NAME:

CS-5340/6340, Fall 2015, Exam #1 Solutions

1. (8 pts) Consider the sentence:

   Three men robbed the bank at midnight with machine guns.

   Assume the **correct** parse is:

   (**S** (**NP** (num Three) (noun men))
     (**VP** (verb robbed)
           (**NP** (art the) (noun bank))
           (**PP** (prep at) (**NP** (noun midnight)))
           (**PP** (prep with) (**NP** (noun machine) (noun guns)))))

   Suppose **your parser** generates the following parse:

   (**S** (**NP** (num Three) (noun men))
     (**VP** (verb robbed)
           (**NP** (art the) (noun bank)
                 (**PP** (prep at) (**NP** (noun midnight))))
           (**PP** (prep with) (**NP** (noun machine))))
     (**VP** (verb guns)))

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

   Correct Constituents (4):
   S, NP(Three men), PP(at midnight), NP(midnight)

   *Missed Constituents (4):*
   VP(robbed the bank at midnight with machine with machine guns), NP(the bank),
   PP(with machine guns), NP(machine guns)

   Wrong Constituents (5):
   VP(robbed the bank at midnight with machine), NP (the bank at midnight), PP(with
   machine), NP(machine), VP(guns)

   Recall = 4/8,   Precision = 4/9

1

2. (12 pts) Answer the questions below based on the four parse trees: A, B, C, D. The answer to each question may be a single parse tree, multiple parse trees, or none of the parse trees. When considering the "meaning" of a parse tree, assume the most common sense meaning of the sentence based on the parse tree structure.
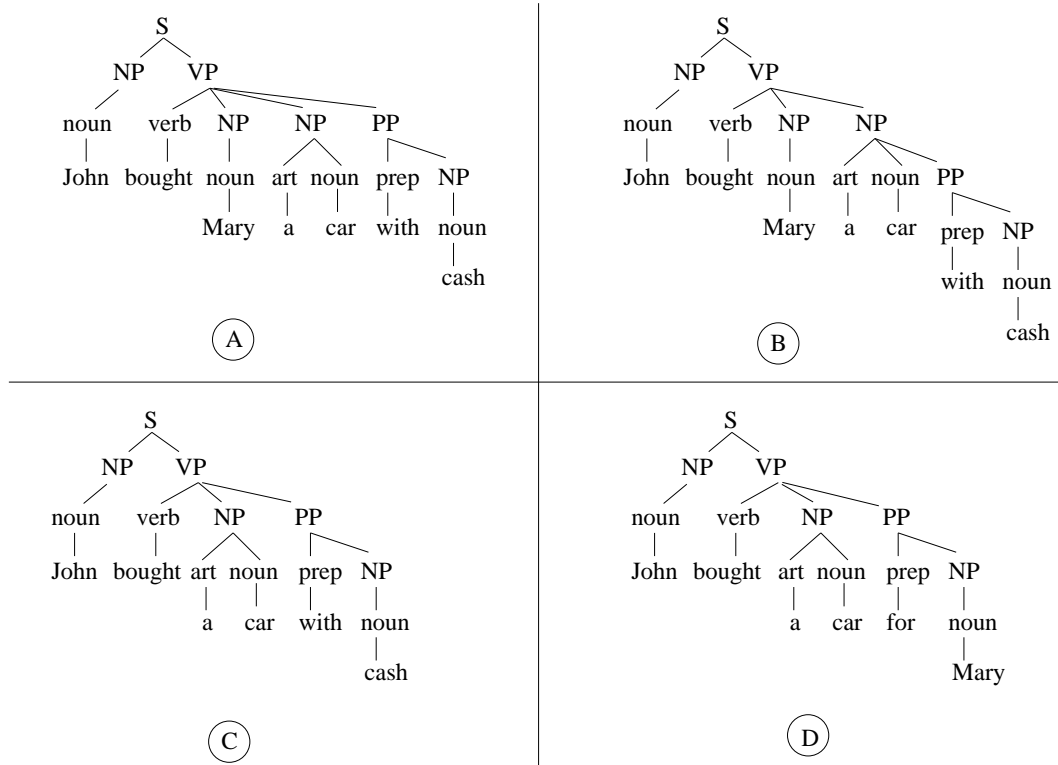
(a) Which (if any) of the parse trees include an intransitive use of the verb *"bought"*?
none

(b) Which (if any) of the parse trees include a ditransitive use of the verb *"bought"*?
A, B

(c) Which (if any) of the parse trees include a passive voice use of the verb *"bought"*?
none

(d) Which (if any) of the parse trees mean that John used cash to buy the car?
A, C

(e) Which (if any) of the parse trees mean that the car contains cash?
B

(f) In which (if any) of the parse trees is "a car" the direct object of "bought"?
A, B, C, D

