

Zachary Harris

3/2/18

Hw3

Task 1.

While looking at the non-terminal symbols in grammar2 and the terminal symbols in the lexicon all the probabilities have been set to 1. At any given level of the parse tree the pcfg only considers the probability of that transition from $A \rightarrow B$ which would be the implementation of unigram probabilities. The product of the all probabilities is will be the probability of the complete parsed sentence.

Task 2.

When running the cfgparser with grammar1 and the lexicon on the example sentences, only 2 of the generated sentences are successfully parsed, while the rest result in failures. When running the combined grammar1, grammar2, and lexicon with the example sentences all of the example sentences were successfully parsed. This is because there exists more rules in the combination of the grammars for the transitions between the non-terminals. By using multiple grammars there become more ways in which complete sentences, even if one of them is shorter with skewed probabilities.

Task 3.

The generated sentences using grammar1 are the most coherent but limited in vocabulary. This is because they have limited rules for the transitioning of words. Although the probabilities for the transitions are more accurate than grammar2 there are not enough of them to account for the whole use of non-terminals in the lexicon.

The sentences generated in grammar2 have a large use of vocabulary but are largely nonsensical. This is because there is a larger list of rules for the transitions between words but the rules are poorly defined with all the probabilities set to 1. The grammar2 can make use of most all of the terminal transitions to words in the lexicon because of the larger rule base.

The sentences generated from the combination of the grammars can make use of a large vocabulary and can still follow the coherency of the first grammar. Because the rules are still poorly defined for grammar2 the ability to use more word in a sentence means the generated sentences will be less understandable but more diverse than using grammar1 alone, but more understandable and less diverse than using grammar 2 alone.

Task 4.

The first part of making the new grammars was to create a new set of rules and probabilities for the non-terminal transitions. This was done using a table example in the book, while the probabilities were adjusted to try and make more coherent sentences given the word list available. The grammar 2 that was given was incorporated to ensure every possible sentence was able to be parsed even if it made no logical sense and raised the cross entropy.

The next step was to create a lexicon from the given words. The lexicon provided was used as a base line and non-terminals were added to allow more options for combinations of words. Adjectives, and

pronouns were added which about half of the Misc category fell into. There were too many Misc categories to go through not enough time to evaluate them all due to obligations for other classes. So the bare minimum was met to get a lower cross entropy than that of the example. When changing the non-terminal ->terminal definitions without adding additional words the generic grammar needed to be updated in ensure it would continue to catch all cases.

Task 5.

9 out of 20 sentences were grammatical.