Probabilistic Context-Free Grammars

COMP3361 — Week 6

Lingpeng Kong

Department of Computer Science, The University of Hong Kong
Many materials from Columbia CS4705 with special thanks!

Probabilistic Context-Free Grammars

R =

S	→ NP VP		1.0
VP	→ Vi		0.4
VP	→ Vt NP		0.4
VP	→ VP PP		0.2
NP	→ DT NN		0.3
NP	→ NP PP		0.7
PP	→ IN NP		1.0
Vi	→ sleeps		1.0
Vt	→ saw		1.0
NN	→ man		0.7
NN	→ woman		0.2
NN	→ telescope		0.1
DT	→ the		1.0
IN	→ with		0.5
IN	→ in		0.5

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5

S

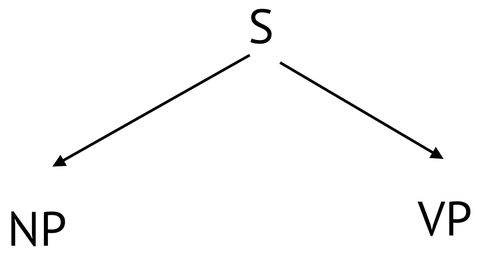
(S)

$P(\boldsymbol{x}, \boldsymbol{y}) =$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5



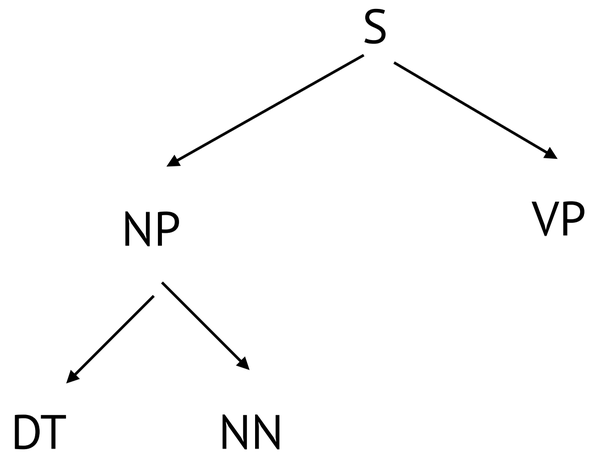
(S (NP VP))

$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5



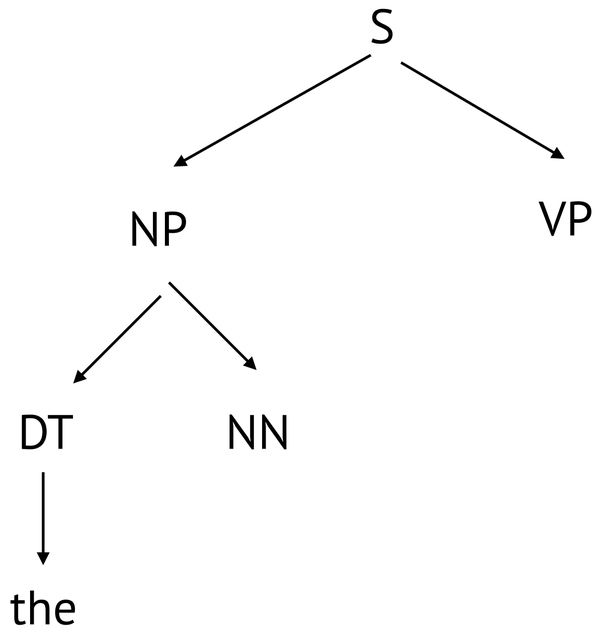
(S (NP (DT NN) VP))

$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5



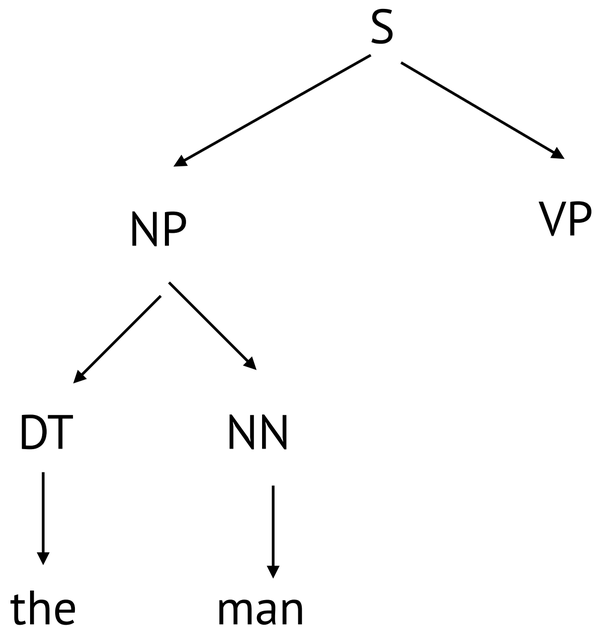
(S (NP ((DT the) NN) VP))

$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times 1 \times$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5



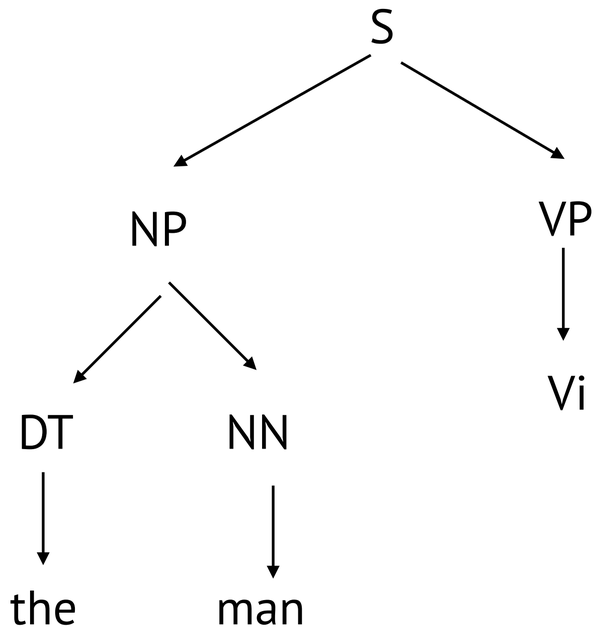
(S (NP ((DT the) (NN man)) VP))

$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times 1 \times 0.7 \times$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5



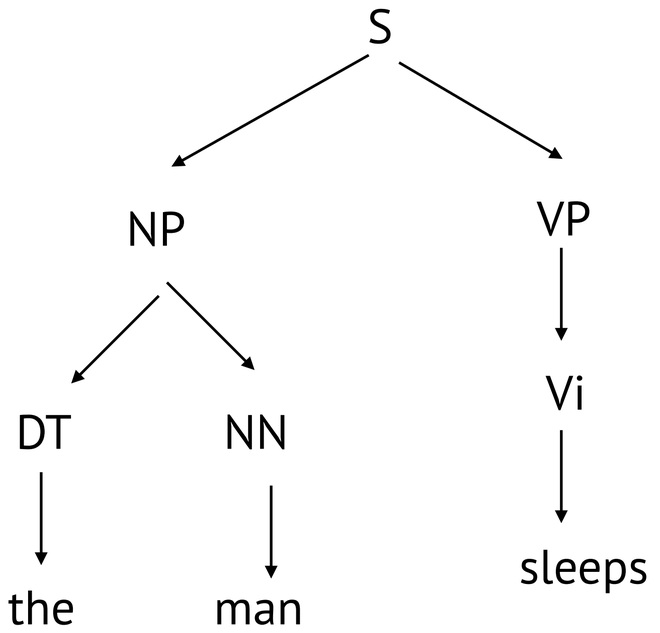
(S (NP ((DT the) (NN man)) (VP (Vi))))

$$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times 1 \times 0.7 \times 0.4 \times$$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps 1.0
				Vt	→	saw 1.0
VP	→	Vi	0.4			
VP	→	Vt NP	0.4	NN	→	man 0.7
VP	→	VP PP	0.2	NN	→	woman 0.2
				NN	→	telescope 0.1
NP	→	DT NN	0.3			
NP	→	NP PP	0.7	DT	→	the 1.0
				IN	→	with 0.5
PP	→	IN NP	1.0	IN	→	in 0.5



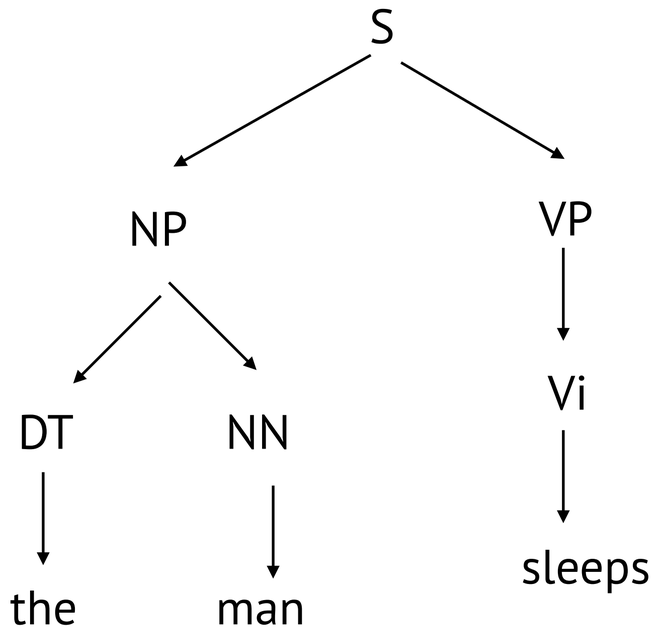
(S (NP ((DT the) (NN man)) (VP (Vi sleeps))))

$$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times 1 \times 0.7 \times 0.4 \times 1$$

