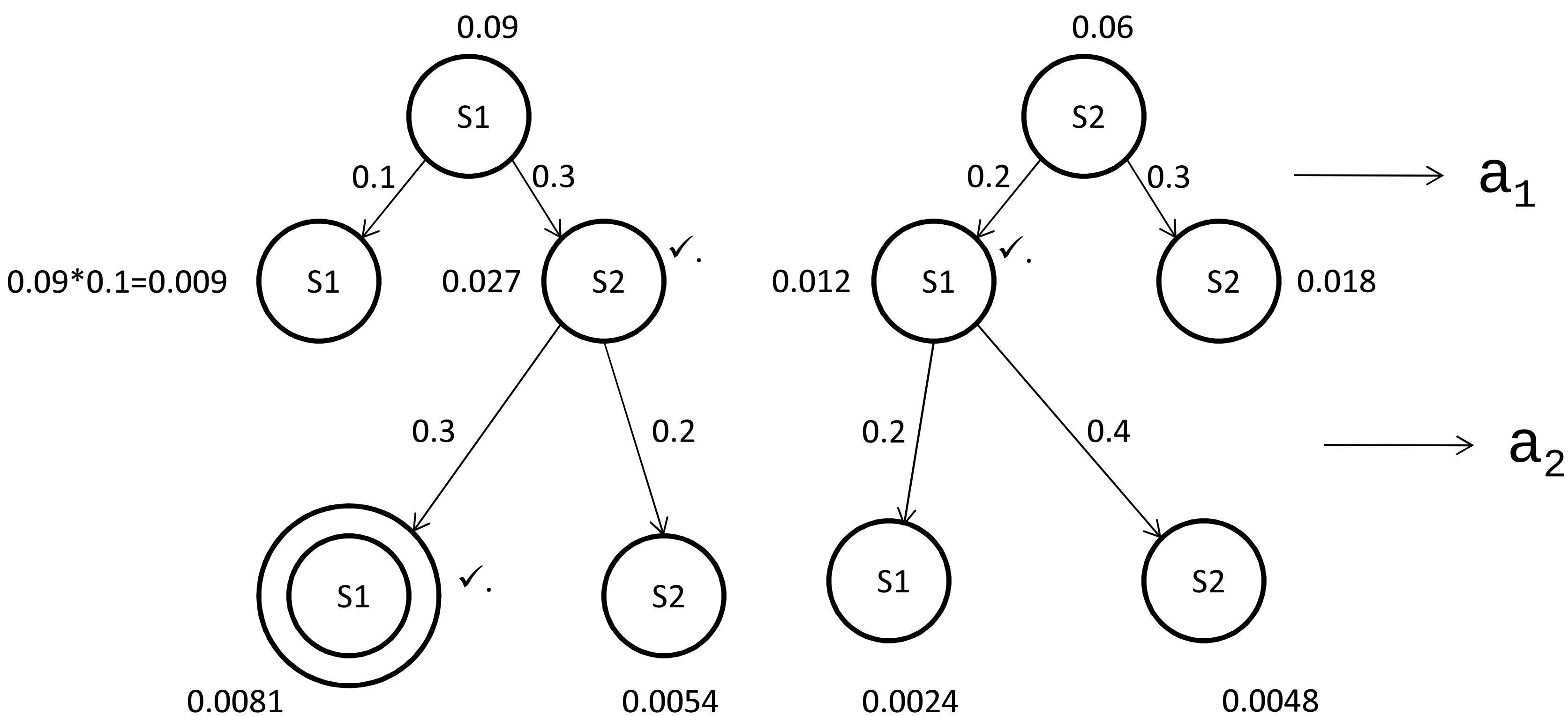


The question here is:
“what is the most likely state sequence given the output sequence seen”

Developing the tree



Tree structure contd…



The problem being addressed by this tree is $S^* = \arg\max_{S} P(S \mid a_1 - a_2 - a_1 - a_2, \mu)$

a1-a2-a1-a2 is the output sequence and μ the model or the machine

Path found:
(working backward)

