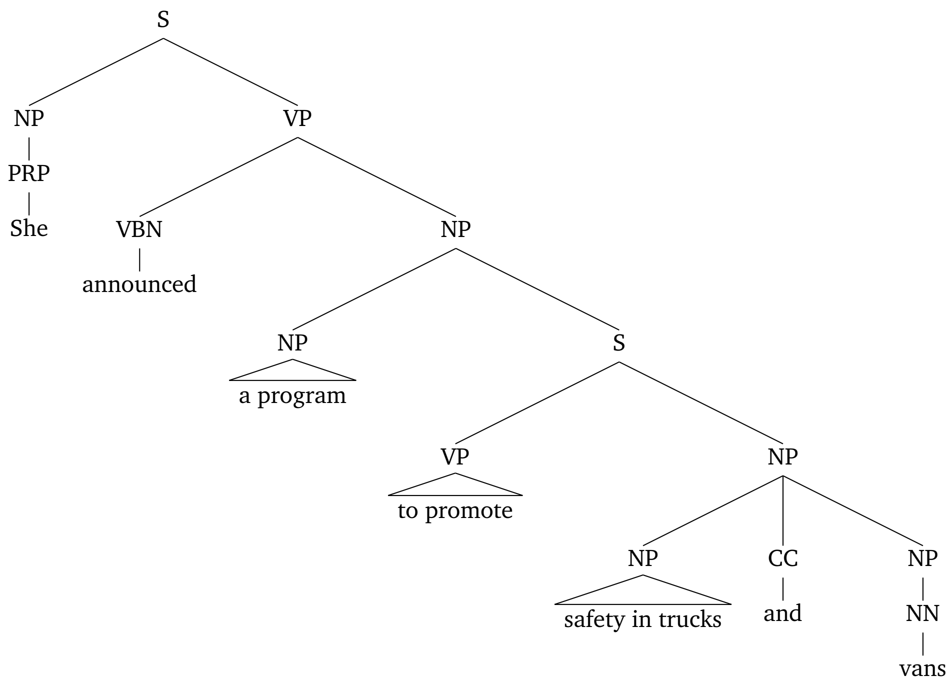
Mining Unstructured Data
Exercises on Constituent Parsing

Context Free Grammars

Exercise 1.

Consider the sentence *She announced a program to promote safety in trucks and vans*
and the following syntactic tree of one of its possible interpretations, in which the program promotes safety in trucks, and also promotes vans:

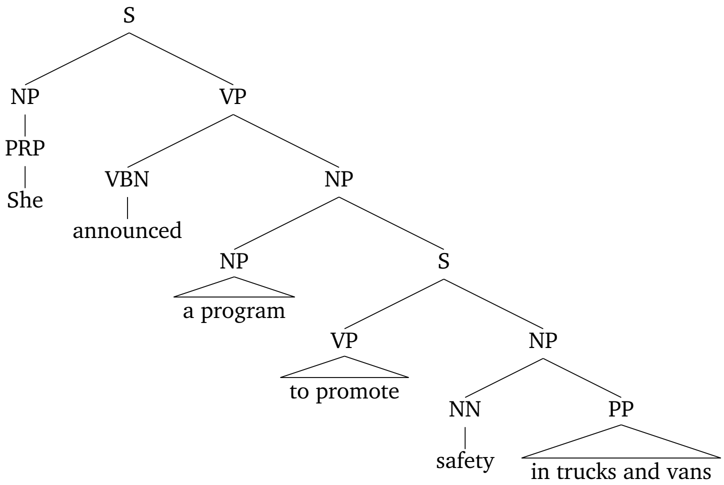


1. Draw the trees for at least three other interpretations for this sentence
2. Draw the trees for at least two interpretations for each of the following sentences
   - *The post office will hold out discounts and service concessions as incentives*
   - *They are hunting lions and tigers*
   - *Monty flies like mosquitoes*

