NAME:

CS-5340/6340, Solutions for Midterm Exam, Fall 2013

1. (8 pts) Consider the following grammar G:

   S $\rightarrow$ NP VP
   NP $\rightarrow$ noun
   VP $\rightarrow$ verb ADJP
   ADJP $\rightarrow$ adj
   noun $\rightarrow$ NLP
   verb $\rightarrow$ is
   adj $\rightarrow$ cool

   The questions below do <u>not</u> involve chart parsing! Simply show the order in which rules would be applied using a general search strategy.

   (a) Show a bottom-up parsing derivation of the sentence *"NLP is cool"* using grammar G.

   One possible solution is:
   NLP is cool
   noun is cool
   NP is cool
   NP verb cool
   NP verb adj
   NP verb ADJP
   NP VP
   S

(b) Show a top-down parsing derivation of the sentence *"NLP is cool"* using grammar G.

One possible solution is:
S
NP VP
noun VP
NLP VP
NLP verb ADJP
NLP is ADJP
NLP is adj
NLP is cool

2. (10 pts) For each sentence below, indicate whether the verb phrase is in *active* or *passive* voice.

- On the way to the park, the man began to whistle.

  Active

- George does not have any brothers or sisters.

  Active

- They have donated thousands of dollars to charity.

  Active

- Mary has been waiting for a kidney transplant.

  Active

- John was thanked for his many contributions to the company.

  Passive

- Many rare dolphins have been found in the Arctic Ocean.

  Passive

- The boys were hiking in the woods for several hours.

  Active

- In Boston, four women were publicly praised for their heroism.

  Passive

- The book about Utah was written by a famous historian.

  Passive

- The girl will be going to Harvard next year.

  Active

3. (10 pts) Circle every base noun phrase (NP) in each sentence below and label it with its syntactic role in the sentence. Choose from the following syntactic roles: **SUBJECT, DIRECT OBJECT (DOBJ), INDIRECT OBJECT (IOBJ)** or **PP** (if the NP is part of a prepositional phrase).

   - Theresa bought a new computer with her credit card.

     (Theresa)/SUBJECT bought (a new computer)/DOBJ with (her credit card)/PP.

   - The man was buried by an avalanche with his dog.

     (The man)/SUBJECT was buried by (an avalanche)/PP with (his dog)/PP.

   - The clerk opened the door for the injured man.

     (The clerk)/SUBJECT opened (the door)/DOBJ for (the injured man)/PP.

   - John sent Mary a package.

     (John)/SUBJECT sent (Mary)/IOBJ (a package)/DOBJ.

   - The woman entered the store with her 3 children.

     (The woman)/SUBJECT entered (the store)/DOBJ with (her 3 children)/PP.

   - Tom hoped that his house would not be robbed.

     (Tom)/SUBJECT hoped that (his house)/SUBJECT would not be robbed.

   - The cars were shipped from Boston to Utah.

     (The cars)/SUBJECT were shipped from (Boston)/PP to (Utah)/PP.

4. (20 pts) Circle every base noun phrase (NP) in each sentence below and label it with the thematic role that best represents the role that it plays with respect to the verb in the sentence. Choose from the following thematic roles: **AGENT, BENEFICIARY, CO-AGENT, CO-THEME, DESTINATION, EXPERIENCER, INSTRUMENT, ORIGIN, RECIPIENT, THEME**.

   - Theresa bought a new computer with her credit card.

     (Theresa)/AGENT bought (a new computer)/THEME with (her credit card)/INSTRUMENT.

   - The man was buried by an avalanche with his dog.

     (The man)/THEME was buried by (an avalanche)/AGENT with (his dog)/CO-THEME.

   - The clerk opened the door for the injured man.

     (The clerk)/AGENT opened (the door)/THEME for (the injured man)/BENEFICIARY.

   - John sent Mary a package.

     (John)/AGENT sent (Mary)/RECIPIENT (a package)/THEME.

   - The woman entered the store with her 3 children.

     (The woman)/AGENT entered (the store)/THEME with (her 3 children)/CO-AGENT.

   - Tom hoped that his house would not be robbed.

     (Tom)/EXPERIENCER hoped that (his house)/THEME would not be robbed.

   - The cars were shipped from Boston to Utah.

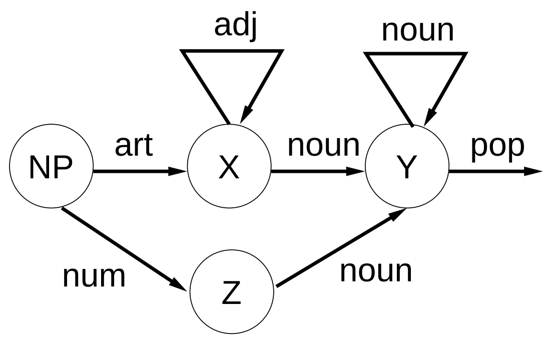
     (The cars)/THEME were shipped from (Boston)/ORIGIN to (Utah)/DESTINATION.