Derivation Example

R =

S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5



(S (NP ((DT the) (NN man)) (VP (Vi sleeps))))

Probabilistic Context-Free Grammars

R =

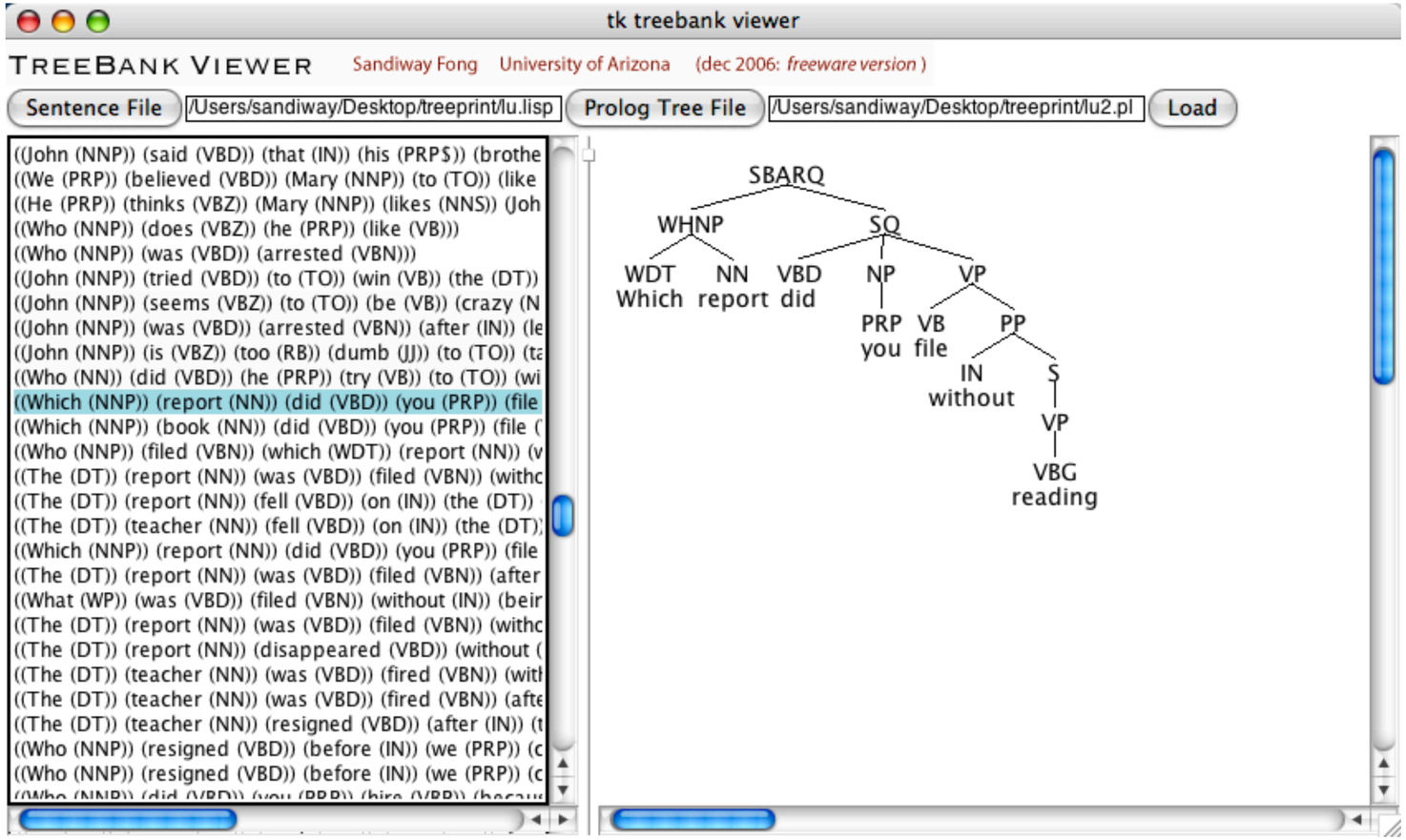
S	→	NP VP	1.0	Vi	→	sleeps	1.0
				Vt	→	saw	1.0
VP	→	Vi	0.4				
VP	→	Vt NP	0.4	NN	→	man	0.7
VP	→	VP PP	0.2	NN	→	woman	0.2
				NN	→	telescope	0.1
NP	→	DT NN	0.3				
NP	→	NP PP	0.7	DT	→	the	1.0
				IN	→	with	0.5
PP	→	IN NP	1.0	IN	→	in	0.5

1. How do we get those weights?

2. How to get the (marginal) probability of this sentence and the most likely parse?

the man saw the woman with a telescope

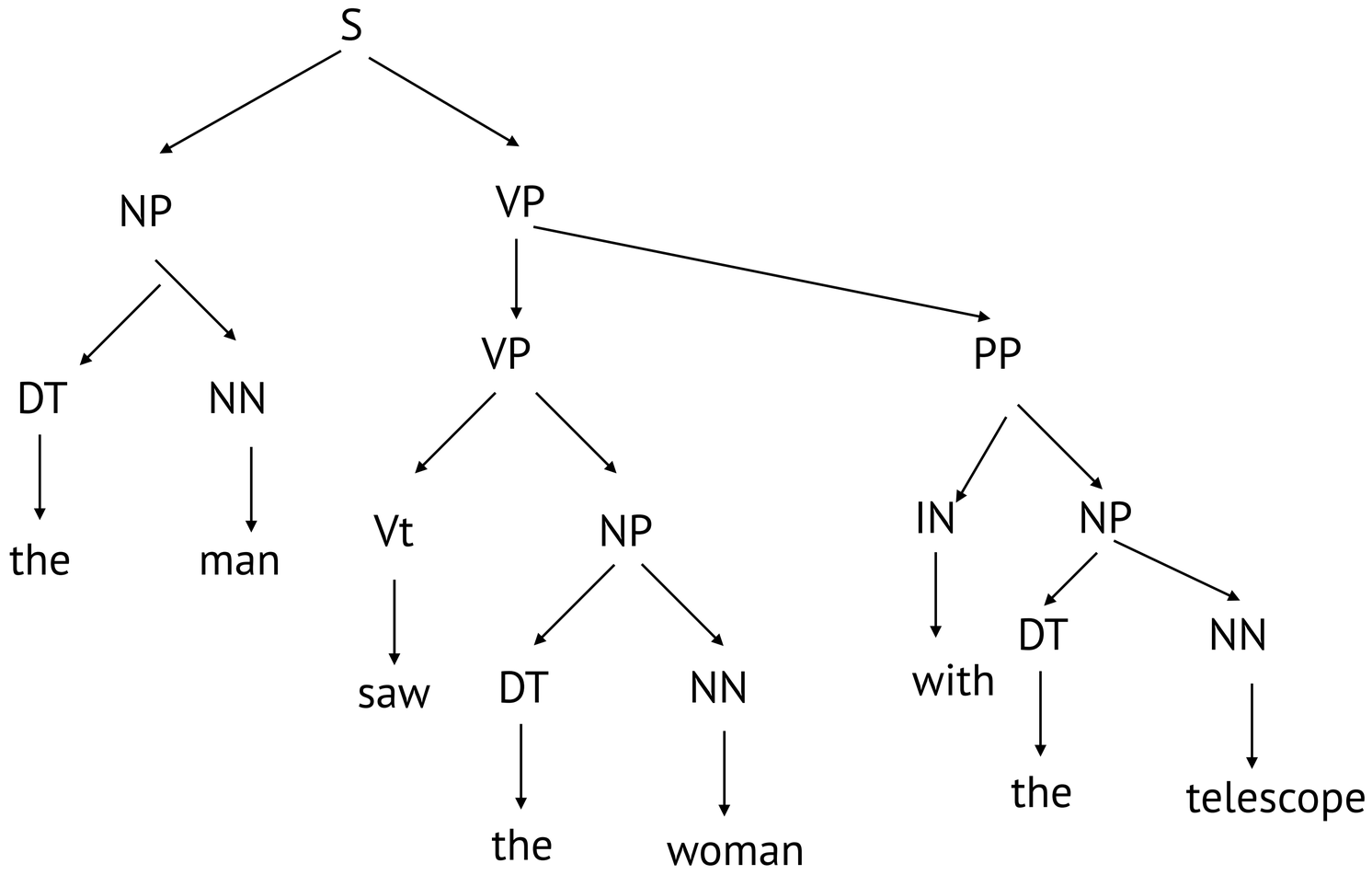
Treebanks (Penn Treebank)



