Question 1 (20 points total)

Q1 Not graded

Consider the following four-line POS-annotated corpus (POS tags are shown above their corresponding words):

```
PRON VB   AUX DT  NN
I    fish for the can

PRON VB  NN
I    can fish

NN   AUX VB
fish can fish

DT  NN  VB PRP DT  NN
the dog is in  the shed

DT  NN  AUX VB  DT  NN
the dog can see the cat
```

1a (10 points): Write down the emission and transition probability tables for a bigram HMM inferred from this corpus using maximum likelihood estimates. You do not need to do any smoothing. You may express the probabilities as fractions instead of decimals if you wish. An extra page is provided in case it is needed.

x \ y	PRON	VB	AUX	DT	NN	PRP	</S>
PRON	0	1	0	0	0	0	0
VB	0	0	1/4	1/4	1/4	1/4	0
AUX	0	2/3	0	1/3	0	0	0
DT	0	0	0	0	1	0	0
NN	0	1/7	2/7	0	0	0	4/7
PRP	0	0	0	1	0	0	0
<S>	2/5	0	0	2/5	1/5	0	0

Transition probability $P(x \to y)$

(continuation of the previous item, if needed)

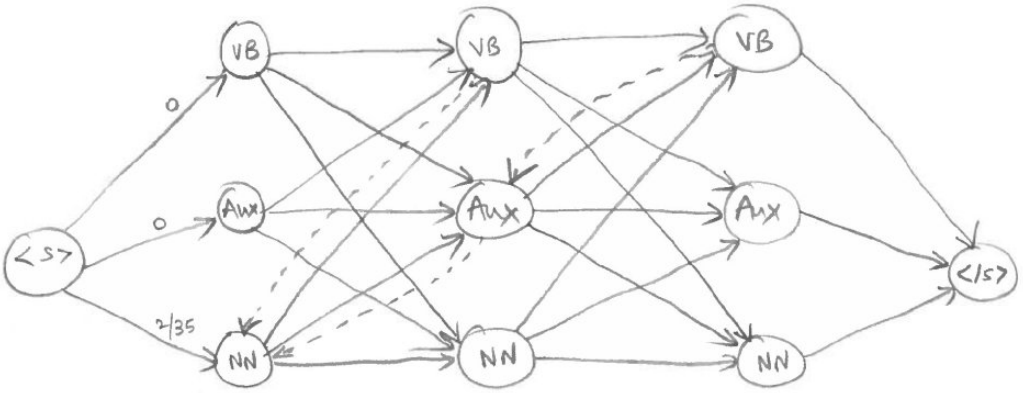
Q2 10

x \ y	I	fish	for	the	can	dog	is	in	shed	see	cat
Pron	1	0	0	0	0	0	0	0	0	0	0
VB	0	2/5	0	0	1/5	0	1/5	0	0	1/5	0
Aux	0	0	1/3	0	2/3	0	0	0	0	0	0
DT	0	0	0	1	0	0	0	0	0	0	0
NN	0	2/7	0	0	1/7	2/7	0	0	1/7	0	1/7
PRP	0	0	0	0	0	0	0	1	0	0	0

emission probability $p(y|x)$

Q3 Not graded

1b (5 points): Using the tables from the previous page, complete the chart to show the use of the Viterbi algorithm to find the most likely POS tag sequence for the sentence "fish can see" and its probability. You may express probabilities as products of fractions or define variables rather than calculating small decimals (e.g. you can write "let $x = 1/3 * 1/5 * 1/9$" and then use x in subsequent equations) but be sure they are defined very clearly on the answer page. Backpointers may be drawn with arrows or with a separate backpointer chart. An extra page is provided in case it is needed.



fish can see

$P(VB|fish) = P(VB|<s>) \times P(fish|VB) = 0$

$P(Aux|fish) = P(Aux|<s>) \times P(fish|Aux) = 0$

$P(NN|fish) = P(NN|<s>) \times P(fish|NN) = \frac{1}{5} \times \frac{2}{7} = \frac{2}{35}$

$P(VB|can) = \max[\ P(VB|VB) \times P(VB|fish) \times P(can|VB),$
$P(VB|Aux) \times P(Aux|fish) \times P(can|Aux),$
$P(VB|NN) \times P(NN|fish) \times P(can|VB)\ ]$

