

NAME:

CS-5340/6340, Midterm Exam, Fall 2013

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

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

The questions below do not 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.

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

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.
- George does not have any brothers or sisters.
- They have donated thousands of dollars to charity.
- Mary has been waiting for a kidney transplant.
- John was thanked for his many contributions to the company.
- Many rare dolphins have been found in the Arctic Ocean.
- The boys were hiking in the woods for several hours.
- In Boston, four women were publicly praised for their heroism.
- The book about Utah was written by a famous historian.
- The girl will be going to Harvard next year.

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.
- The man was buried by an avalanche with his dog.
- The clerk opened the door for the injured man.
- John sent Mary a package.
- The woman entered the store with her 3 children.
- Tom hoped that his house would not be robbed.
- The cars were shipped from Boston to Utah.

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.

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

- The clerk opened the door for the injured man.

- John sent Mary a package.

- The woman entered the store with her 3 children.

- Tom hoped that his house would not be robbed.

- The cars were shipped from Boston to Utah.

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

(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

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:

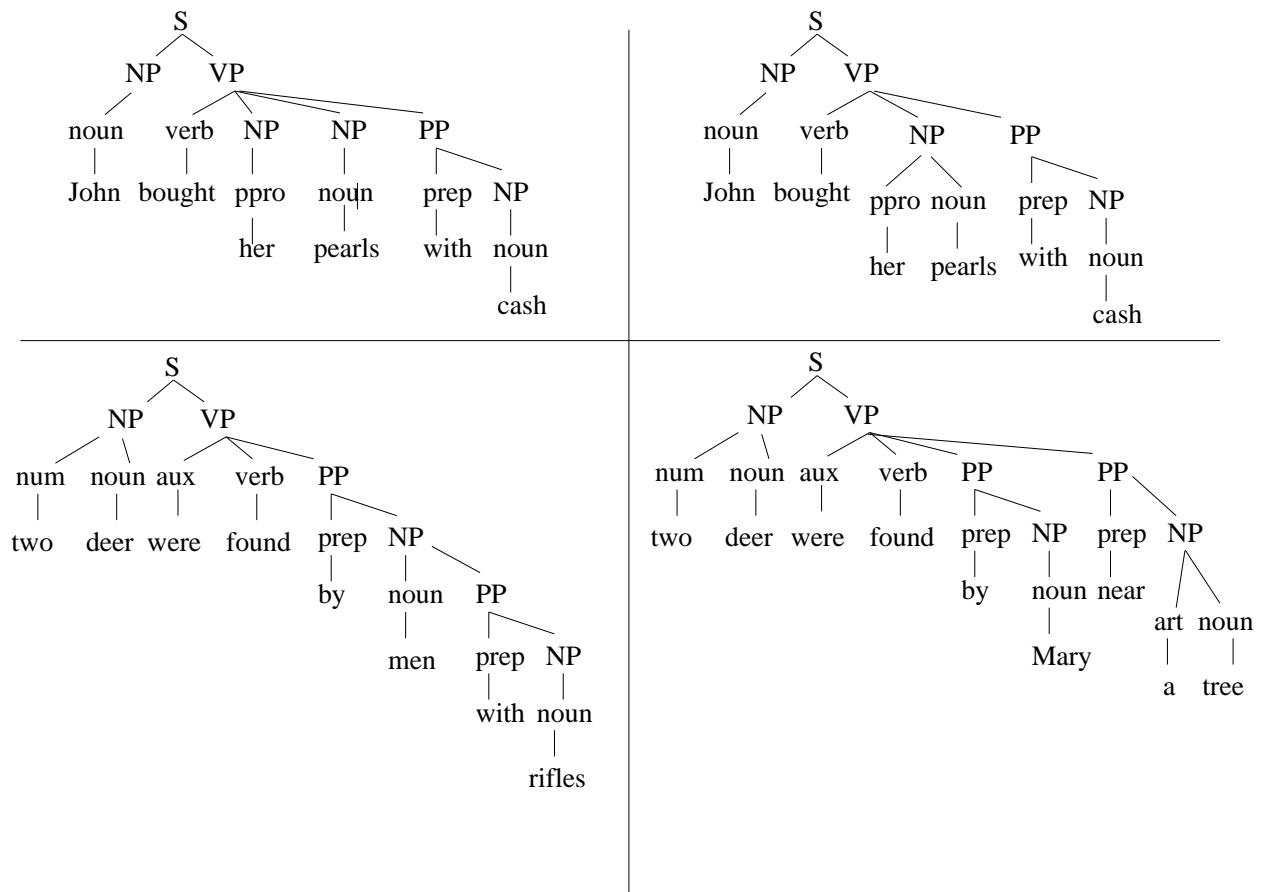
(**S** (**NP** (adj natural) (noun language))
 (**VP** (verb processing))
 (**VP** (verb is)
 (**NP** (art a) (adj fun) (noun topic))
 (**PP** (prep to) (**NP** (noun study))))))

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!**

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 \rightarrow John”.

Leave your probabilities in fractional form!



Grammar Rule	Probability

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.
 - (b) Top-down chart parsing would be better for shallow parsing than bottom-up chart parsing.
 - (c) RTNs use a *bottom-up* parsing strategy.
 - (d) $P(X \mid Y)$ should be the same as $P(X)$ if X and Y are independent of each other.
 - (e) A morphological analyzer can automatically produce features that are useful for parsing.
 - (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.
 - (g) When training and evaluating machine learning systems, cross-validation is especially useful if the amount of annotated training data is small.
 - (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.
 - (i) Context-free grammars can recognize sentence structures that regular grammars cannot recognize.
 - (j) A lexical N-gram model uses part-of-speech tags.

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			
pen	400			
pencil	98			
eraser	2			
chalk	0			

CS-5340/6340, Midterm Exam, Fall 2013

NAME:

Problem	Points Possible	Points Received
<i>1</i>	8	
<i>2</i>	10	
<i>3</i>	10	
<i>4</i>	20	
<i>5</i>	12	
<i>6</i>	8	
<i>7</i>	12	
<i>8</i>	20	
<i>9*</i>	12	
<i>TOTAL</i>		

** means CS-6340 students only!*