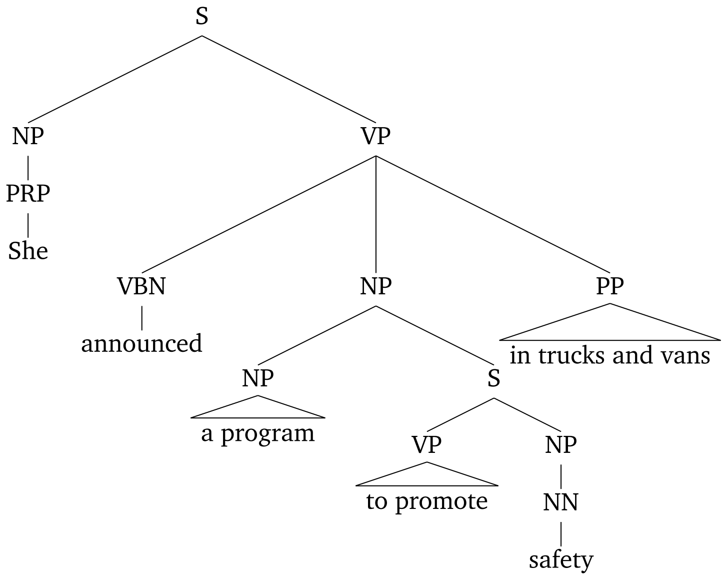
SOLUTION

1. Find three interpretations:

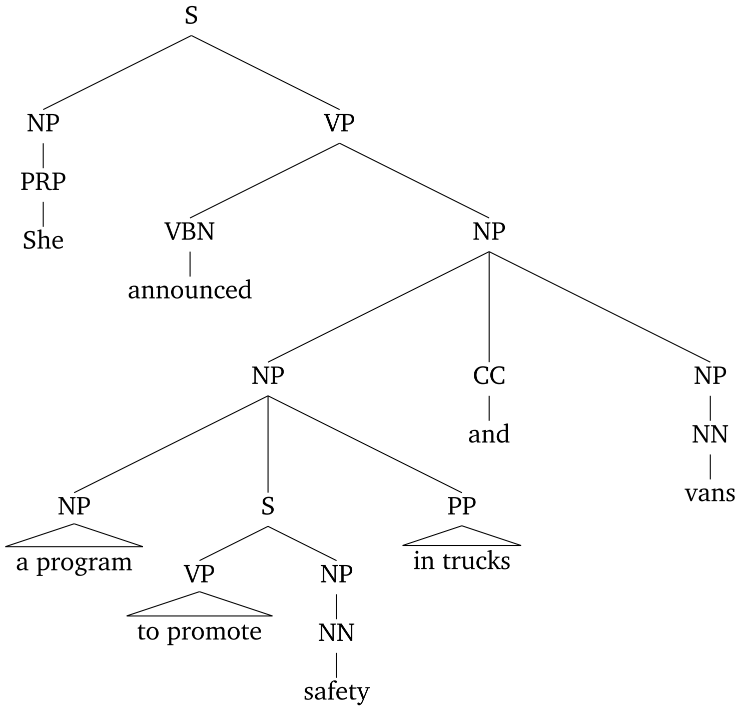
   Interpretation 1: The announced program promotes safety in both trucks and vans.

   Interpretation 2: The program is announced in trucks and vans.

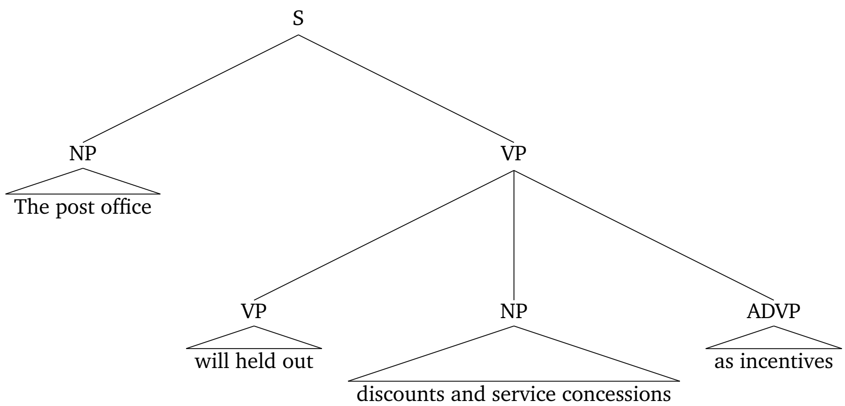
   

Interpretation 3: The announced program promotes safety in trucks. Vans are also announced.