$= \max\left[0, 0, \frac{1}{7} \times \frac{2}{35} \times \frac{1}{5}\right] = \frac{2}{35 \times 35}$

(continuation of the previous item, if needed)

Q4 5

$P(Aux|can) = \max[0, 0, P(Aux|NN) \times P(can|Aux) \times P(NN|fish)]$

$= \frac{2}{7} \times \frac{2}{3} \times \frac{2}{35} = \frac{8}{21 \times 35}$

$P(NN|can) = \max[0, 0, P(can|NN) \times P(NN|NN) \times P(NN|fish)]$

$= \frac{1}{7} \times 0 = 0$

$P(VB|see) = \max[P(see|VB) \times P(VB|VB) \times P(VB|can),$
$P(see|VB) \times P(VB|Aux) \times P(Aux|can),$
$P(see|VB) \times P(VB|NN) \times P(NN|can)]$

$= \max[0, \frac{1}{5} \times \frac{2}{3} + \frac{8}{21 \times 35}, \frac{1}{5} \times \frac{1}{7} \times 0]$

$= \frac{16}{15 \times 21 \times 35}$

$P(Aux|see) = \max[P(Aux|VB) \times P(see|Aux) \times P(VB|can),$
$P(Aux|Aux) \times P(see|Aux) \times P(Aux|can),$
$P(Aux|NN) \times P(see|Aux) \times P(NN|can)]$

$= \max[\frac{1}{4} \times 0, 0, 0] = 0$

$P(NN|see) = \max[P(NN|VB) \times P(see|NN) \times P(VB|can),$
$P(NN|Aux) \times P(see|NN) \times P(Aux|can),$
$P(NN|NN) \times P(see|NN) \times P(NN|can)]$

$= \max[\frac{1}{4} \times 0, 0, 0] = 0$

∴ Most likely Pos tag sequence is

fish	can	see
NN	Aux	VB

Probability is $\frac{16}{15 \times 21 \times 35}$

Q5 2

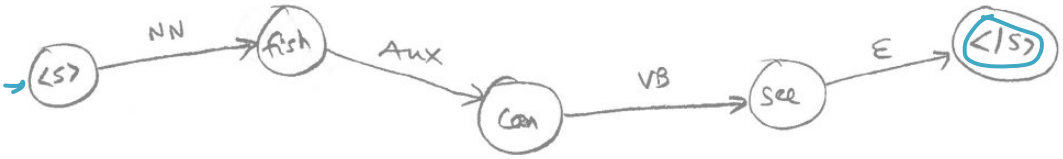
1c (2 points): What is the most likely tag sequence for "fish can see" and what is the joint probability of observing this sentence and its most likely tag sequence, according to the bigram HMM model (you can use variables and products of fractions, as in the previous steps)?

Most likely tag sequence for "fish can see" is NN Aux VB.

