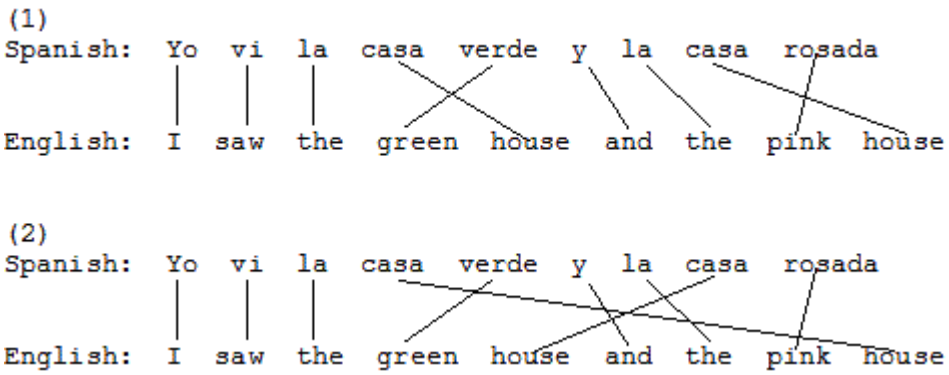
CS224N: Natural Language Processing
Spring 2008
Weekly Quiz

Quiz 3

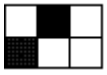
1: Consider the following sentence pairs:



Which of the following is true?

- ☐ The Model 1 scores of (1) and (2) are the same
- ☐ The Model 2 scores of (1) and (2) are the same
- ☐ Both of A and B are true
- ☐ Neither A nor B is true

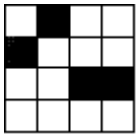
2: How many phrases can you extract from the following alignment?



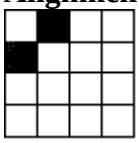
- ☐ <2
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ >5

3: What happens when the output of alignment merging is more sparse? Here is a concrete example for you to think about:

Alignment 1: (denser)



Alignment 2: (sparser)



- You can extract more (or at least the same number of) phrases from the denser alignment
- You can extract more (or at least the same number of) phrases from the sparser alignment
- All of the above
- None of the above

4: Which of the following statements is true?

- The so-called phrase-based statistical MT system does not use linguistic phrases.
- Naive Bayes classifiers make a strong conditional independence assumption that the probability of observing the conjunction of attributes is equal to the product of the individual probabilities given the class is known.
- The independence assumption that Naive Bayes is making is too strong, therefore it almost always fails to predict the right class.
- A and B are true
- B and C are true
- None of the above

5: Which of the following POS parses is most correct for the given sentence:

He was a very judicious person.

- He/NN was/VBZ a/AT very/RB judicious/JJ person/NN.
- He/NP was/VBZ a/AT very/QL judicious/JJ person/NNS.
- He/PRP was/VBD a/AT very/QL judicious/JJ person/NN.
- He/PRP was/VBD a/AT very/RB judicious/JJ person/NNS.

6: Which of the following is true about log-linear models for MT?

- By computing probability as an exp(sum()) over features instead of a products of individual probabilities, they avoid the underflow problem of other MT models.
- Their framework allows for the easy inclusion of arbitrary features.
- They unfortunately require the use of a reverse translation model.

- All of the above.
- None of the above.

7: Which of the following is typically not a useful feature for performing NER:

- Character affixes
- Hapax legoma
- Gazeteers
- Word shape
- Stemmed lexical items
- None of the above

8: Which of the following is not needed to implement the Viterbi algorithm:

- Emission probabilities of outputs versus states
- Prior probabilities over the states and outputs
- Transition probabilities of states versus states
- None of the above

9: For the example of the Viterbi algorithm in Fig. 5.18 of J&M (p. 148), we can calculate the maximum transition probabilities for each of the remaining states in the "want" column, i.e., before multiplying in the emission probabilities p("want"|NN), p("want"|TO), etc. Then which of the following will not be one of the maximum pre-emission probabilities assigned to one of the three remaining states, NN, TO, and PPSS:

- 0.250e-5
- 0.350e-5
- 1.975e-5
- 3.000e-5
- None of the above

10: Which of the following task is considered as an information extraction task

- Temporal analysis: figuring out when the events in a text happened and how they relate to each other in time.
- Named entity recognition: detect and classify all proper names in a text
- Reference resolution: link and cluster mentions of named entities into sets that correspond to the entities behind the mentions
- All of the above