

CSCI 5832 Spring 2009 Quiz 2

Name: _____

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work. _____.

1. **(5 points)** What is your favorite kind of food?
2. Assume in the context of HMMs, that S refers to a sequence of states, O refers to a sequence of observations, and that M is a particular HMM model. **Name** the algorithms that compute the following values:
 - a) **(5 points)** $\operatorname{argmax} P(S|O, M)$ **Viterbi**
 - b) **(5 points)** $P(O| M)$ **Forward**
3. **(5 points)** Describe one kind of evidence that is used to justify the presence of a constituent (i.e., a particular non-terminal category) in a grammar.

One kind is external evidence. That is evidence of the form "this kind of constituent can always be followed by blah blah blah. Or "this constituent is always preceded by "foo".

4. Consider the following (inelegant) grammar rules from the Penn Treebank:

VP \rightarrow VBD PP
VP \rightarrow VBD PP PP
VP \rightarrow VBD PP PP PP
VP \rightarrow VBD PP PP PP PP
...

- a) **(5 points)** Give two rules that can be used to replace these four (and all the ensuing ... rules).

VP \rightarrow VBD PP
VP \rightarrow VP PP

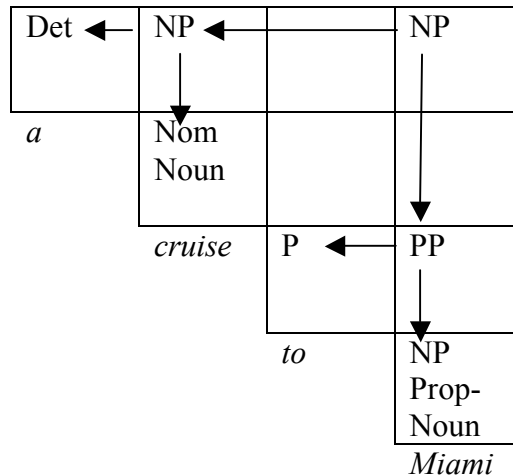
- b) **(5 points)** Is the grammar with your new rules weakly equivalent to the original set of rules?

Depends. If you take the Treebank set to represent a finite set of following PPs then no, since the proposed rules accept more strings (strings with a potentially infinite set of PPs).

5. (10 points) True or False: The following phrase is ambiguous given the attached grammar: *False, there's only one possible tree given this grammar. See the attached page.*

Book the flight and hotel for Boston

6. (15 points) Using the attached grammar, apply the CKY algorithm to the phrase *a cruise to Miami*. Show the completed table.



7. (5 points) Assuming that W stands for the words in a sentence and T for a parse tree, give a succinct mathematical equation for what the probabilistic CKY algorithm does.

$$\text{Argmax } P(T|W)$$

8. (10 points) What problem with probabilistic grammars does the technique of **parent annotation** address? Be specific.

The "bag of rules" basic model doesn't take into account where in a tree a given rule is being used. The parent annotation approach replaces a constituent with the name of the constituent concatenated with its parent node. This makes the rules specific to the immediately dominating tree that it occurs in.

9. (5 Points) Define the notions of **recall** and **precision** in the context of parser evaluation.

In the context of parser evaluation, we have parser guesses and gold-standard reference answers. Recall refers to ratio of the number of correct sub-trees (rule expansions) found in the guesses to the number of sub-trees in the reference answers (i.e, what percent of the trees that I should have found did I actually find). Precision refers to the ratio of the number of correct sub-trees found in the guesses to the number of sub-trees found overall. (i.e. what percent of the su-trees I found were right).

VP → Verb NP

Verb → book

NP → NP PP

NP → Proper-Noun

NP → Det Nominal

Nominal → Noun

Nominal → Nominal CONJ

CONJ → and Nominal

Nominal → Nominal Noun

PP → P NP

Det → *a, an, the*

Noun → *flight, hotel, cruise*

Proper-Noun → *Miami, Boston*

P → *in, to, for*