$S_1 \quad S_2 \quad S_1 \quad S_2 \quad S_1$

$a_1 \quad a_2 \quad a_1 \quad a_2$

Problem statement: Find the best possible sequence

$$S^* = \arg\max_{S} P(S \mid O, \mu)$$

$$where,\ S \rightarrow \text{State Seq},\ O \rightarrow \text{Output Seq},\ \mu \rightarrow \text{Model or Machine}$$

$$\text{Model or Machine} = \{S_0, S, A, T\}$$

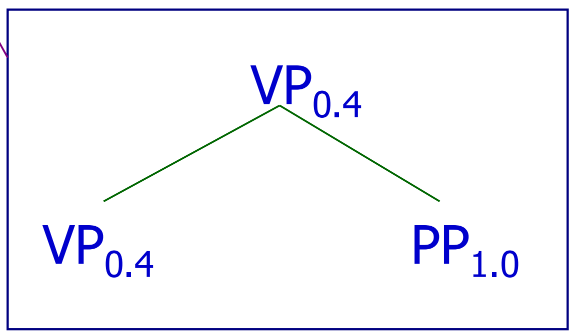
- S_0 → Start symbol
- S → State collection
- A → Alphabet set
- T → Transitions

T is defined as $P(S_i \xrightarrow{a_k} S_j) \quad \forall_{i,\ j,\ k}$

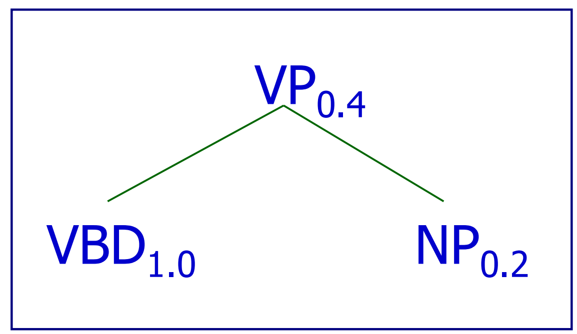
Calculation of $\delta_i(p,q)$

$$\delta_{VP}(3,7) = P(\text{sprayed the building with bullets} \mid VP_{3,7}, G)$$
$$= \max\{P(VP \to VP\ PP) * \delta_{VP}(3,5) * \delta_{PP}(6,7),$$
$$P(VP \to VBD\ NP) * \delta_{VBD}(3,3) * \delta_{NP}(4,7)\}$$
$$= \max\{0.6 * 1.0 * 0.3, 0.4 * 1.0 * 0.015\} = 0.18$$

This rule is chosen



0.6 * 1.0 * 0.3 = 0.18



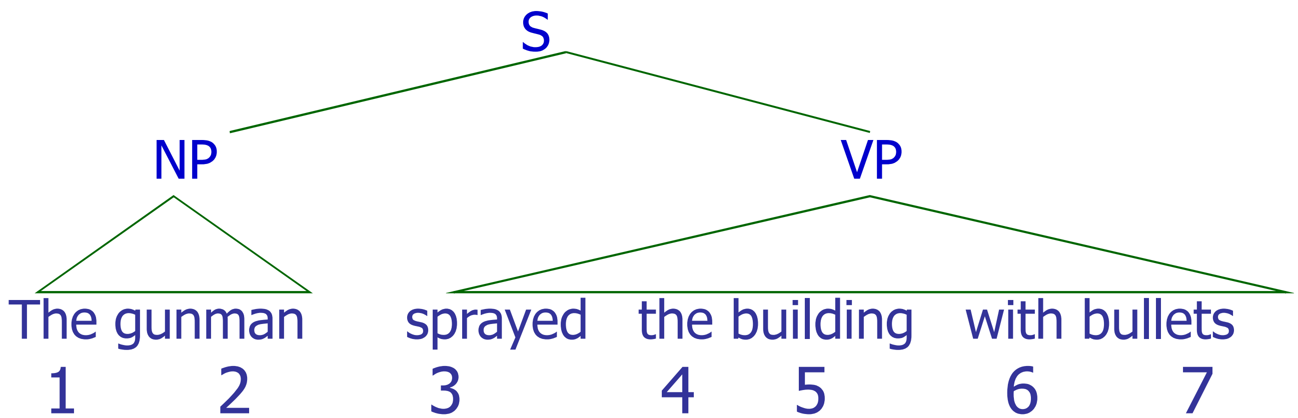
0.4 * 1.0 * 0.015 = 0.06

Viterbi-like Algorithm

- Base case: $\delta_i(k,k) = \beta_i(k,k)$
- Induction :
  - $\psi_i(p,q)$ stores
    - RHS of the rule selected
    - Position of splitting
  - Example : $\psi_{VP}(3,7)$ stores VP, PP and split position = 5 because VP $\rightarrow$ VP PP is the rule used.
- Backtracing : Start from $\psi_1(1,7)$ and $\delta_1(1,7)$ and backtrace.

