Question 1.

Read the news story in the box
("Green-Eyed Monster " by Corey Hajim,
http://www.fortune.com/fortune/smallbusiness/offhours/articles/0,15114,1133509,00.html?promoid=cnn)

> Doug Lambrecht is on a mission: to convince diners of the benefits of real wasabi. Most sushi restaurants, both in the U.S. and Japan, do not serve the genuine article. That green stuff next to your spicy tuna roll is usually a combination of horseradish, mustard extract, and food coloring. Genuine wasabi is expensive (the plants are hard to cultivate) and tastes sweeter, with less concentrated heat. Lambrecht's company, Real Wasabi , based in Hilton Head, S.C., which he launched with two partners a year ago, aims to spread the green truth.
>
> An independent investment advisor, Lambrecht, 54, has always loved wasabi, to the point where he once put a wasabi vanity plate on his green BMW. After tasting the real thing seven years ago, he started researching the plant's putative health benefits. Wasabi has antibacterial qualities (which is why Japanese chefs first paired it with raw fish—to protect diners from microbes if the fish had gone bad), and some holistic health experts claim that it strengthens the immune system, reduces mucus, fights cancer, and detoxifies the liver and digestive system.

a (5 points) Identify and name at least two RST relationships in the story. For each of them, specify its nucleus/nuclei and/or satellite(s).

b (5 points) Discuss what heuristic(s) indicate that the word "it" (in the phrase "… that it strengthens …" in the last sentence of the story) should associate with "wasabi" and not with some other potential antecedent such as "raw fish".

c (5 points) Draw a dependency tree for the first sentence of the second paragraph ("An independent …"). You don't need to labels of the edges.

d (5 points) Show an instance of a long distance syntactic dependency within a sentence.

e (5 points) Discuss how a feature-based grammar can be used to ensure agreement in long-distance relationships. Use the sentence in part d. to illustrate your approach.

f (5 points) Find an instance in which a bigram HMM tagger will tag a given word with the wrong part of speech while a trigram HMM tagger would get that same word right.

g (5 points) Represent the sentence "Real Wasabi will prevent you from getting sick if you eat bad fish" in first order logic.

Question 3.

Give an example and/or description for each of the following phenomena and/or techniques:

a (5 points) unification

b (5 points) reification

c (5 points) interlingua

d (5 points) lexicalized grammar

Question 15.

Order the following syntactic features in decreasing order by salience (according to the Lappin/Leass algorithm for anaphora resolution): direct object (accusative), indirect object, subject, recency.

Question 30.

Give three problems that arise in language generation which don't exist in language understanding (parsing).

Question 38.

One of the main problems in discourse analysis is pronoun resolution (interpretation). List five types of syntactic and/or semantic preferences in pronoun interpretation. Give an example of each.

Question 39.

Give five examples of noun phrases with different syntactic structure. Explain why they are all noun phrases. You can refer to the principle of constituency.

Question 54.

Define the following terms and give appropriate examples.

2. Transformation-based part-of-speech tagging

3. Phrasal verbs

4. Lexicon

5. Anaphoric relations (or anaphora)

Question 55.

In the sentence "I want to book a flight to Los Angeles.", the word "book" should be tagged as a verb and not as a noun, although in English it is more often used as a noun than a verb. Describe what knowledge a part of speech tagger needs to have to be able to correctly tag "book" as a verb in the example sentence.

Question 72.

Consider the following regular grammar for NP:

(DT) (CD) JJ* (VBG) NN* NN

Where NN is either NNS or NNP.

Write an equivalent CFG.

Then convert the CFG to CNF (Chomsky Normal Form)

Question 73.

Why are “mildly context-sensitive grammars” like Tree Adjoining Grammars, Inversion Grammars, and CCG so important in NLP?

Question 97.

Represent the following sentences using event logic:

a (3 points) When the phone rang, Jeremy had eaten.

b (3 points) When the phone rang, Jeremy was eating.

Question 201

Part #4 — 25 points

Consider a PCFG with the following rules

```
S → V N
S → D N
D → a
D → the
N → dog
V → saw
V → like
```

and the following parameters:

$q(\texttt{S} \rightarrow \texttt{V N}) = 0.6$
$q(\texttt{S} \rightarrow \texttt{D N}) = 0.4$
$q(\texttt{D} \rightarrow \texttt{a}) = 0.2$
$q(\texttt{D} \rightarrow \texttt{the}) = 0.8$
$q(\texttt{N} \rightarrow \texttt{dog}) = 1$
$q(\texttt{V} \rightarrow \texttt{saw}) = 0.6$
$q(\texttt{V} \rightarrow \texttt{like}) = 0.4$

For any sentence x, define $\mathcal{T}(x)$ to be the set of parse trees for x, and define

$$p(x) = \sum_{t \in \mathcal{T}(x)} p(t)$$

where $p(t)$ is the probability of parse tree t under the PCFG shown above.

Question 7 (10 points) List all sentences x such that $p(x) > 0$, where $p(x)$ is defined through the above PCFG. For each sentence, write down its probability.

Question 203 PP

Part #1 — 10 points

We define a PCFG where the non-terminal symbols are $\{S, A, B\}$, the terminal symbols are $\{a, b\}$, and the start non-terminal (which is always at the root of the tree) is S. The PCFG has the following rules:

Rule	Probability
$S \rightarrow AB$	0.2
$S \rightarrow AA$	0.3
$S \rightarrow BS$	0.5
$A \rightarrow a$	0.7
$A \rightarrow b$	0.3
$B \rightarrow a$	0.2
$B \rightarrow b$	0.8

Question 1 — 10 points

For the input string *abab*, draw two possible parse trees. Show how to calculate the probabilities of those trees.

Question 207

1. **(5 points)** Consider the use of the word "*dumped*" in the following examples

   a) *Boulder officials **dumped** the use of the weed killer RoundUp in public spaces.*

   *b) Authorities suspect that someone **dumped** an unknown quantity of gas into the sewer system,*

   Which ***one*** of the following terms would best describe the relationship between the senses of the word *dumped* in these two sentences?

   i. Homonymy

   ii. Polysemy

   iii. Inflection

iv. None of the above

Question 208

6. Consider the following Dr. Seuss rhyme. See extra sheet.

One fish two fish red fish blue fish black fish blue fish

a) **(5 points)** Show a table with the bigram counts (as in the text) for this corpus. Given this table, give the **P(fish|two)** and **P(black|fish).**

Question 212

Consider the following bilingual (French-English) corpus.

maison blanche
white house

blanche
white

Considering only the following three alignment types:

| | | X

Start with a uniform distribution for $t(e_i|fj)$. Show the values of $t(e_i|fj)$ after two iterations of the EM algorithm.