

# Phrase Structures

Computational Linguistics

Emory University

Jinho D. Choi



EMORY  
UNIVERSITY



# Context-Free Grammar

Computer Science  
Backus–Naur form

Linguistics  
Phrase structure grammar

$$G = (N, \Sigma, P, S)$$

$\Sigma$  : a finite set of **terminals** (word tokens).

$N$  : a finite set of **non-terminals** (POS/phrase/clause tags).

$P$  : a finite set of **production rules**, where  $N \rightarrow (N \cup \Sigma)^*$ .

$S$  : a **start symbol** representing the whole sentence, where  $S \in N$ .



# Context-Free Grammar

“I bought a car”

$\Sigma$  : a finite set of **terminals** (word tokens).

$\Sigma = \{I, \text{bought}, a, \text{car}\}$

$N$  : a finite set of **non-terminals** (POS/phrase/clause tags).

$N = \{\text{PRP}, \text{VBD}, \text{DT}, \text{NN}, \text{NP}, \text{VP}\}$

$P$  : a finite set of **production rules**, where  $N \rightarrow (N \cup \Sigma)^*$ .

$P = \{ \begin{array}{ll} \text{PRP} \rightarrow I, & \text{NP} \rightarrow \text{PRP}, \\ \text{VBD} \rightarrow \text{bought}, & \text{NP} \rightarrow \text{DT NN}, \\ \text{DT} \rightarrow a, & \text{VP} \rightarrow \text{VBD NP}, \\ \text{NN} \rightarrow \text{car}, & \text{S} \rightarrow \text{NP VP} \end{array} \}$



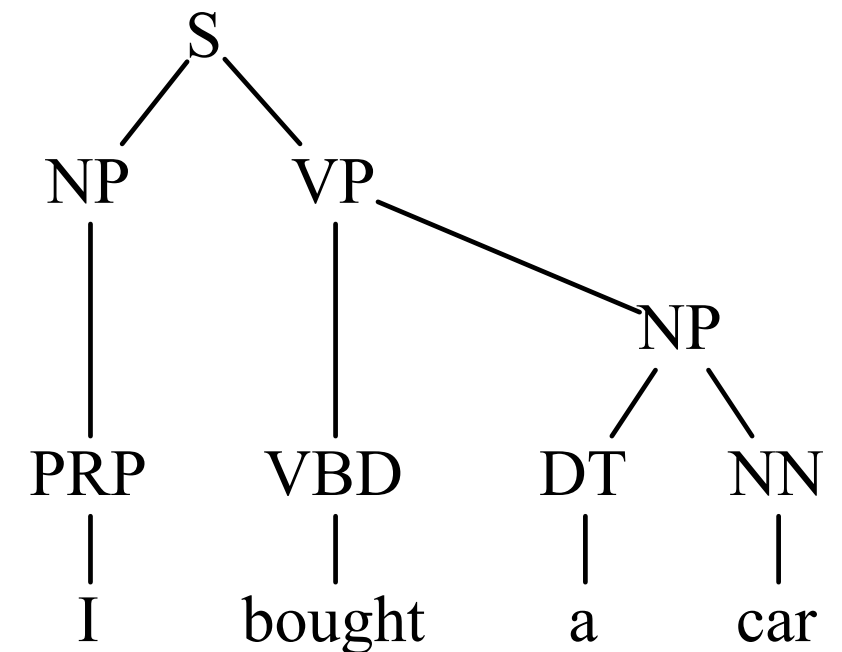
# Phrase Structures

“I bought a car”

$\Sigma = \{I, \text{bought}, a, \text{car}\}$

$N = \{PRP, VBD, DT, NN, NP, VP\}$

$P = \{PRP \rightarrow I, \quad NP \rightarrow PRP, \\ VBD \rightarrow \text{bought}, \quad NP \rightarrow DT \text{ } NN, \\ DT \rightarrow a, \quad VP \rightarrow VBD \text{ } NP, \\ NN \rightarrow \text{car}, \quad S \rightarrow NP \text{ } VP\}$



## Exercises

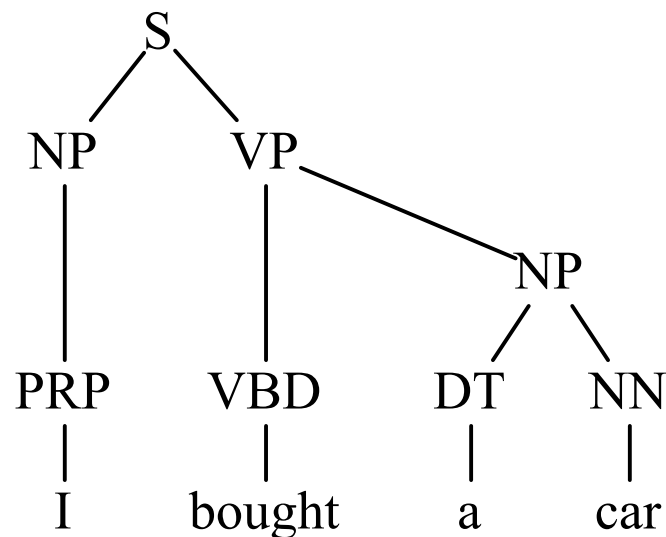
“I bought you a car”

“I bought a car yesterday”

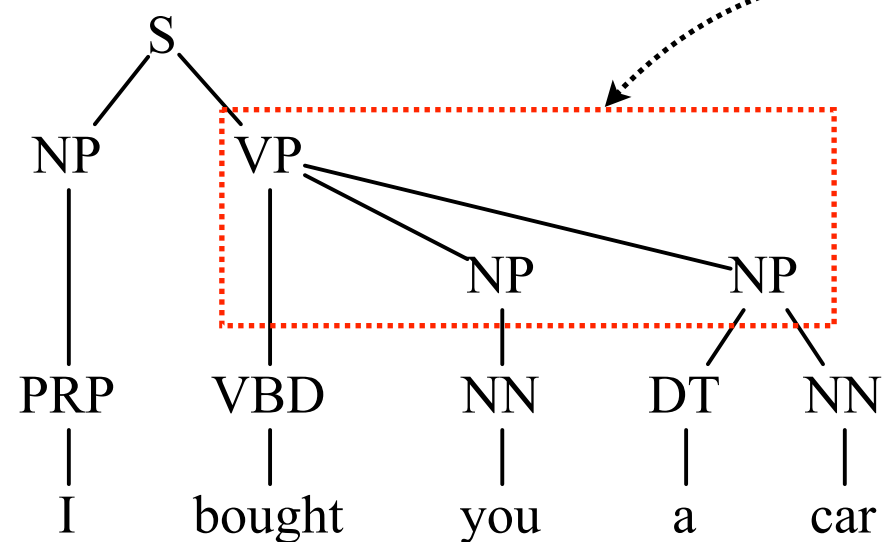


# Phrase Structures

“I bought a car”

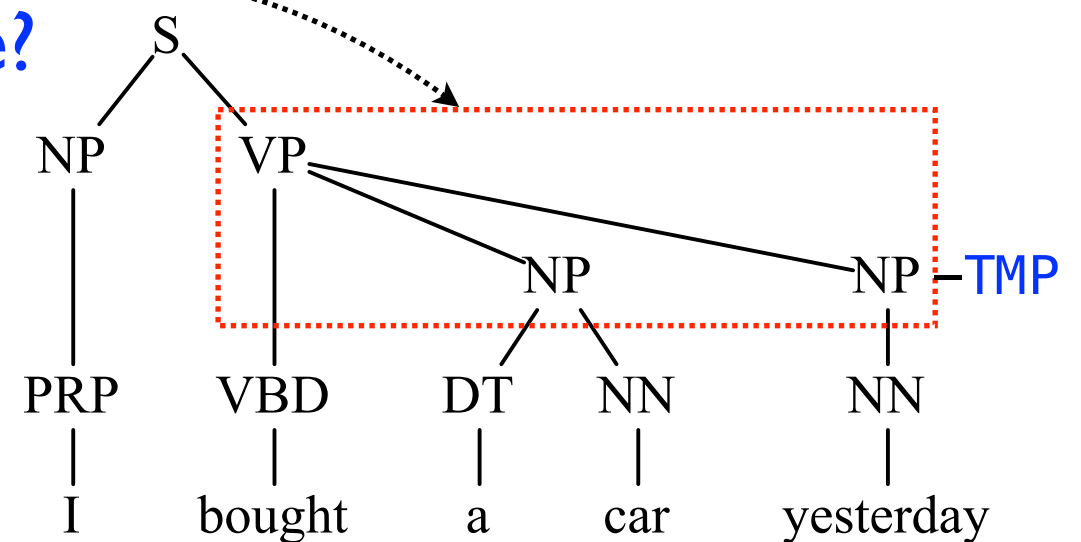


“I bought you a car”



same  
structure?