Joint probability is $\frac{16}{15 \times 21 \times 35}$

Q6 1

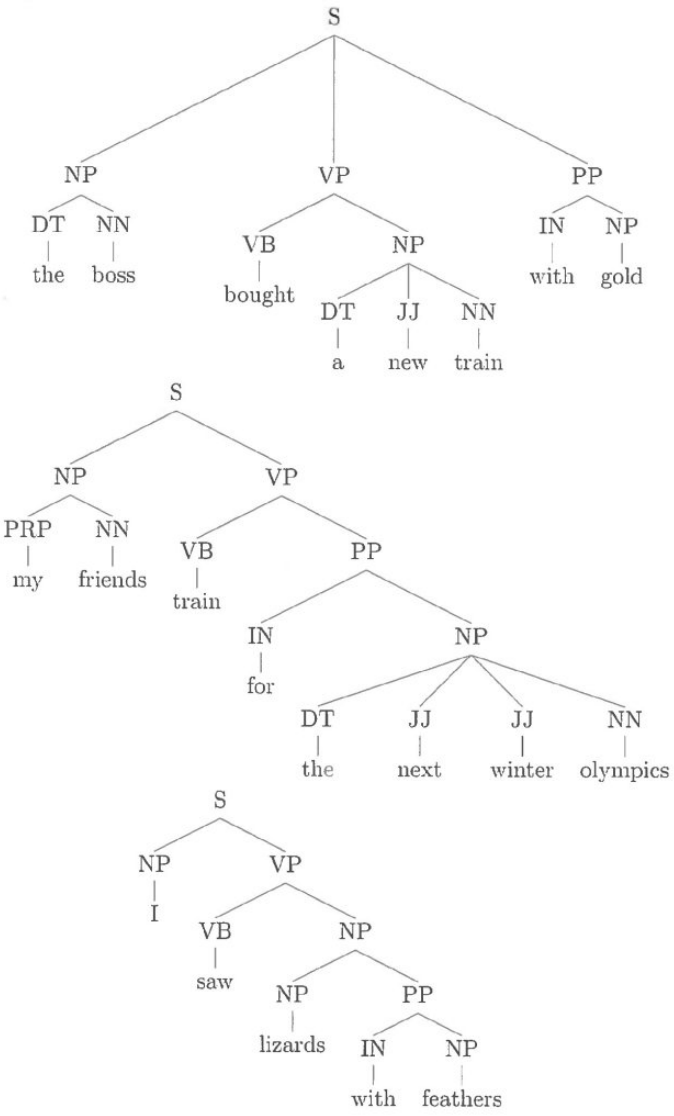
1d (3 points): Draw an FSA that represents the non-zero-probability POS tag sequences allowed by the bigram model. Include POS tag labels on the arcs but do not include weights. (You may find it convenient to label the states but this is not strictly necessary). Use ϵ to label any epsilon arcs you include.



Question 2 (20 points total)

Q7 Not graded

Consider the following corpus of constituent trees:



Q8 5

2a (5 points): Write down the CFG rules that can be extracted from this corpus.

ay use pipe notation to save space, i.e. to write the ruleset

```
A -> B
A -> C D E
```

you can simply write

```
A -> B | C D E
```

S → NP VP PP | NP VP
NP → DT NN | gold | PRP NN | I | lizards | feathers
VP → VB NP | VB PP
PP → IN NP
DT → the | a
NN → boss | train | olympics | friends
VB → bought | train | saw
JJ → new | next | winter
IN → with | for
NP → DT JJ JJ NN | NP PP
PRP → my

Q9 3

2b (3 points): Identify the *nonterminal* rules that are not in Chomsky Normal Form (CNF) and write down the replacement rules obtained by right-branching and Markovizing, such that no sibling information is retained (i.e. retain parent label only)

1. S → NP VP PP

Replacement: S → NP V'

V' → VP PP

2. NP → DT JJ JJ NN

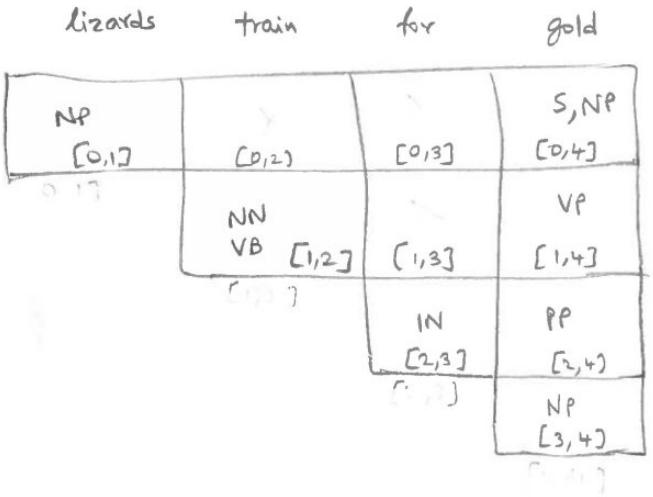
Replacement: NP → DT K'

J' → JJ NN

K' → JJ J'

Q10 10

2c (10 points): Write down all the states that will be built as a result of a CKY ...e for the sentence "lizards train for gold". States should have the form [X, i, j] ...cating words i through j are covered by at least one subtree rooted in X. Use ...r the lowest index. Show your work. An extra page is provided in case it is needed.



PP → IN NP
[PP, 2, 4]

VP → VB PP
[VP, 1, 4]

S → NP VP
[S, 0, 4]

NP → NP PP
[NP, 0, 4]

(continuation of the previous item, if needed)

Q11 **Not graded**

Q12 2

2d (2 points): Write down all the parse trees for "lizards train for gold" that are licensed by the CNF grammar.