3. (18 pts) Imagine that you have a tiny Treebank (texts with human generated parses) that consists of the two parse trees below. Reverse-engineer a probabilistic context-free grammar (PCFG) from this Treebank. You only need to create grammar nules for the S, NP, VP, and PP non-terminals (i.e., you don't need to create rules from the part-of-speech tag nodes). *Please leave your probabilities in fractional form!*
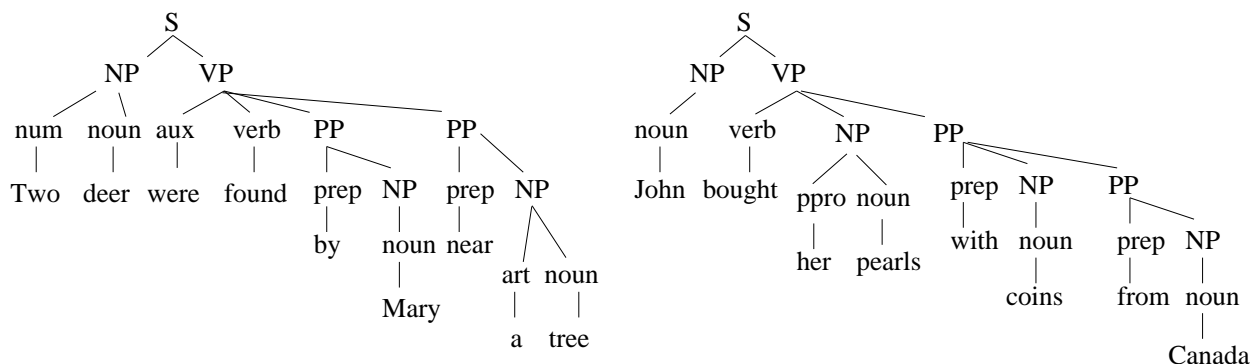
The correct PCFG is:

Grammar Rule	Probability
S → NP VP	2/2
NP → num noun	1/7
NP → noun	4/7
NP → art noun	1/7
NP → ppro noun	1/7
VP → aux verb PP PP	1/2
VP → verb NP PP	1/2
PP → prep NP	3/4
PP → prep NP PP	1/4

4. (14 pts) Imagine that you have a perfect shallow parser that identifies noun phrase (NP) chunks using an **IOB** tagging scheme. Label each word in the sentences below (which are all quotes from the comedian Steven Wright) with an **I**, **O**, or **B** tag corresponding to the NP chunking. (Here we mean simple or "base" NPs, in contrast to complex NPs that might include embedded constituents.)

If you don't know what IOB tagging is, then circle each distinct noun phrase and you will get half credit.

(a) Everywhere is within walking distance if you have the time.

Everywhere/B is/O within/O walking/B distance/I if/O you/B have/O the/B time/I.

(b) On the other hand you have different fingers.

On/O the/B other/I hand/I you/B have/O different/B fingers/I.

(c) Hard work pays off in the future. Laziness pays off now.

Hard/B work/I pays/O off/O in/O the/B future/I. Laziness/B pays/O off/O now/O.

(d) If you were going to shoot a mime, would you use a silencer?

If/O you/B were/O going/O to/O shoot/O a/B mime/I, would/O you/B use/O a/B silencer/I?

5. (12 pts) Consider the grammar below:

$$NP \rightarrow art\ NP1$$
$$NP \rightarrow ppro\ NP1$$
$$NP1 \rightarrow num\ NP1$$
$$NP1 \rightarrow NP2$$
$$NP2 \rightarrow adj\ NP2$$
$$NP2 \rightarrow adj\ NP3$$
$$NP3 \rightarrow noun\ NP3$$
$$NP3 \rightarrow noun$$

State whether the grammar **CAN** or **CANNOT** produce each part-of-speech tag sequence below.

(a) art num noun
   CANNOT

(b) art num num noun
   CANNOT

(c) art adj noun
   CAN

(d) art noun
   CANNOT

(e) art num adj noun
   CAN

(f) art num num adj adj noun noun
   CAN

(g) art adj num noun
   CANNOT

(h) art adj adj noun noun
   CAN

(i) art ppro num noun
   CANNOT

(j) ppro num num adj noun noun
   CAN

(k) ppro noun noun noun
   CANNOT

(l) ppro adj adj adj
   CANNOT

6. (16 pts) Answer each question below as **True** or **False** (no explanation is necessary):

(a) Recursive transition networks (RTNs) can perform shallow parsing.
   TRUE

(b) In chart parsing, you can change the parsing strategy from depth-first search to breadth-first search by changing the way that the agenda operates.
   TRUE

(c) In chart parsing, you can change the parsing strategy from top-down parsing to bottom-up parsing by changing the way that the agenda operates.
   FALSE

(d) An NLP system that identfies phrases describing a vehicle would be an information extraction system.
   TRUE

(e) An NLP system that identifies phrases that describe the victims killed by hurricanes in the U.S. last year would be an information extraction system.
   TRUE

(f) An NLP system that identifies news articles about politics would be an information extraction system.
   FALSE

(g) Number agreement can help a parser decide on the most likely parse trees.
   TRUE

(h) Subcategorization frames can help a parser decide on the most likely parse trees.
   TRUE

7. (10 pts) Imagine that you have a collection of 500 documents and humans have assigned the correct shallow parsing labels for every sentence. Let's call this your "annotated corpus" for shallow parsing. Suppose you have developed a machine learning (ML) classifier that performs shallow parsing, and you decide to evaluate its performance using cross-validation on the annotated corpus.