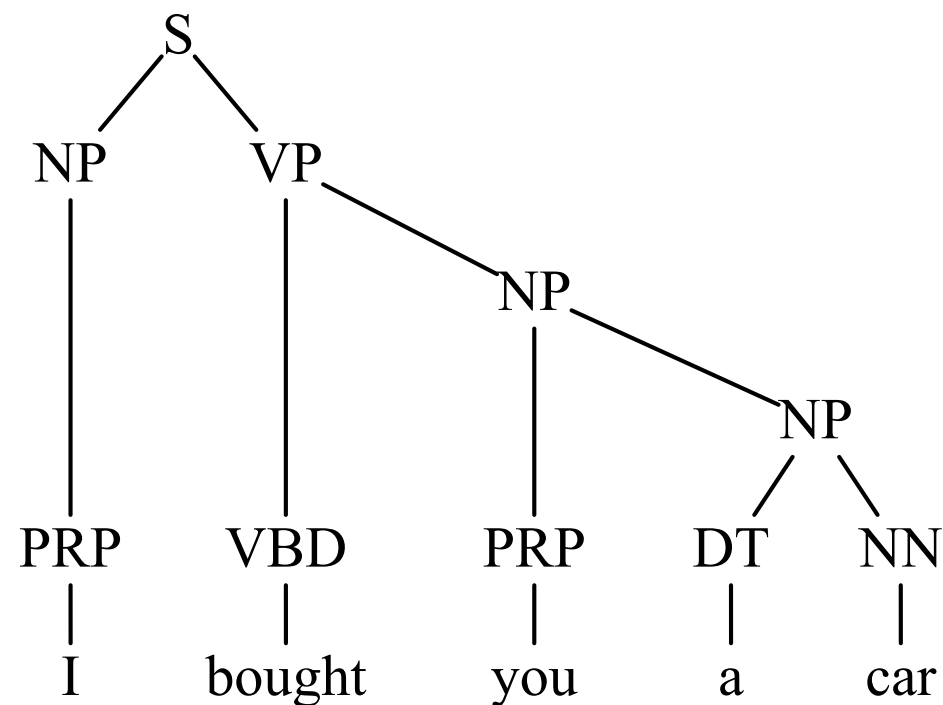
“I bought a car yesterday”



# Phrase Structure Rules

All **siblings** on the **right-hand side** should convey meaningful **relations**.

“I bought you a car”



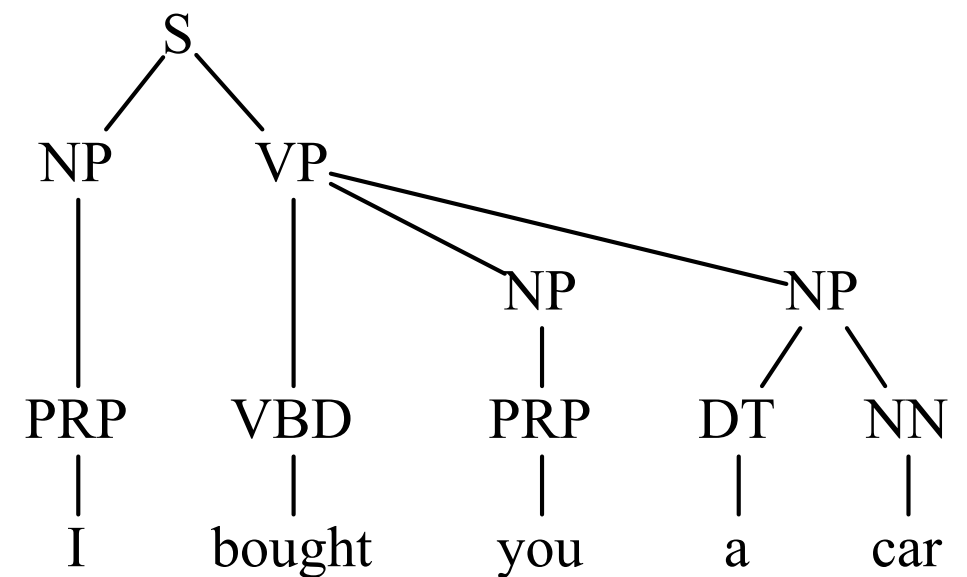
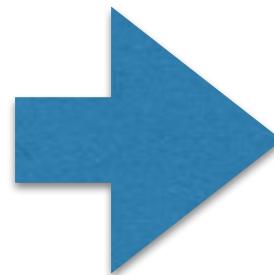
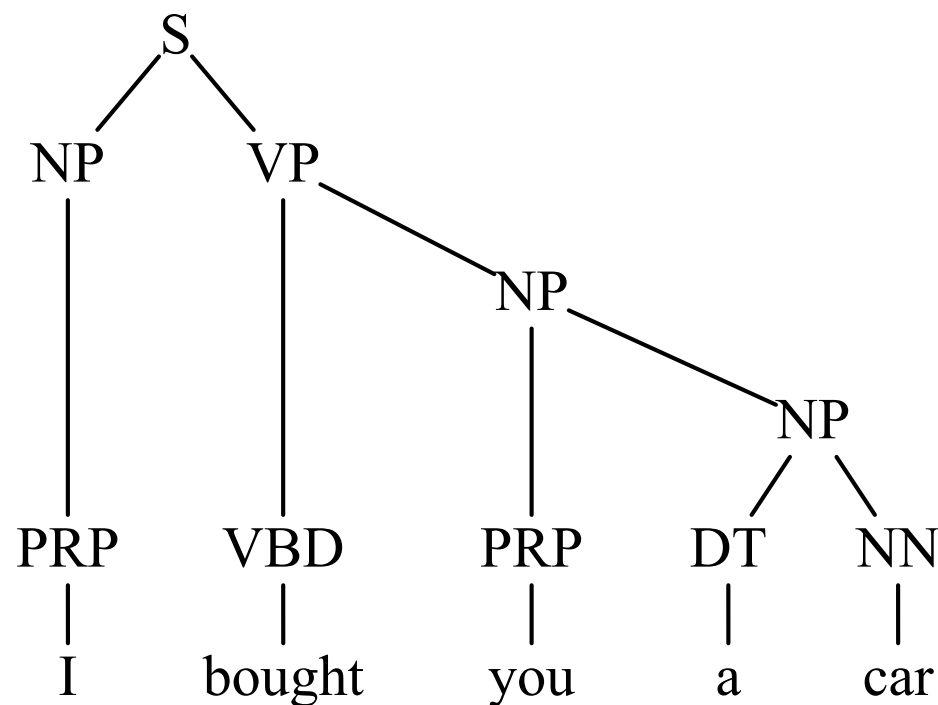
$S \rightarrow NP \ VP$      $\leftarrow$  **NP** is the subject of (the head of) **VP**  
 $VP \rightarrow VBD \ NP$      $\leftarrow$  **NP** is the object of **VBD**?  
 $NP \rightarrow PRP \ NP$      $\leftarrow$  ?



# Phrase Structure Rules

All **siblings** on the **right-hand side** should convey meaningful **relations**.

