Questions for Flipped Classroom Session of COMS 4705 Week 4, Fall 2014. (Michael Collins)

Question 1 Consider a context-free grammar with the following rules (assume that *S* is the start symbol):

S → NP VP
NP → DT NN
NP → NP PP
PP → IN NP
VP → VB NP
DT → the
NN → man
NN → dog
NN → cat
NN → park
VB → saw
IN → in
IN → with
IN → under

Question 1a: How many parse trees are there under this grammar for the sentence *the man saw the dog in the park*?

Question 1b: How many parse trees for *the man saw the dog in the park with the cat*?

Question 1c: Define C_n to be the n'th *Catalan number*. This is defined (see Wikipedia for a full description) as

$$C_n = \frac{(2n)!}{(n+1)!n!}$$

It can be shown that C_n is the number of binary-branching trees with $n+1$ leaves.

Now consider a sentence that is grammatical under the above context-free grammar, and has exactly k prepositions following the verb, and 0 prepositions before the verb (a preposition is any word with the tag *IN*). How many parse trees will this sentence have?

Question 2 Consider the following grammar (assume that *S* is the start symbol):

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

The grammar has the problem that it generates ungrammatical strings, 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

How would you modify the grammar so that all of the sentences that it generates are grammatical?

Hint: the new grammar should contain the following rules, which distinguish between singular and plural NPs:

NP-s $\rightarrow$ DT NN
NP-p $\rightarrow$ DT NNS

Question 3 Consider the following grammar (assume that *S* is the start symbol):

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

An infinite number of sentences are grammatical under this grammar, for example

the man saw the dog
the man saw the dog with the telescope
the man saw the dog with the telescope under the dog
...

We will define the *language* of the context-free grammar to be the set of all grammatical sentences under the grammar.

Now consider a hidden Markov model (HMM), which defines a distribution $p(x_1 \dots x_n, y_1 \dots y_{n+1})$ over sentences $x_1 \dots x_n$ paired with tag sequences $y_1 \dots y_{n+1}$. For a given HMM, define the *language* of the HMM to be the set of sentences $x_1 \dots x_n$ such that

$$\max_{y_1 \cdots y_{n+1}} p(x_1 \dots x_n, y_1 \dots y_{n+1}) > 0$$

I.e., there must be at least one tag sequence $y_1 \dots y_{n+1}$ that gives a probability $p(x_1 \dots x_n, y_1 \dots y_{n+1}) > 0$.

Question: Write down a bigram HMM such that the language of the HMM is the same as the language of the context-free grammar given above.

Question 4 Consider the following grammar (assume that *S* is the start symbol):

S $\rightarrow$ NP VP
NP $\rightarrow$ DT NN
VP $\rightarrow$ Vt NP
VP $\rightarrow$ Vdt NP NP
NP $\rightarrow$ NP RELC
RELC $\rightarrow$ WH S-GAP
S-GAP $\rightarrow$ VP
S-GAP $\rightarrow$ NP VP-GAP
VP-GAP $\rightarrow$ Vt
VP-GAP $\rightarrow$ Vdt NP
WH $\rightarrow$ that
DT $\rightarrow$ the
NN $\rightarrow$ man
NN $\rightarrow$ dog
NN $\rightarrow$ cat
NN $\rightarrow$ park
Vt $\rightarrow$ saw
Vdt $\rightarrow$ gave

Question 4a: Show parse trees for the sentences *the man that saw the dog saw the cat* and *the man that the cat saw saw the dog*

Question 4b: Give a sentence that is grammatical under the above grammar, and which has the trigram *saw saw saw*. Show the parse tree for the sentence.

Question 4c: Assume that we add the following rules to the grammar, so that the sentence *the man said the cat saw the dog* can be parsed correctly:

VP $\rightarrow$ V3 S
V3 $\rightarrow$ said

What rules would you add to the grammar so that the following sentence can be parsed?:

the dog that the man said the cat saw saw the park