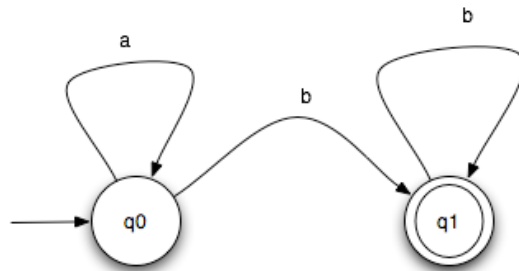CSCI 5832 Spring 2007 Quiz 1

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) True or False:** You've heard the statement "Winter in Boulder is not
   normally like this" too many times.
2. **(10 points)** Describe two **kinds** of evidence used to determine the parts of speech
   of words (use the back). I'm not talking here about tagging per se. I'm concerned
   with the basic nature parts of speech. Hint: don't say *meaning*.

3. **(10 points)** Give a regular expression for the language *L1* that consists of strings
   that begin **with at least 2 a's** and then continue (and end) with 0 or more **b's.**

4. *(10 points)* Consider the following
   FSA. Describe the relationship
   between the language accepted by
   this machine and *L1*.

5. *5.* Consider the string *aaaabbb* as a training corpus.

   a) **(5 points)** Give a table of counts
   for the letter **bigrams** that occur
   given this string (eg. assume that
   row_a, col_b contains the count for
   the bigram "ab".

	a	b
a		
b		

   b) **(10 points)** Give the table of
   counts that results from smoothing
   the original counts using Good-
   Turing smoothing. Assume that
   counts 3 and greater are reliable and don't need smoothing.

	a	b
a		
b		