4.5 million words of American English

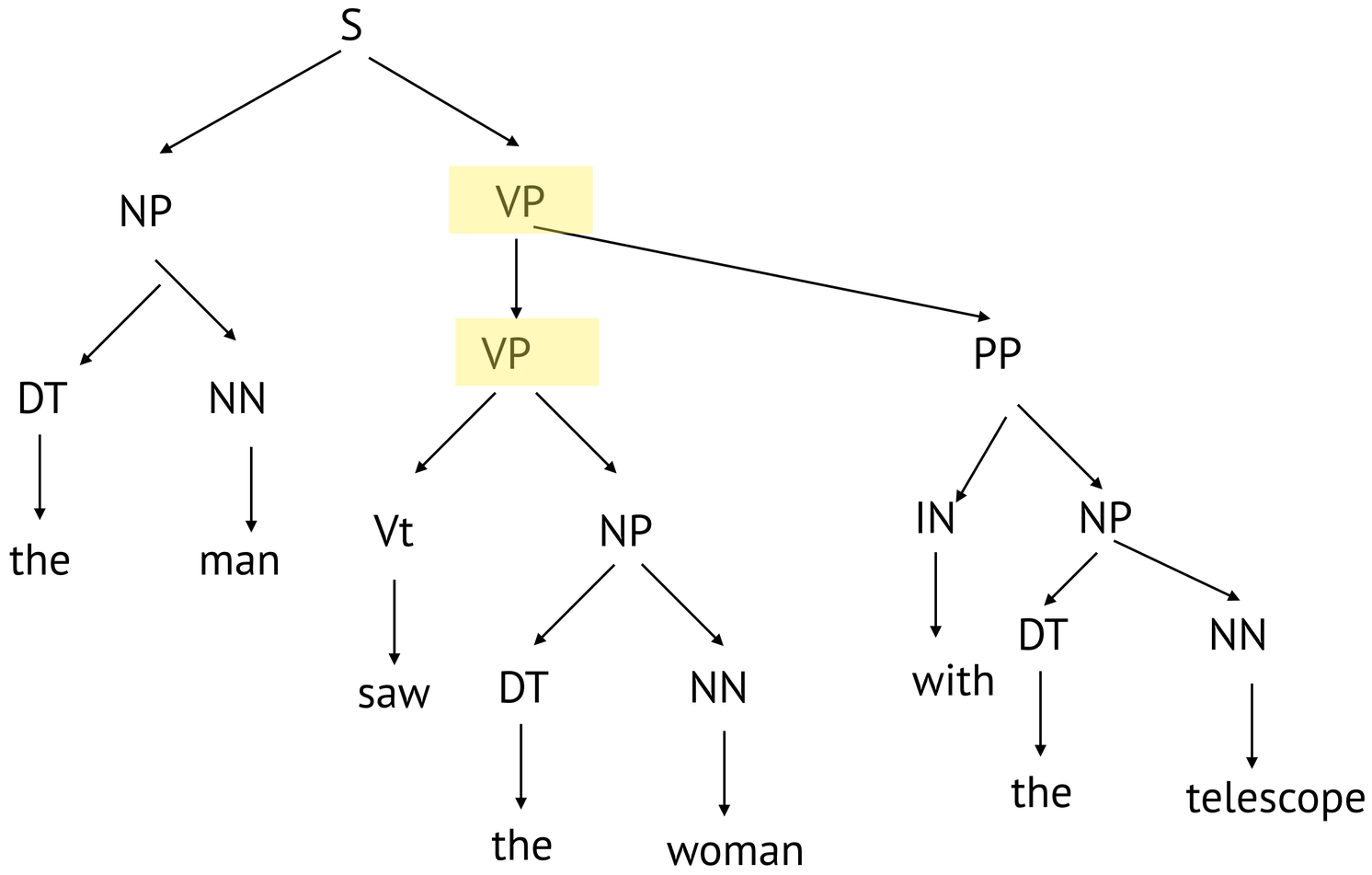
Mitch Marcus

Estimating the Probabilities from the Treebank



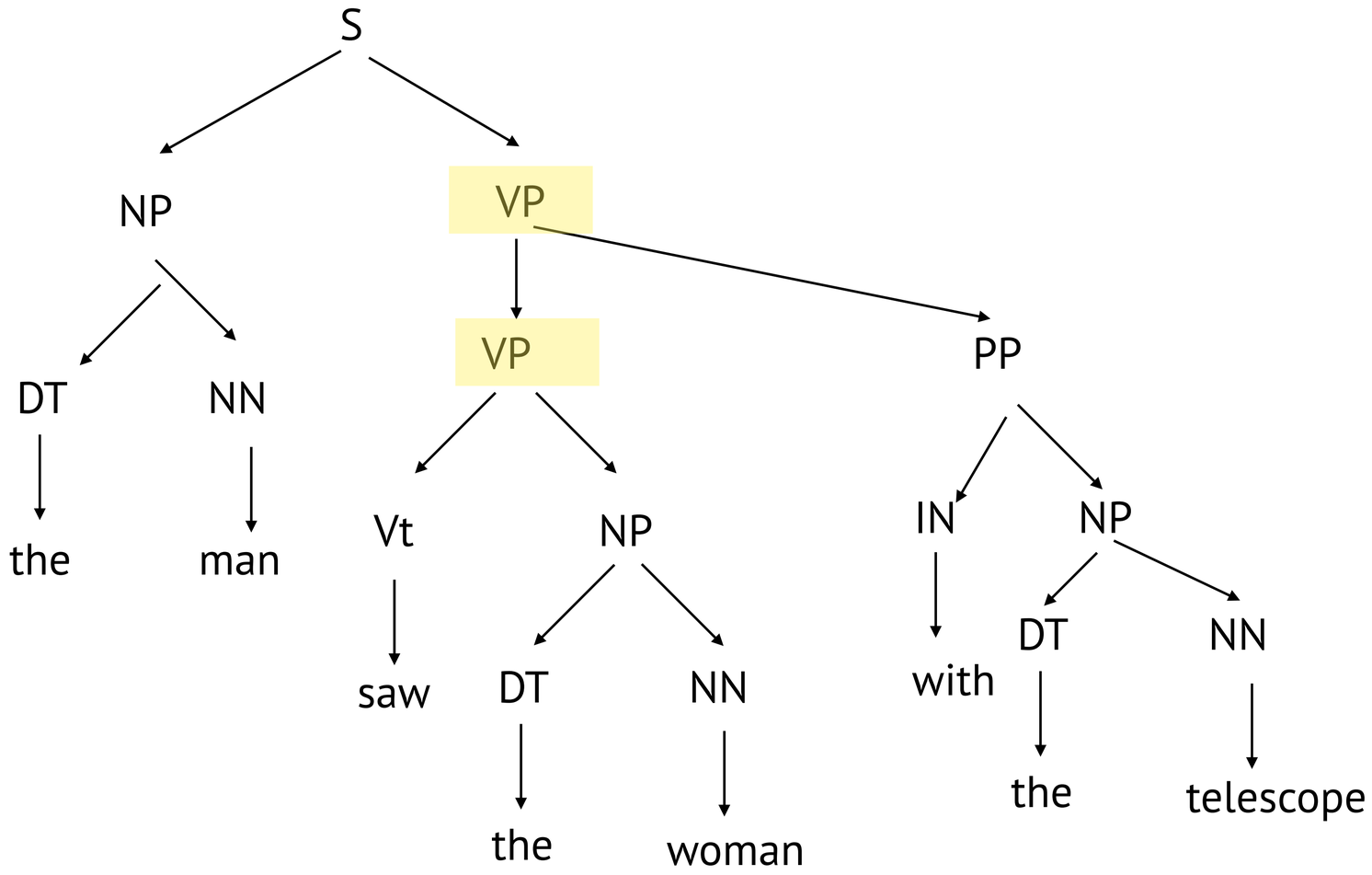
$$p(\mathrm{VP} \rightarrow \mathrm{VP\ PP}) = ?$$

Estimating the Probabilities from the Treebank



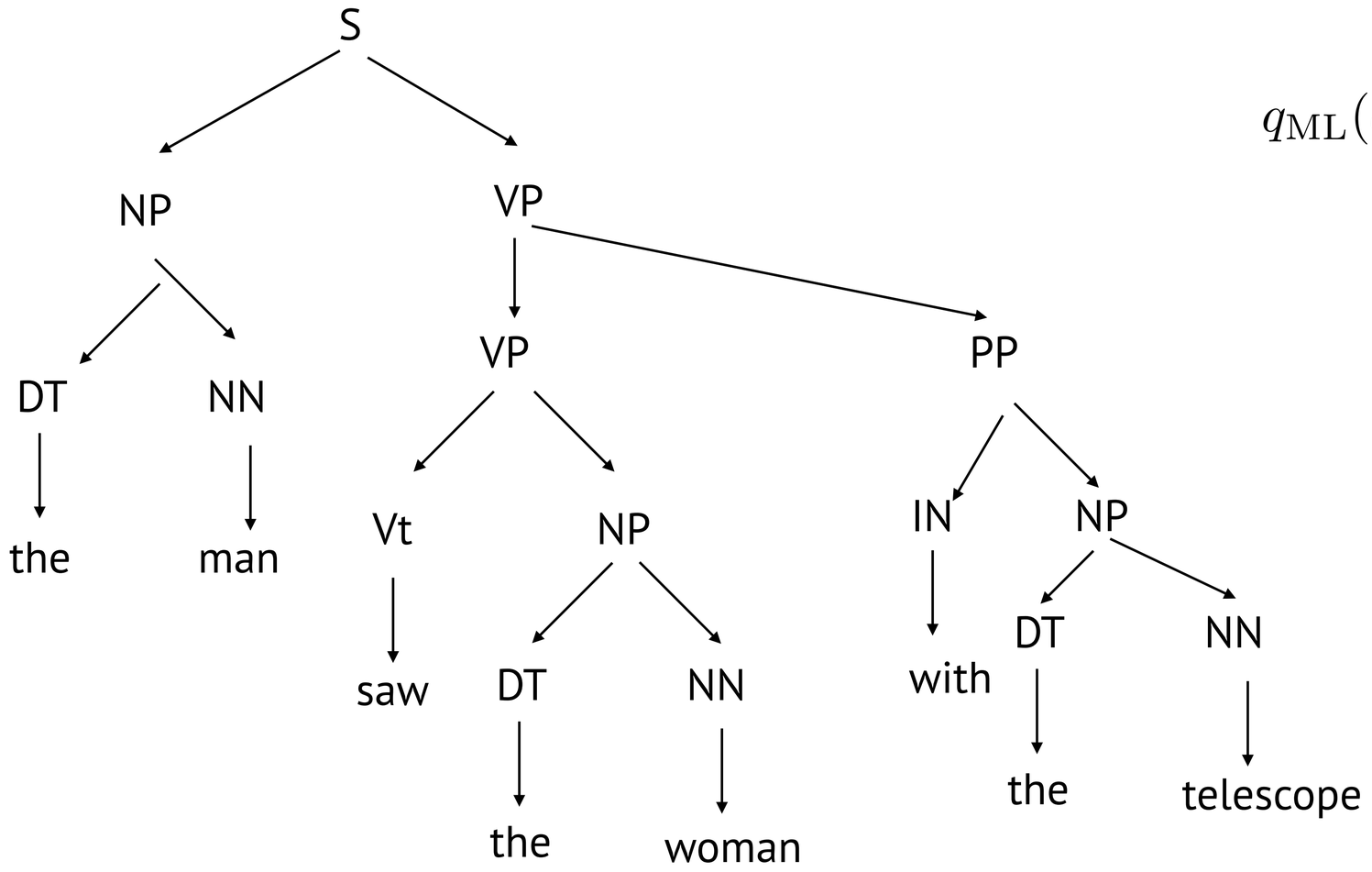
$$p(\mathrm{VP} \rightarrow \mathrm{VP\ PP}) = ?$$

