Phrase Structures

Computational Linguistics

Emory University

Jinho D. Choi

EMORY
UNIVERSITY

Context-Free Grammar

Computer Science
Backus–Naur form

Linguistics
Phrase structure grammar

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

Σ : 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"

Σ : a finite set of terminals (word tokens).

Σ = {I, bought, a, car}

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

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

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

```
P = {PRP → I,        NP → PRP,
     VBD → bought,   NP → DT NN,
     DT  → a,        VP → VBD NP,
     NN  → car,       S → NP VP}
```

EMORY UNIVERSITY

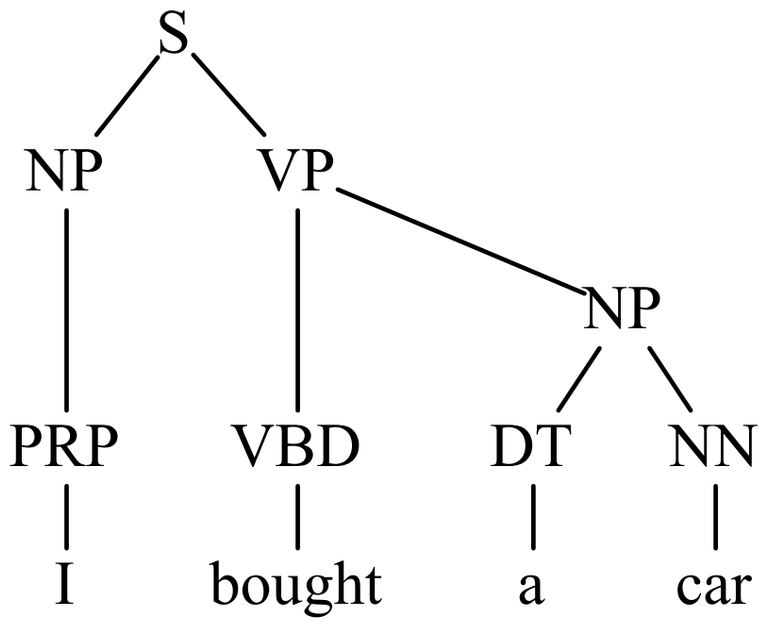
Phrase Structures

"I bought a car"

Σ = {I, bought, a, car}

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

P = {PRP → I, NP → PRP,
VBD → bought, NP → DT NN,
DT → a, VP → VBD NP,
NN → car, S → NP VP}



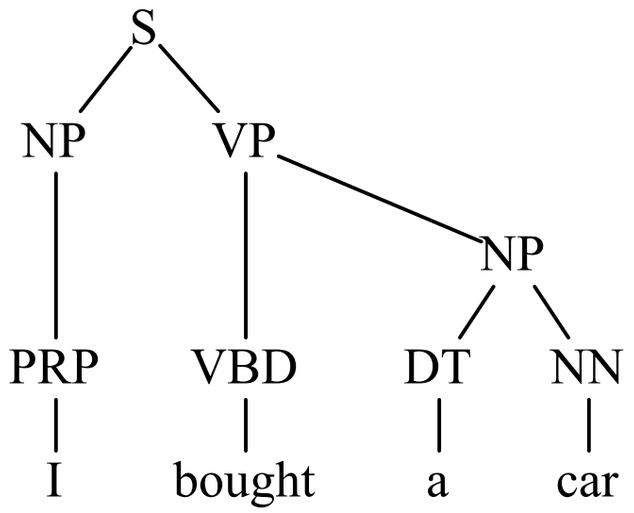
Exercises

"I bought you a car"

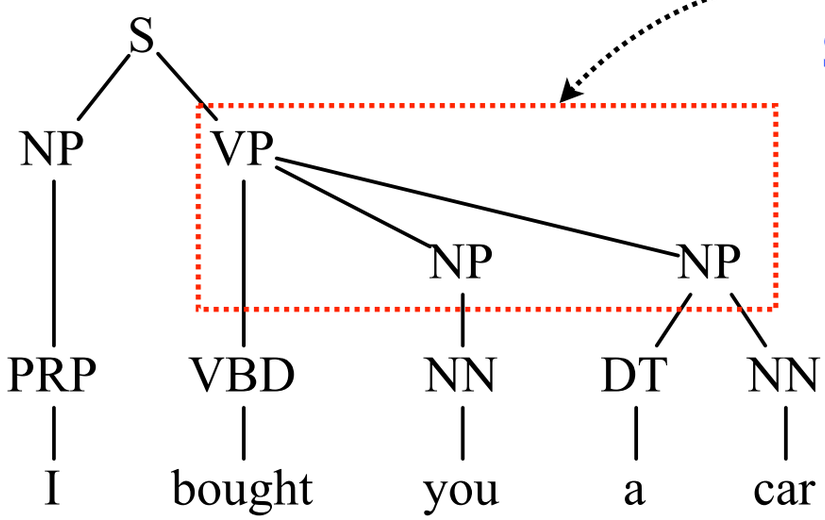
"I bought a car yesterday"

Phrase Structures

"I bought a car"

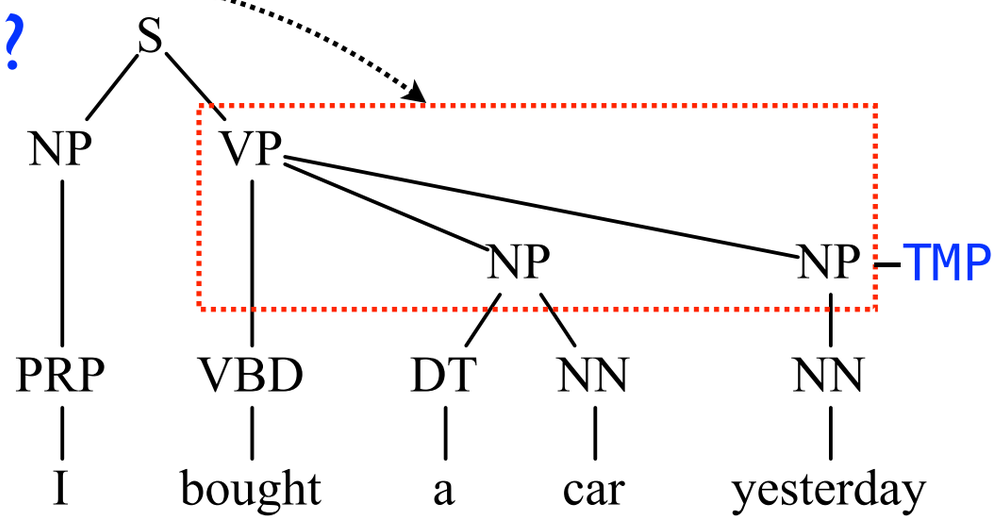


"I bought you a car"



same structure?