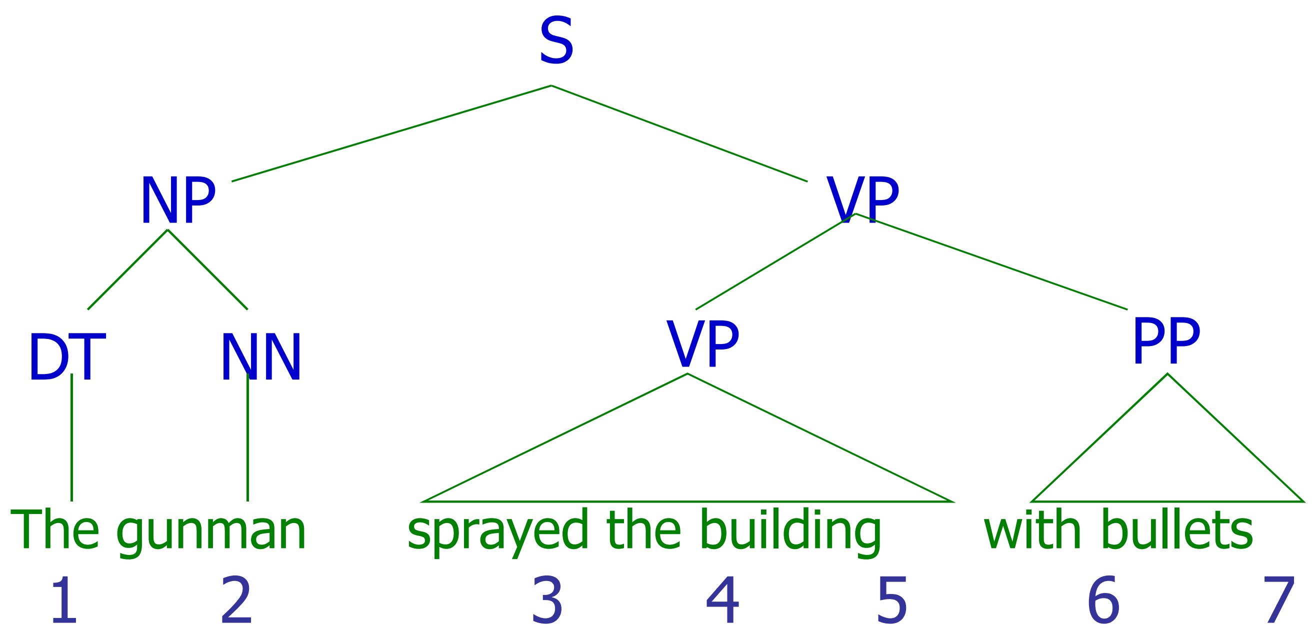
Example

- $\psi_1(1,7)$ records S $\rightarrow$ NP VP & split position as 2



- $\psi_{NP}(1,2)$ records NP $\rightarrow$ DT NN & split position as 1
- $\psi_{VP}(3,7)$ records VP $\rightarrow$ VP PP & split position as 5