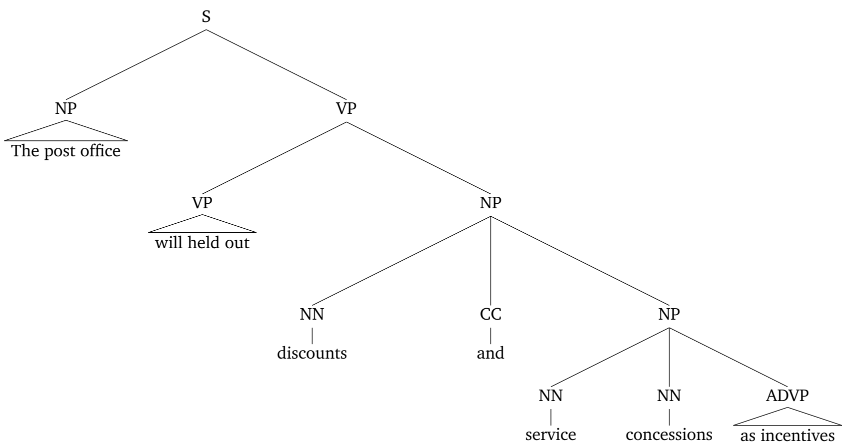


2. Find two interpretations for each sentence

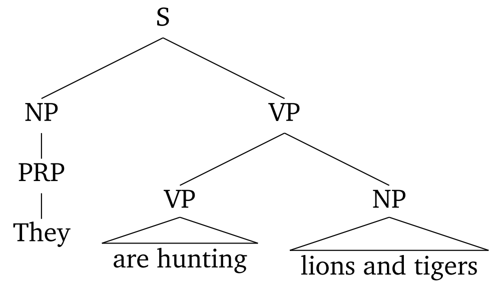
Sentence 1, Interpretation 1: Discounts and concessions are held out as incentives



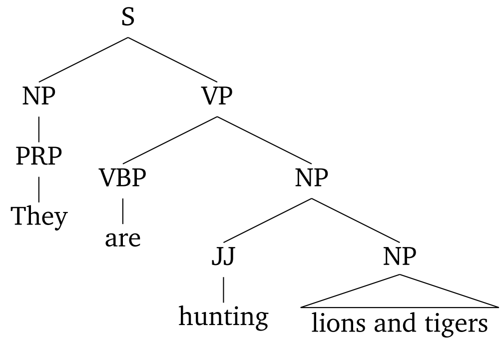
Sentence 1, Interpretation 2: Discounts and concessions are held out. Concessions look like incentives



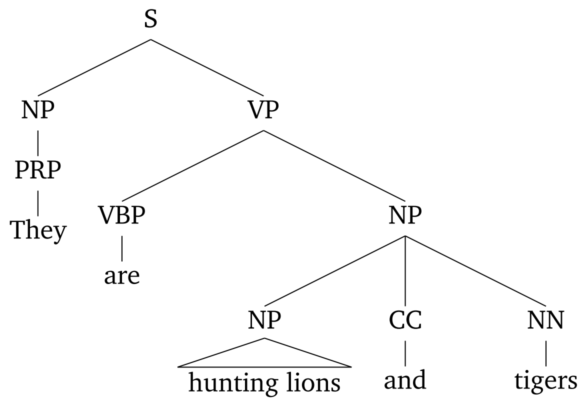
Sentence 2, Interpretation 1: Someone is hunting big felines.



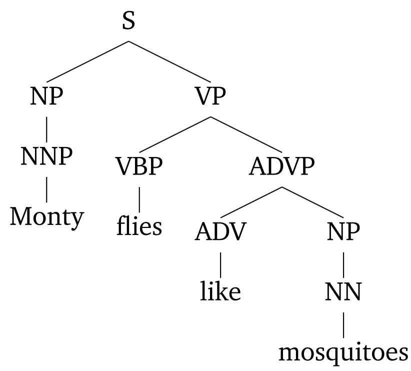
Sentence 2, Interpretation 2: Those animals are big felines that hunt.



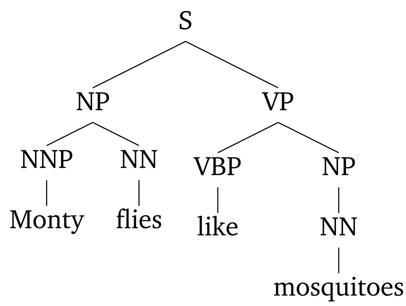
Sentence 2, Interpretation 3: Those animals are lions that hunt, and also tigers



Sentence 3, Interpretation 1: Someone named Monty moves through the air in the same way than mosquitoes do

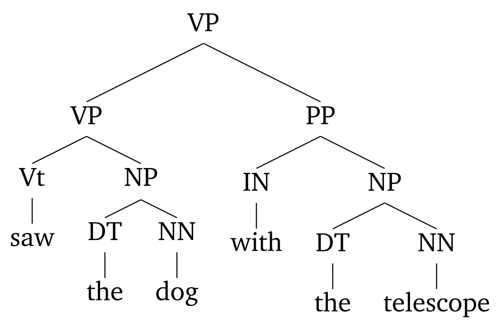


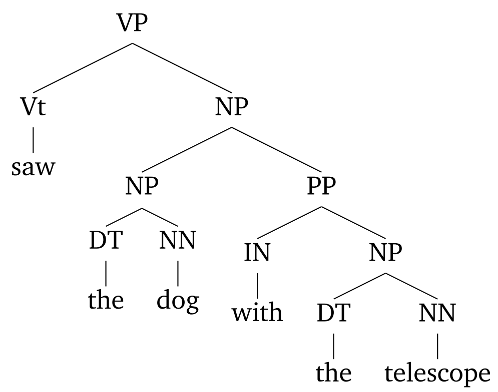
Sentence 3, Interpretation 2: Flies from a place named Monty are fond of mosquitoes



Exercise 2.