"I bought a car yesterday"

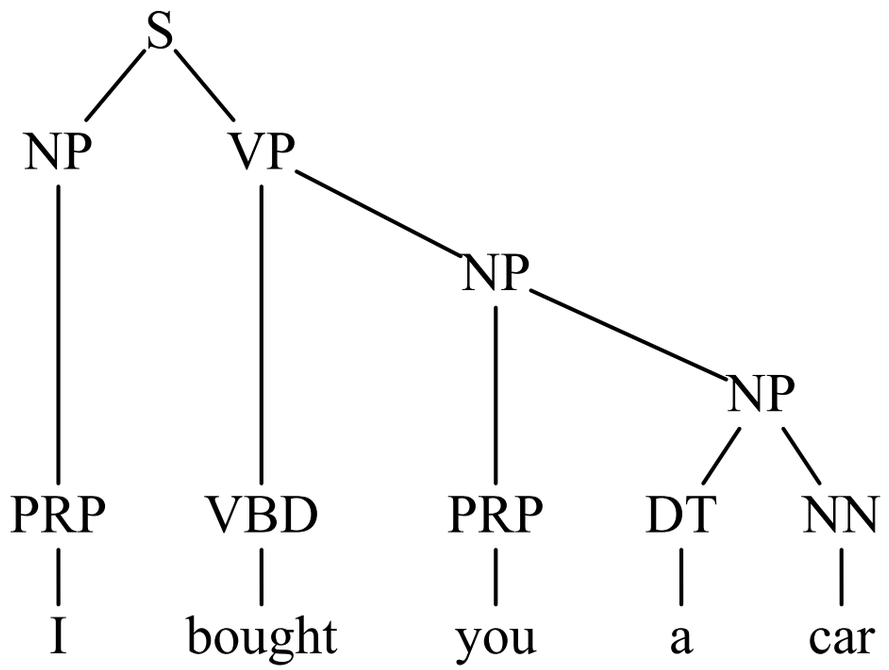


EMORY UNIVERSITY

Phrase Structure Rules

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

"I bought you a car"



S → NP VP ← NP is the subject of (the head of) VP

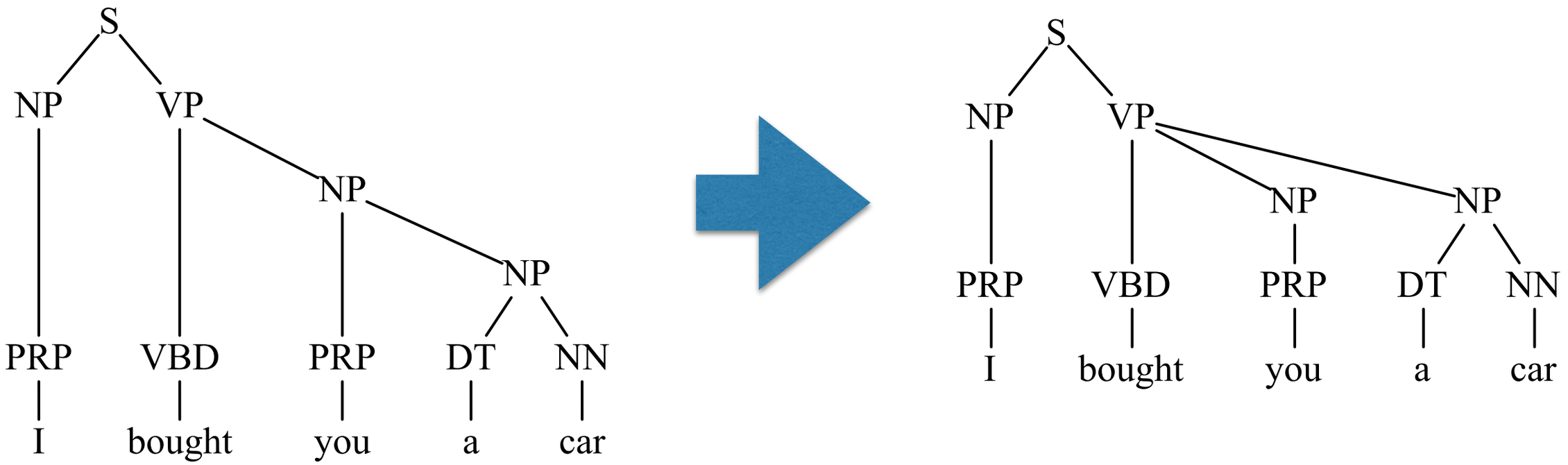
VP → VBD NP ← NP is the object of VBD?

NP → PRP NP ← ?

Phrase Structure Rules

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

"I bought you a car"