Example

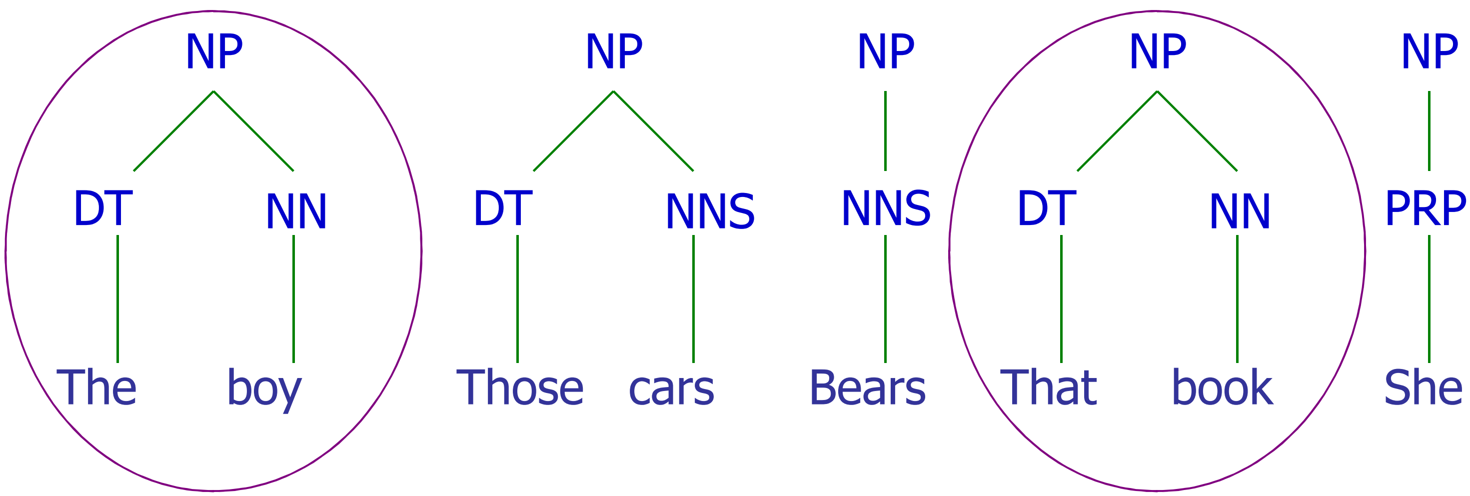

S
NP
VP
DT
NN
VP
PP
The gunman
sprayed the building
with bullets
1
2
3
4
5
6
7

Grammar Induction

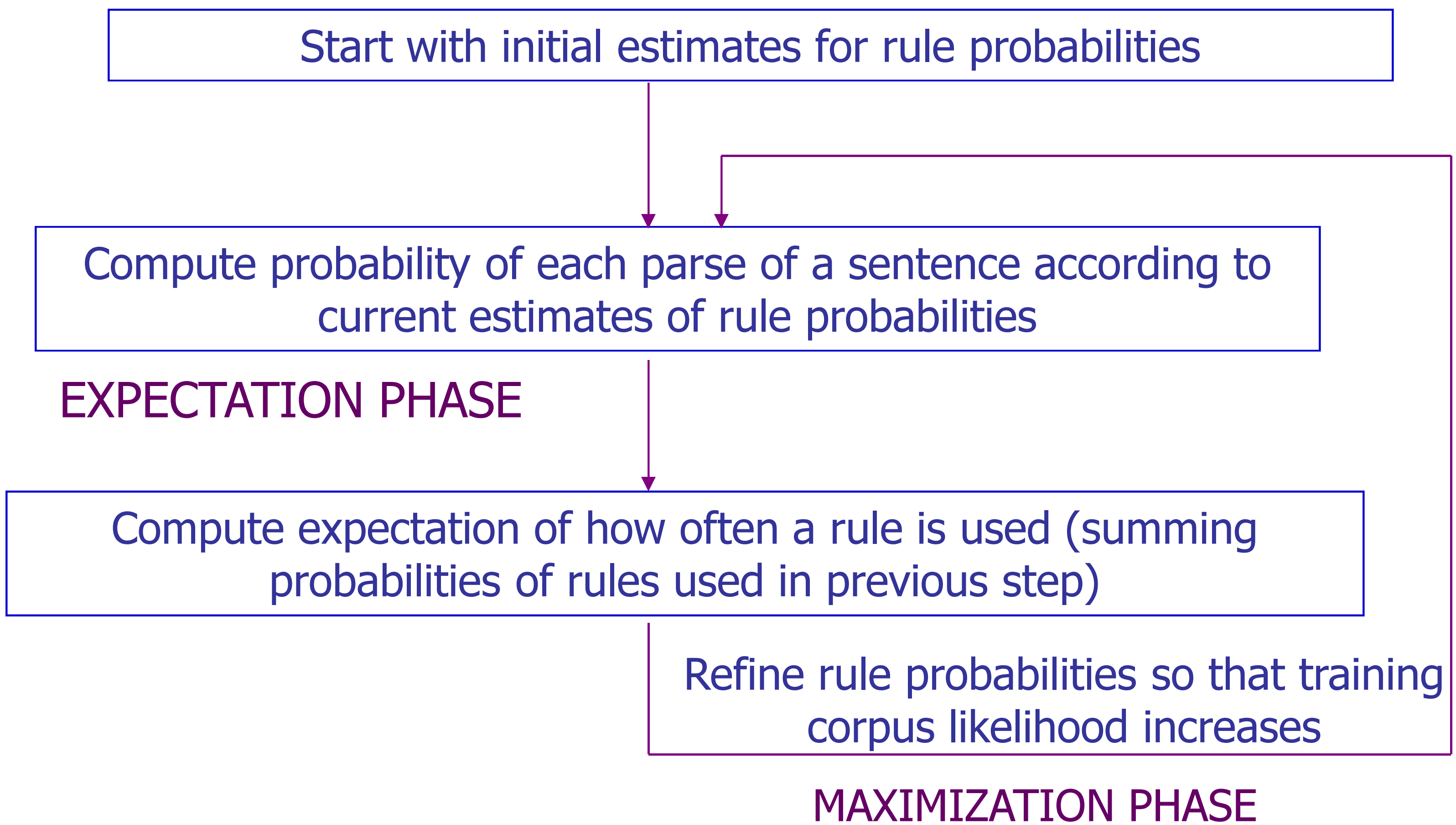
- Annotated corpora like Penn Treebank
- Counts used as follows:

$$P(\text{NP} \rightarrow \text{DT NN}) = \frac{\#\ \text{NP} \rightarrow \text{DT NN is used}}{\#\ \text{An NP rule is used}} = \frac{2}{5}$$

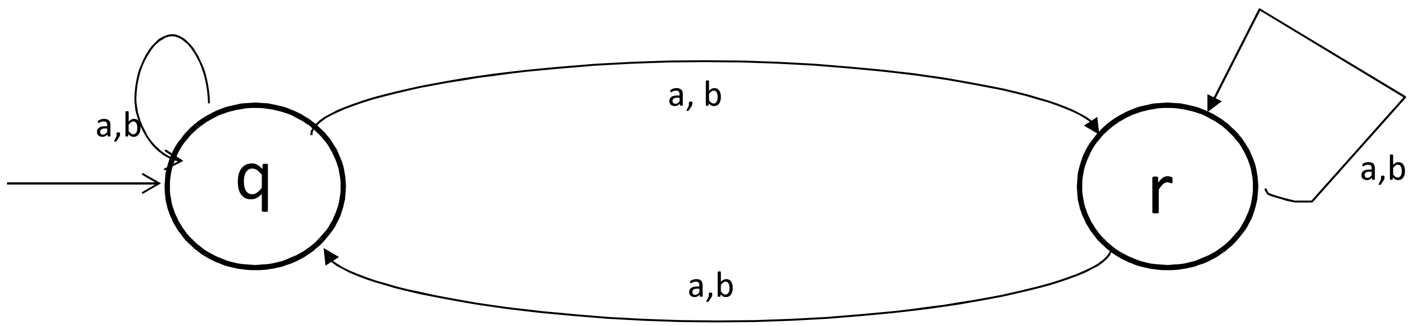
- Sample training data:



Grammar Induction for Unannotated Corpora: EM algorithm



Baum-Welch algorithm: counts



String = abb aaa bbb aaa

Sequence of states with respect to input symbols

o/p seq

State seq

$$\rightarrow q \xrightarrow{a} r \xrightarrow{b} q \xrightarrow{b} q \xrightarrow{a} r \xrightarrow{a} q \xrightarrow{a} r \xrightarrow{b} q \xrightarrow{b} q \xrightarrow{b} q \xrightarrow{a} r \xrightarrow{a} q \xrightarrow{a} r$$

Calculating probabilities from table

$$P(q \xrightarrow{a} r) = 5/8$$

$$P(q \xrightarrow{b} r) = 3/8$$

$$P(s^i \xrightarrow{w_k} s^j) = \frac{c(s^i \xrightarrow{w_k} s^j)}{\sum_{l=1}^{T}\sum_{m=1}^{A} c(s^i \xrightarrow{w_m} s^l)}$$

T=#states

A=#alphabet symbols

Table of counts

Src	**Dest**	**O/P**	**Count**
q	r	a	5
q	q	b	3
r	q	a	3
r	q	b	2

Now if we have a non-deterministic transitions then multiple state seq possible for the given o/p seq (ref. to previous slide's feature). Our aim is to find expected count through this.

