1. In a discriminative model, where $x$ is an observed variable, and $y$ is a hidden variable, the following probability distribution is optimized:

   (c) $P(y|x)$ [72% correct]

2. For probabilistic context-free grammars, the relative frequency estimate of the rule probabilities is the:

   (a) maximum likelihood estimate [52% correct]: we are talking about a standard PCFG here, which is a generative model, so we get the MLE when we estimate the parameters using relative frequency. To get the MCLE we’d need a discriminative training procedure. The Bayesian estimate for PCFGs is actually quite involved, we haven’t discussed it in the lectures.

3. The following is not a property of discriminative parsing:

   (d) it involves a stack and priority queue [68% correct]: This is property of Roark’s incremental parser. You could imagine a discriminative parser with stack and priority queue, but it’s not a necessary property of it.

4. The purpose of the loss function in the Johnson and Charniak discriminative parser is:

   (b) to compute the probability of the best parses, relative to the gold-standard parses [12% correct]: The parameter learning algorithm adjusts the feature weights (using the loss function), not the loss function itself. (And there is a choice of different learning algorithms for MaxEnt, not covered in the lecture.)

5. In parsing, what is strong incrementality?

   (c) the parser computes only connected analyses, i.e., partial analyses cannot contain unconnected nodes [40% correct]

6. What is a language model?

   (b) a model that assigns a probability to a string of words [80% correct]

7. Roark’s incremental parser is using the following parsing algorithm:

   (b) top-down [40% correct]: The answer is in the title of Roark’s paper! So that means that 60% you haven’t read it :-). Seriously, please revise parsing strategies based on Jurafsky and Martin or your notes from Inf2a/FNLP/ANLP.

8. In Roark’s parser, what is the look-ahead probability?

   (c) the probability that the stack $A_n^0$ expands to the next word $w_i$ by applying an arbitrary number of grammar rules [72% correct]
9. Why is it necessary to interpolate the prefix probabilities returned by an incremental parser with those of an n-gram model?

(d) all of the above [36% correct]: This was a bit of a humorous question, but it is indeed impossible to publish a paper on language modeling without comparing to (and interpolating with) an n-gram model. As explained in the lecture, n-gram models don’t require annotated training data, so you can simply improve their performance by training on a larger amount of text.

10. In the following sentence, what is the frame evoking element?
[Google] [snapped up] [YouTube] [for $1.65 billion.]

(b) snapped up [100% correct]

11. Which of the following are sub-tasks of semantic role labeling (tick all that apply)?

(a) identifying frame evoking elements [84% correct]
(c) labeling frame elements [76% correct]
(d) finding the boundaries of frame elements [60% correct]


(b) Bayes classifier [28% correct]: they compute the probability of a class given the features, this is equivalent to a Bayes classifier (it would be naive Bayes if conditional independence between all features was assumed, but they don’t do this); linear interpolation and backoff are used to smooth the probability estimates.