S → NP VP

VP → VBD NP

NP → PRP NP

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

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

2nd NP is the direct object of VBD

NP → DT NN

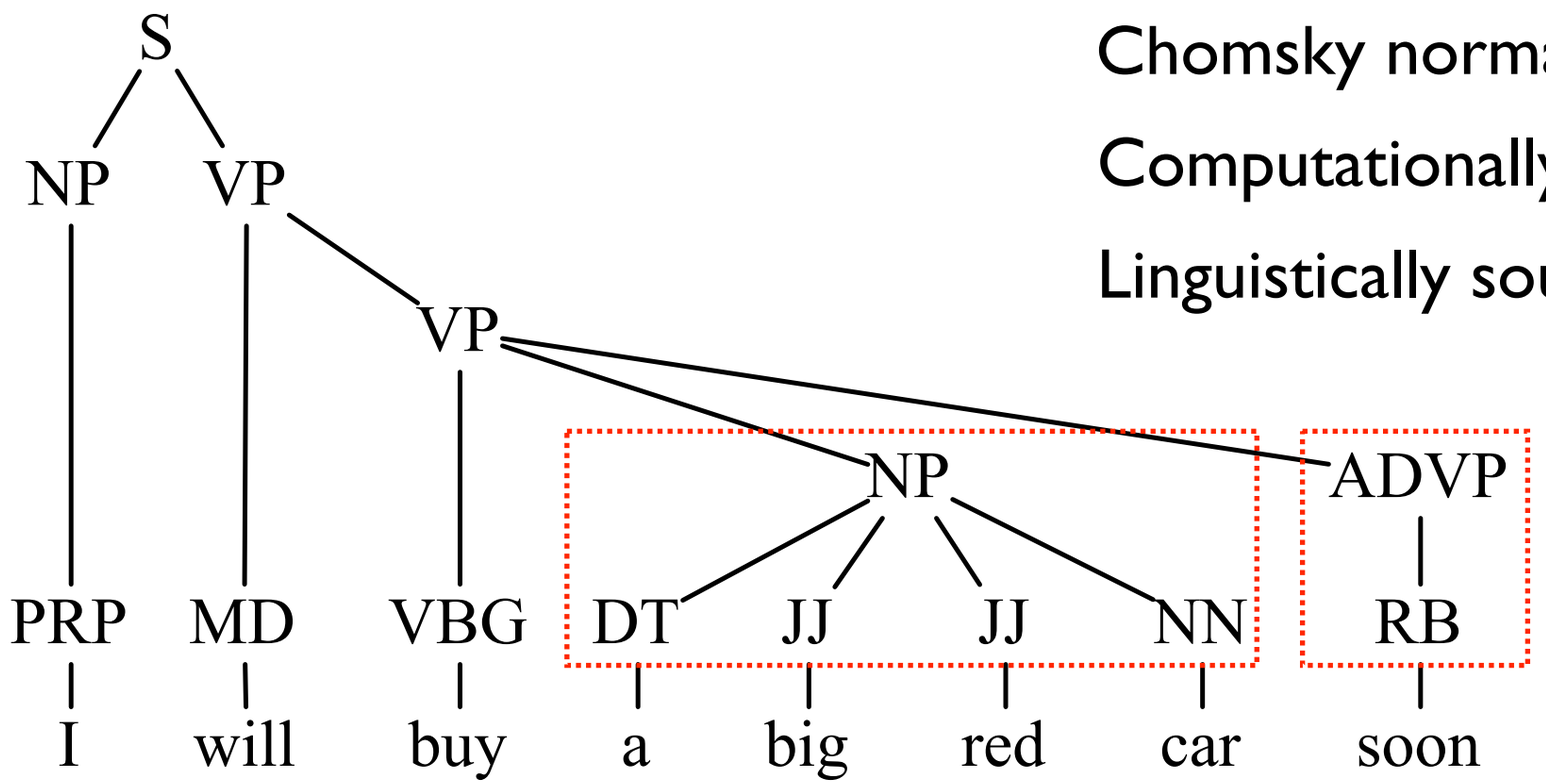
DT is the determiner of NN ↵

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"



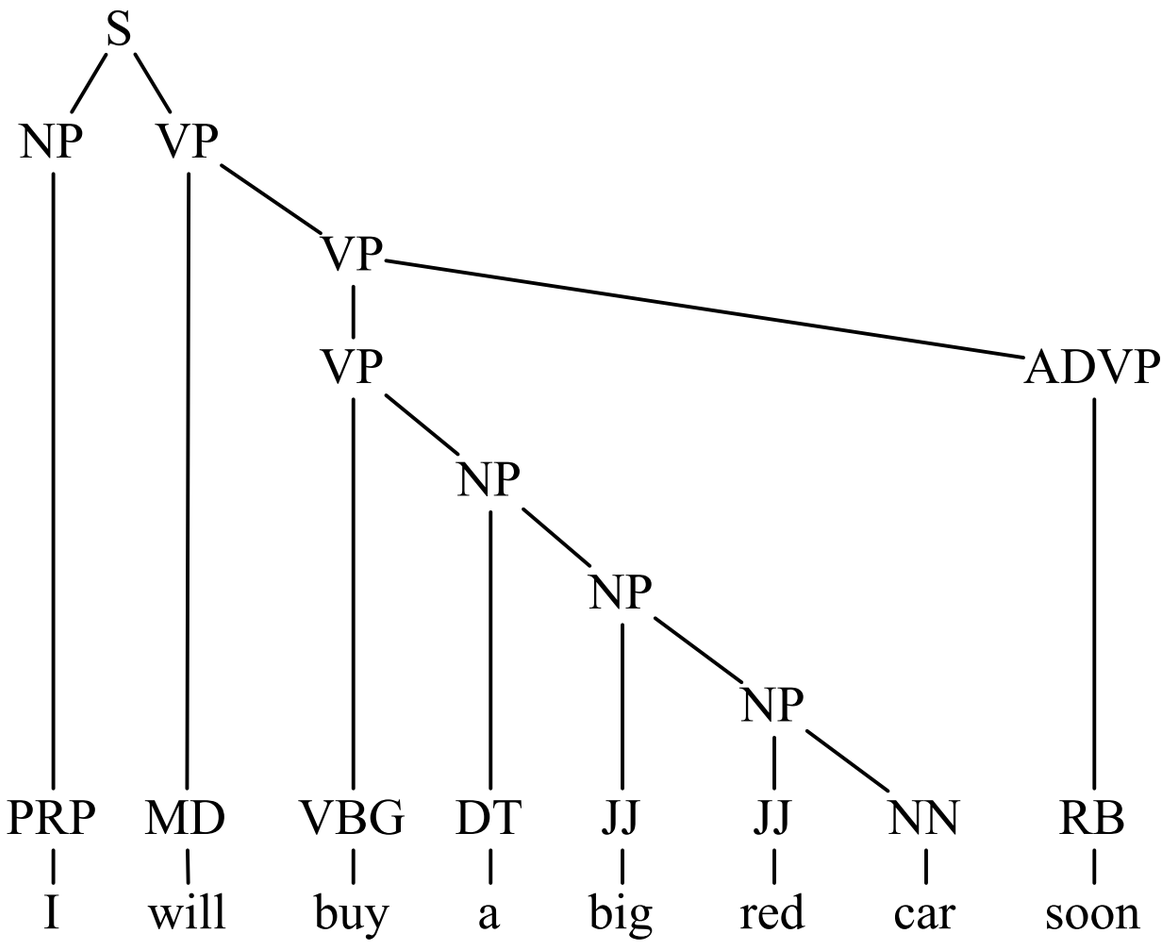
Chomsky normal form?

Computationally consistent?

Linguistically sound?

Chomsky Normal Form

"I will buy a big red car soon"



Non-lexicalized

```
S  → NP VP
VP → MD NP
VP → VP ADVP
VP → VBG NP
NP → DT NP|JJ NP|JJ NN
```

recursive

Lexicalized

```
NP → PRP → I
MD → will
VBG → buy
DT → a
JJ → big|red
NN → car
ADVP → RB → soon
```

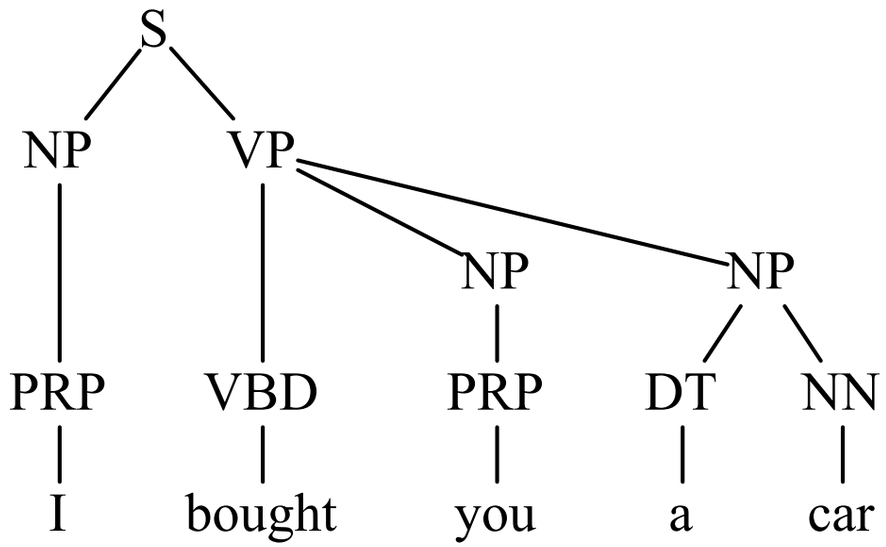
unary

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.