Say we have the phrase *saw the dog with the telescope* and we are given the gold parse tree (left) and the predicted parse tree (right):





What are the precision and recall of this predicted parse tree?

SOLUTION

The gold tree has 11 nonterminal nodes, each expanded using certain rule.

The predicted tree has 11 nonterminal nodes, each expanded using a rule that may be or not the right one. In particular, the predicted rules $\mathrm{VP} \to \mathrm{Vt}\ \mathrm{NP}$ and $\mathrm{NP} \to \mathrm{NP}\ \mathrm{PP}$ are not in the gold tree. All the other rules are found in the gold tree, so they are right.

Precision is the number of right rules divided by the number of predicted rules (i.e. size of the predicted tree). So in this case, it is 9/11 = 81.8%.

Recall is the number of right rules divided by the number of expected rules (i.e. size of the gold tree). So in this case, it is also 9/11 = 81.8% (since the size of both trees was the same, which shouldn't necessarily happen)

Exercise 3.

Consider the following CFG:

S $\rightarrow$ NP VP	DT $\rightarrow$ the	NN $\rightarrow$ park
NP $\rightarrow$ DT NN	NN $\rightarrow$ man	VB $\rightarrow$ saw
NP $\rightarrow$ NP PP	NN $\rightarrow$ dog	IN $\rightarrow$ with
PP $\rightarrow$ IN NP	NN $\rightarrow$ cat	IN $\rightarrow$ under
VP $\rightarrow$ VB NP		

1. How many parse trees are there under this grammar for the sentence *the man saw the dog in the park* ?
2. How many parse trees are there under this grammar for the sentence *the man saw the dog in the park with the cat* ?

SOLUTION

1. The sentence *the man saw the dog in the park* has a unique analysis in this grammar, where the dog is in the park. This is because the grammar does not have a rule such as e.g. VP $\rightarrow$ VP PP that allows the PP *in the park* to be attached to the verb *saw*.
2. The sentence *the man saw the dog in the park with the cat* has two analysis under this grammar: One where the cat is with the dog, and another where the cat is with the park.

Exercise 4.

Consider the following CFG:

S → NP VP	DT → the	NNS → cats
NP → DT NN	NN → man	NNS → parks
NP → DT NNS	NN → dog	VB → see
NP → NP PP	NN → cat	VB → sees
PP → IN NP	NN → park	IN → in
VP → VB NP	NNS → dogs	IN → with
VP → VP PP		

This grammar overgenerates incorrect English sentences, such as:

the dog see the cat
the dog in the park see the cat
the dog in the park see the cat in the park
the dogs sees the cat
the dogs in the park sees the cat
the dogs in the park sees the cat in the park

1. Modify the grammar so that all generated sentences respect third-person subject-verb agreement rules for English

SOLUTION

The rule joining the subject and the verb of the sentences is S → NP VP, so we need to alter this rule to allow only the combination of singular NP with third person VP, and plural NP with non-third person VP. For this, we need different rules for singular/plural NP and for third/non-third person VP.

Thus, the top rule S → NP VP needs to be replaced with:
S → NPs VPs
S → NPp VPp

All the NP rules must distinguish both kinds of noun phrases, replacing them with:
NPs → DT NN
NPp → DT NNS
NPs → NPs PP
NPp → NPp PP

Finally, the rules for verb phrases must also distinguish both cases:
VBs → sees
VBp → see
VPs → VBs NP
VPp → VBp NP
VPs → VPs PP
VPp → VPp PP

To avoid an explosion of rules, we can keep a generic NP to be used for noun phrases after the verb or inside a PP:
NP → NPs
NP → NPp

Probabilistic Context Free Grammars

Exercise 5.

Using the following PCFG in CNF:

$S \rightarrow NP\ VP$	1.0	$P \rightarrow$ with	1.0
$NP \rightarrow NP\ PP$	0.4	$V \rightarrow$ saw	1.0
$PP \rightarrow P\ NP$	1.0	$NP \rightarrow$ astronomers	0.1
$VP \rightarrow V\ NP$	0.7	$NP \rightarrow$ ears	0.18
$VP \rightarrow VP\ PP$	0.3	$NP \rightarrow$ saw	0.04
		$NP \rightarrow$ stars	0.18
		$NP \rightarrow$ telescopes	0.1

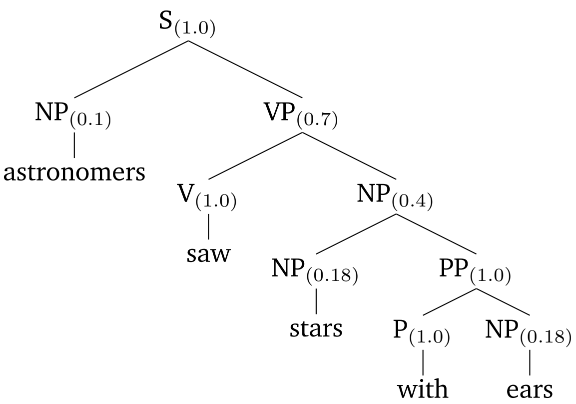
Work with the sentence: *astronomers saw stars with ears*

- How many correct parses are there for this sentence?
- Write them, along with their probabilities.