(a) If you use 5 folds (partitions) for the cross-validation, how many documents will your ML classifier be given for training in each of the 5 experiments?

400

(b) If you use 10 folds (partitions) for the cross-validation, how many documents will your ML classifier be given for training in each of the 10 experiments?

450

(c) If you only had 30 documents in your annotated corpus. Would it be better to use a **small** number of folds (say, 3 folds) or a **large** number of folds (say, 10 folds)?

Larger. The greater the number of folds, the more data can be used for training. For example with 3 folds, you'd have 20 documents for training. But with 10 folds you'd have 27 documents for training, which is 35% more training data.

(d) If you only had 30 documents in your annotated corpus, would it be better to split your corpus into a single training set and a single test set, or to use cross-validation?

Cross-validation because you'll be able to test on ALL the data. With a single training/test split, you'd have much less test data. You can also maximize the amount of training data if you do cross-validation with a large number of folds.

(e) If your ML classifier is overly aggressive and labels nearly every word as being in a noun phrase, will its recall or precision be higher for noun phrase labels?

Recall. The overly aggressive nature of the classifier will produce many false hits, which will be penalized when measuring precision. And since nearly all words will be labeled as being in NPs, most of the true NPs words should be identified, producing high recall.

8. (10 pts) Use the following tables of probabilities to answer this question:

P(space | V)	.10
P(space | N)	.30
P(rocks | V)	.20
P(rocks | N)	.40

P(N | ϕ)	.80
P(V | ϕ)	.20
P(N | N)	.25
P(N | V)	.30
P(V | N)	.50
P(V | V)	.10

Assume that there are 2 possible part-of-speech tags: NOUN (N) and VERB (V). The network below would be used by the Viterbi algorithm to find the most likely sequence of part-of-speech tags for the sentence *"Space rocks"*:
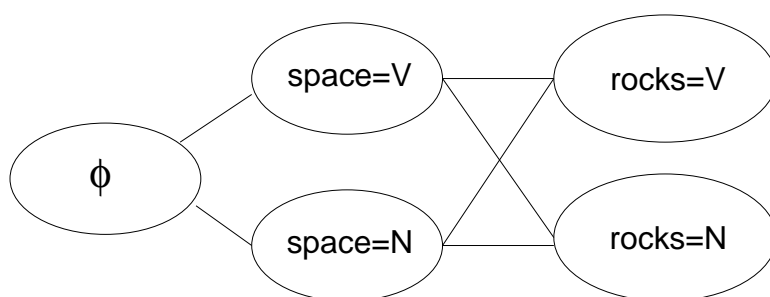
Using the Viterbi algorithm, compute the probability for each of the following nodes in the network. Show both <u>the equation</u> and <u>the numbers that would be filled in</u>. For (a) and (b), please compute a specific value (it should be simple enough that you will not need a calculator). You do not need to compute a specific value for (c) and (d) – it is sufficient to just show the equation with the numbers filled in.

(a) P(space=V)

$P(V \mid \phi) * P(space \mid V) = .20 * .10 = .02$

(b) P(space=N)

$P(N \mid \phi) * P(space \mid N) = .80 * .30 = .24$

(c) P(rocks=V)

$P(rocks \mid V) *$
MAX $(P(V \mid V)*.02, P(V \mid N)*.24)$
$= .20 *$ MAX $(.10 * .02, .50 * .24)$
$= .20 *$ MAX$(.002, .12) = .024$

8

(d) P(rocks=N)

$P(rocks \mid N)$ *
MAX ($P(N \mid V)$*.02, $P(N \mid N)$*.24)
= .40 * MAX (.30 * .02, .25 * .24)
= .40 * MAX(.006, .06) = .024

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

9. (10 pts) Fill in the table below with morphology rules that would derive the following words from the specified root form in a linguistically sensible way. Some of these words will require a *multi-step derivation*. Make sure that you list the rule required for <u>each</u> step of the derivation as a separate row in the table.

  (a) geologist (root = "geology")

  (b) leafiness (root = "leaf")

  (c) unreadable (root = "read")

  (d) insanity (root = "sane")

Derived Word	Affix	Replacement Chars	POS of root word	POS of derived word
geologist	ist	y	noun	noun
leafiness	iness	y	adj	noun
leafy	y	-	noun	adj
unreadable	un	-	adj	adj
readable	able	-	verb	adj
insanity	ity	e	adj	noun
insane	in	-	adj	adj

NOTE: For "insanity", we also accepted morphology rules for "ity" as a suffix (adj → noun) to create "sanity" from "sane", and "in" as a prefix (noun → noun) to create "insanity" from "sanity". The "in" prefix rule is more commonly for adjectives (e.g., "insubordinate") but also occurs with some nouns (e.g., "inattention").

We did not accept derivations for "unreadable" that included the morphology rule "un" as a prefix (verb → verb) to create "unread" from "read" because the action of "unreading" is not really sensible here.