**LING 571 – HW3**

This homework comprises of several parts. First, a Probabilistic Context-Free Grammar was trained from 514 sentences taken from the Air Travel Information System Domain. The program reads in the sentences and by the means of induction, processed them into PCFG in CNF. The approach adopted here

Using the PCFG created earlier, the program then generates a Probabilistic CKY table. The approach taken here is, every [i][j] contains a linked-list of all possible parses with each parse being associated with a probability. For each sentence, when the program exits at [0][lens] (where lens is the length of sentence), the probabilities would have been calculated for all completed parses. The best parse (“TOP” non-terminal with the highest probability) will be return and the program then backtracks to reconstruct the best parsed sentence in bracketed form.

Attempts to improve the results were made in the final part of the program. The primary objective was to improve the processing performance with minimal or no loss in accuracy. The approach taken was, instead of looping through all possible parses, those with lower probabilities at every [i][j] were pruned. With lesser parses at each [i][j] to process and probabilities to calculate, some efficiencies should be expected.

*HW3 Submitted By:*

*Wee Teck Tan*