Estimating the Probabilities from the Treebank



$$p(\mathrm{VP} \rightarrow \mathrm{VP\ PP}) = \frac{1}{2}$$

Estimating the Probabilities from the Treebank

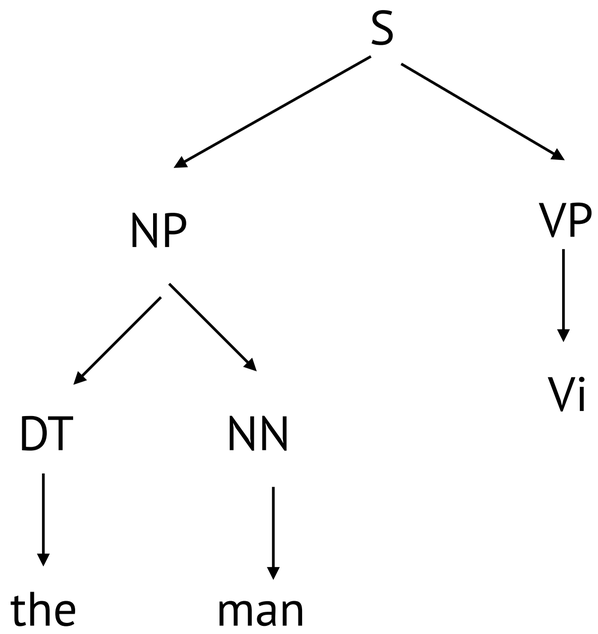


$$q_{\mathrm{ML}}(\alpha \rightarrow \beta) = \frac{\mathrm{Count}(\alpha \rightarrow \beta)}{\mathrm{Count}(\alpha)}$$

MLE estimation!

Parsing with PCFGs

Parsing: $\arg\max_{\boldsymbol{y} \in \boldsymbol{Y}} p(\boldsymbol{x}, \boldsymbol{y})$



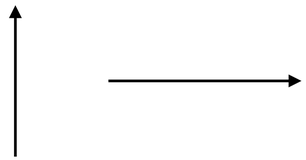
(S (NP ((DT the) (NN man)) (VP (Vi sleeps))))

$$P(\boldsymbol{x}, \boldsymbol{y}) = 1 \times 0.3 \times 1 \times 0.7 \times 0.4 \times 1$$

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

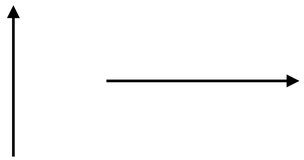
Find the lowest cost path from bottom left corner to the upper right corner



Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

Find the lowest cost path from bottom left corner to the upper right corner

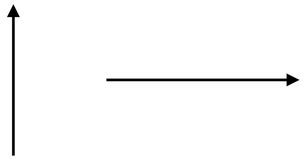


2 + 2 + 5 + 4 + 2 + 9 + 7 = 31

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

Find the lowest cost path from bottom left corner to the upper right corner



2 + 2 + 6 + 1 + 1 + 1 + 7 = 20

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

2			

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)			
4 (d)			
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)			
4 (d)			
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)			
4 (d)	10 (l)		
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)			
9 (d)	13 (l)		
4 (d)	10 (l)	11 (l)	
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)	15 (d)		
9 (d)	13 (l)	14 (d)	
4 (d)	10 (l)	11 (l)	12 (l)
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

17 (d)	15 (d)	23 (d)	
9 (d)	13 (l)	14 (d)	13 (d)
4 (d)	10 (l)	11 (l)	12 (l)
2	7 (l)	11 (l)	13 (l)

Dynamic Programming

8	2	9	7
5	4	3	1
2	6	1	1
2	5	4	2

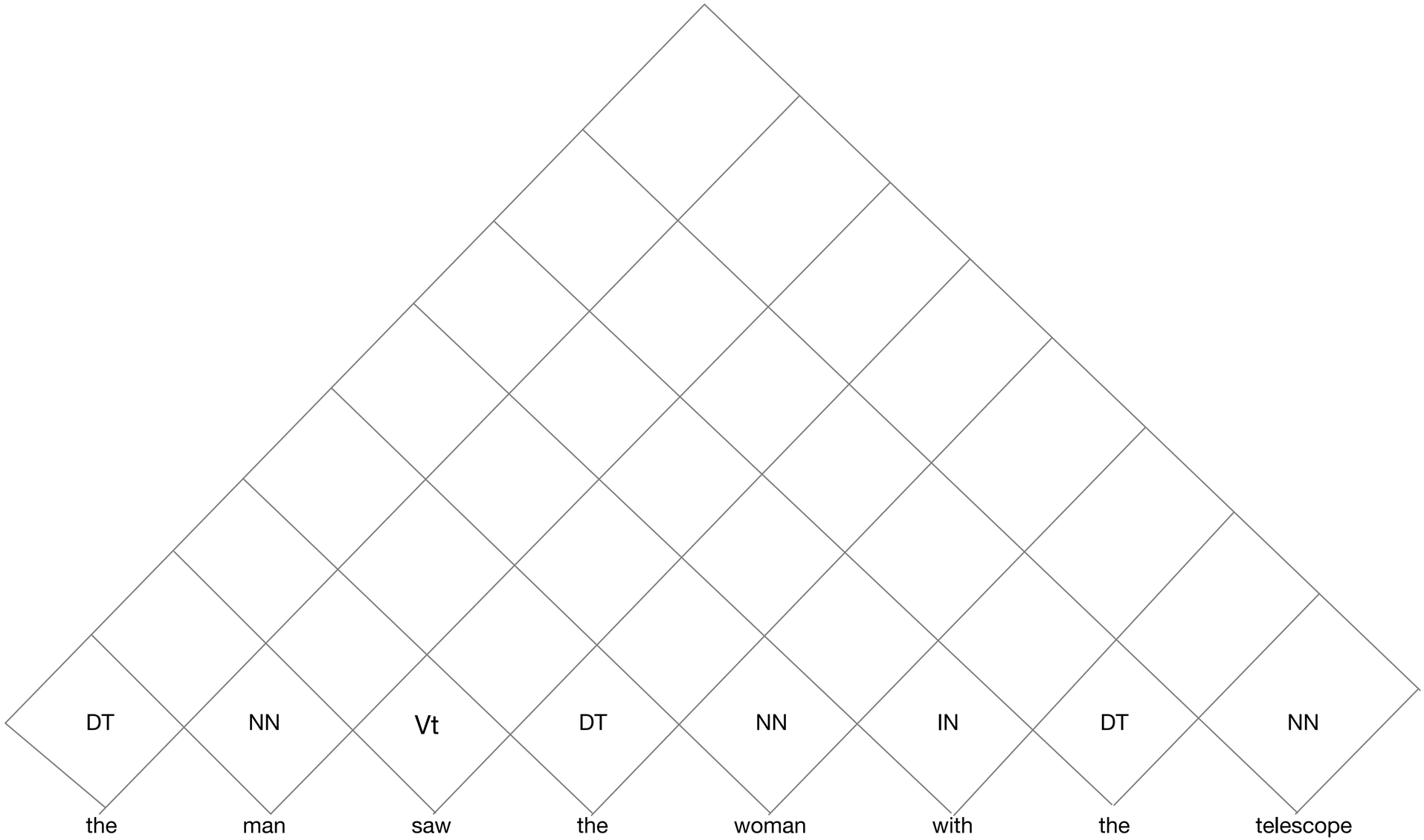
17 (d)	15 (d)	23 (d)	20 (d)
9 (d)	13 (l)	14 (d)	13 (d)
4 (d)	10 (l)	11 (l)	12 (l)
2	7 (l)	11 (l)	13 (l)