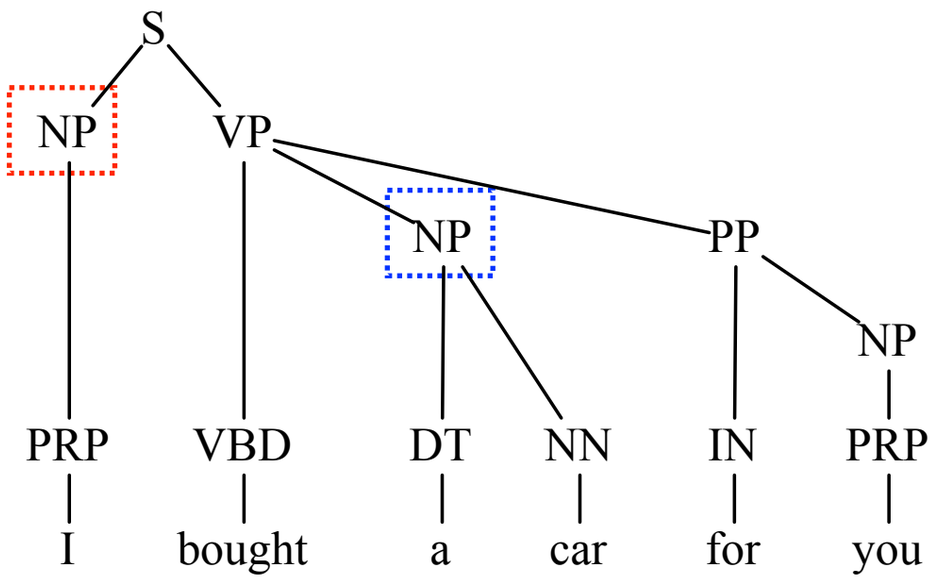

```
((S (NP (PRP I))
    (VP (VBD bought)
        (NP (PRP you))
        (NP (DT a)
            (NN car)))))
```

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

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

Passive Construction

"I bought a car for you"

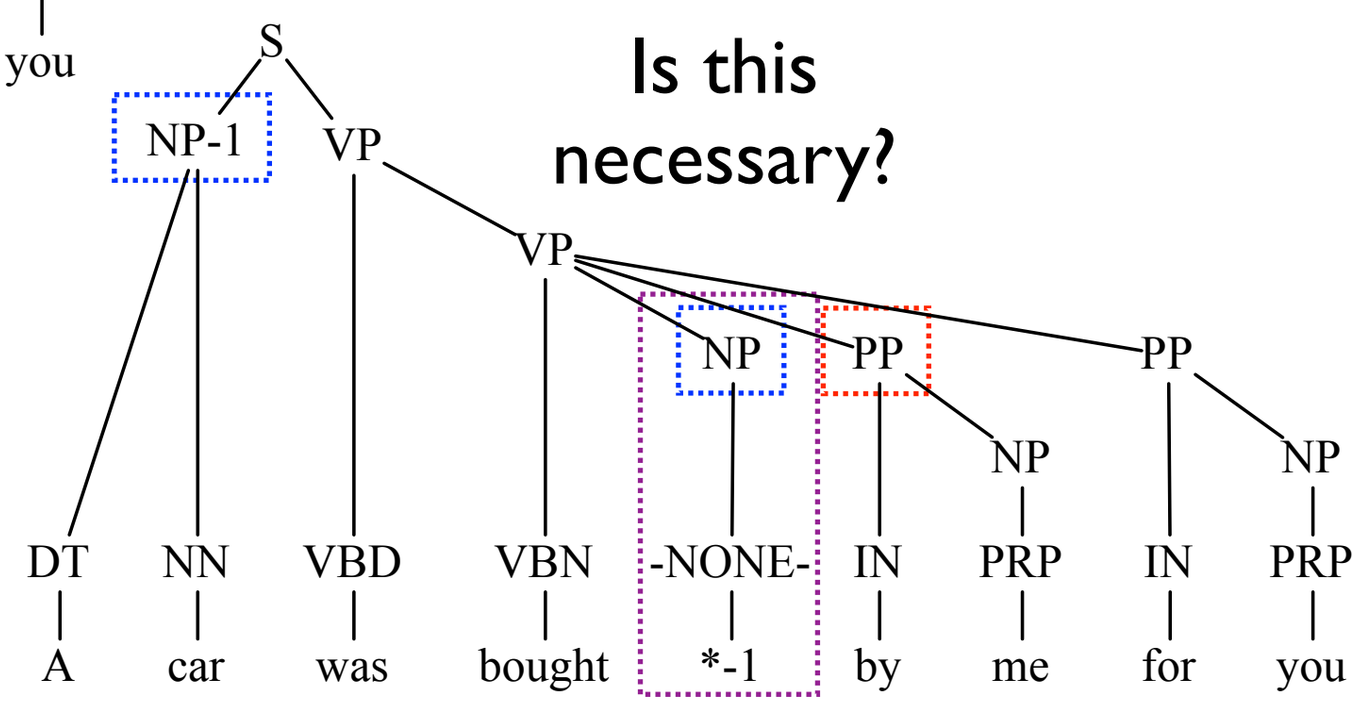


"A car was bought by me for you"

