NLP Assignment - 4

Report | Kaustav Vats (2016048)

Assumptions and Observations

1. Eliminate start symbol from RHS. This case is not present in given grammar, no such grammar would be given which has this case.
2. Eliminate null - This case is not present in given grammar, no such grammar would be given which has this case.
3. Eliminate terminals from RHS if they exist with other terminals or non-terminals = This case is not present in given grammar, no such grammar would be given which has this case.

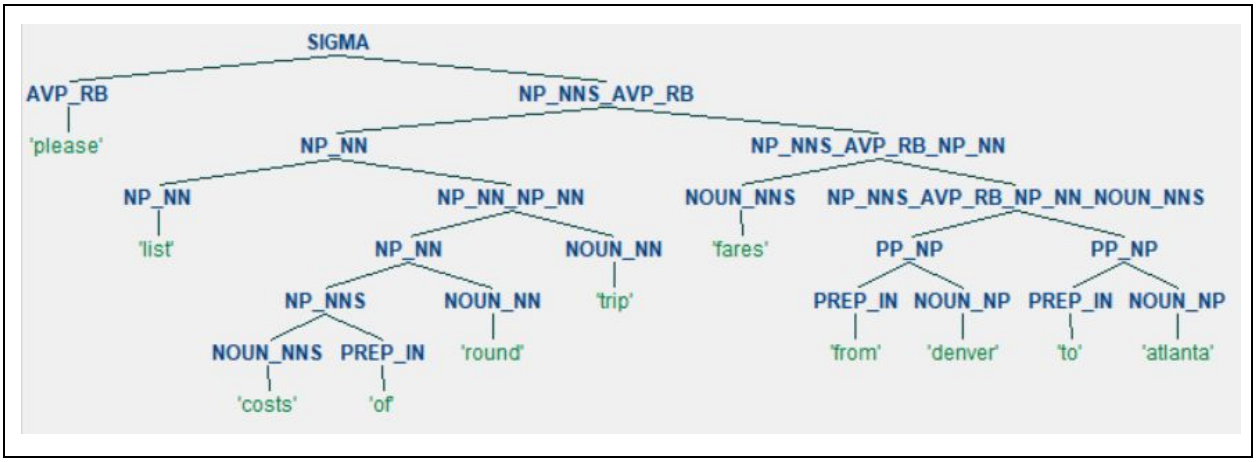
Description of CFG 2 CNF converter:

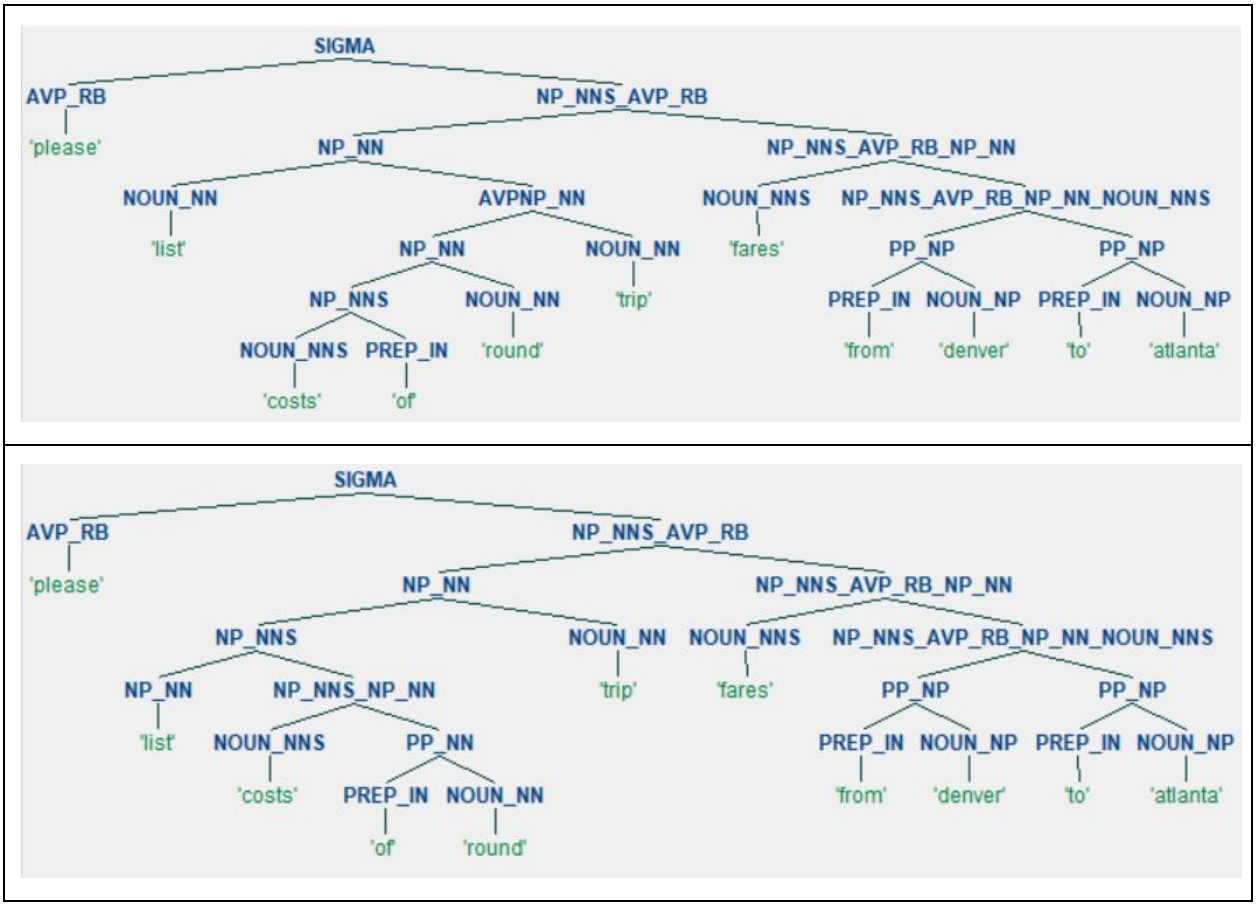
1. Large rule elimination- Modified rules like A -> B C D to A -> NR0 D and NR0 -> B C
2. Eliminating Unit rules - A -> B, B->C changes to A -> C
3. Some optimization like removing duplicate ex - A-> B C | B C | D changes to A -> B C | D

To convert CFG to CNF:

1. Created my own implementation of converting CFG to CNF by few set of rules based on observations made from atis grammar.
2. Also used inbuilt CNF converter present in NLTK Grammar Library.

Created Node class to store parsed sentence in tree form.
For every sentences, ran CKY algorithm to find all parse tree and also showed count of no of parse trees.
Below are the parse trees for this sentence -> please list costs of round trip fares from denver to atlanta. [Sentence 91]





Some of the files description-

1. CKY-Parser.py: Contains all the code from the beginning,
2. CNF.txt: Contains CNF form of atis.cfg
3. Parse Count.txt: Contains no of parse trees in two columns: a.
    a. First column contains all parse tree count
    b. Second column contains parse count from my implementation.

For better understanding of the code, open jupyter notebook. Lots of comments are mentioned in the code.

To test for custom grammar and custom sentence. Changes are only required in 2nd cell of the jupyter notebook.