Parsing with PCFGs

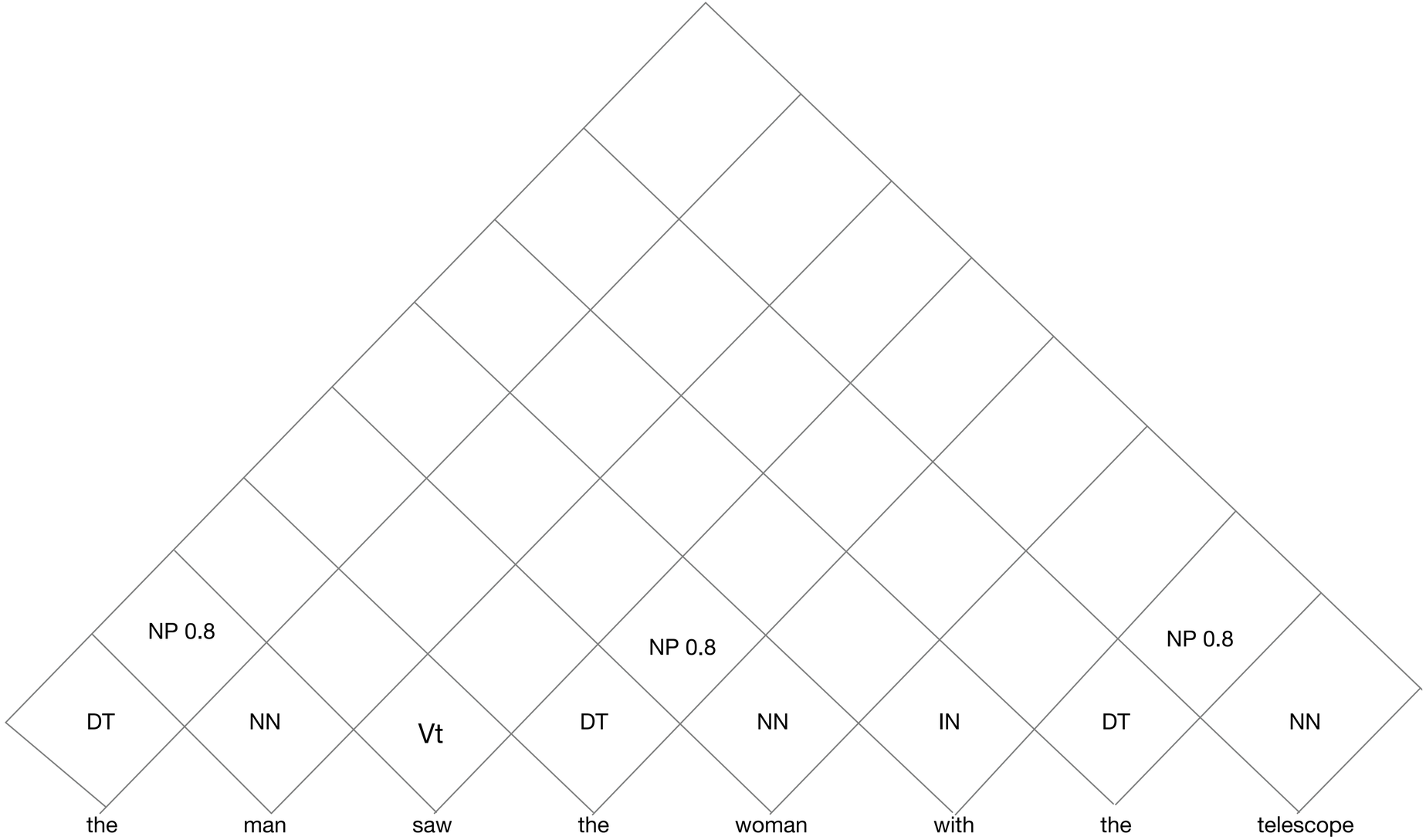
R =

S	→	NP VP	1.0	Vt	→	saw	1.0
				NN	→	man	0.1
VP	→	Vt NP	0.8	NN	→	woman	0.1
VP	→	VP PP	0.2	NN	→	telescope	0.3
				NN	→	dog	0.5
NP	→	DT NN	0.8				
NP	→	NP PP	0.2	DT	→	the	1.0
PP	→	IN NP	1.0	IN	→	with	0.6
				IN	→	in	0.4