5. (12 pts) For each grammar below, draw a recursive transition network (RTN) that recognizes exactly the same language as the grammar.

(a) The following grammar with initial state NP defines noun phrase structures:

NP $\rightarrow$ art NP1
NP $\rightarrow$ num NP2
NP1 $\rightarrow$ adj NP1
NP1 $\rightarrow$ NP2
NP2 $\rightarrow$ noun
NP2 $\rightarrow$ noun NP2

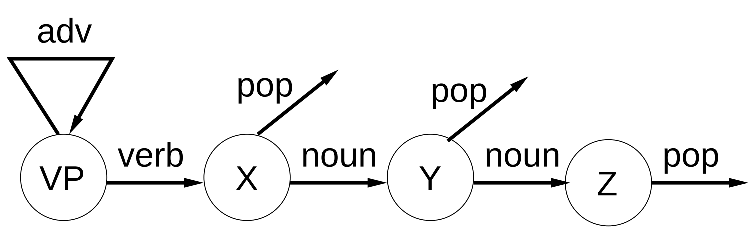
One possible solution is:



(b) The following grammar with initial state VP defines verb phrase structures:

VP $\rightarrow$ A1 VP1
VP $\rightarrow$ VP1
A1 $\rightarrow$ adv
A1 $\rightarrow$ adv A1
VP1 $\rightarrow$ verb
VP1 $\rightarrow$ verb NPS
NPS $\rightarrow$ noun
NPS $\rightarrow$ noun noun

One possible solution is:



6. (8 pts) Consider the sentence:

Natural language processing is a fun topic to study.

Assume the correct parse is:

(**S** (**NP** (adj natural) (noun language) (noun processing))
  (**VP** (verb is)
    (**NP** (art a) (adj fun) (noun topic)
      (**VP** (inf to) (verb study)))))

Suppose your parser generates the following parse:

(**S** (**NP** (adj natural) (noun language))
  (**VP** (verb processing))
  (**VP** (verb is)
    (**NP** (art a) (adj fun) (noun topic))
    (**PP** (prep to) (**NP** (noun study)))))

Compute the recall and precision accuracy of your parser for this sentence. Do NOT count part-of-speech tags when measuring accuracy. Only count the higher-level syntactic constituents (**S**, **NP**, **VP**, **PP** constituents), which are shown in **boldface**. **Leave your answers in fractional form!**

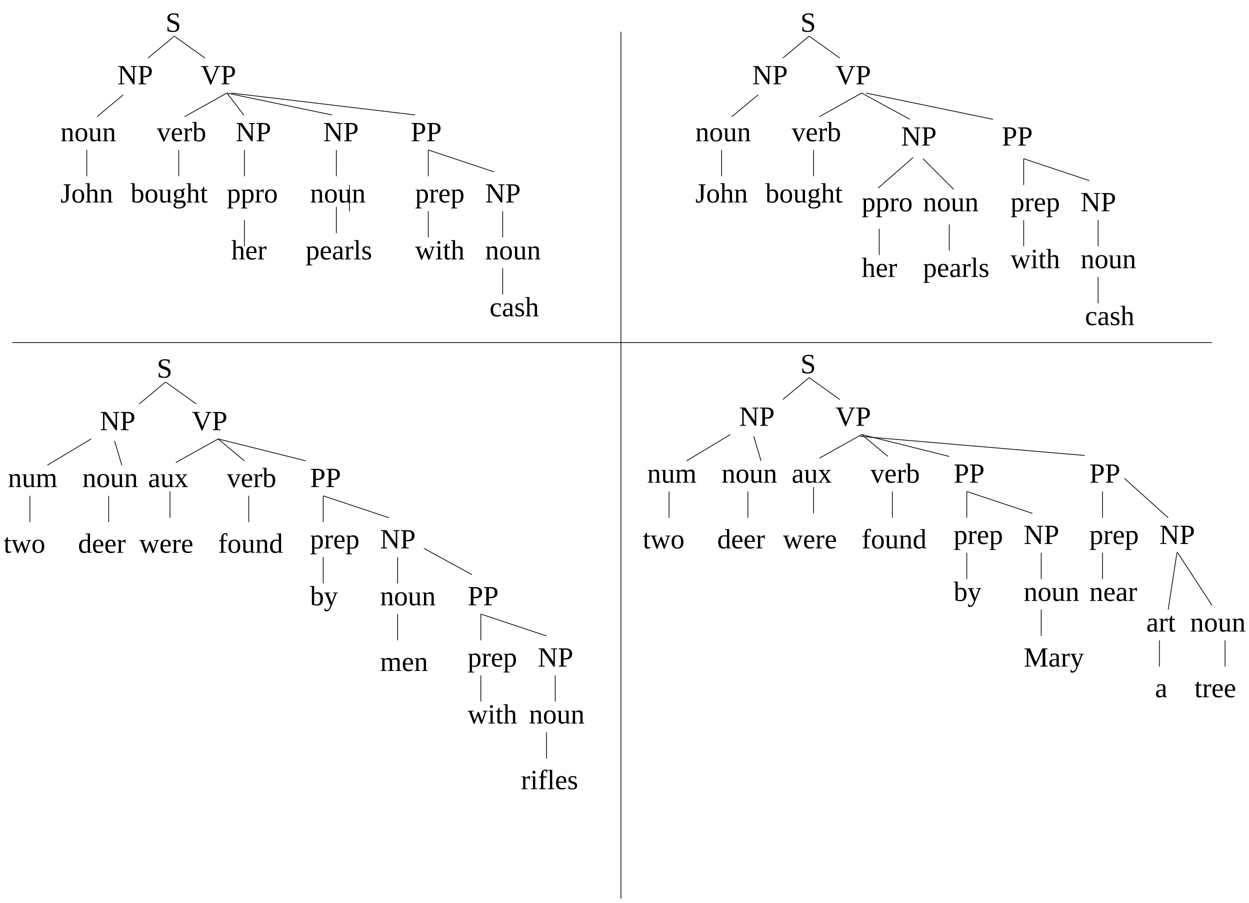
To compute recall we note that there are 5 target constituents in the correct parse, of which 2 were correctly identified by your parser. The correct constituents are the S and first VP. So Recall = 2/5.