:

S → NP V' | NP VP

V' → VP PP

NP → DT NN | PRP NN | I | gold | lizards | feather | NP PP | DT K'

VP → VB NP | VB PP

PP → IN NP

DT → the | a

NN → boss | train | olympics | friends

VB → bought | train | saw

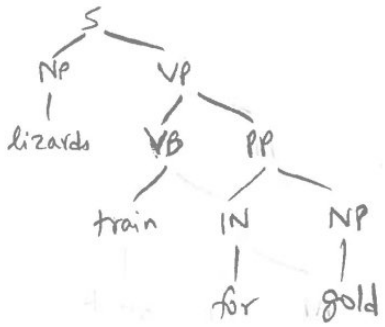
JJ → new | next | winter

IN → with | for

J' → JJ NN

K' → JJ J'

PRP → my



Q13 3

Question 3 **Multiple Choice (2 points each; 20 total): For each question circle all answers that apply. Zero, one, or more than one may apply in each case.**

3a When do Laplace smoothing and additive smoothing α give the same results?

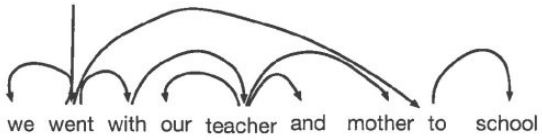
1. When either empty training data **or** $\alpha = 1$
2. Only when both empty training data **and** $\alpha = 1$
3. Sentences only one word long
4. Each word type appears exactly once

3b $P(A) = \frac{1}{2}$; $P(A \cap B) = \frac{1}{4}$; $P(B) = \frac{3}{10}$; **What is** $P(A|B)$?

1. 5/6
2. 1
3. 1/4
4. 1/2

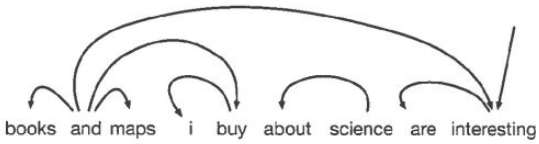
For each of the next four questions, given a dependency tree, indicate which properties are exhibited

3c



1. Projective
2. Non-Projective
3. Labeled
4. Functional Head Annotation Standard
5. Coordinating Conjunction Annotation Standard

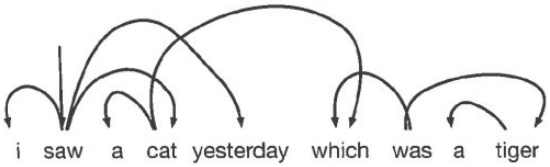
3d



Q14 4

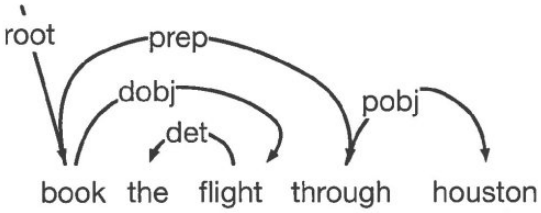
1. Projective
2. Non-Projective
3. Labeled
4. Functional Head Annotation Standard
5. Coordinating Conjunction Annotation Standard

3e



1. Projective
2. Non-Projective
3. Labeled
4. Functional Head Annotation Standard
5. Coordinating Conjunction Annotation Standard

3f



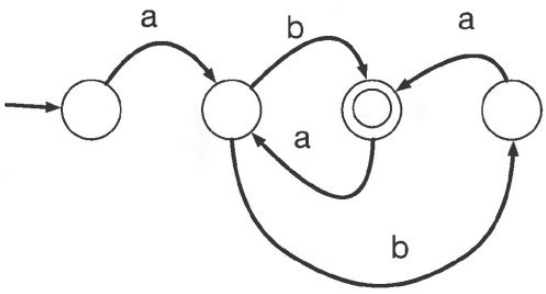
1. Projective
2. Non-Projective
3. Labeled
4. Functional Head Annotation Standard
5. Coordinating Conjunction Annotation Standard

Q15 2

3g In Arc-Eager labeled dependency parsing with a label set of size N, how many different actions can be taken at step $i > 1$?

