NLP - First Exam Fall 2013

Answer 3 of the following 5 questions for 10 points each.

Q1
Hidden Markov models are used successfully to improve POS tagging. Discuss in a few paragraphs the difficulties that make their application to WSD less successful.

Given the following Definite Clause Grammar:

```
s-->np, vp.
vp-->v, o₁, p, o₂.
np-->d, n.
o₁-->d, n₁.
o₂-->d, n₂.
n-->[thief].
n-->[thieves].
d-->[the].
v-->[gives].
v-->[give].
n₁-->[bitcoin].
n₁-->[bitcoin].
n₁-->[diamond].
n₁-->[diamond].
p-->[to].
n₂-->[charity].
n₂-->[policeman].
n₂-->[policemen].
```

Q2
add arguments that ensure singular-plural attribute matching.

Q3
add arguments that build a syntax tree after recognizing or generating a sentence.
Q4
You have the following information from a corpus:

- The corpus contains a total of
  - 3,000 words,
  - 500 unique words,
  - 1,000 unique bigrams.
- "I will" occurs 150 times
- "I" occurs 400 times
- "will" occurs 250 times

What is the bigram probability for "I will" (conditional probabilities, no smoothing)? Explain how you computed the result, what formulas you have used.

Q5
Given the following probabilities

\[
\begin{align*}
P(\text{sleep|NN}) &= 0.0005 \\
P(\text{sleep|VB}) &= 0.001 \\
P(\text{NN|DT}) &= 0.4 \\
P(\text{VB|DT}) &= 0.01
\end{align*}
\]

and assuming we know that "the" is a DT, what is the most likely tag for the word "sleep" in the text "The sleep of the reason brings to life the monsters"? Explain why.