“I bought you a car”



$S \rightarrow NP \ VP$

$VP \rightarrow VBD \ NP$

$NP \rightarrow PRP \ NP$

**NP** is the subject of (the head of) **VP** →

1st **NP** is the indirect object of **VBD** →

2nd **NP** is the direct object of **VBD**

$S \rightarrow NP \ VP$

$VP \rightarrow VBD \ NP \ NP$

$NP \rightarrow DT \ NN$

**DT** is the determiner of **NN** ←



EMORY  
UNIVERSITY

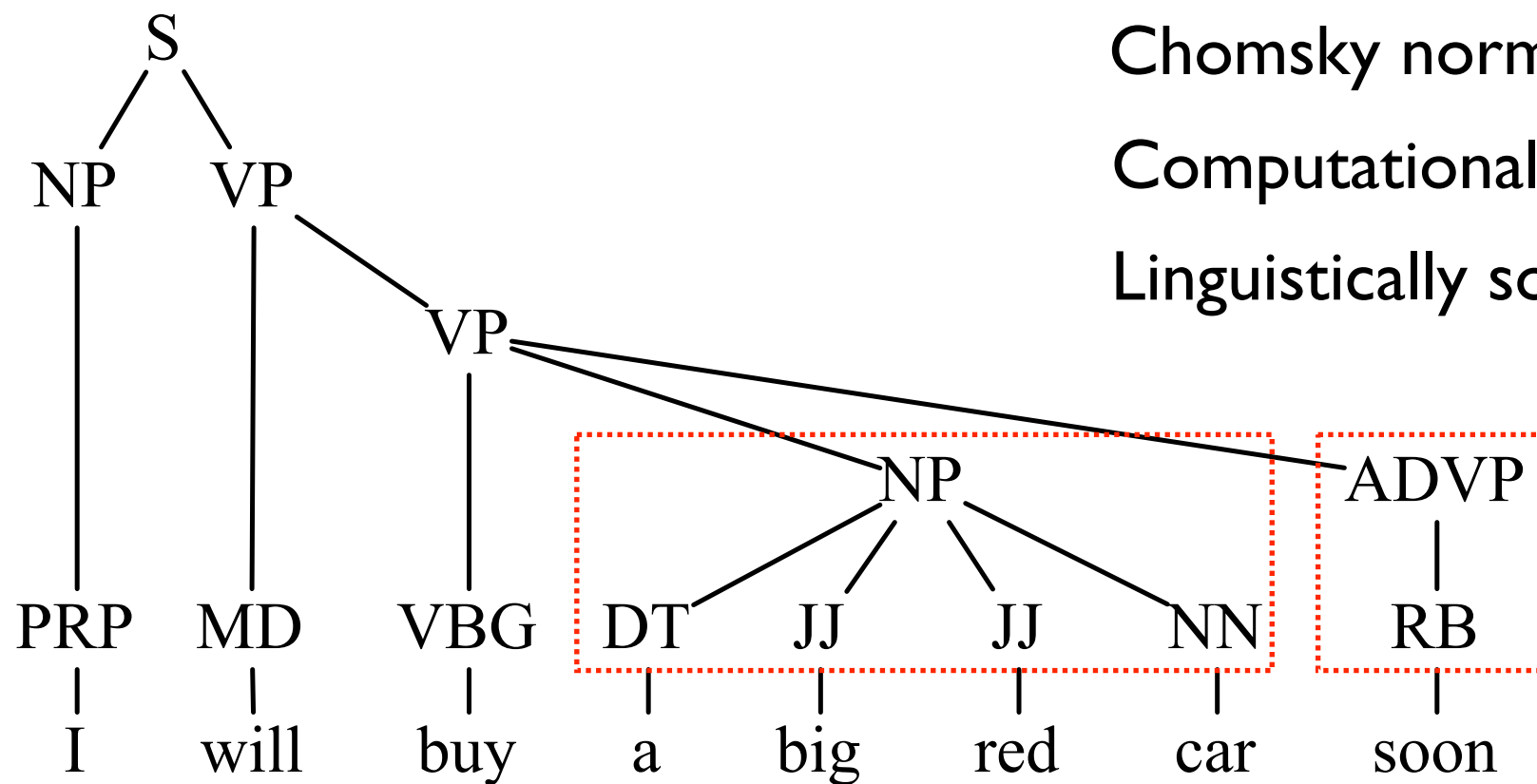


# Chomsky Normal Form

All production rules are  $A \rightarrow BC$  or  $A \rightarrow \alpha$  ( $A, B, C \in N, \alpha \in \Sigma$ ).

Why consider **CNF**?

“I will buy a big red car soon”



Chomsky normal form?

Computationally consistent?

Linguistically sound?





# Chomsky Normal Form

“I will buy a big red car soon”