1. $N + 2$
2. N^i
3. $N^i + 2$
4. $4N$
5. 3

3h What is the regular expression for the language accepted by this NFSA?



1. $ab(ab)+$
2. $ab(ab)*$
3. $ab(a|b)*$
4. $(ab|aba)+$

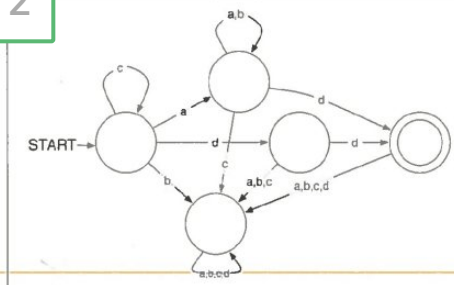
3i Assume your language model for English $P(w_1, w_2, ..., w_n)$ uses a trigram model. The independence assumption for trigrams assumes which of the following are independent?

1. w_4 =the and w_1=coffee
2. w_{23} and w_{25}
3. w_3=blue and w_4=jeans
4. w_3 and w_6

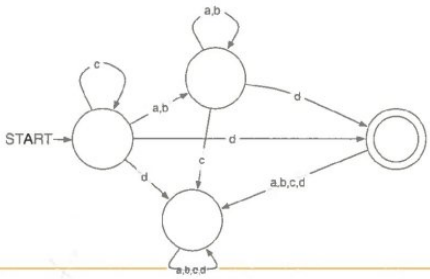
Q16 2

3i Which of these FSAs represents the regexp $(a|b)^*c^*d$?

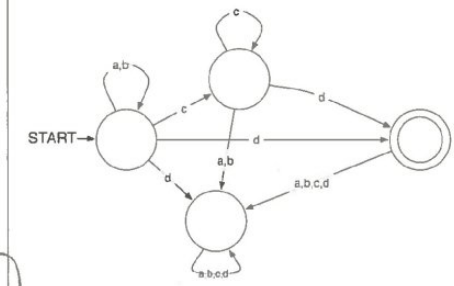
1



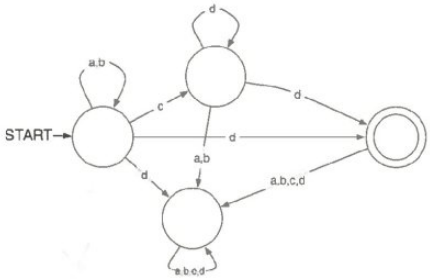
3



2



4



Question 4 Short Answers

Q17 4 0 points total) **Ramesh wants to calculate the probability of a 70-word sentence. He is using a smoothed bigram model trained on the ClueWeb data set and has estimated probabilities for out-of-vocabulary words. However, when he multiplies all the bigram probabilities together Ramesh's language model program still reports the probability of this sentence is zero.**

1. Why might this be so?

2. What steps can Ramesh take to get a non-zero probability under this model that can be used in downstream tasks?

3. Why are these steps justified?

4b (5 points total) Suppose I have a dictionary with 5000 words in it and I decide to generate a 5-word "sentence" by choosing each word from the dictionary with uniform probability.

1. What is the size of Ω (the sample space) in this experiment?

2. If E is the event "my sentence starts with the word *the*", how many outcomes are there in E?

3. What is $P(E)$?

Q18 0

4. Let A be the event "my sentence ends with the word *the*". Are A and E mutually exclusive? If so, explain why. If not, give an example of an outcome that belongs to both A and E.

5. Are A and E independent? Explain your answer.

Q19 0

4c (5 points total) The following questions involve the following terms: *discourse*, *lexemes*, *morphology*, *pragmatics*, *semantics*, *syntax*

1. Place these terms in order, from least ambiguous (for NLP purposes)/shallowest to most ambiguous/deepest

2. Choose three of the above terms and, for each, provide an example of ambiguity where it might be difficult for an NLP system to correctly analyze that property in a piece of text. An example for *orthography* might be "recognizing a handwritten "n" vs. "m". Give your example for the first chosen term here:

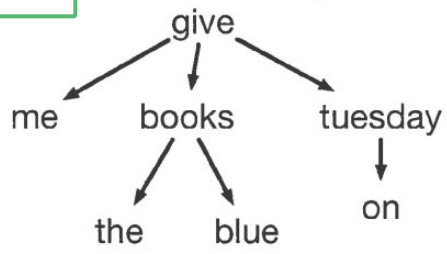
3. Give your example for the second chosen term here:

4. Give your example for the third chosen term here:

Question 5 (20 points total)

Q20 10

Consider the following dependency tree for the sentence "give me the blue books on tuesday":



5a (10 points): Complete the table below, which shows the stack, available symbol list, parsing decision, and relations added in an oracle arc-standard unlabeled parse of the sentence. Some entries have been added for you. An extra page is provided in case it is needed. You may wish to begin there or on scratch paper before filling out the table.

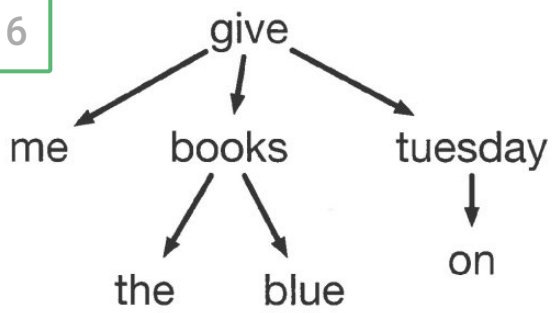
step	stack	symbols	action	relations added
1	[root]	give me the blue books on tuesday	S	
2	[root] give	me the blue books on tuesday	S	
3	root, give, me	the blue books on tuesday	R	give → me
4	root, give	the blue books on tuesday	S	
5	root, give, the	blue books on tuesday	S	
6	root, give, the, blue	books on tuesday	S	
7	[root] give the blue books	on tuesday	L	blue ← books
8	root, give, the, books	on tuesday	L	the ← books
9	root, give, books	on tuesday	R	give → books
10	root, give	on tuesday	S	
11	root, give, on	tuesday	S	
12	root, give, on, tuesday		L	on ← tuesday
13	root, give, tuesday		R	give → tuesday
14	root, give		R	root → give
15	[root]		done	

(continuation of the previous item, if needed)

Q21 **Not graded**

Q22 6

The next question uses the same dependency tree as the last one, so it is provided again for convenience:



5b (10 points): Complete the table below, which shows the stack, available symbol list, parsing decision, and relations added in an oracle arc-eager unlabeled parse of the sentence. Some entries have been added for you. An extra page is provided in case it is needed. You may wish to begin there or on scratch paper before filling out the table.

step	stack	symbols	action	relations added
1	[root]	give me the blue books on tuesday	R	root → give
2	root, give	me the blue books on tuesday	R	give → me
3	root, give, me	the blue books on tuesday	R	
4	root, give	the blue books on tuesday	S	
5	root, give, the	blue books on tuesday	S	
6	[root] give the blue	books on tuesday	L	blue ← books
7	root give the	books on tuesday	L	the ← books
8	root give	books on tuesday	R	give → books
9	root give books	on tuesday	S	
10	root give books on	tuesday	L	on ← tuesday
11	root give books on tuesday		R	
12	root give books on		R	give → tuesday
13			reduce	
14				
15	[root]		done	

(continuation of the previous item, if needed)

Q23 **Not graded**

Question 6 (10 points total) (due to Michael Collins)

Q24 0

In this question we consider a very simple setting, where every sentence is of length 2 (not including the STOP symbol): that is, every sentence is of the form u, v where $u \in \mathcal{V}$ and $v \in \mathcal{V}$ for some vocabulary $\mathcal{V}$. We define X_1 to be the random variable (RV) corresponding to the first word in the sentence, and X_2 to be the RV corresponding to the second word.

6a (5 points)

In our first model, we assume that for any u, v,

$$P(X_1 = u, X_2 = v) = P(X_1 = u) \times P(X_2 = v)$$

i.e. the two random variables are *independent*. For this model, prove that

$$\sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} P(X_1 = u, X_2 = v) = 1$$

6b (5 points)

Q25 0

In econd model, we assume that for any u, v,

$$P(X_1 = u, X_2 = v) = P(X_1 = u) \times P(X_2 = v | X_1 = u)$$

i.e. the two random variables are not independent. For this model, prove that

$$\sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} P(X_1 = u, X_2 = v) = 1$$