**Computational Linguistics**
**LING 409**

**Spring 2013**

**MIDTERM (60 pts)**

**Question 1 (5 points)**

True/False: the Brill tagger is an example of unsupervised learning. Please explain your answer in one or two sentences. Really, please do not write a book on any of these – remember, I have to read them all!

**Question 2 (5 points)**

What (maximum likelihood) formula is used to estimate trigram probabilities $P(w_n|w_{n-1},w_{n-2})$ using a corpus. Don't worry about smoothing or backoff.

**Question 3 (5 points)**

Give 2 examples of sentences that fit this part of speech sequence:

NN RB VBZ AT NNS PPS VBZ

**Question 4 (5 points)**

True or false: The more completely your system can memorize the training data the better it will perform in actual usage. Briefly explain your answer (250 words max!).

**Question 5 (5 points)**

Give an example of each of the following instances of morphological derivation: verb → noun; adjective → noun; adjective → adverb; adjective → verb.

**Question 6 (10 points)**

Calculate the bigram probability[1] of the sentence "I want a British lunch", given the following unigram and bigram counts:

Bigram counts:

| | **<s>** | **I** | **want** | **a** | **British** | **lunch** |
|---|---|---|---|---|---|---|
| **<s>** | 0 | 859 | 357 | 452 | 22 | 57 |
| **I** | 0 | 8 | 1087 | 0 | 0 | 10 |
| **want** | 0 | 3 | 0 | 60 | 80 | 120 |
| **a** | 0 | 0 | 0 | 2 | 42 | 205 |
| **British** | 0 | 0 | 10 | 0 | 3 | 72 |
| **lunch** | 0 | 4 | 0 | 0 | 4 | 0 |

Unigram counts:

| | |
|---|---|
| **<s>** | 10,000 |
| **I** | 3,437 |
| **want** | 1,215 |
| **a** | 6,342 |
| **British** | 359 |
| **lunch** | 2,768 |

**Question 7 (5 points)**

Which of the following statements about two arbitrary deterministic finite-state automata A and B is the most accurate?

(a) The concatenation of A and B is a finite-state automaton.

(b) There exists at least one string (including the empty string ε) which is accepted by both A and B.

(c) Both (a) and (b) are true.

(d) Neither (a) nor (b) are true.

(e) If (a) above is true, then it follows that (b) above is also true.

---

1 You are not required to do all the calculations. Just show the formula with the counts (i.e., leave your answers in fractional form)

**Question 8 (5 points)**

Which of the followings statements about the automaton shown in Figure 1 is/are true?

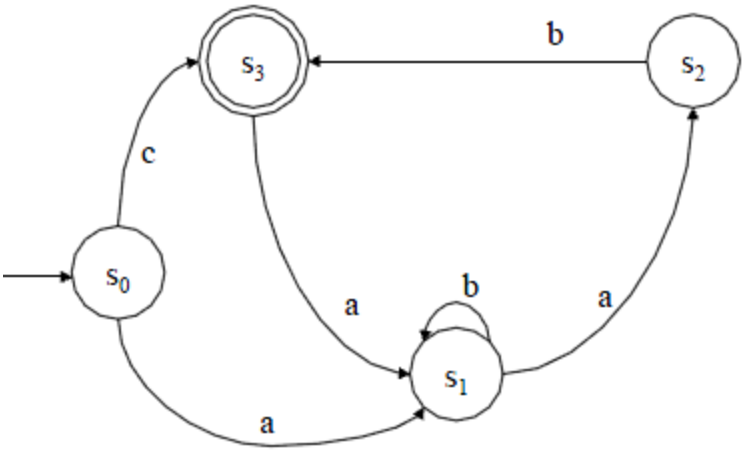


Figure 1: Automaton for Question 8.

(a) All strings that it accepts have an even number of "a"'s.

(b) All strings that it accepts end in "b".

(c) All strings that it accepts must contain at least one "c".

(d) Both (a) and (b) above.

(e) Both (a) and (c) above.

**Question 9 (10 points)**

Using a chart, take an APPLE and make some PIE from it. What is the total cost (edit distance)? Assume equal costs for the three basic edit operations.

**Question 10 (5 points)**

Imagine we initialized the Ice Cream HMM with the following parameters. Which statement is most true about these initial settings. Please explain your answer in a sentence or two.

| | p(…\|C) | p(…\|H) | p(…\|START |
|---|---|---|---|
| p(1\|…) | 0.20 | 0.20 | |
| p(2\|…) | 0.60 | 0.60 | |
| p(3\|…) | 0.20 | 0.20 | |
| p(C\|…) | 0.45 | 0.45 | 0.50 |
| p(H\|…) | 0.45 | 0.45 | 0.50 |
| p(STOP\|…) | 0.10 | 0.10 | 0.00 |

(a) The model would learn the weather pattern in a few iterations.

(b) The model would learn the weather pattern in a massive number of iterations.

(c) The model will never learn the weather pattern.

(d) The model will never converge on a solution.

(e) Both (c) and (d) above.