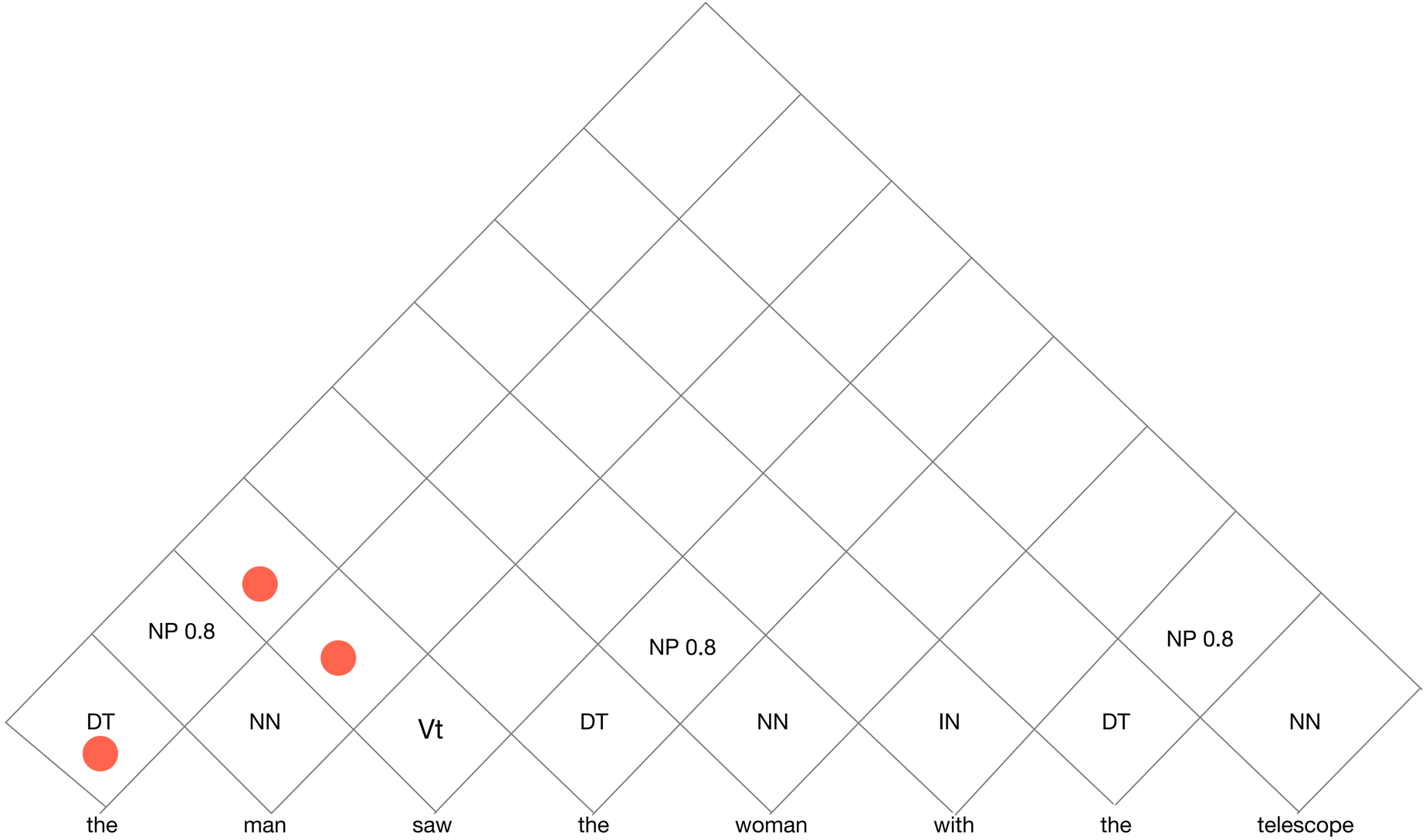
Parsing with PCFGs



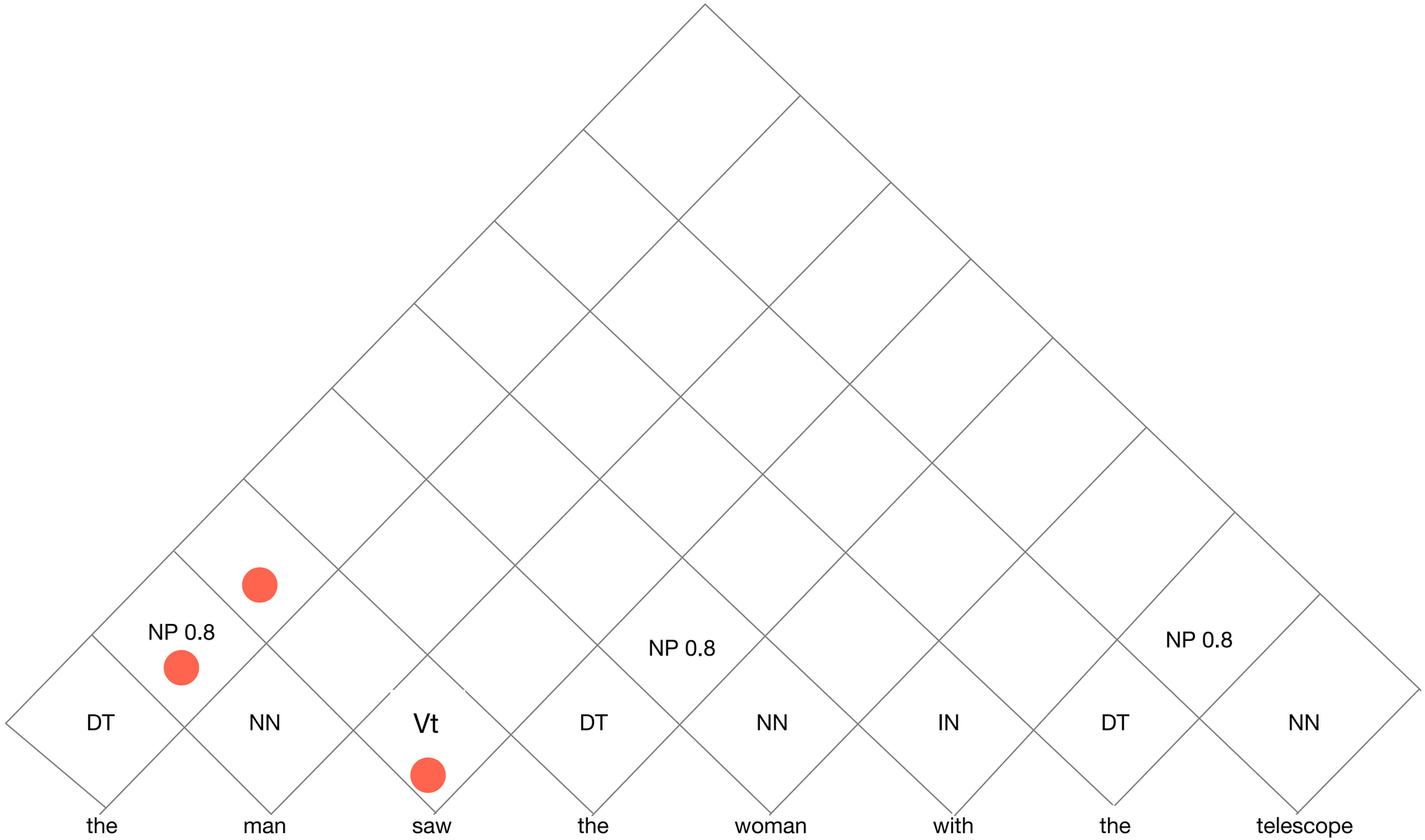
Parsing with PCFGs



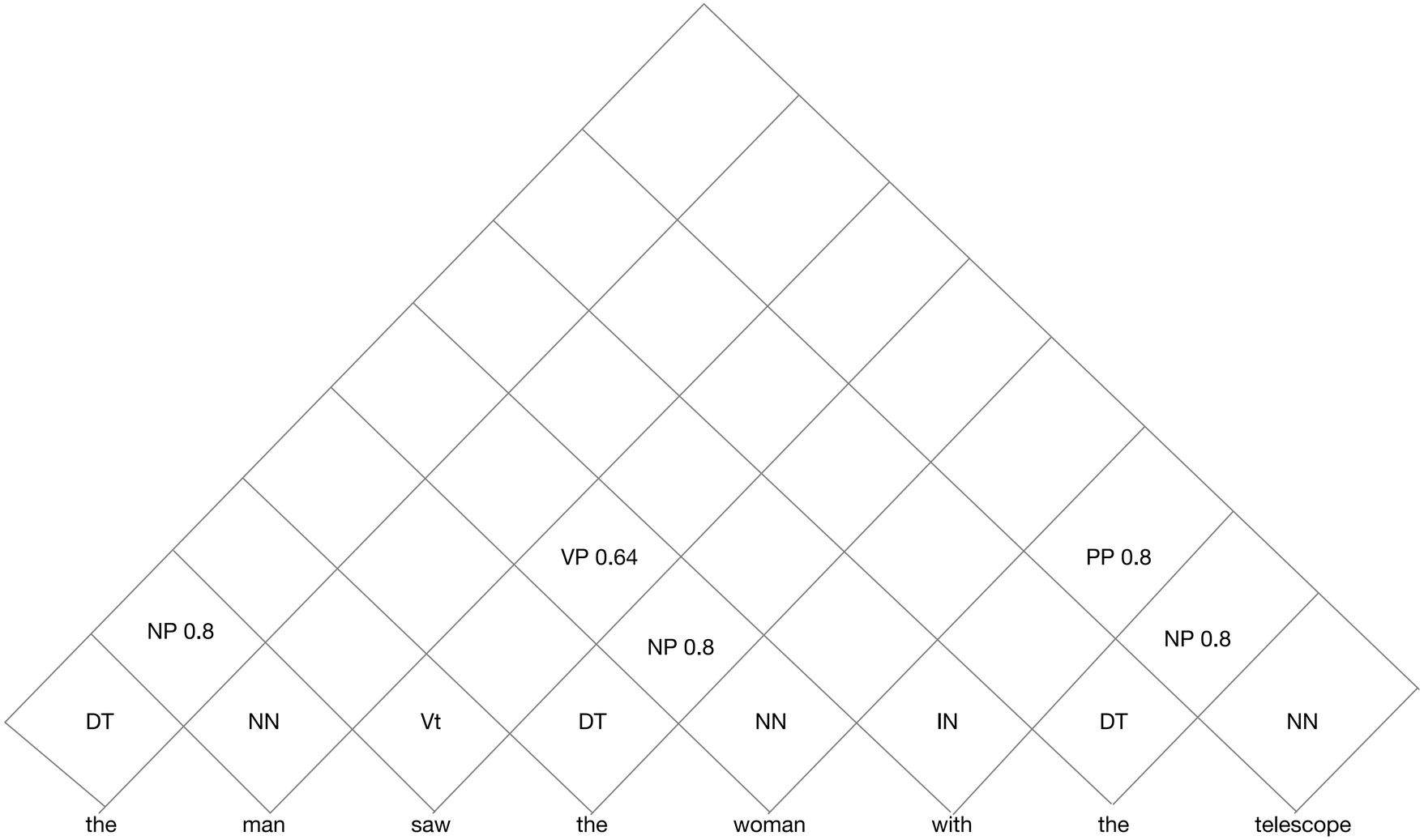
Parsing with PCFGs



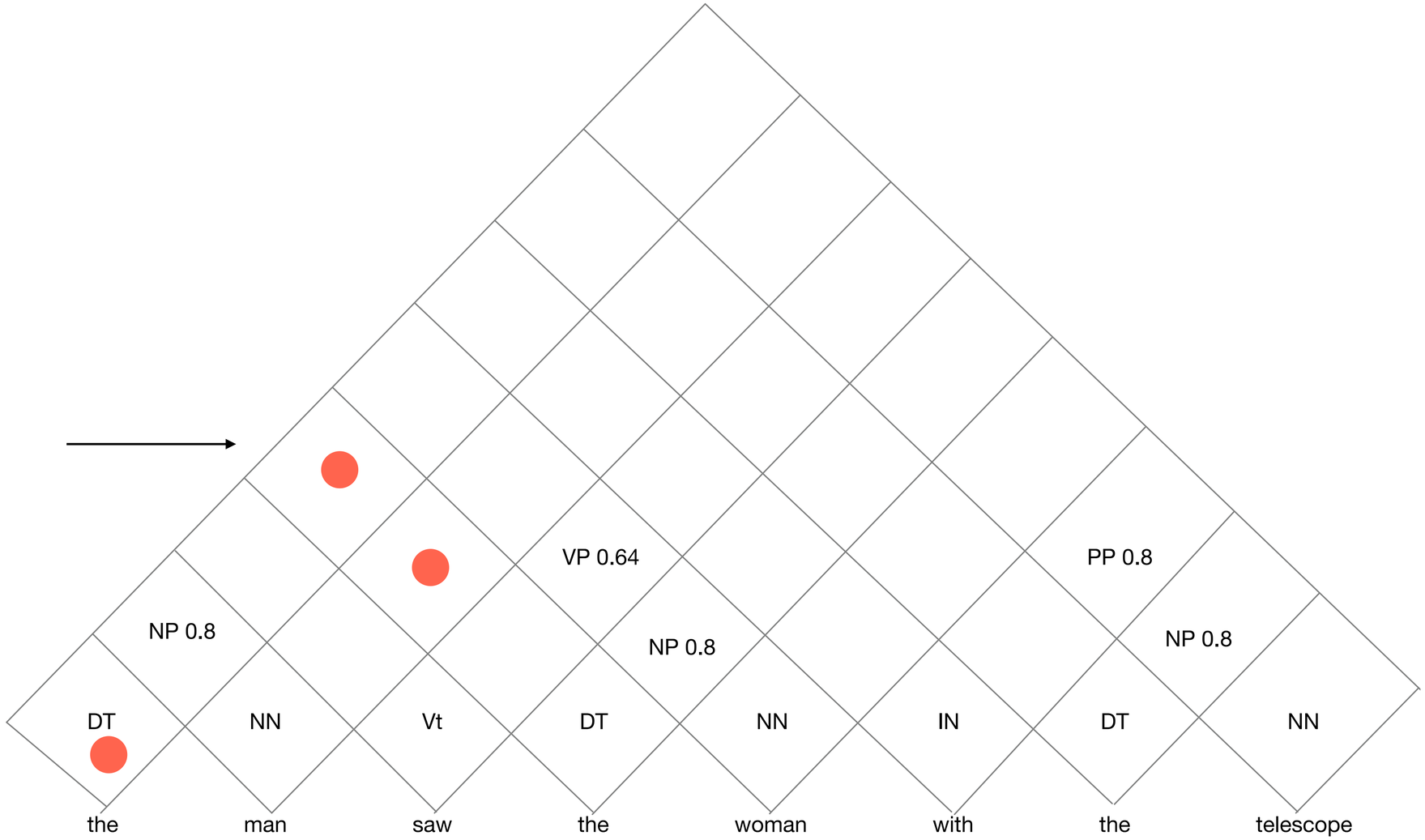
Parsing with PCFGs