NP Movement

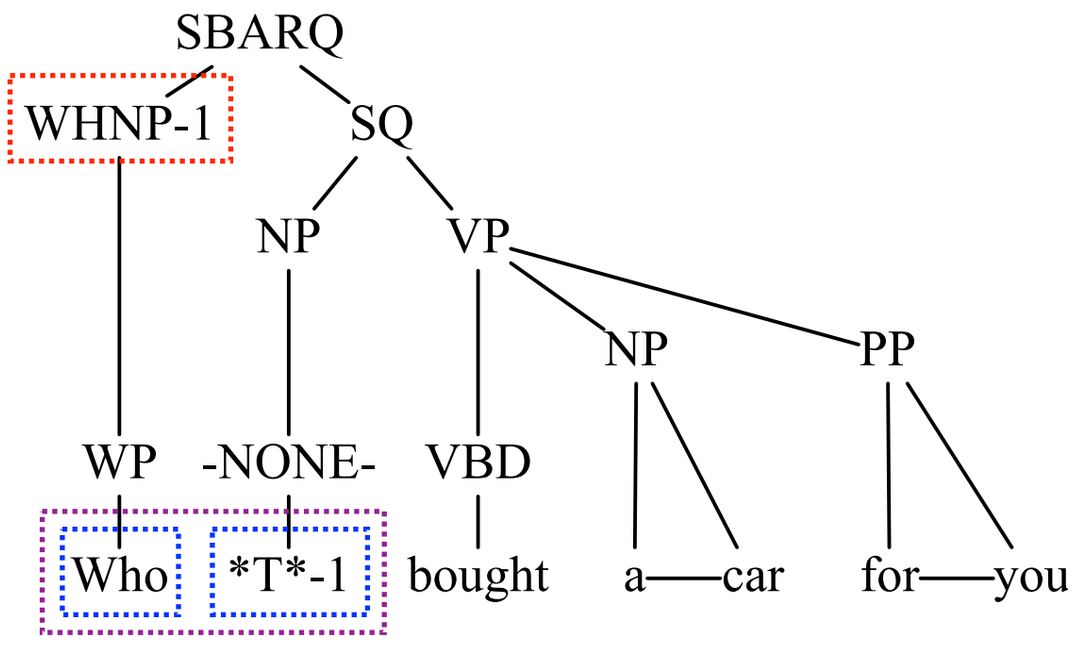
Did NP really move?

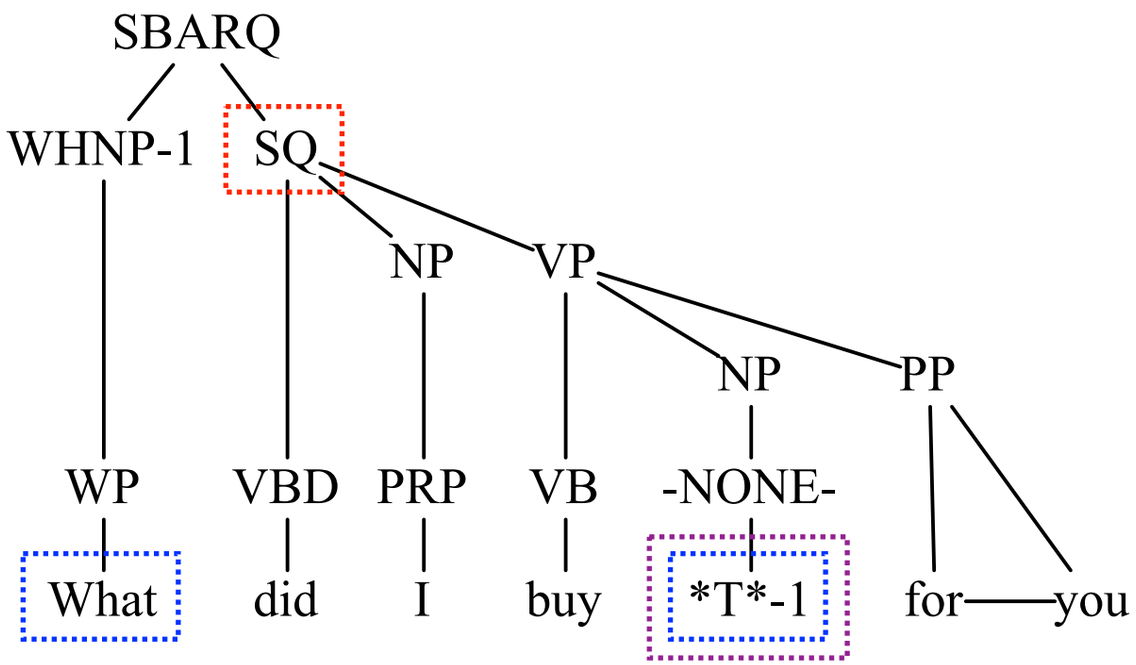
Is this necessary?



What about automatic parsing?

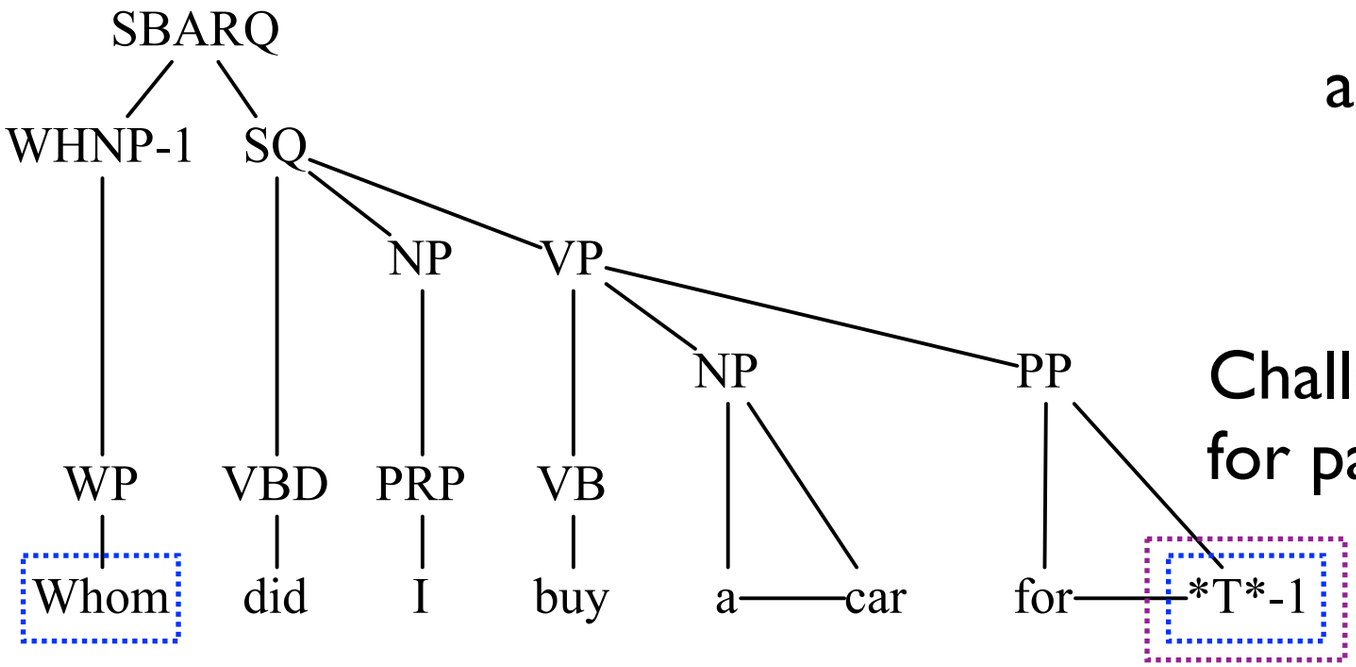
Wh-Questions





Is this necessary?