## Non-lexicalized

$S \rightarrow NP \ VP$

$VP \rightarrow MD \ NP$

$VP \rightarrow VP \ ADVP$

$VP \rightarrow VBG \ NP$  **recursive**

$NP \rightarrow \boxed{DT \ NP | JJ \ NP} | JJ \ NN$

## Lexicalized

$\boxed{NP \rightarrow PRP \rightarrow I}$  **unary**

$MD \rightarrow will$

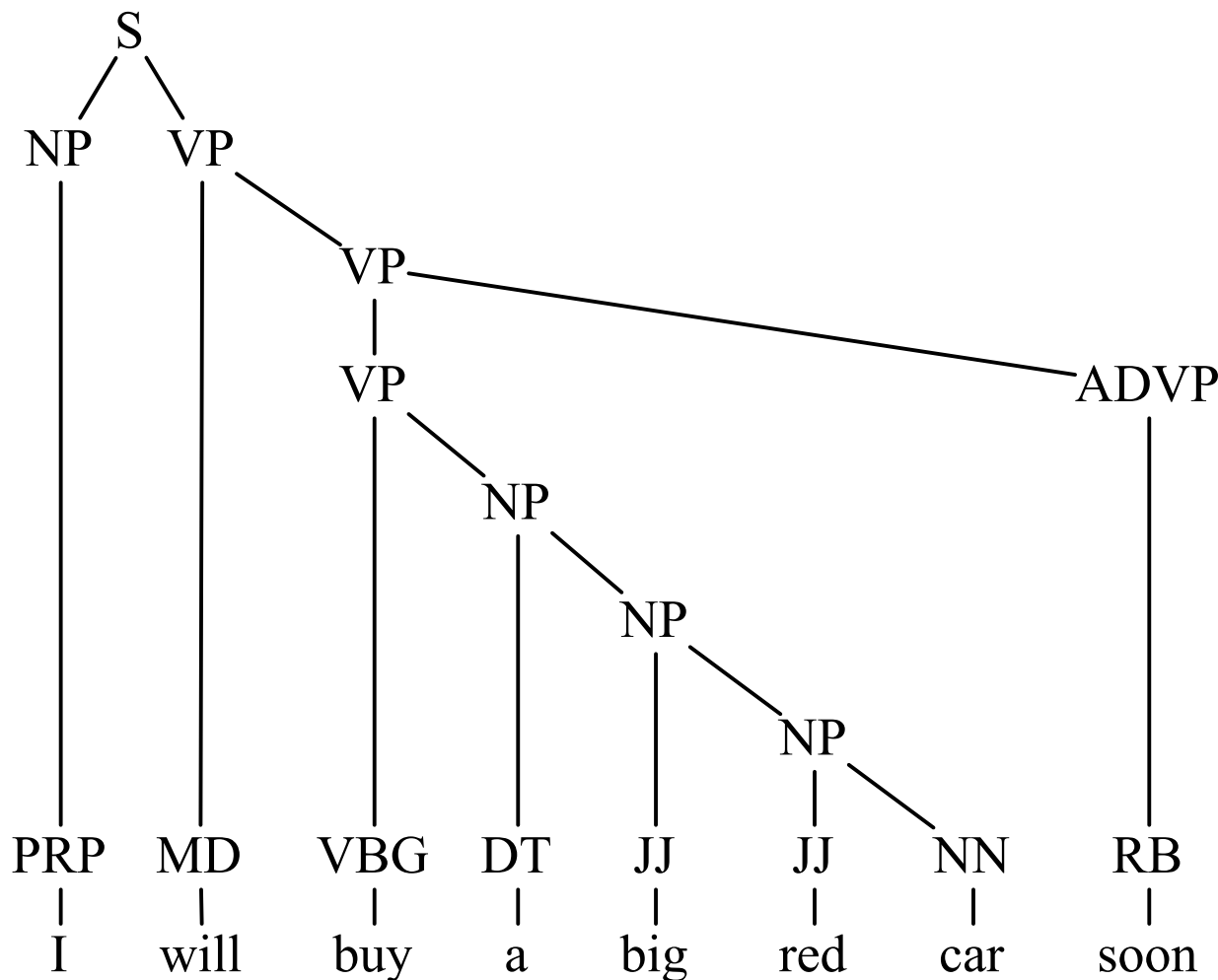
$VBG \rightarrow buy$

$DT \rightarrow a$

$JJ \rightarrow big | red$

$NN \rightarrow car$

$\boxed{ADVP \rightarrow RB \rightarrow soon}$

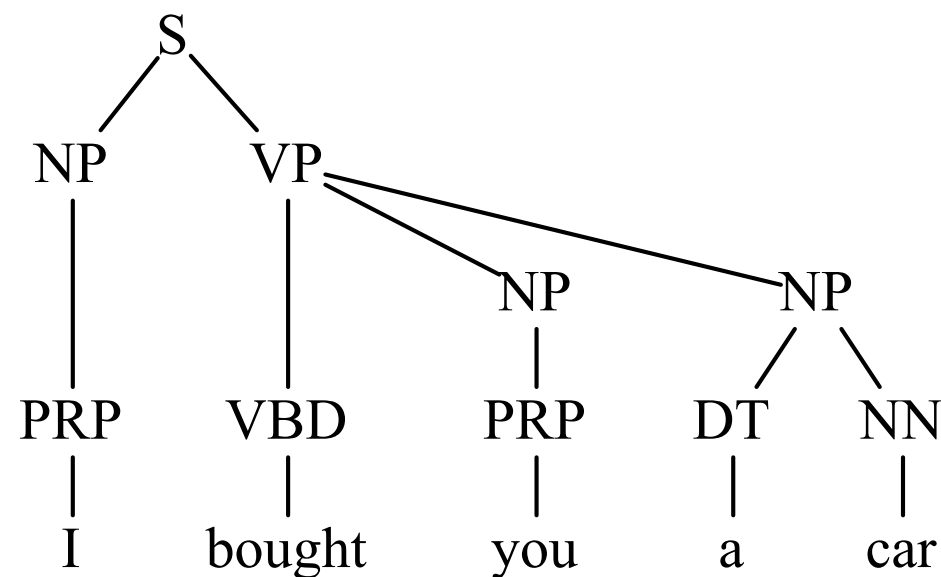


# Penn Treebank

A corpus containing 1M sentences from [Wall Street Journal](#) articles.

Each sentence is parsed into **phrase structure** trees.

Each tree is annotated in **parenthetical** notation.



```
((S (NP (PRP I))
    (VP (VBD bought)
        (NP (PRP you))
        (NP (DT a)
            (NN car))))))
```

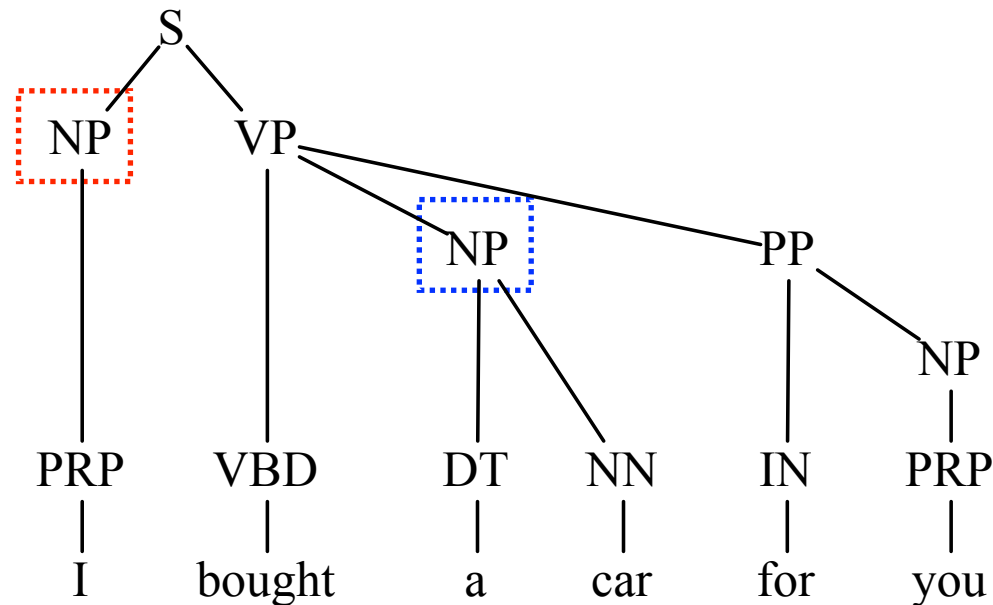
<https://www.cis.upenn.edu/~treebank/>

<http://web.mit.edu/6.863/www/PennTreebankTags.html>



# Passive Construction

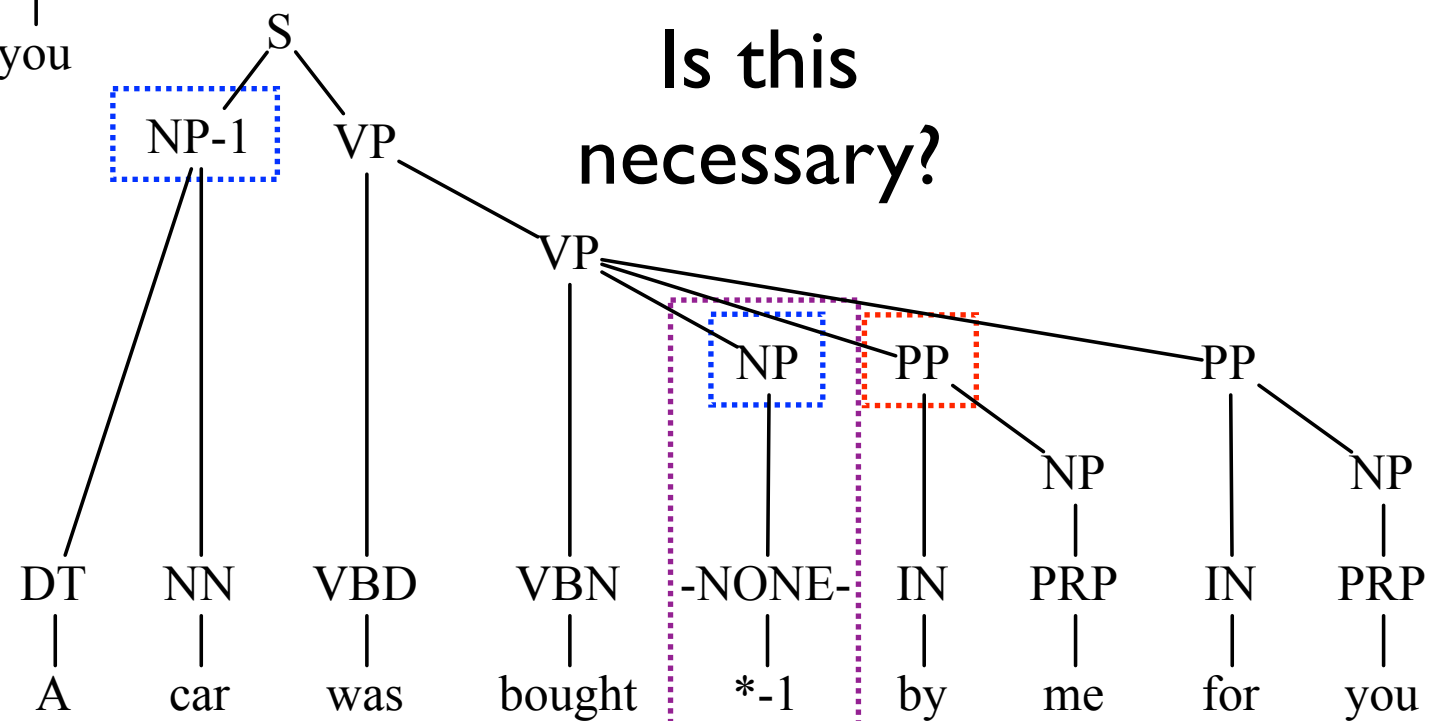
“I bought a car for you”