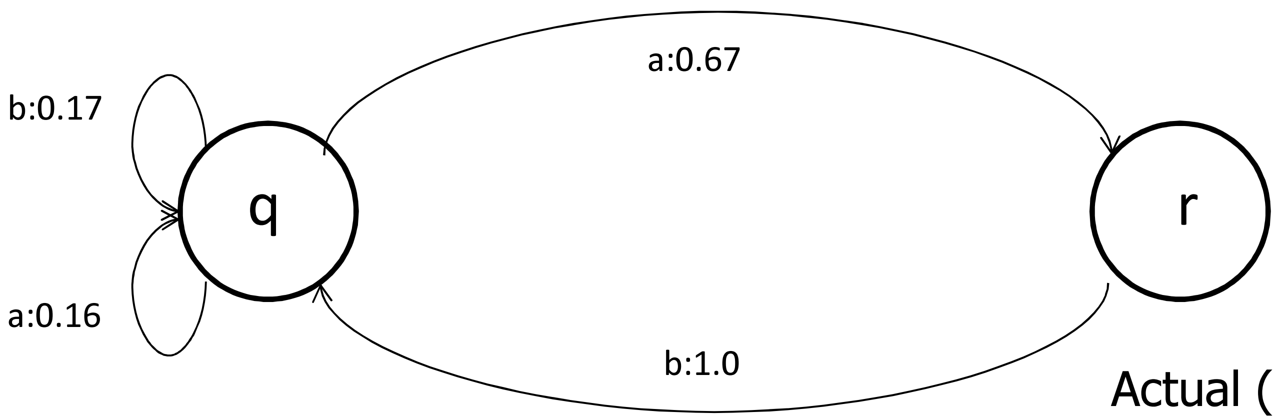
Interplay Between Two Equations

$$P(s^i \xrightarrow{W_k} s^j) = \frac{c(s^i \xrightarrow{W_k} s^j)}{\sum_{l=0}^{T} \sum_{m=0}^{A} c(s^i \xrightarrow{W_m} s^l)}$$

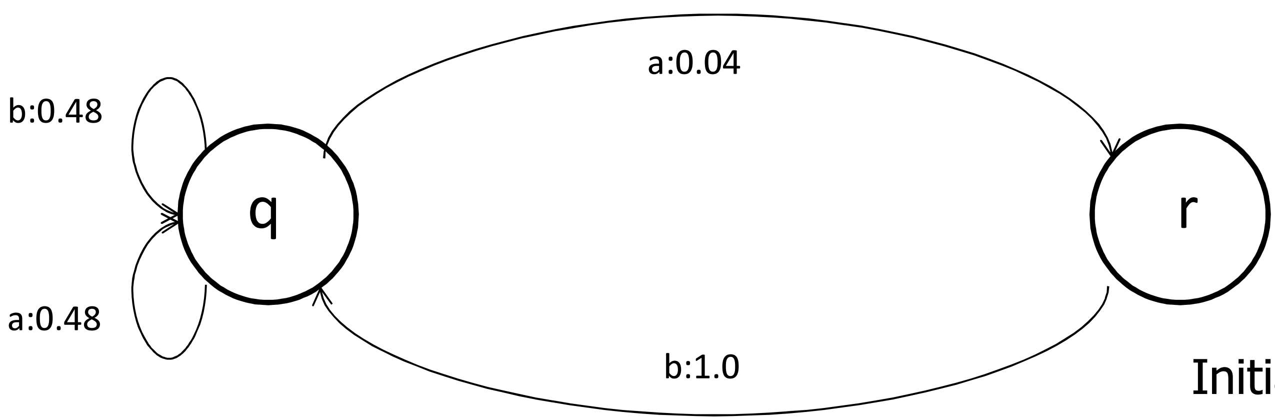
$$C(s^i \xrightarrow{W_k} s^j) = \sum_{s_{0,n+1}} P(S_{0,n+1} \mid W_{0,n}) \times n(s^i \xrightarrow{W_k} s^j, S_{0,n+1}, w_{0,n})$$