SOLUTION

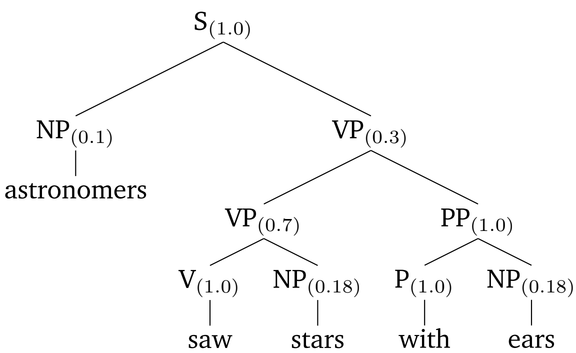
There are two possible parsers for this sentence according to the given grammar:

Option 1: (the stars had ears)



Probability: $1.0 \times 0.1 \times 0.7 \times 1.0 \times 0.4 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.00091$

Option 2: (Astronomers had their ears while watching the stars –or used ears to watch them)



Probability: $1.0 \times 0.1 \times 0.3 \times 0.7 \times 1.0 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.00068$

Exercise 6.

Given the following PCFG:

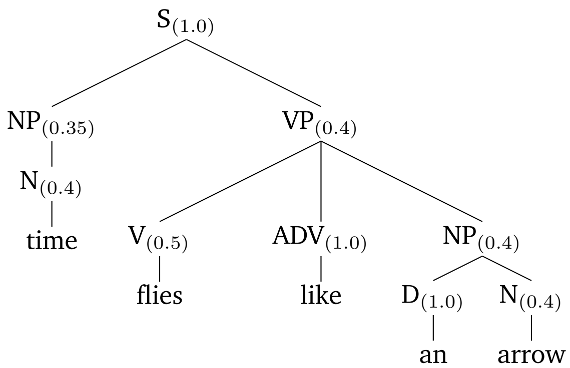
S → NP VP	1.0	N → time	0.4
NP → N N	0.25	N → flies	0.2
NP → D N	0.4	N → arrow	0.4
NP → N	0.35	D → an	1.0
VP → V NP	0.6	ADV → like	1.0
VP → V ADV NP	0.4	V → flies	0.5
		V → like	0.5

and the sentence *time flies like an arrow*

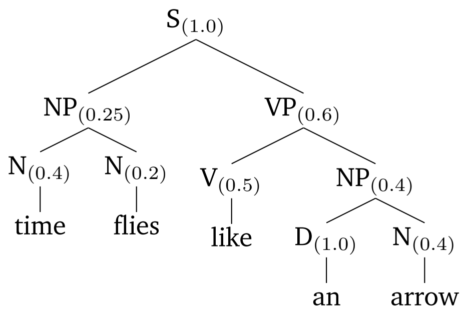
1. Write two parse trees that this grammar generates for this sentence
2. Compute the probability of each tree.
3. Convert the grammar to CNF and emulate the behaviour of the CKY algorithm on this sentence. Provide the final chart with all the information involved.

SOLUTION

1. Option 1: (time goes by so fast that reminds of an arrow)



Option 2: (Alien 4-dimensional flies are fond of arrows)



2. Option 1 probability: $1.0 \times 0.35 \times 0.4 \times 0.4 \times 0.5 \times 1.0 \times 0.4 \times 1.0 \times 0.4 = 0.00448$
   Option 2 Probability: $1.0 \times 0.25 \times 0.4 \times 0.2 \times 0.6 \times 0.5 \times 0.4 \times 1.0 \times 0.4 = 0.00096$

3. (a) Conversion of the grammar to CNF:

   Chomsky Normal Form requires that all rules have a right hand side with exactly two non-terminals, or exactly one terminal. Rules NP → N and VP → V ADV NP violate this condition, so we need to transform them.

   On the one hand, we shrink rule NP → N with those which produce N, so with N → time, N → flies y N → arrowand finally we include the following rules:

   NP → time $0.35 \times 0.4 = 0.14$
   NP → flies $0.35 \times 0.2 = 0.07$
   NP → arrow,¡ $0.35 \times 0.4 = 0.14$

   We remove the original rule not in CNF, but we keep the other 3 rules as they are used by other rules (ex: NP → N N).