To compute precision we note that your parser produced 7 constituents, of which 2 were correct. So Precision = 2/7.

7. (12 pts) Assume that the parse trees below are a (tiny!) manually parsed text corpus. Generate a probabilistic context-free grammar from these parse trees. Fill in the provided table with all of the grammar rules that appear in these parse trees and the probability that would be assigned to each rule (based on all 4 parse trees together).

   NOTE: You should not include grammar rules that map from part-of-speech tags to words. That is, you can ignore rules that have a specific word on the right-hand side, such as "noun → John".

   Leave your probabilities in fractional form!



Grammar Rule	**Probability**
S $\rightarrow$ NP VP	4/4
NP $\rightarrow$ noun	7/13
VP $\rightarrow$ verb NP NP PP	1/4
NP $\rightarrow$ ppro	1/13
PP $\rightarrow$ prep NP	6/6
VP $\rightarrow$ verb NP PP	1/4
NP $\rightarrow$ ppro noun	1/13
NP $\rightarrow$ num noun	2/13
VP $\rightarrow$ aux verb PP	1/4
NP $\rightarrow$ noun PP	1/13
VP $\rightarrow$ aux verb PP PP	1/4
NP $\rightarrow$ art noun	1/13

8. (20 pts) Answer *true* or *false* for each question below. You do not need to explain your answer.

   (a) The Viterbi algorithm can be used to find the 2 best part-of-speech tag sequences for a sentence.

   false

   (b) Top-down chart parsing would be better for shallow parsing than bottom-up chart parsing.

   false

   (c) RTNs use a *bottom-up* parsing strategy.

   false

   (d) P(X | Y) should be the same as P(X) if X and Y are independent of each other.

   true

   (e) A morphological analyzer can automatically produce features that are useful for parsing.

   true

   (f) During chart parsing, two different grammars that recognize exactly the same language will always result in the same number of entries on the chart.

   false

   (g) When training and evaluating machine learning systems, cross-validation is especially useful if the amount of annotated training data is small.

   true

   (h) A text classifier that only produces answers when it is highly confident that its answers are correct usually has high recall but low precision.

   false

   (i) Context-free grammars can recognize sentence structures that regular grammars cannot recognize.

   true

   (j) A lexical N-gram model uses part-of-speech tags.

   false

IMPORTANT: Question #9 is for CS-6340 students ONLY!

9. (12 pts) The table below shows 5 words with their frequency values based on an imaginary corpus. Assume that these 5 words are the only words in your vocabulary. First, give the *unsmoothed probability* estimate for each word (unigram) based on the provided frequency counts. Second, use Add-Two Smoothing to compute the *smoothed* frequency count and *smoothed* probability estimate for each word. Add-Two Smoothing works exactly the same way as Add-One Smoothing, except that a value of 2 is added during smoothing instead of 1.

   You should not need a calculator! Just plug in the appropriate numbers and show the resulting equation or fraction.

Word	**Unsmoothed Frequency**	**Unsmoothed Probability**	**Smoothed Frequency**	**Smoothed Probability**
table	500	500/1000	502 * (1000/1010)	502/1010
pen	400	400/1000	402 * (1000/1010)	402/1010
pencil	98	98/1000	100 * (1000/1010)	100/1010
eraser	2	2/1000	4 * (1000/1010)	4/1010
chalk	0	0/1000	2 * (1000/1010)	2/1010