No. of times the transitions $s^i \xrightarrow{w_k} s^j$ occurs in the string

Illustration



Actual (Desired) HMM



Initial guess

One run of Baum-Welch algorithm: *string ababb*

$\in \rightarrow a$	$a \rightarrow b$	$b \rightarrow a$	$a \rightarrow b$	$b \rightarrow b$	$b \rightarrow \in$	P(path)	$q \xrightarrow{a} r$	$r \xrightarrow{b} q$	$q \xrightarrow{a} q$	$q \xrightarrow{b} q$
q	r	q	r	q	q	0.00077	0.00154	0.00154	0	0.00077
q	r	q	q	q	q	0.00442	0.00442	0.00442	0.00442	0.00884
q	q	q	r	q	q	0.00442	0.00442	0.00442	0.00442	0.00884
q	q	q	q	q	q	0.02548	0.0	0.000	0.05096	0.07644
Rounded Total →						0.035	0.01	0.01	0.06	0.095
New Probabilities (P) → State sequences							0.06 =(0.01/(0.01+0.06+0.095)	1.0	0.36	0.581

* ε is considered as starting and ending symbol of the input sequence string. Through multiple iterations the probability values will converge.

Computational part (1/2)

$$C(s^i \xrightarrow{w_k} s^j) = \sum_{s_{0,n+1}} [P(S_{0,n+1} | W_{0,n}) \times n(s^i \xrightarrow{w_k} s^j, S_{0,n+1}, W_{0,n})]$$

$$= \frac{1}{P(W_{0,n})} \sum_{s_{0,n+1}} [P(S_{0,n+1}, W_{0,n}) \times n(s^i \xrightarrow{w_k} s^j, S_{0,n+1}, W_{0,n})]$$

$$= \frac{1}{P(W_{0,n})} \sum_{t=0,n} \sum_{s_{0,n+1}} [P(S_t = s^i, W_t = w_k, S_{t+1} = s^j, S_{0,n+1}, W_{0,n})]$$

$$= \frac{1}{P(W_{0,n})} \sum_{t=0,n} [P(S_t = s^i, W_t = w_k, S_{t+1} = s^j, W_{0,n})]$$

$$S0 \xrightarrow{w_0} S1 \xrightarrow{w_1} S1 \xrightarrow{w_2} \ldots\ Si \xrightarrow{w_k} Sj \ldots \rightarrow Sn\text{-}1 \xrightarrow{w_{n-1}} Sn \xrightarrow{w_n} Sn+1$$

Computational part (2/2)

$$\sum_{t=0}^{n} P(S_t = s^i, S_{t+1} = s^j, W_t = w_k, W_{0,n})$$

$$= \sum_{t=0}^{n} P(W_{0,t-1}, S_t = s^i, S_{t+1} = s^j, W_t = w_k, W_{t+1,n})$$

$$= \sum_{t=0}^{n} P(W_{0,t-1}, S_t = s^i) P(S_{t+1} = s^j, W_t = w_k \mid W_{0,t-1}, S_t = s^i) P(W_{t+1,n} \mid S_{t+1} = s^j)$$

$$= \sum_{t=0}^{n} F(t-1, i) P(S_{t+1} = s^j, W_t = w_k \mid S_t = s^i) B(t+1, j)$$

$$= \sum_{t=0}^{n} F(t-1, i) P(S_{t+1} = s^j, W_t = w_k \mid S_t = s^i) B(t+1, j)$$

$$= \sum_{t=0}^{n} F(t-1, i) P(s^i \xrightarrow{w_k} s^j) B(t+1, j)$$

$$S0 \xrightarrow{w_0} S1 \xrightarrow{w_1} S1 \xrightarrow{w_2} \ldots Si \xrightarrow{w_k} Sj \ldots \rightarrow Sn\text{-}1 \xrightarrow{w_{n-1}} Sn \xrightarrow{w_n} Sn+1$$

