1. (True or False). Wordnet contains relations that constitute a class inheritance hierarchy over the words.

2. Give examples of three different types of structural ambiguities and say why they are ambiguous.

3. In each of the following sentences, identify the semantic roles selecting from *agent, patient, theme, experiencer, stimulus, goal, recipient, benefactive, source, instrument, location, temporal*. Justify your choice.

The company wrote me a letter.
Jack opened the lock with a paper clip.

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

VP → VBD PP
VP → VBD PP PP
VP → VBD PP PP PP
VP → VBD PP PP PP PP
...

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

5. **(5 points) True/False:** Given a sentence S, the following equation adequately describes the computation performed by the probabilistic CKY algorithm given in the text.

$$P(S) = \sum_{T s.t. S=Yield(T)} P(T)$$

6. Assuming the grammar below, show the parse tree for the sentence *The big yellow dog sat under the house.*

S-> NP VP
VP -> VP PP
VP -> verb NP
VP -> verb
NP -> DET NOM
NOM -> ADJ NOM
NOM -> NOUN
PP -> PREP NP
DET -> the
ADJ -> big
ADJ -> yellow
NOUN -> dog

VERB -> sat
PREP -> under
NOUN -> house

7. Show how you would have to modify the grammar above to handle the sentence *The dog in the white hat ran under the house.*

8. **Hidden Markov Models:** You are given the sentence below and the tables of probabilities show in Table 3a (this page) and Table 3b (next page).

I promise to back the bill.

- **a.** Describe how the probabilities would be obtained using the Penn Treebank.

- **b.** A hidden markov model includes states, observations, transition probabilities, observation likelihoods. Describe what each one of these would correspond to when using an HMM for POS tagging.

- c. Given the sentence ``I *promise to back the bill.*'' show how you would compute the probability of ``*back*'' as a verb versus the probability of ``*back*'' as a noun using the probabilities in Tables 3a and 3b using the Viterbi algorithm. You are given the values for the third column of the Viterbi table which correspond to observation 3 or ``to''. They are VB: 0, TO: .00000018, NN: 0, PPSS: 0. Thus, you will show two computations both of which will use these values. You do not need to do the arithmetic; just show the formula that would be computed.

	I	***promise***	***to***	***back***
VB	0	.0093	0	.00008
TO	0	0	.99	0
NN	0	.0085	0	.00068
PPSS	.37	0	0	0

Table 3a: Observation Likelihoods

	VB	**TO**	**NN**	**PPSS**
<s>	.019	.0043	.041	.067
VB	.0038	.035	.047	.0070
TO	.83	0	.00047	0
NN	.0040	.016	.087	.0045
PPSS	.23	.00079	.0012	.00014

Table 3b: Tag transition probabilities. The rows are labeled with the conditioning event. Thus, P(VB|<s>) = .019.

9. Consider the sentences *President George Bush has re-invigorated the economy by providing a bail-out program for failing Wall Street firms*. and *President George Bush has caused a disastrous economic situation by failing to provide regulations on Wall Street firms.* You'd like to compute the likelihood of these sentences given a corpus of NY Times, Wall Street Journal and the New York Post gathered over the last year. You develop a bi-gram language model. Describe how you would: 1. Build the language model, 2. Compute the likelihood of these sentences and 3. Evaluate your language model.

10. Given the following three-layer neural network, where $A_{in}$is the input to the network; □ is the gold outputs; Z_i is the i^th hidden state; A_i is the i^th activation, $b_i$is the i^th bias term, f_i is the i^th activation function…

$$
\begin{aligned}
Z_1 &= W_1 A_{in} + b_1 \\
A_1 &= f_1(Z_1) \\
Z_2 &= W_2 A_1 + b_2 \\
A_2 &= f_2(Z_2) \\
Z_{out} &= W_{out} A_2 + b_{out} \\
A_{out} &= f_{out}(Z_{out}) \\
Loss &= f_{loss}(A_{out}, y)
\end{aligned}
$$

Suppose you know the following: Your inputs are one-hot word vectors with a vocabulary of 15,000; your batch size is 200; the hidden dimension of the first layer is 512; the hidden dimension of the second layer is 256; and your outputs are vectors of length 10. What are the dimensions of the following?

A_{in}

W_1

b_1

A_1

W_2

b_2

W_{out}

A_{out}