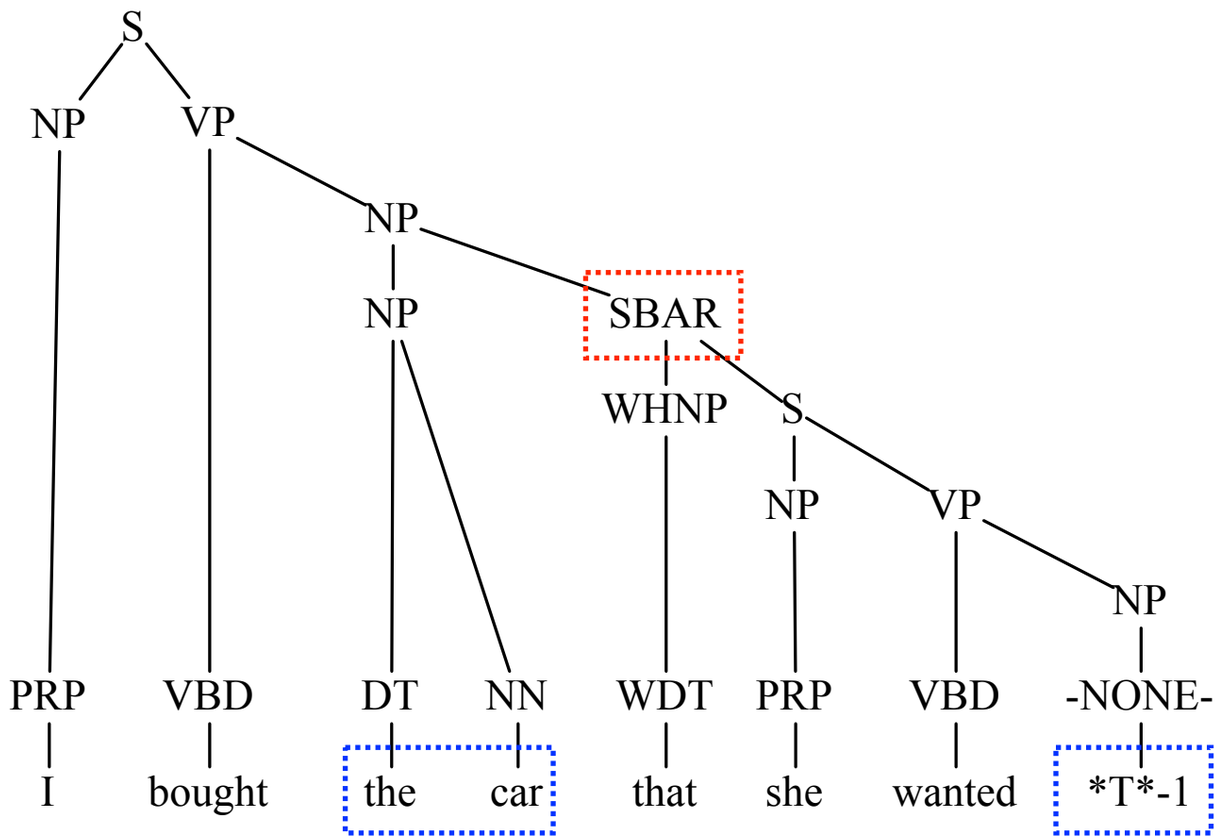
What about this?



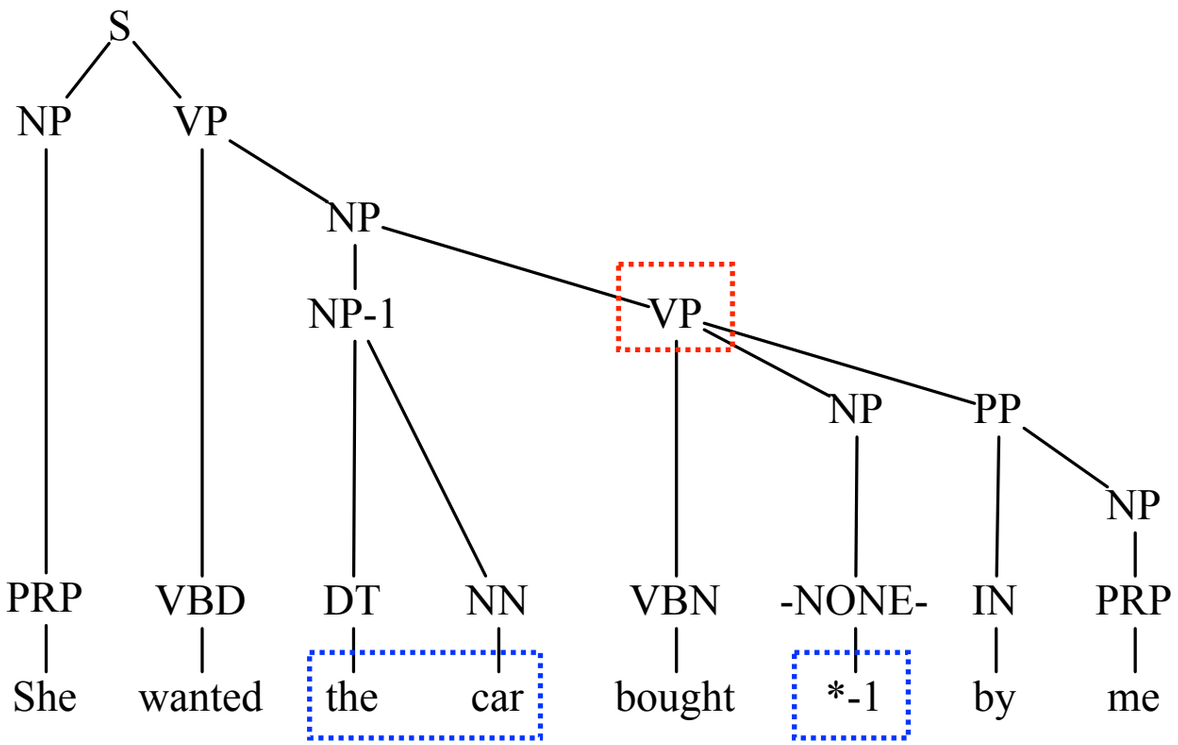
Challenging for parsing?