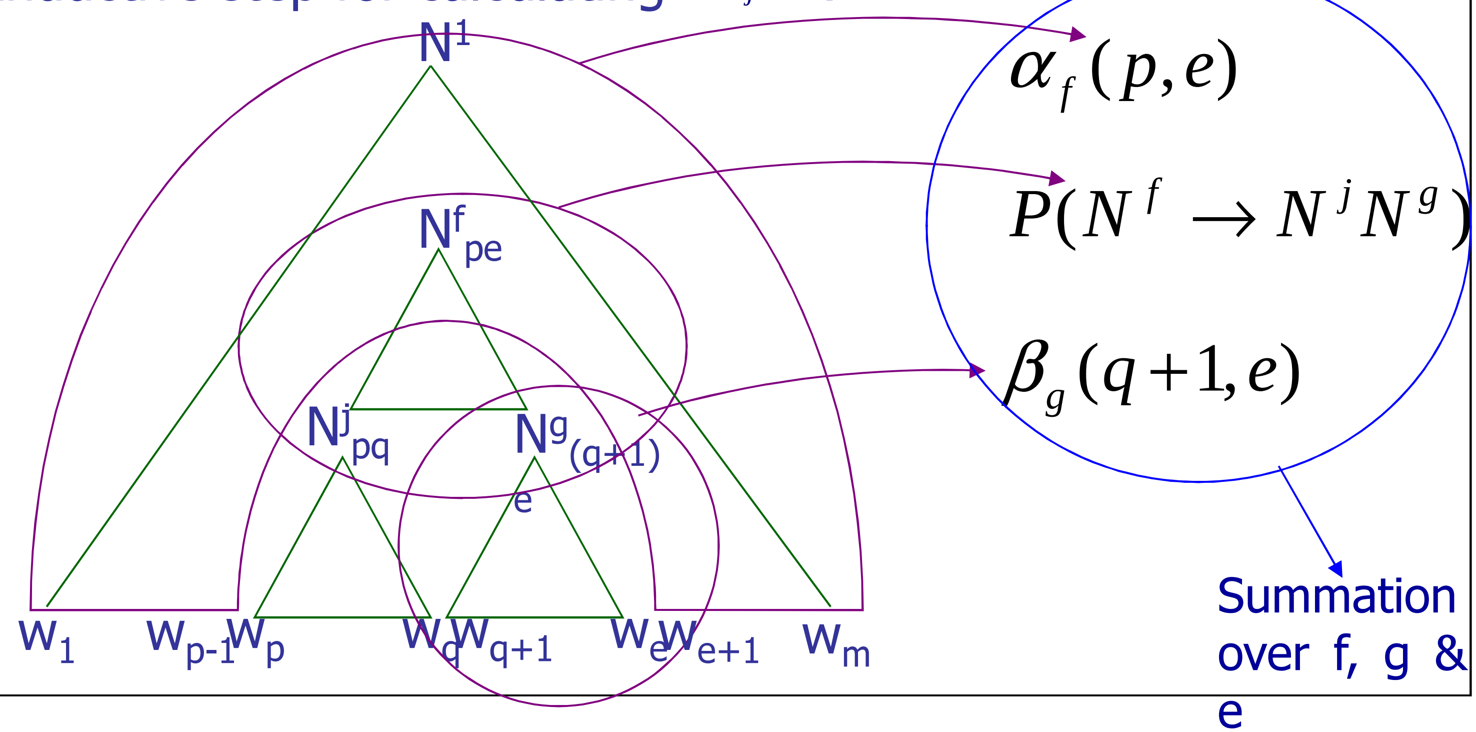
Outside Probabilities $\alpha_j(p,q)$

Base case:

$$\alpha_1(1,m) = 1 \text{ for start symbol}$$

$$\alpha_j(1,m) = 0 \text{ for } j \neq 1$$

Inductive step for calculating $\alpha_j(p,q)$:



Probability of a Sentence

$$P(w_{1m}, N_{pq} \mid G) = \sum_j P(w_{1m} \mid N^j_{pq}, G) = \sum_j \alpha_j(p,q)\beta_j(p,q)$$

- Joint probability of a sentence w_{1m} and that there is a constituent spanning words w_p to w_q is given as:

$$P(\text{The gunman....bullets}, N_{4,5} \mid G)$$

$$= \sum_j P(\text{The gunman...bullets} \mid N^j_{4,5}, G)$$

$$= \alpha_{NP}(4,5)\beta_{NP}(4,5)$$

$$+ \alpha_{VP}(4,5)\beta_{VP}(4,5) + \ldots$$