Homework 6	CMSC 723
Parsing	
Due: October 28, 2013	

1 Context-Free Grammars (30 points)

Consider the following context free grammar:

Intermediate Rules	Terminal Rules
S → NP VP	D → a | the
NP → (D) NOM	V → slept | disappeared | loved | gave | relied
VP → V (NP) (NP)	N → cat | dog | roof | man
NOM → N	P → in | on | with
NOM → NOM PP	CONJ → and | or
VP → VP PP	
PP → P NP	
X → X CONJ X	

We designate the start state to be S, and X in the last rule stands for any part of speech (thus the last rule allows NP → NP CONJ NP but not NP→ NP CONJ VP).

1. Give a grammatical English sentence that has one an only one valid interpretation in this grammar. Draw the tree corresponding to it.

2. Give a grammatical English sentence that has more than one valid interpretation in this grammar. Draw two trees, and discuss whether the meaning in English is ambiguous.

3. Give a sentence that has a valid interpretation in this grammar but is not grammatical.

4. Give a grammatical English sentence using only these terminals that does not have a valid interpretation in this grammar. Why doesn't it have a valid interpretation?

5. Can you give an upper bound for the number of unique sequences of terminals this grammar admits? If not, why not?

2 Penn Treebank (30 points)

In this section, let's consider the 10% sample of the Penn Treebank included in NLTK.

1. What is the minimum and maximum height of sentences in the Treebank? Give an example tree for both. How does the depth correlate with sentence length?

2. In the following sentence:

```
( (S
    (NP-SBJ-1
      (NP (NNP Rudolph) (NNP Agnew) )
      (, ,)
      (UCP
        (ADJP
          (NP (CD 55) (NNS years) )
          (JJ old) )
        (CC and)
        (NP
          (NP (JJ former) (NN chairman) )
          (PP (IN of)
            (NP (NNP Consolidated) (NNP Gold) (NNP Fields) (NNP PLC) ))))
      (, ,) )
    (VP (VBD was)
      (VP (VBN named)
        (S
          (NP-SBJ (-NONE- *-1) )
          (NP-PRD
            (NP (DT a) (JJ nonexecutive) (NN director) )
            (PP (IN of)
              (NP (DT this) (JJ British) (JJ industrial) (NN conglomerate) ))))))
    (. .) ))
```

   what does "(-NONE- *-1)" mean? Explain both in terms of this sentence specifically and what it means in general linguistically.

3. In order to build a PCFG, we need to estimate the probability of all of the production rules. What is the probability of each of the following production rules? This is a rare case where we actually want the MLE estimate, as zeros rule out possible parse trees. How many unique non-terminal productions are there? How about terminal productions?

Production Rule	Probability
NP → DT JJ NNS	
VP → VB NP	
ADJP → JJ PP	
VP → MD VP	
NN → 'stock'	
IN → 'like'	
IN → 'on'	
IN → 'with'	
IN → 'about'	
IN → 'over'	

3 Probabilistic Context-Free Grammars (40 points)

Consider the following sentence and the following PCFG:

time flies like an arrow

Intermediate Rules		Terminal Rules	
S → NP VP	0.8	N → time	0.5
S → VP	0.2	N → flies	0.3
VP → V NP	0.5	N → arrow	0.2
VP → V PP	0.3	V → time	0.3
VP → VP PP	0.2	V → flies	0.3
NP → Det N	0.3	V → like	0.4
NP → N	0.3	P → like	1.0
NP → N N	0.2	Det → an	1.0
NP → NP PP	0.2		
PP → P NP	1.0		

1. What are the four possible parse trees for this sentence?
2. Draw a CKY chart parse for the sentence and determine the most likely parse of this sentence. Show your work.