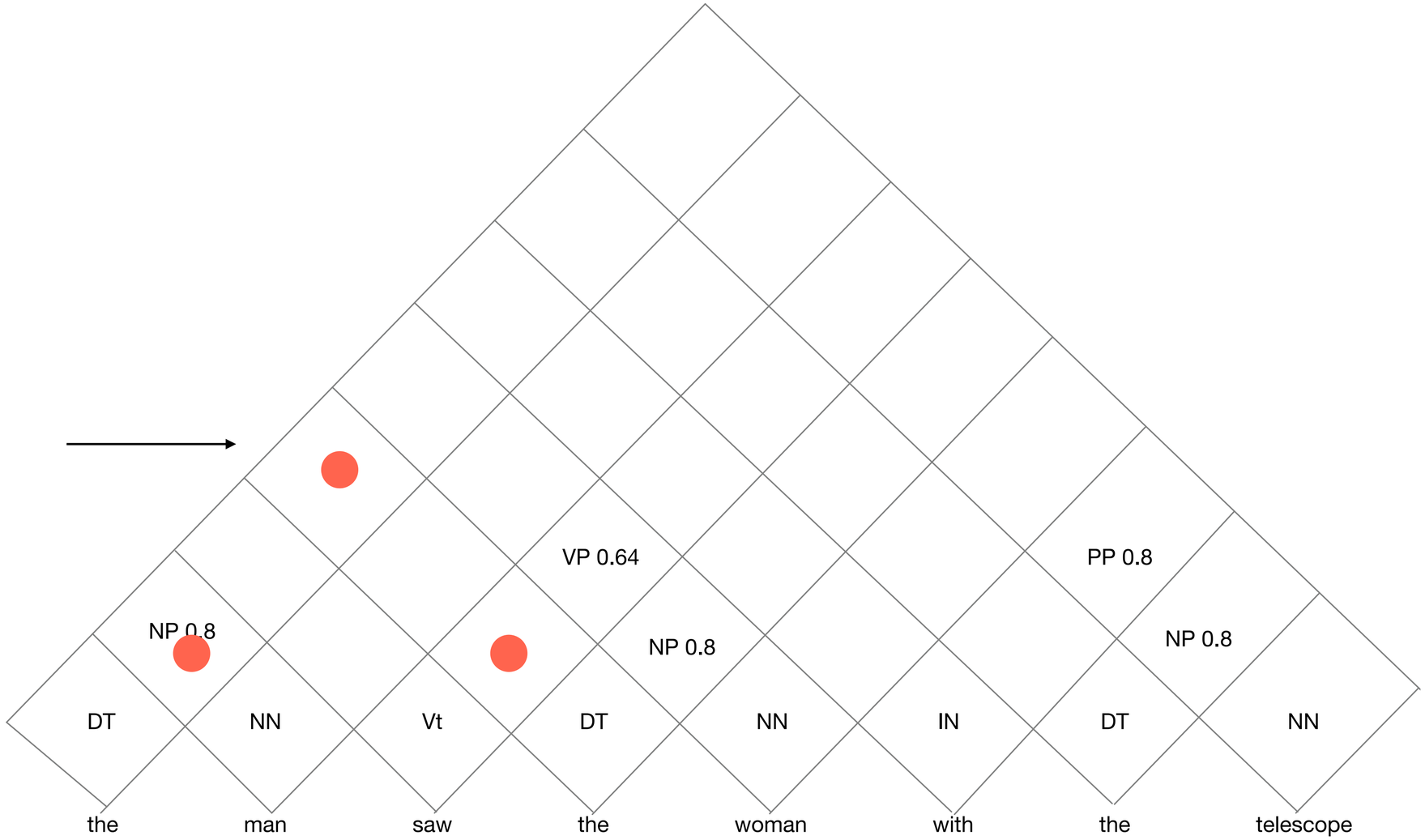
Parsing with PCFGs



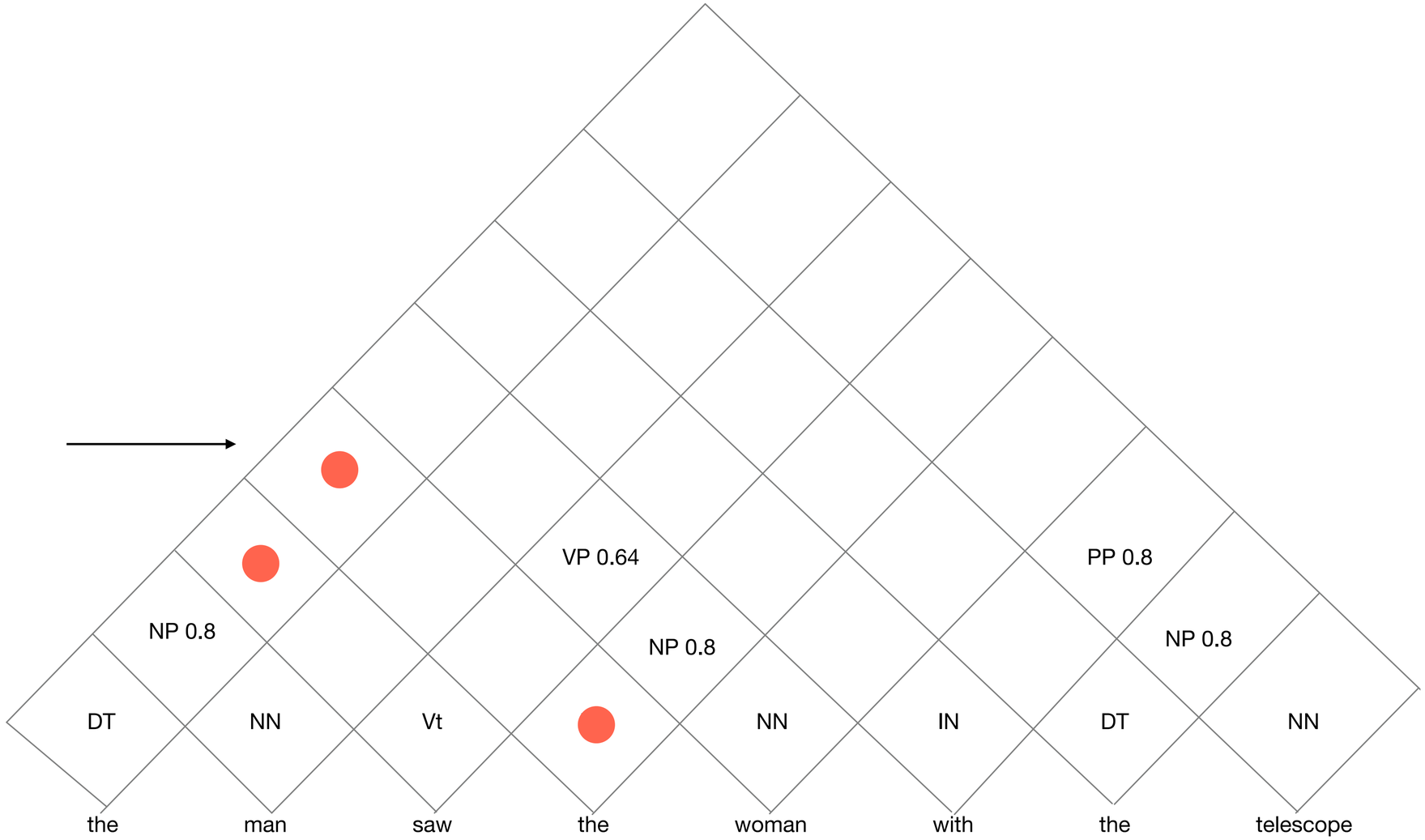
Parsing with PCFGs



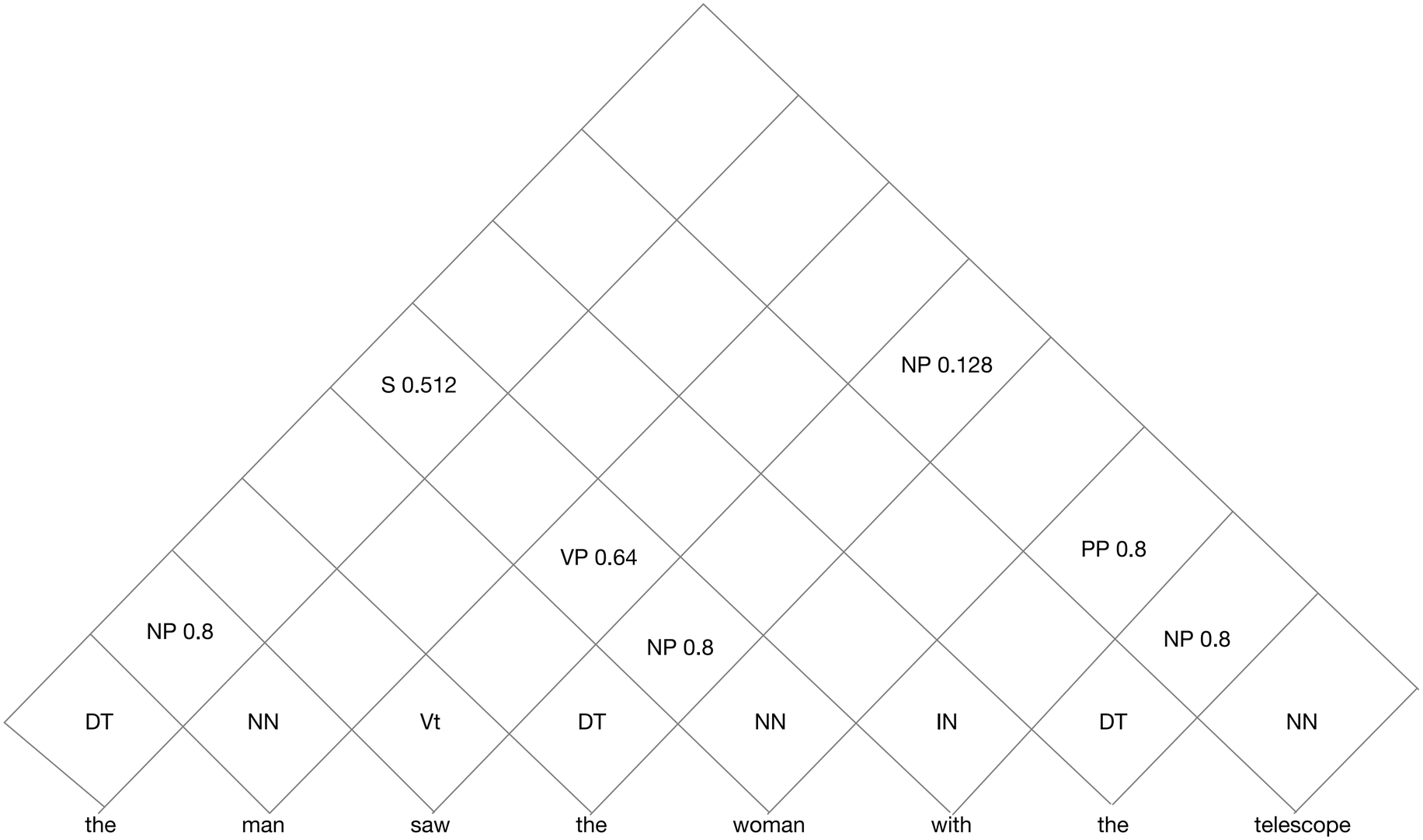
Parsing with PCFGs



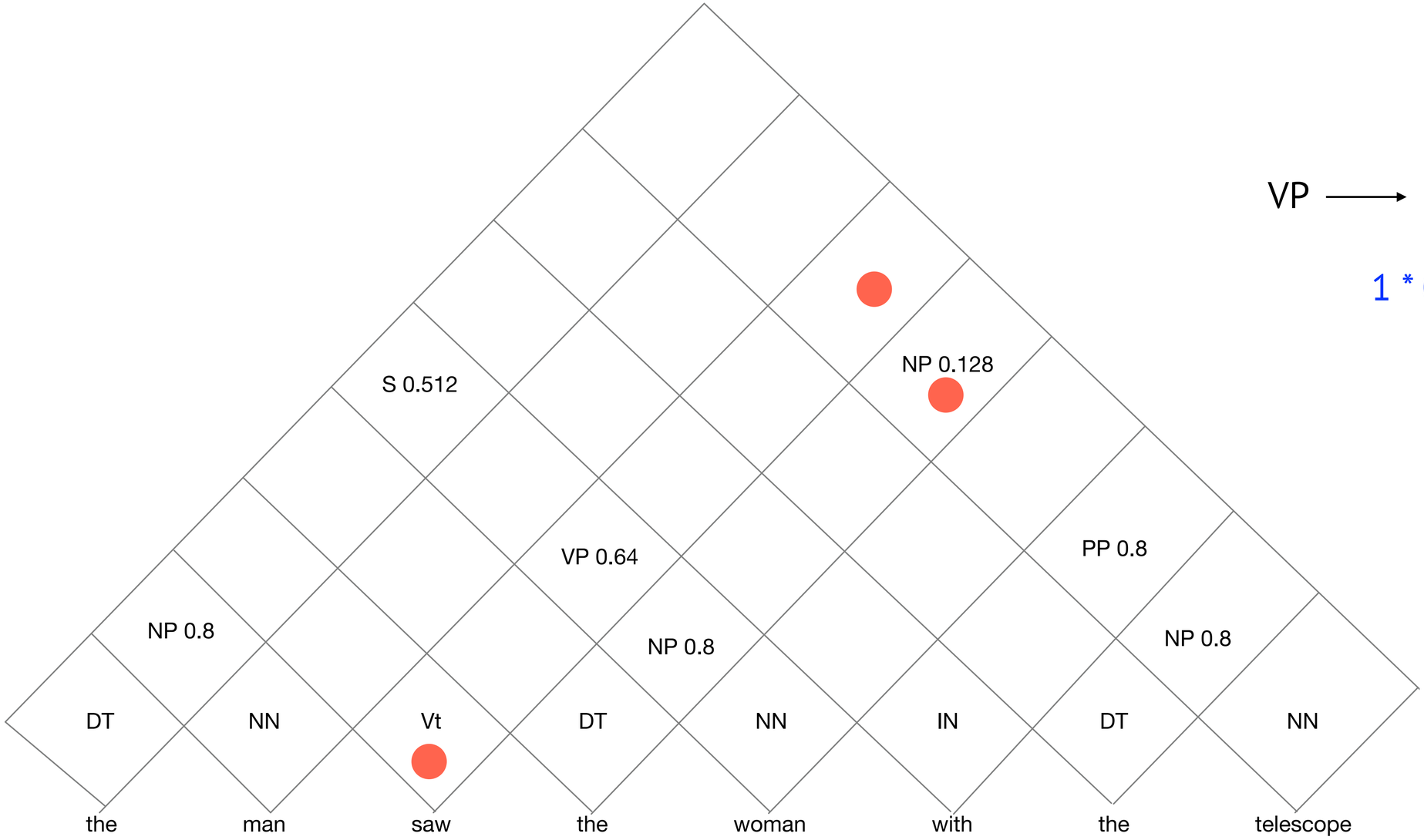