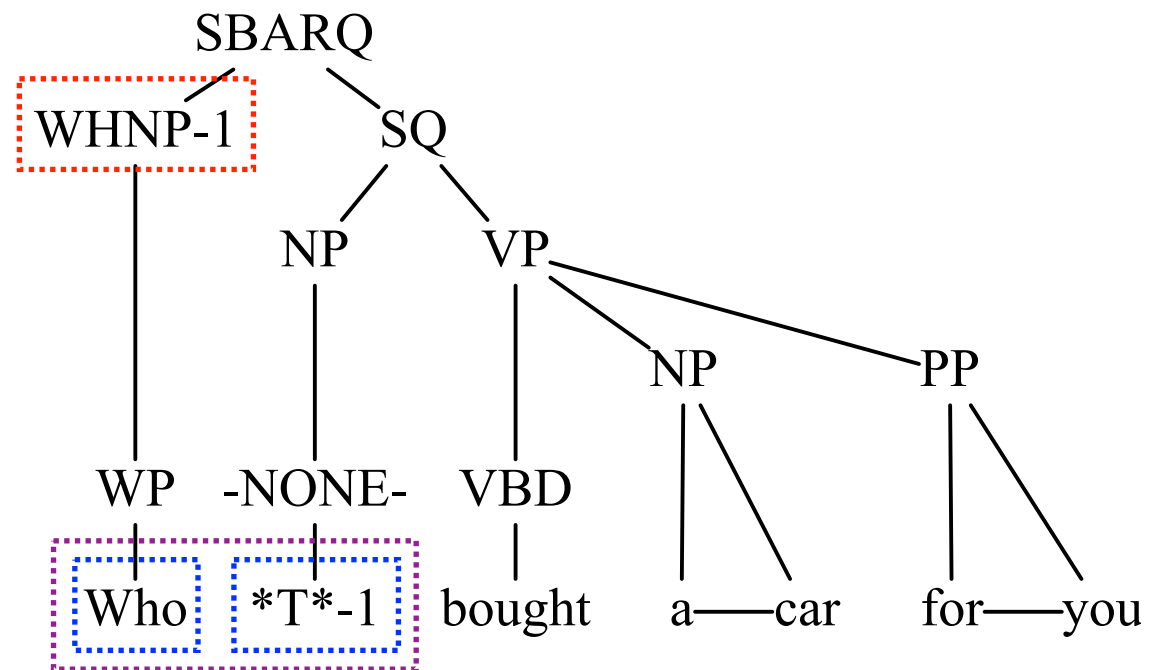
“A car was bought by me for you”

**NP Movement**  
Did NP really move?

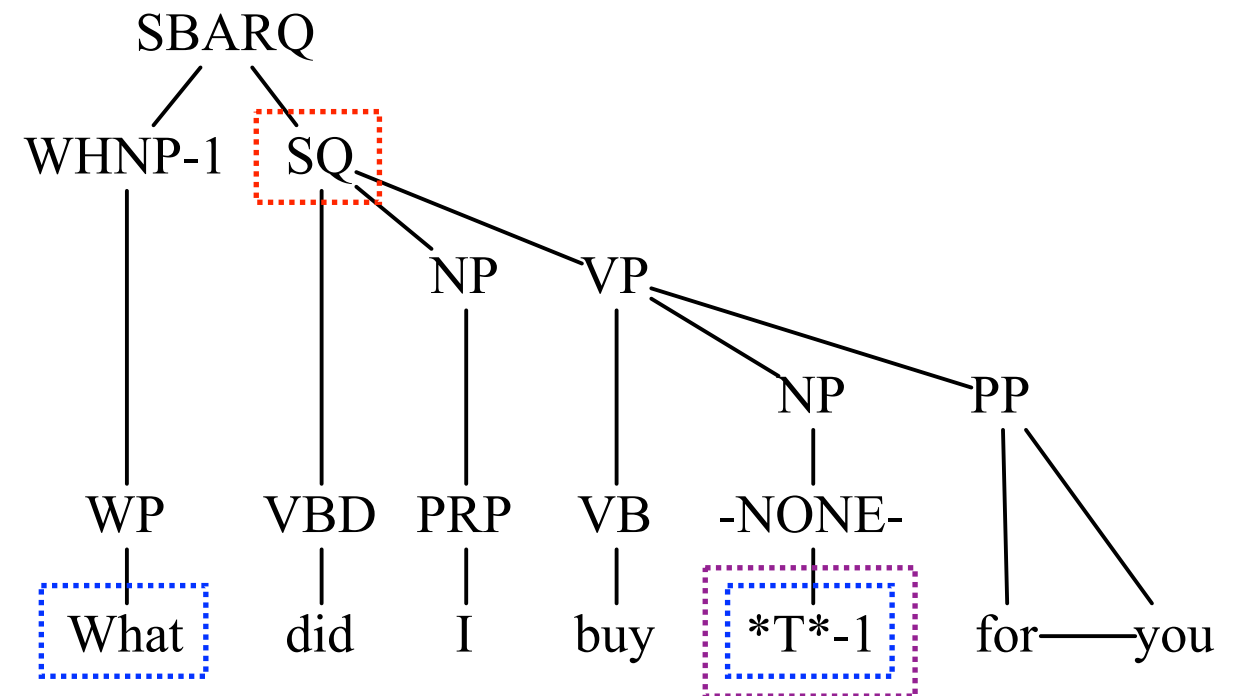
What about  
automatic parsing?



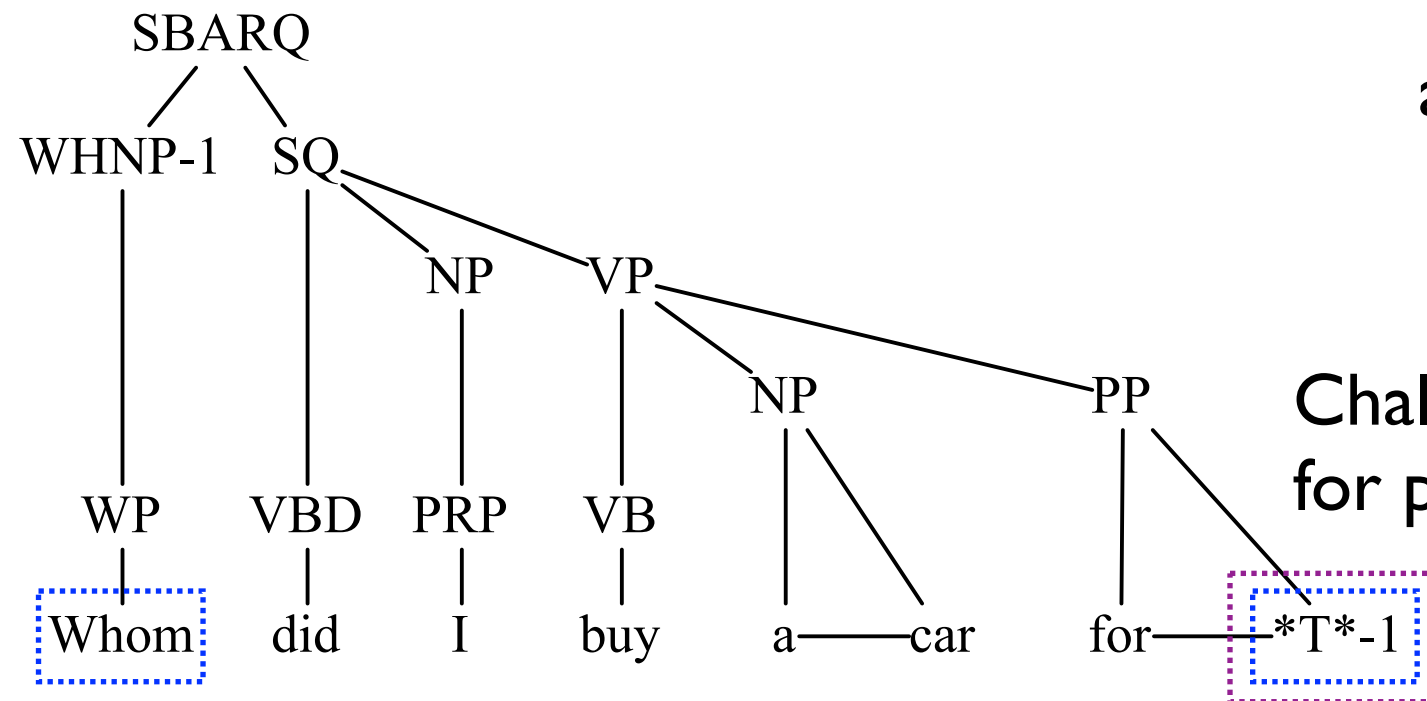
# Wh-Questions



Is this  
necessary?



What  
about this?

