Computational Linguistics I *CMSC723, Fall 2010*

Midterm Exam

Name: Solution

UID: ______________________

RIGHT NOW: Write your name and UID in the top right corner of every page!!! **This is worth 5 points!**

Test Information

This test has two parts: **true/false** and **short answer**.

There is **no partial credit** for the true/false questions, so don't bother explaining your answers.

Grading (do not write in this section)

Name+uID	_____ / 5
1: True/False	_____ / 20
2: Ambiguity	_____ / 25
3: HMMs	_____ / 25
4: Grammars	_____ / 25
Total	_____ / 100

1 True/False (20%)

Please write (T or F) in the parentheses to the left. No partial credit. Each is worth 2 points. Not necessarily balanced.

(F) If I estimate a left-to-right bigram language model and separately estimate a right-to-left bigram language model, and apply them to any test sentence, they will always give the same probability.

(T) Smoothing is a technique for taking probability mass from frequent events and "redistributing" it to low count and unseen events.

(T) If I estimate a standard trigram language model *without* smoothing and apply it to a sentence that contains at least one out of vocabulary term (aka unknown work), the language model will assign zero probability to this sentence.

(F) Regular languages are at least as expressive as context free languages.

(F) $p(a \mid b) = p(b \mid a)p(b)/p(a)$

(F) $p(a, b \mid c, d) = p(a \mid c, d)p(b \mid a)$

(F) Parsing with context free grammars can always be done in time quadratic in the length of the input.

(T) Converting a feature-based grammar to a standard context free grammar can result in a blow-up in the size of the grammar that is exponential in the number of features.

(F) In practice, the run time of a treebank parser usually has more to do with the length of the input than the size of the grammar.

(T) Any HMM can be converted into a completely right-branching CFG (i.e., all the rules are of the form "X → a B" where "a" is a terminal, and B could either be a terminal or a non-terminal).

2 Linguistic Ambiguity (25%)

Below there are five types of linguistic ambiguity and a "keyword" for each of them. Construct an example sentence for each that has the appropriate type of ambiguity with respect to the keyword. Write two intepretations for the example you constructed. (Each is worth 5 points.) If you cannot come up with a single sentence that has more than one interpretation, you may give two sentences each with a different interpretation, but you will lose one point for doing so.

Part of speech / can

I can punch

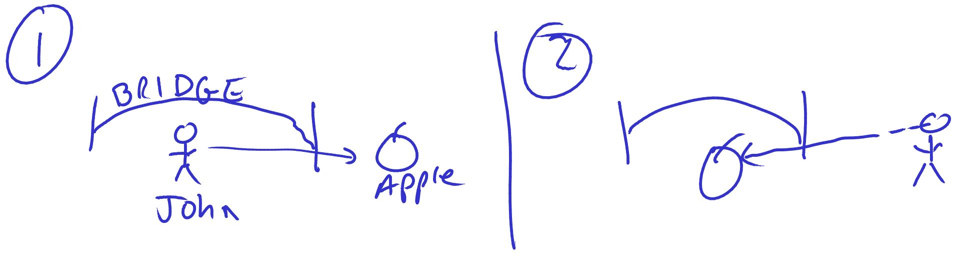
→ Aux: I am able to punch (someone)

→ Verb: I put (Hawaiian) punch in cans.

PP attachment / under

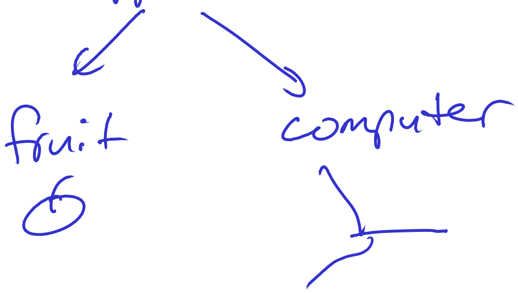
vs.

John saw the apple under the bridge.



Word sense / apple

I love apples



Pragmatic / *your choice*

"Can you pass the pepper?"

"Yes, I am able to."

"Here's the pepper."

Discourse / *your choice*

① Mary went to a barbeque.

② They only had vegetarian dishes.

if ③ = "Mary was pissed" then 1 → 2 is contrastive.

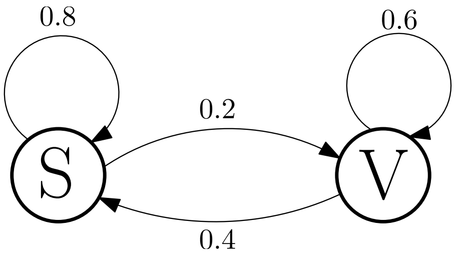
if ③ = "Mary had been excited to go to a veg. BBQ for months" then 1 → 2 is background info.