   On the other hand, we split rule VP → V ADV NP to get two new rules in CNF:

   VP → V ADVP 0.4
   ADVP → ADV NP 1.0

   The resulting gramar in CNF is:

S $\to$ NP VP	1.0	N $\to$ time	0.4
NP $\to$ N N	0.25	N $\to$ flies	0.2
NP $\to$ D N	0.4	N $\to$ arrow	0.4
NP $\to$ time	0.14	D $\to$ an	1.0
NP $\to$ flies	0.07	ADV $\to$ like	1.0
NP $\to$ arrow	0.14	V $\to$ flies	0.5
VP $\to$ V NP	0.6	V $\to$ like	0.5
VP $\to$ V ADVP	0.4		
ADVP $\to$ ADV NP	1.0		

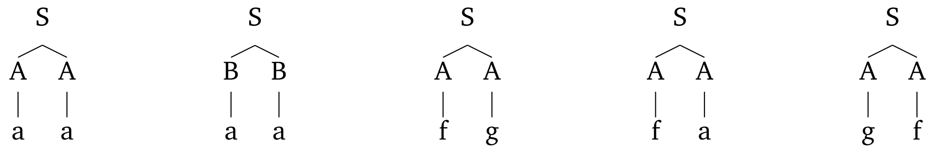
(b) CKY chart:

				15 0.00448 S $\to$ $N_{11}VP_{25}$ $(1.0 \times 0.14 \times 0.032)$ 0.00096 S $\to$ $NP_{12}VP_{35}$ $(1.0 \times 0.02 \times 0.048)$
			14	25 0.0033 S $\to$ $N_{11}VP_{25}$ $(1.0 \times 0.07 \times 0.048)$ 0.032 VP $\to$ $V_{22}ADVP_{35}$ $(0.4 \times 0.5 \times 0.16)$
		13	24	35 0.16 ADVP $\to$ $ADV_{33}NP_{45}$ $(1.0 \times 1.0 \times 0.16)$ 0.048 VP $\to$ $V_{33}NP_{45}$ $(0.6 \times 0.5 \times 0.16)$
	12 0.02 NP $\to$ $N_{11}N_{22}$ $(0.25 \times 0.4 \times 0.2)$	23	34	45 0.16 NP $\to$ $D_{44}N_{55}$ $(0.4 \times 1.0 \times 0.4)$
11 0.14 NP $\to$ time 0.4 N $\to$ time	22 0.07 NP $\to$ flies 0.2 N $\to$ flies 0.5 V $\to$ flies	33 1.0 ADV $\to$ like 0.5 V $\to$ like	44 1.0 D $\to$ an	55 0.4 N $\to$ arrow
time	flies	like	an	arrow

(Blue rule in cell 35 indicates the most likely subtree selected in that cell)

Exercise 7.

Consider that you have as a training corpus a treebank containing the following trees. Each tree was observed the number of times indicated below it.



75 10 325 8 428

1. What PCFG would one get from this treebank (using MLE)?
2. Given the obtained grammar:
   - What is the most likely parse of the string *a a*?
   - Is this a reasonable result? Discuss why.

SOLUTION

1. The given collection of training trees, taking into account the number of repetitions of each, will produce the following counts of rule applications:

S → A A	$1 \times 75 + 0 \times 10 + 1 \times 325 + 1 \times 8 + 1 \times 428 = 836$
S → B B	$0 \times 75 + 1 \times 10 + 0 \times 325 + 0 \times 8 + 0 \times 428 = 10$
S → anything	$1 \times 75 + 1 \times 10 + 1 \times 325 + 1 \times 8 + 1 \times 428 = 846$
B → a	$0 \times 75 + 2 \times 10 + 0 \times 325 + 0 \times 8 + 0 \times 428 = 20$
B → anything	$0 \times 75 + 2 \times 10 + 0 \times 325 + 0 \times 8 + 0 \times 428 = 20$
A → a	$2 \times 75 + 0 \times 10 + 0 \times 325 + 1 \times 8 + 0 \times 428 = 158$
A → f	$0 \times 75 + 0 \times 10 + 1 \times 325 + 1 \times 8 + 1 \times 428 = 761$
A → g	$0 \times 75 + 0 \times 10 + 1 \times 325 + 0 \times 8 + 1 \times 428 = 753$
A → anything	$2 \times 75 + 0 \times 10 + 2 \times 325 + 2 \times 8 + 2 \times 428 = 1672$

   Thus, the MLE probability for each rule would be:

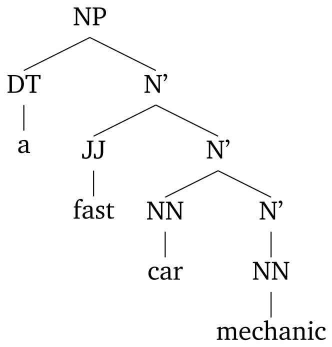
$P(\text{S} \to \text{A A})$	=	$P(\text{AA}\|\text{S})$	=	#(S → A A)/#(S → anything)	=	$836/846 = 0.988$
$P(\text{S} \to \text{B B})$	=	$P(\text{BB}\|\text{S})$	=	#(S → B B)/#(S → anything)	=	$10/846 = 0.012$
$P(\text{B} \to \text{a})$	=	$P(\text{a}\|\text{B})$	=	#(B → a)/#(B → anything)	=	$20/20 = 1.000$
$P(\text{A} \to \text{a})$	=	$P(\text{a}\|\text{A})$	=	#(A → a)/#(A → anything)	=	$158/1672 = 0.095$
$P(\text{A} \to \text{f})$	=	$P(\text{f}\|\text{A})$	=	#(A → f)/#(A → anything)	=	$761/1672 = 0.455$
$P(\text{A} \to \text{g})$	=	$P(\text{g}\|\text{A})$	=	#(A → g)/#(A → anything)	=	$753/1672 = 0.450$

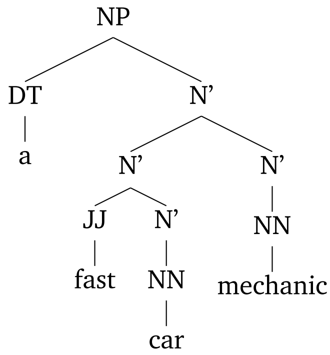
2. The input sequence *a a* can be derived by the obtained grammar in only two ways, which correspond to the first two trees in the training data.
   The first tree has probability: $P(\text{S} \to \text{A A}) \times P(\text{A} \to \text{a}) \times P(\text{A} \to \text{a}) = 0.988 \times 0.095 \times 0.095 = 0.009$
   The second tree has probability: $P(\text{S} \to \text{B B}) \times P(\text{B} \to \text{a}) \times P(\text{B} \to \text{a}) = 0.012 \times 1.000 \times 1.000 = 0.012$
   So, the most likely parse tree is the second one.

   The first tree appears 75 times in the training data, while the second one occurs only 10 times, so one would expect the probability for the former to be higher. However, we do not compute the tree probability by counting how many times the whole tree occurs, but we just approximate it by multiplying the individual rule probabilities. Thus, the fact that B produces a with much higher probability than A is biassing the result.
   This is due to the reduced amount of training data combined with the use of MLE: Since we do not perform any smoothing to consider the possibility of B producing other symbols, we are overestimating the probability of the rule $\text{B} \to \text{a}$.

Exercise 8.

Consider the two following parse trees:





Discuss whether the following statements are true or false and why:

1. The two parse trees receive the same probability under any PCFG
2. The first parse tree receives higher probability if $P(\text{N}' \rightarrow \text{NN N}') > P(\text{N}' \rightarrow \text{N}' \text{ N}')$
3. The first parse tree receives higher probability if $P(\text{N}' \rightarrow \text{NN N}') > P(\text{N}' \rightarrow \text{N}' \text{ N}') + P(\text{N}' \rightarrow \text{NN})$

SOLUTION

1. False, since the set of rules used in each tree differ: left tree uses rule $\text{N}' \rightarrow \text{NN N}'$ while the tree on the right uses the rule $\text{N}' \rightarrow \text{N}' \text{ N}'$. Also, the former uses rule $\text{N}' \rightarrow \text{NN}$ once but the latter uses it twice. So, assuming Q is the product of probabilities of rules shared by both trees, the first tree has probability $Q \times P(\text{N}' \rightarrow \text{NN N}')$, and the second has probability $Q \times P(\text{N}' \rightarrow \text{N}' \text{ N}') \times P(\text{N}' \rightarrow \text{NN})$, which are not necessarily equal.
2. True, since under this condition $Q \times P(\text{N}' \rightarrow \text{NN N}') > Q \times P(\text{N}' \rightarrow \text{N}' \text{ N}')$. If we multiply the right hand side term by $P(\text{N}' \rightarrow \text{NN})$ which is smaller than 1, the difference will increase.
3. False, since probabilities are multiplied, not added. The first tree would have higher probability if $P(\text{N}' \rightarrow \text{NN N}') > P(\text{N}' \rightarrow \text{N}' \text{ N}') \times P(\text{N}' \rightarrow \text{NN})$

Exercise 9.