Relative Clause

I bought the car that she wanted

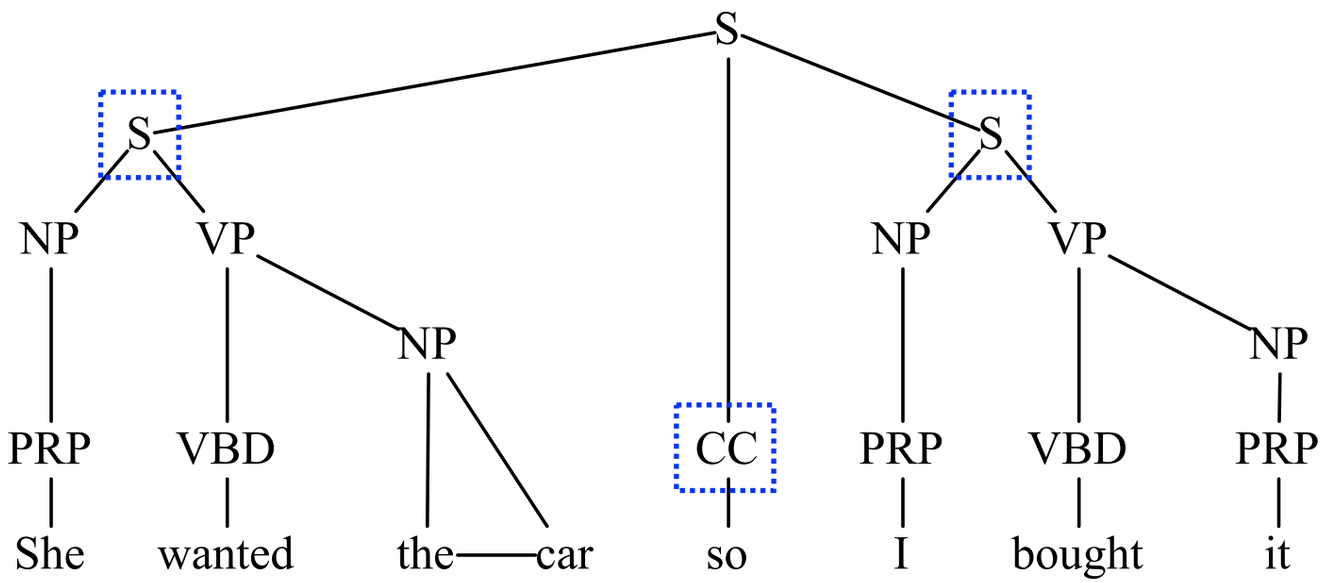


She wanted the car bought by me

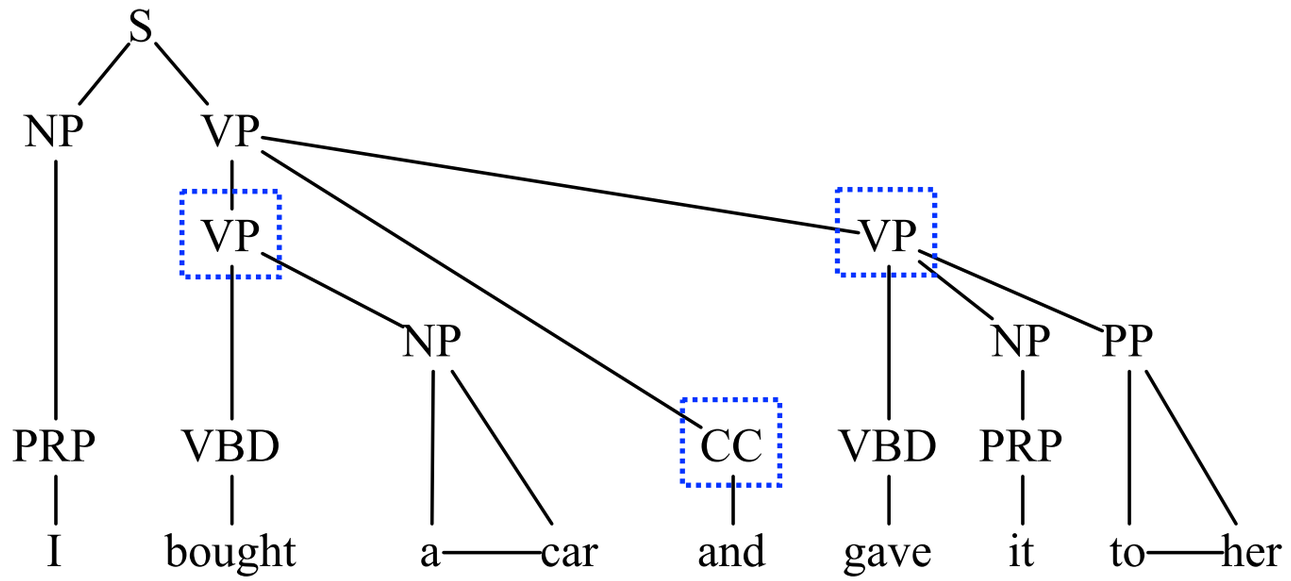


Coordination

She wanted the car so I bought it.



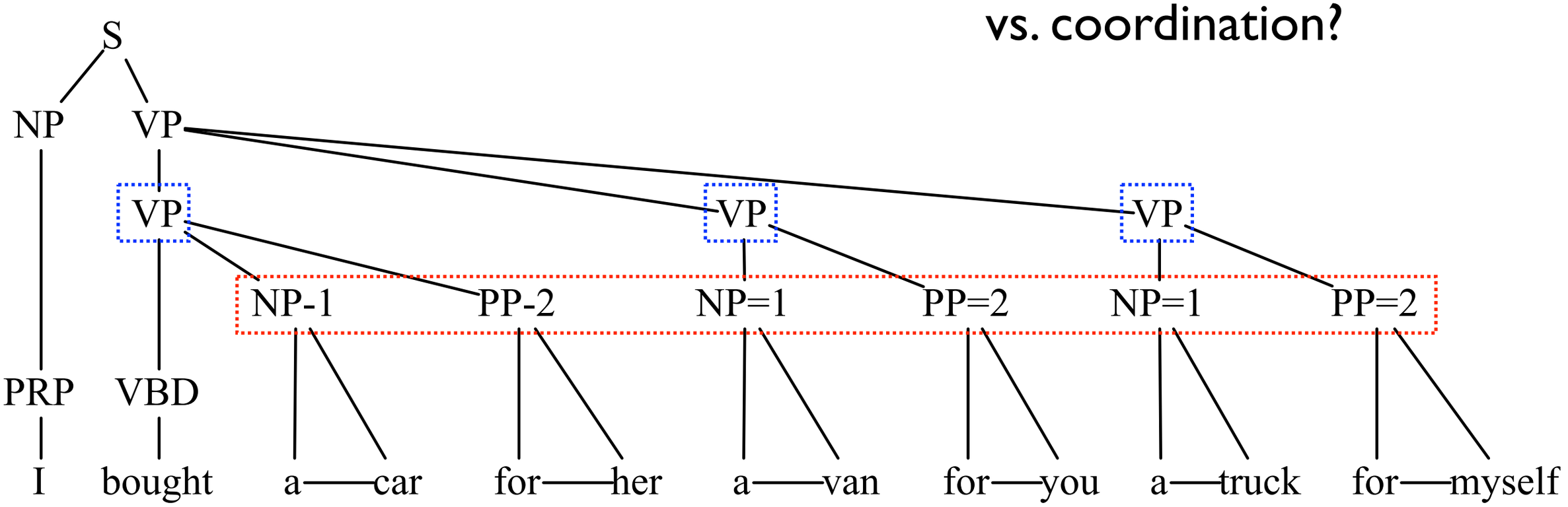
I bought a car and gave it to her.