3 Hidden Markov Models (25%)

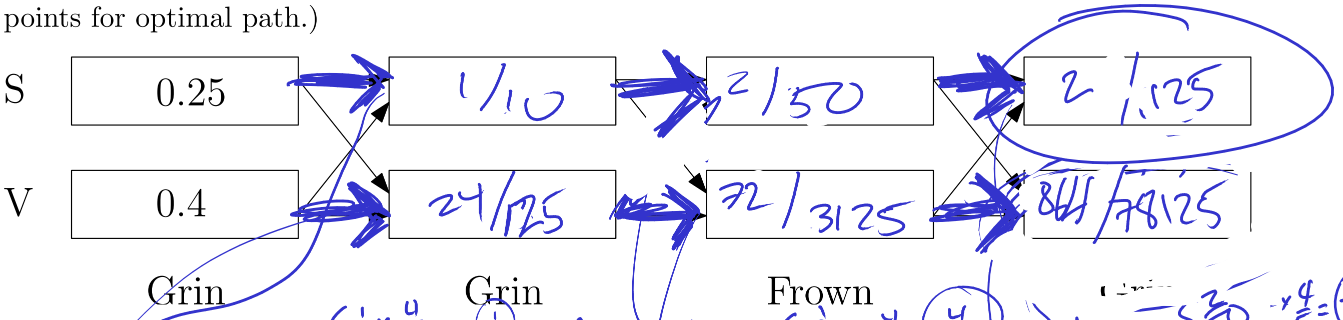
Consider the HMM below. In this world, every time step (say every few minutes), you can either be **S**tudying or playing **V**ideo games. You're also either **G**rinning or **F**rowning while doing the activity.

E	$p(E \mid X = S)$
grin	0.5
frown	0.5

E	$p(E \mid X = V)$
grin	0.8
frown	0.2



(15 pts) Suppose that we believe that the initial state distribution is 50/50. We observe: Grin, Grin, Frown, Grin. Run the *Viterbi algorithm* by filling in the values of the lattice below. Emphasize the back pointers by thickening the edges in the lattice. (You may use blank space for intermediate work, but please write your final answer in the lattice.) What is the most likely path for this sequence of observations? (Two points for each value, four points for correct back pointers and two points for optimal path.)



$\frac{1}{2} \times \max \begin{cases} \frac{1}{4} \times \frac{4}{5} = \frac{1}{5} \\ \frac{2}{5} \times \frac{2}{5} = \frac{4}{25} \end{cases} = \frac{1}{10}$

$\frac{4}{5} \times \max \begin{cases} \frac{1}{4} \times \frac{1}{5} = \frac{1}{20} \\ \frac{2}{5} \times \frac{3}{5} = \frac{6}{25} \end{cases} = \frac{24}{125}$

$\frac{1}{2} \times \max \begin{cases} \frac{1}{10} \times \frac{4}{5} = \frac{4}{50} \\ \frac{24}{125} \times \frac{3}{5} = \frac{48}{625} \end{cases}$

$\frac{1}{5} \times \max \begin{cases} \frac{1}{10} \times \frac{1}{5} = \frac{1}{250} \\ \frac{24}{125} \times \frac{3}{5} = 72/3125 \end{cases}$

$\frac{1}{2} \times \max \begin{cases} \frac{2}{50} \times \frac{4}{5} = \frac{2}{125} \\ \frac{72}{3125} \times \frac{2}{5} = 72/15625 \end{cases}$

$\frac{4}{5} \times \max \begin{cases} 2/50 \times \frac{1}{5} = 2/625 \\ 72/3125 \times \frac{3}{5} = 864/78125 \end{cases}$

path = S S S S

in decimal, ≈

.25 → .1 → .04 → .016

.4 → .192 → .023 → .011

(10 pts) Suppose that we *didn't* know the emission probabilities or transition probabilities for this HMM. Instead, we had to estimate them from data. Consider the following data set:

S→G: 3 V→G: 8
S→F: 8 V→F 1

S→S: 6 V→S: 4
S→V: 5 V→V: 4

```
                             1 1 1 1 1 1 1 1 1 1 2
 time: 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
       ---------------------------------------
state: S S V V V S S S S S V S V V S V S S V V
  obs: G F G G F F F F G F G G G G F G F F G G
```