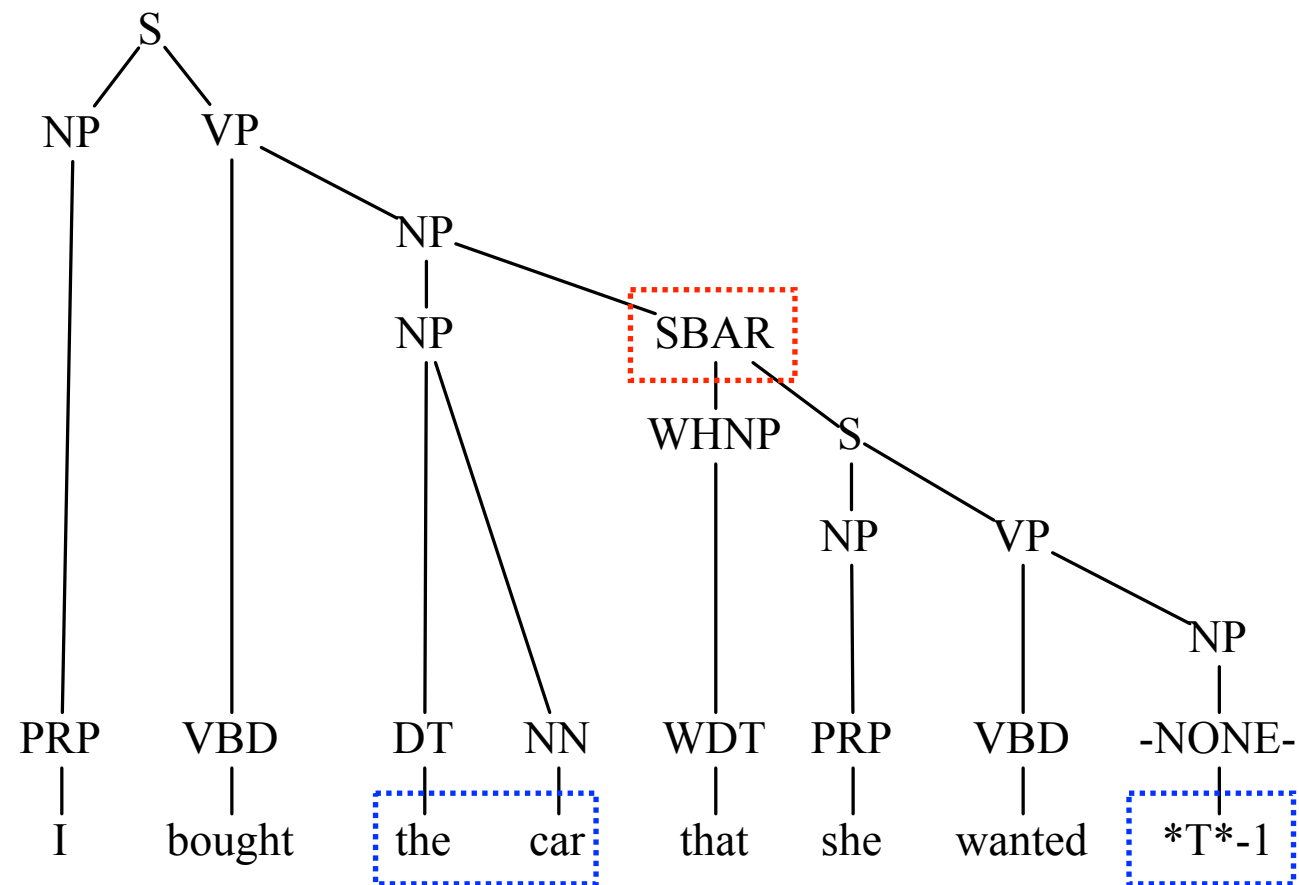


Challenging  
for parsing?