Gapping Relation

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

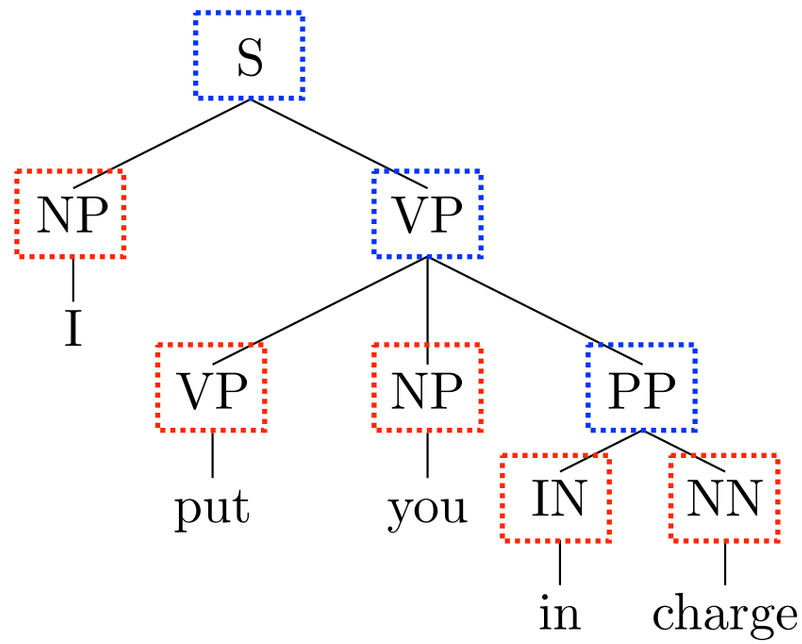
vs. coordination?



Bracketing Scores

Evaluate the automatically parsed trees.

Gold

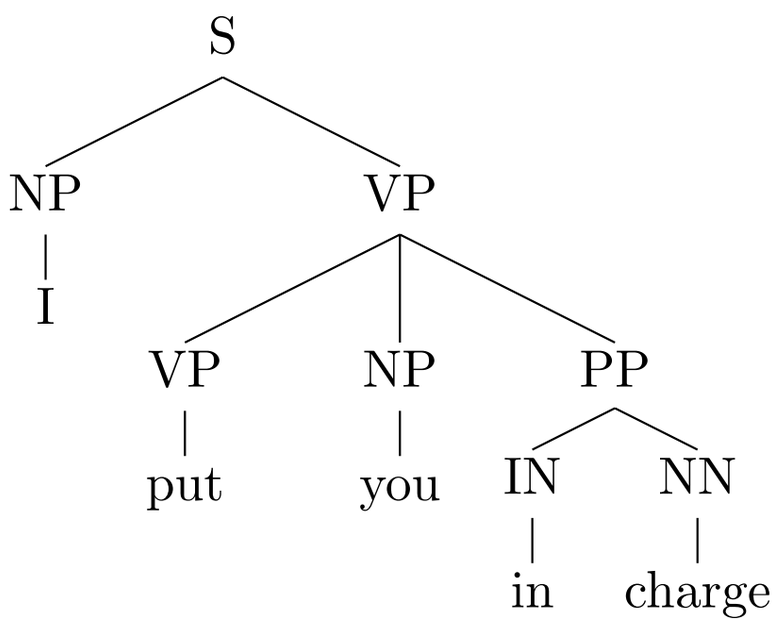


Gold: S(0:5), VP(1:5), PP(3:5), NP(0:1), VP(1,2), NP(2,3), IN(3,4), NN(4,5)

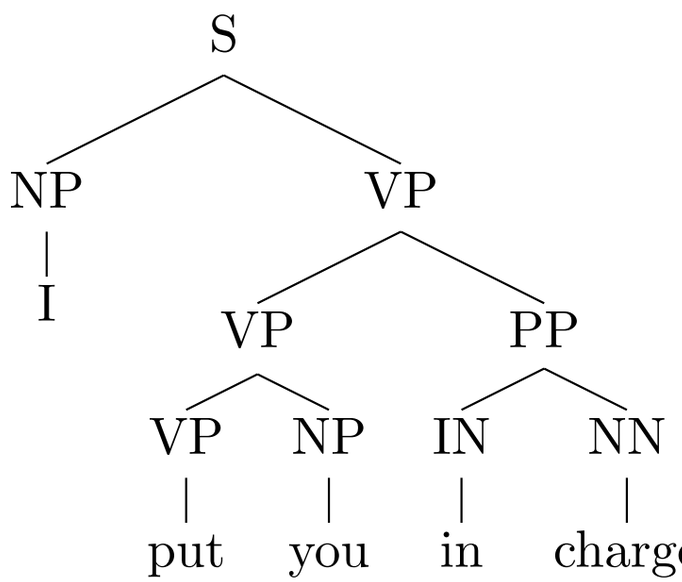
Bracketing Scores

Evaluate the automatically parsed trees.

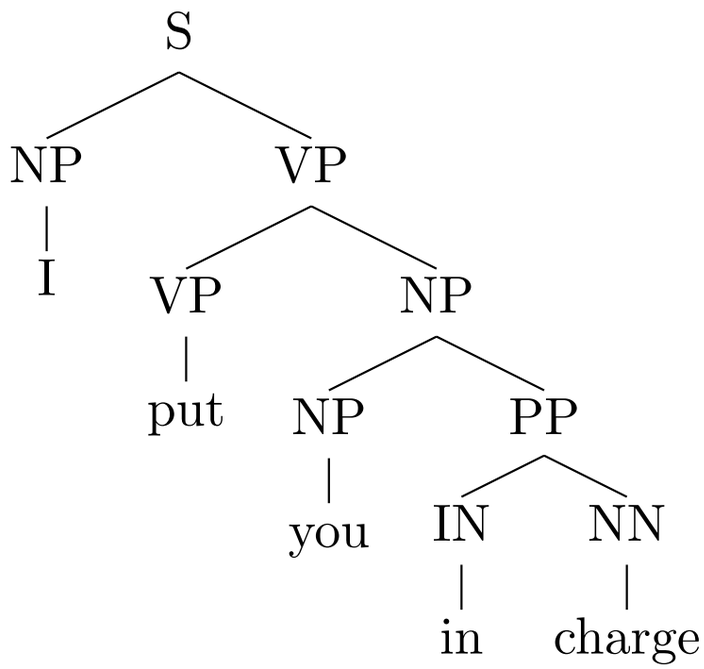
Gold



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)

EMORY UNIVERSITY

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
VP → VP PP 0.2
VP → VP NP 0.5
VP → put 0.3

NP → NP PP 0.2
NP → I 0.3
NP → you 0.3
NP → charge 0.2

PP → IN NP 1.0
IN → in 1.0

	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		