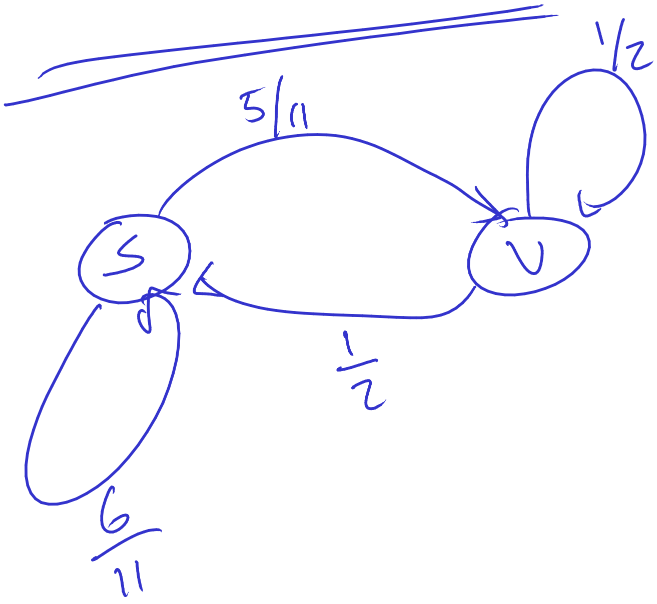
(Sorry for the lame "image".)

Based on this data, estimate the emission probabilities and the transition probabilities for this HMM. Write them out in the same form as values I gave at the beginning of this part of the exam. **Do this first with no smoothing and them with add-λ smoothing with $\lambda = 0.1$.** (Five points for each, three of which are for transitions and two of which are for emissions.)

$\lambda = 0$:

E	x=S
G	3/11
F	8/11

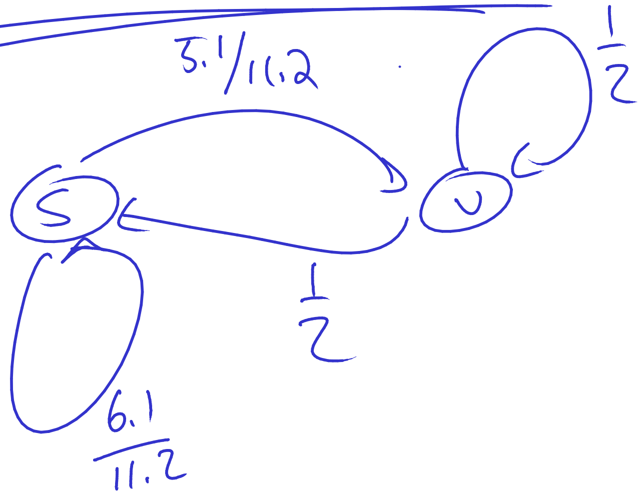
E	x=V
G	8/9
F	1/9



$\lambda = 0.1$

E	x=S
G	3.1/11.2
F	8.1/11.2

E	x=V
G	8.1/9.2
F	1.1/9.2



4 Context-Free Parsing (25%)

(5 pts) Write down a context-free grammar that will accept exactly strings of the form $a^n b^n c^m d^m$ or tell me that it cannot be done. (I.e., members of this language are strings like "aabbccccdddd" or "aaabbbcd" but not "aabbbcd" because the counts don't match.)

$S \rightarrow X\ Y$

$X \rightarrow a\ b$

$X \rightarrow a\ X\ b$

$Y \rightarrow c\ d$

$Y \rightarrow c\ Y\ d$

(5 pts) Write down a context-free grammar that will accept exactly strings of the form $a^n b^m c^m d^n$ or tell me that it cannot be done.

Can't be done!

(15 pts) Consider the following grammar:

```
S -> A B (0.8)  |  B A (0.2)
A -> a   (0.5)  |  b A (0.5)
B -> B A (0.6)  |  b B (0.2)  |  B -> b a (0.2)
```

If we ran a CKY parsing algorithm on the sentence "b a b a", we would find the *most likely tree* for this sentence. This is analogous to the *Viterbi* algorithm. However, we might want to instead compute the *total probability* of this sentence by *summing* over all possible trees. This is analogous to the *forward* algorithm. We can do this essentially by replacing the "max" function in CKY with a "sum" function. That is, instead of storing the probability of *best* NP that spans (5,10), you store the sum of probabilities of *all* NPs that span (5,10). Then, you extract the sum of all "S"es that span the whole sentence.

Construct a chart and run this "inside" algorithm on the string "b a b a". Hint: there should be more than one way to derive this string under the given grammar. If you want to round values, round them to the nearest two significant decimals (i.e., round "0.293" to "0.29", or "0.0000843" to "0.000084"). Note that you do not need to store "backpointers" in this case, since it doesn't make sense. In each cell, just write the probability of any constituent that can go there. If you want to show your work, you may do it below, or on scratch paper, but please attach the scratch paper to your submission.

