

Homework #3

Use the online version of the Stanford Statistical Parser to parse the following sentences:

<http://nlp.stanford.edu:8080/parser/>

- (a) *the bartender prepared a drink and a snack for the customer*
- (b) *the bartender prepared a drink with tequila and orange juice*
- (c) *the customer ordered the drink of the day and a tosti*

Make sure you understand the bracketed notation produced by the parser. Transform the bracketed parse output for each sentence to a parse tree (you can ignore the node ROOT). You will need these trees to answer some of the questions below.

1. Give a CFG that contains the rules required to derive the parse trees produced by the Stanford parser for these three sentences (ignore the node ROOT and consider S the start symbol). 1 points
2. Recall that a sentence is syntactically ambiguous with respect to a grammar if the grammar can assign more than one parse tree to it. According to the CFG you have given in (1), are these sentences syntactically ambiguous? If so, how many parse trees are possible for each sentence and what type of structural ambiguity does each sentence exhibit? Justify your answer carefully. 3 points
3. Although these three sentences may be structurally ambiguous, we have clear intuitions regarding their correct structure given their most plausible meaning. In fact, as you may have experienced when answering (2), for humans it is often difficult to realise that a sentence can be structurally ambiguous because usually one single structure stands out as obvious. In contrast, computers have a hard time dealing with ambiguity. 4 points

While the syntactic structure that the Stanford parser assigns to the first sentence makes a lot of sense, you should have noticed that the syntactic structure that the parser outputs for the second and third sentences is not in line with their most likely meaning.

The CFG you have given in (1) should be able to assign the correct parse trees to each of the sentences. Give these three correct trees and consider them to be a small treebank. Convert your CFG into a PCFG by augmenting each rule with a probability estimated from this treebank. Indicate how you have calculated the probability of each rule.

4. Use your PCFG to calculate the probability of the parse tree given by the Stanford parser for the sentence *the bartender prepared a drink with tequila and orange juice* and the probability of the correct parse tree you have given for this sentence in (3) above. Indicate how you have calculated the probabilities. Are these probabilities different? If so, could the difference explain why the Stanford parser selects the intuitively incorrect syntactic structure for this sentence? 2 points
5. Consider the correct trees in the treebank you have constructed in (3) to be the gold standard 4 points

¹Assignment created by Dr. Tejaswini Deoskar

parse trees for these sentences. Evaluate the output of the Stanford parser for that sentence against this gold standard. Report precision, recall, and F1. Indicate how you have calculated each of these measures.