Consider the following PCFG

S → NP VP	1.0	VP → sleeps	1.0
NP → DT NBAR	1.0	DT → the	1.0
NBAR → NN	0.7	NN → mechanic	0.1
NBAR → NBAR NBAR	0.3	NN → car	0.2
		NN → metal	0.7

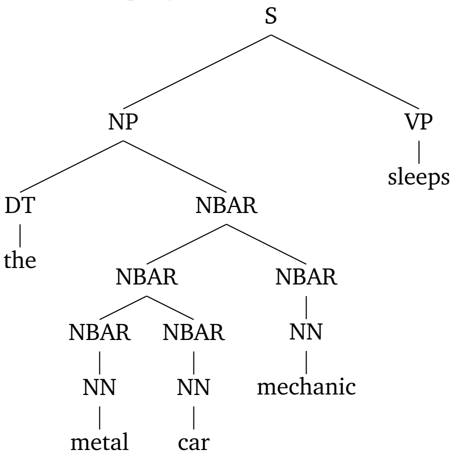
1. What is the parse tree with highest probability for the sentence *the metal car mechanic sleeps* ?

2. Modify the grammar above so that the sentece *the human language technology rules* has two interpretations (one about *human language* and another about *human technology*). Draw the trees for both interpretations, and point out which is the most likely.

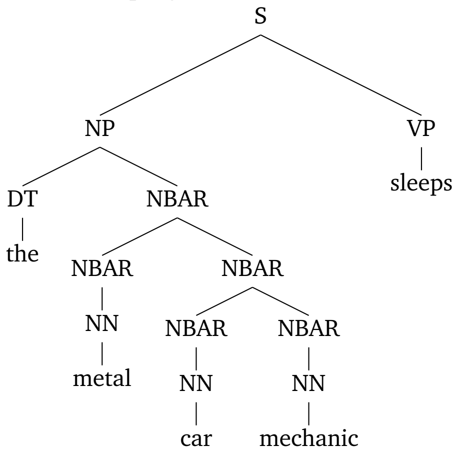
SOLUTION

1. This grammar produces two possible trees for the given sentence:

   Option 1: (the mechanic that works on metal cars is sleeping)

   Option 2: (the metal mechanic that works on cars is sleeping)

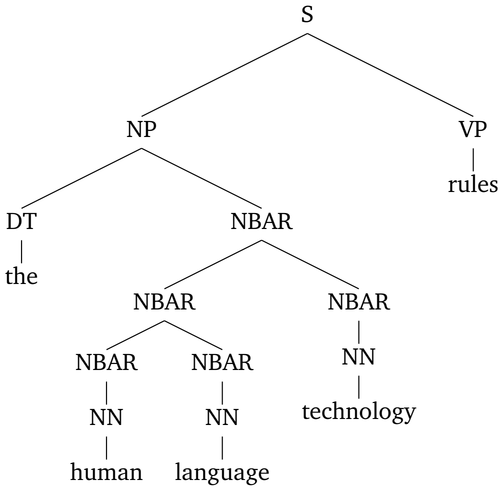
   Both trees use once the rules S → NP VP and NP → DT NBAR, twice the rule NBAR → NBAR NBAR, and three times the rule NBAR → NN. Also, the rules producing the leaves are also the same. The only difference is the order in which the rules are applied. Thus, the probabilities of both trees are identical, and there is not one single best tree, but two.

2. The grammar already allows the ambiguous structure for the sequence NN NN NN. We only need to add the new words to the grammar, and fix the probabilities. Rule probabilities are invented, since we do not have training data, but we need to ensure that the rules for the same non-terminal symbol add up to 1. New rules are highlighted in red. Redistributed probabilities are shown in blue.

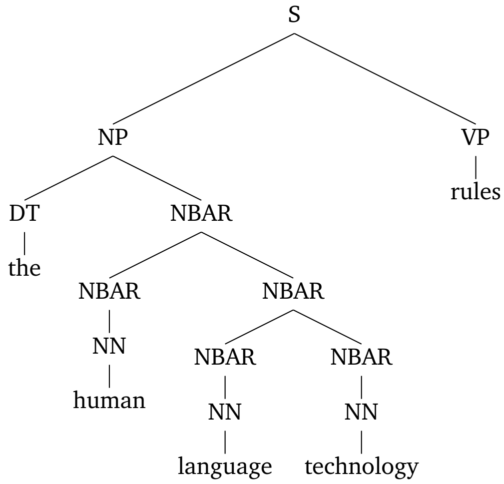
S → NP VP	1.0	VP → sleeps	0.5	VP → rules	0.5
NP → DT NBAR	1.0	DT → the	1.0		
NBAR → NN	0.7	NN → mechanic	0.1	NN → language	0.2
NBAR → NBAR NBAR	0.3	NN → car	0.2	NN → technology	0.2
		NN → metal	0.2	NN → human	0.1

   Possible trees for the new sentence are :

Option 1: (technology that deals with human language is cool)



Option 2: (human technology that deals with language is cool)



Again, since both trees have the same rules, they have exactly the same probability.