$P(S) = 0.05$

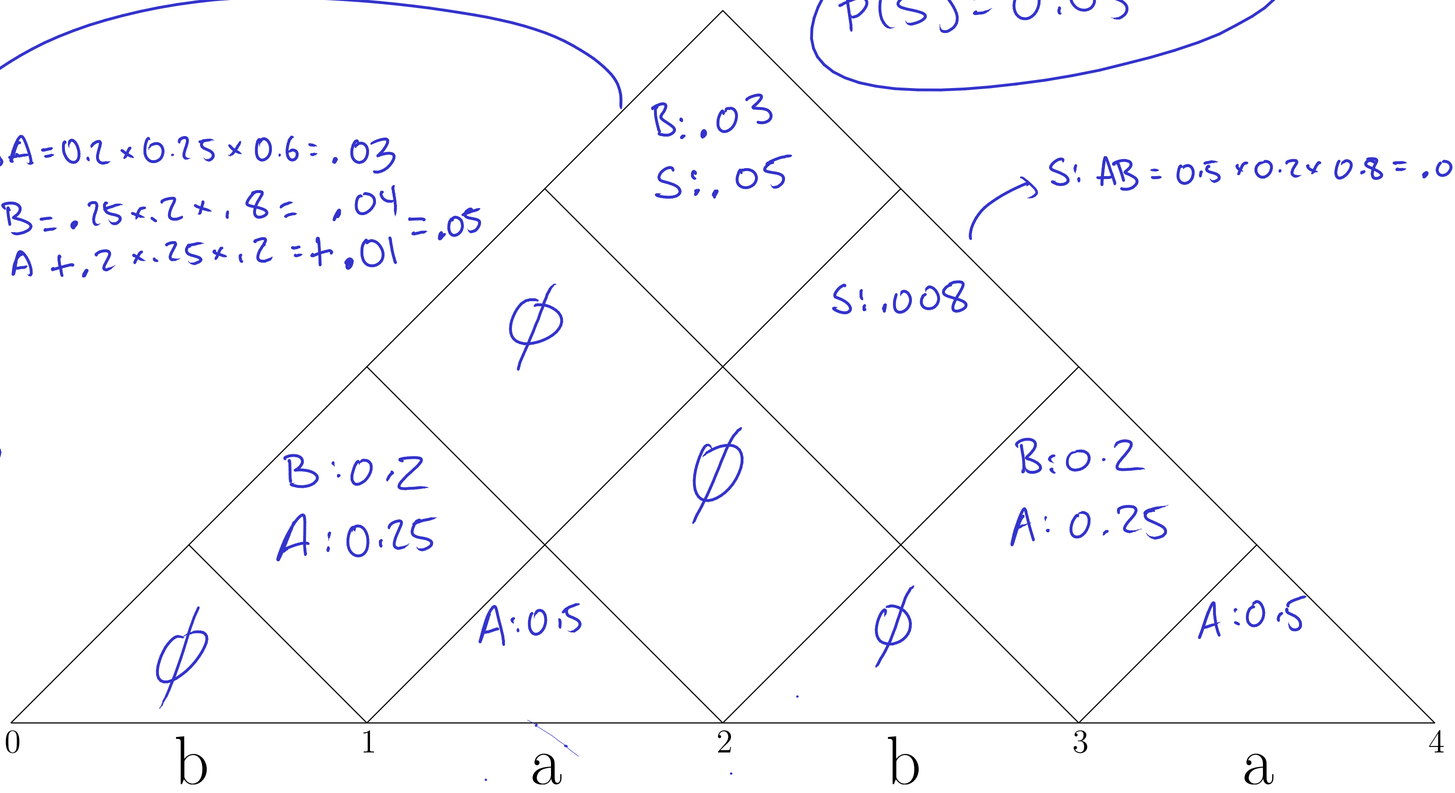
$B: BA = 0.2 \times 0.25 \times 0.6 = .03$

$S: AB = .25 \times .2 \times .8 = \ .04 \ = .05$

$+ BA + .2 \times .25 \times .2 = + .01$

Key step

$S: AB = 0.5 \times 0.2 \times 0.8 = .008$


B: .03
S: .05
∅
S: .008
B: 0.2
A: 0.25
∅
B: 0.2
A: 0.25
∅
A: 0.5
∅
A: 0.5
0
b
1
a
2
b
3
a
4