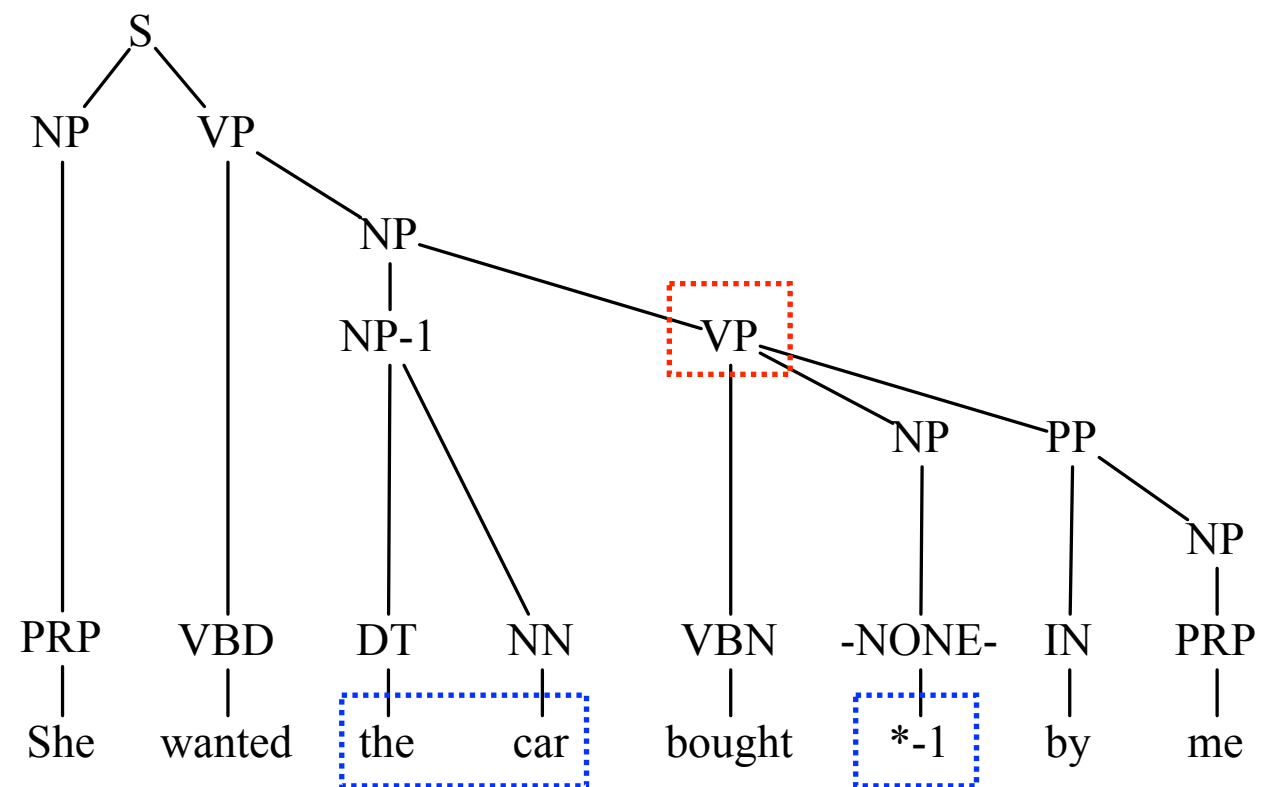


# Relative Clause

I bought the car that she wanted

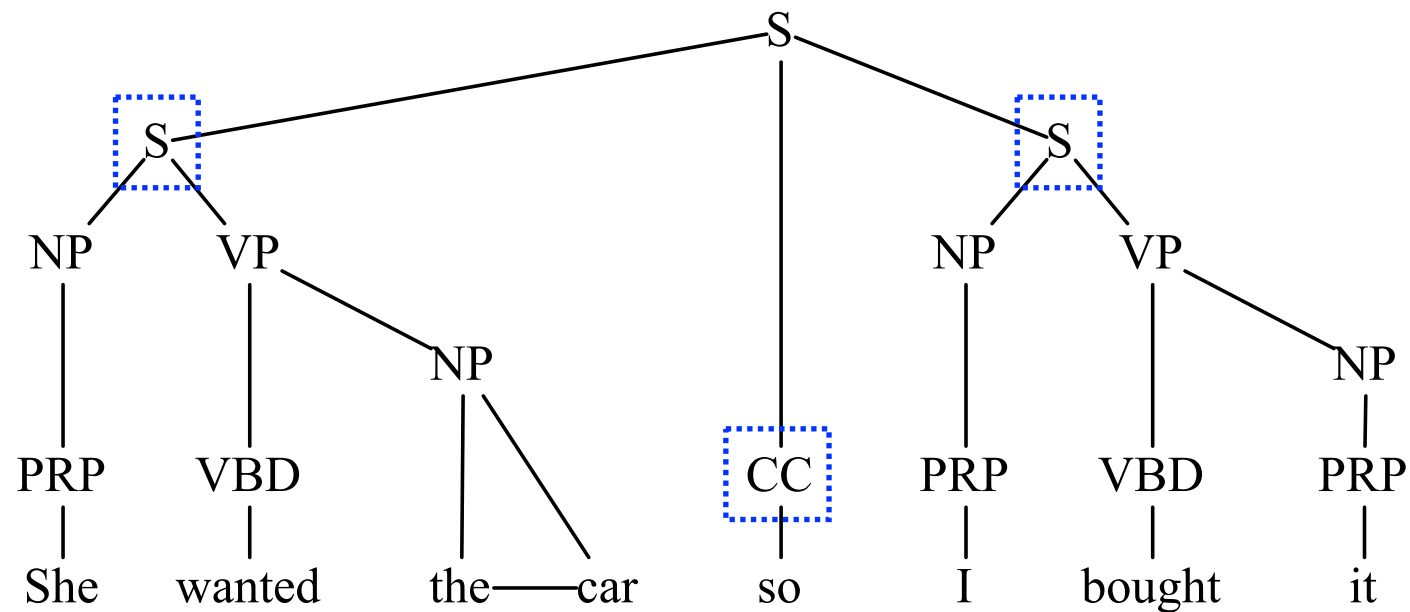


She wanted the car bought by me

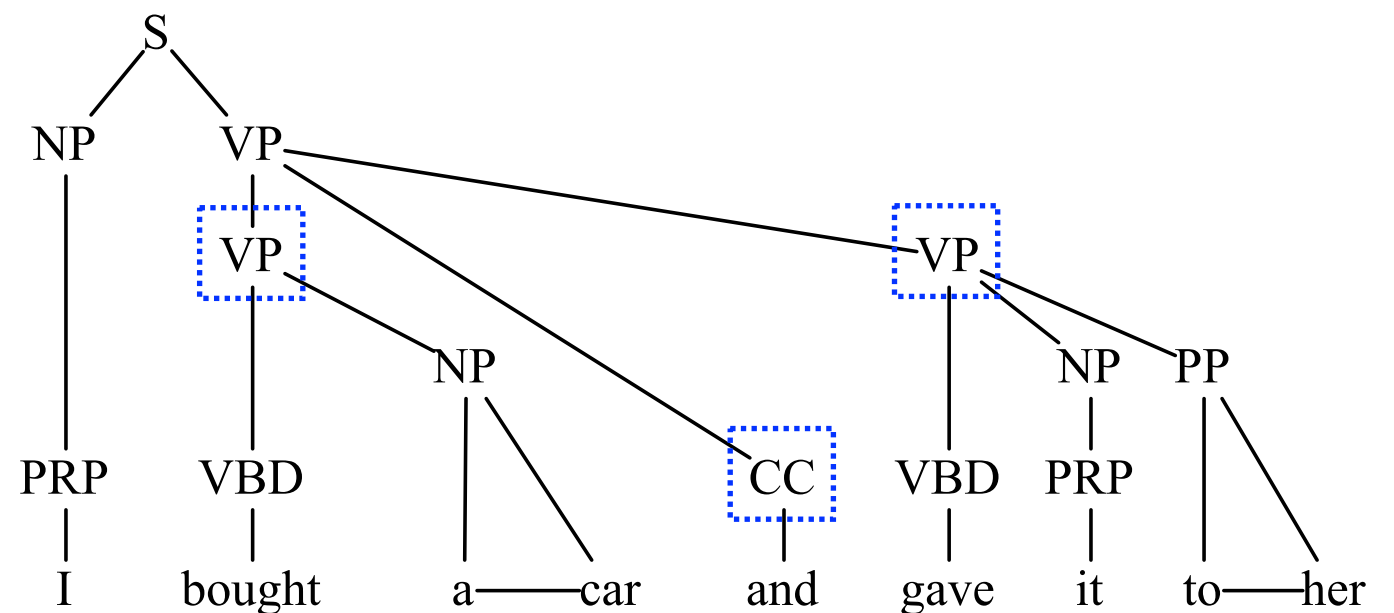


# Coordination

She **wanted** the car **so** I **bought** it.



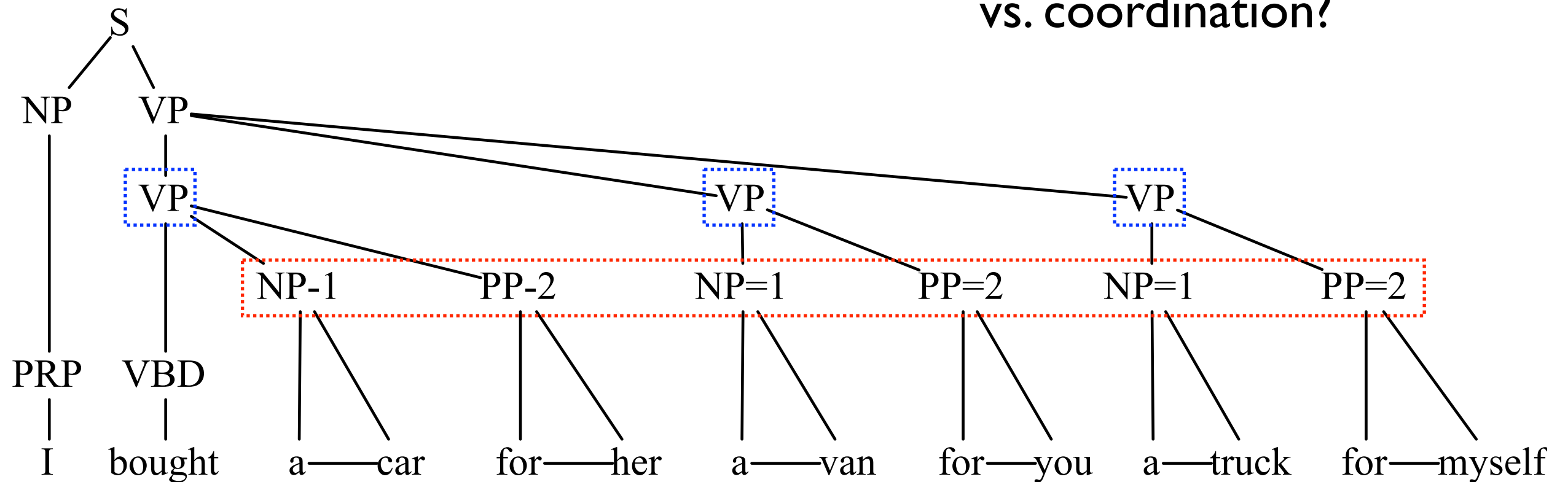
I **bought** a car **and** **gave** it to her.



# Gapping Relation

I bought a car for you , a van for you , and a truck for myself

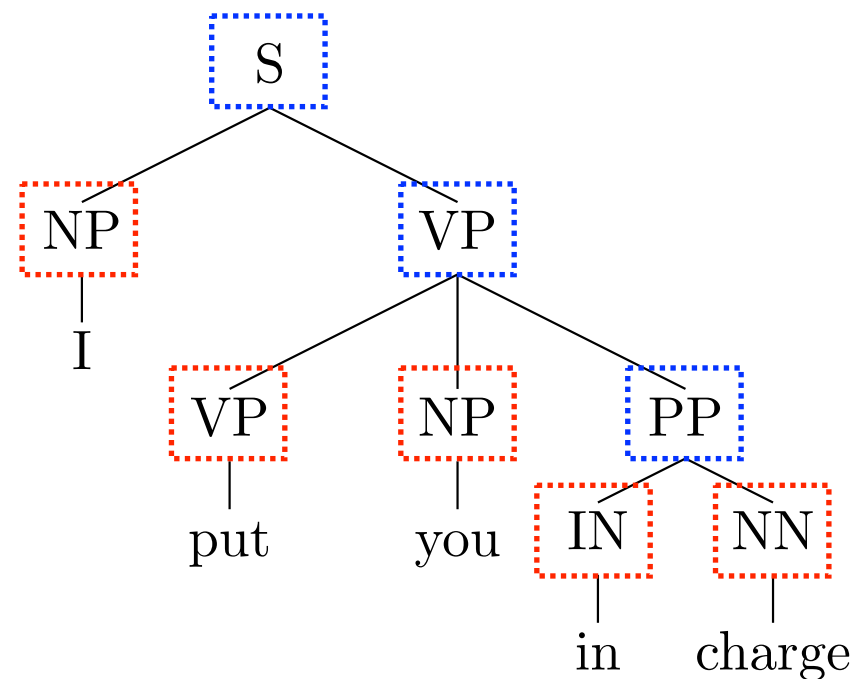
vs. coordination?



# Bracketing Scores

Evaluate the **automatically** parsed trees.