Parsing with PCFGs



Parsing with PCFGs



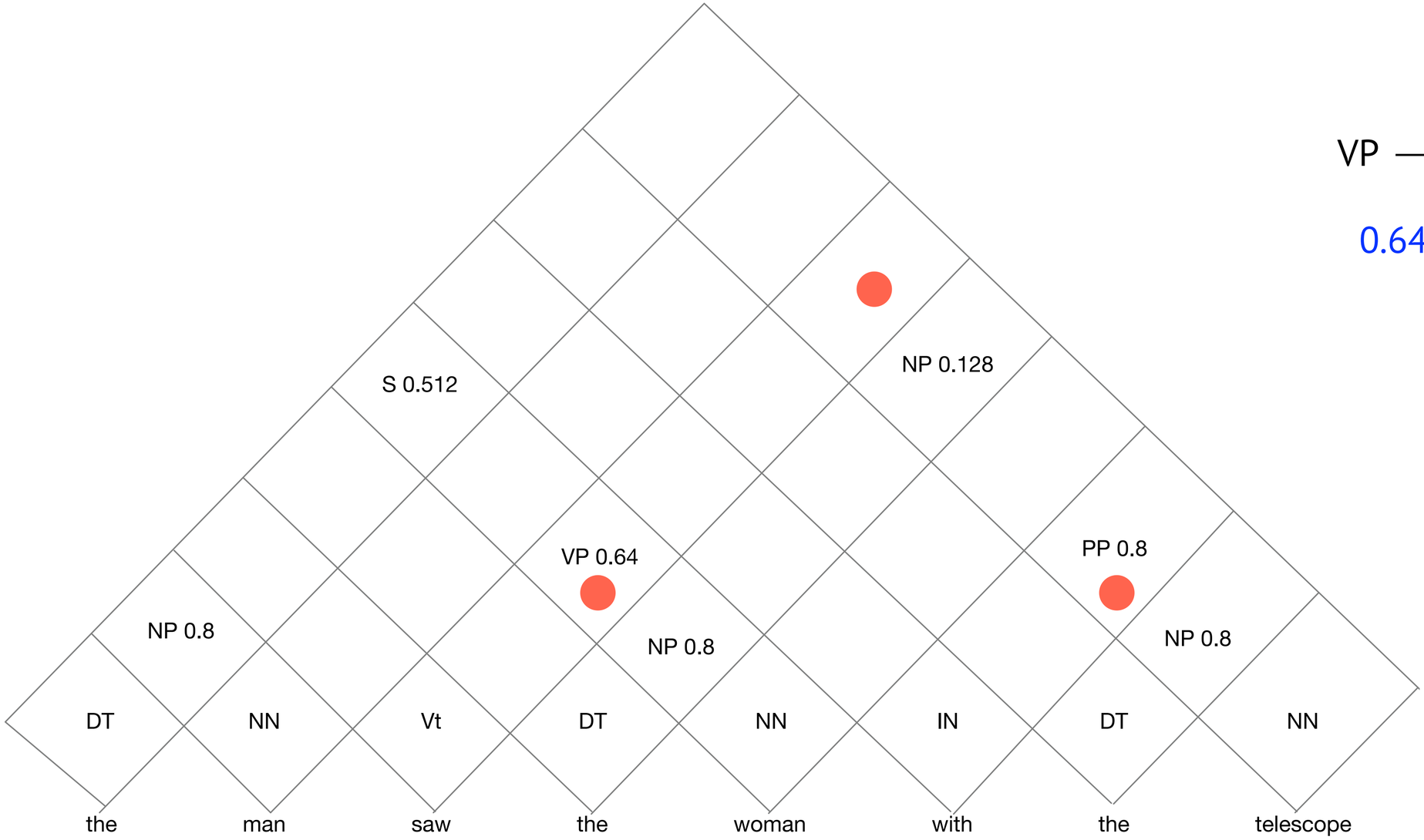
Parsing with PCFGs



VP ⟶ Vt NP 0.8

1 * 0.128 * 0.8 = 0.1024

Parsing with PCFGs



VP ⟶ VP PP 0.2

0.64 * 0.8 * 0.2 = 0.1024

Parsing with PCFGs (CYK Algorithm)