Gold



Gold: S(0:5), VP(1:5), PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)

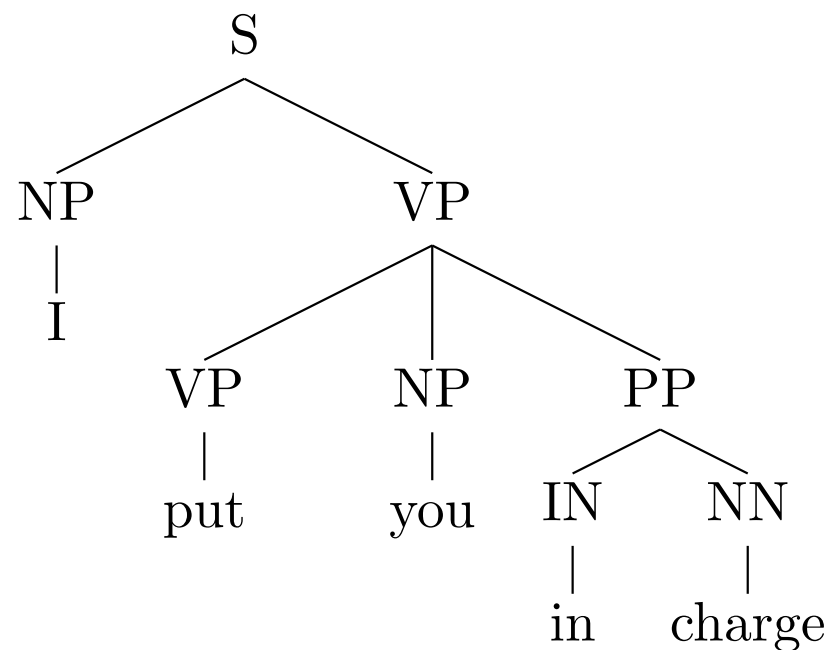




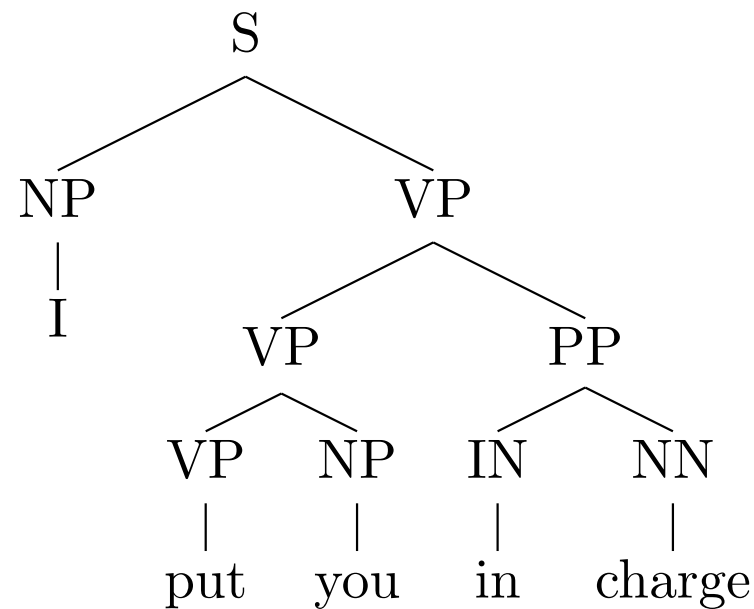
# Bracketing Scores

Evaluate the **automatically** parsed trees.

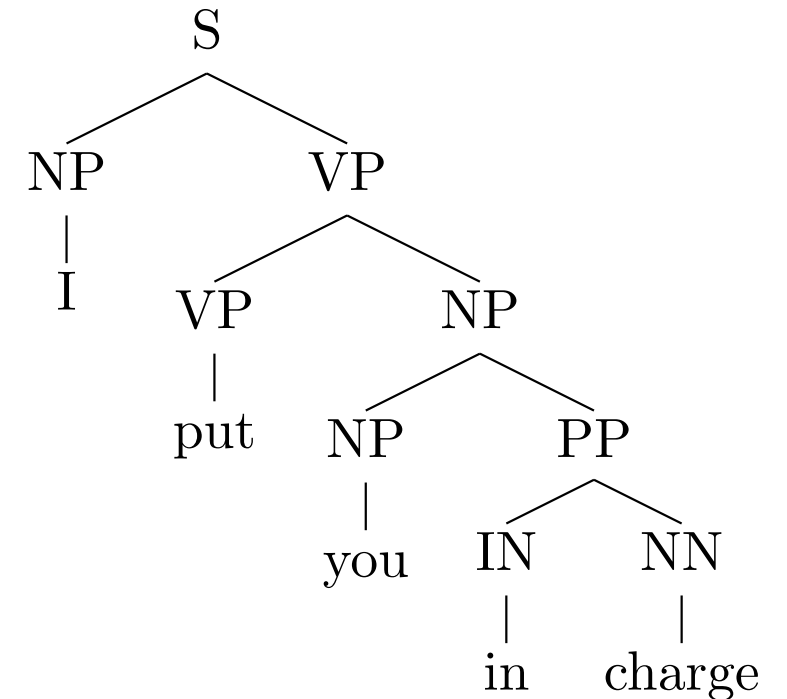
Gold



Parse 1



Parse 2



**Gold:** S(0:5), VP(1:5), PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)

**Parse 1:** S(0:5), VP(1:5), **VP(1:3)**, PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)

**Parse 2:** S(0:5), VP(1:5), **NP(2:5)**, PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)



# Bracketing Scores

Precision

$$\frac{\# \text{ of correct brackets}}{\# \text{ of parsed brackets}} = 8/9$$

Recall

$$\frac{\# \text{ of correct brackets}}{\# \text{ of gold brackets}} = 8/8$$

F1-Score

$$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Gold: S(0:5), VP(1:5), PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)

Parse 1: S(0:5), VP(1:5), VP(1:3), PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)

Parse 2: S(0:5), VP(1:5), NP(2:5), PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)



# I put you in charge

	1	2	3	4	5
0					
1					
2					
3					
4					

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3				
1		VP → put 0.3			
2			NP → you 0.3		
3				IN → in 1.0	
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09			
1		VP → put 0.3			
2			NP → you 0.3		
3				IN → in 1.0	
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09			
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09			
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09			
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		





# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09			
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP 1.0      NP → NP PP 0.2      PP → IN NP 1.0  
 VP → VP PP 0.2      NP → I 0.3      IN → in 1.0  
 VP → VP NP 0.5      NP → you 0.3  
 VP → put 0.3      NP → charge 0.2



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP 1.0      NP → NP PP 0.2      PP → IN NP 1.0  
 VP → VP PP 0.2      NP → I 0.3      IN → in 1.0  
 VP → VP NP 0.5      NP → you 0.3  
 VP → put 0.3      NP → charge 0.2



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP 1.0      NP → NP PP 0.2      PP → IN NP 1.0  
 VP → VP PP 0.2      NP → I 0.3      IN → in 1.0  
 VP → VP NP 0.5      NP → you 0.3  
 VP → put 0.3      NP → charge 0.2



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		





# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		VP → VP PP .2*.045*.2 =.0018
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		VP → VP NP .5*.3*.012 =.0018
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		VP → VP PP .2*.045*.2 =.0018
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		VP → VP PP .2*.045*.2 =.0018
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP 1.0      NP → NP PP 0.2      PP → IN NP 1.0  
 VP → VP PP 0.2      NP → I 0.3      IN → in 1.0  
 VP → VP NP 0.5      NP → you 0.3  
 VP → put 0.3      NP → charge 0.2



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		VP → VP PP .2*.045*.2 =.0018
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP 1.0      NP → NP PP 0.2      PP → IN NP 1.0  
 VP → VP PP 0.2      NP → I 0.3      IN → in 1.0  
 VP → VP NP 0.5      NP → you 0.3  
 VP → put 0.3      NP → charge 0.2



# I put you in charge

	1	2	3	4	5
0	NP → I 0.3	S → NP VP 1*.3*.3 =.09	S → NP VP 1*.3*.045 =.0135		S → NP VP 1*.3*.0018 =.00054
1		VP → put 0.3	VP → VP NP .5*.3*.3 =.045		VP → VP PP .2*.045*.2 =.0018
2			NP → you 0.3		NP → NP PP .2*.3*.2 =.012
3				IN → in 1.0	PP → IN NP 1*1*.2 =.2
4					NP → charge 0.2

S → NP VP	1.0	NP → NP PP	0.2	PP → IN NP	1.0
VP → VP PP	0.2	NP → I	0.3	IN → in	1.0
VP → VP NP	0.5	NP → you	0.3		
VP → put	0.3	NP → charge	0.2		

