NLP ASSIGNMENT 4

Here I have first converted the given CGF into CNF and then created the dictionary for the grammar.

The file convert_grammar.py converts the grammar from CFG to CNF.

The file create_dict.py creates a dictionary of the grammar and also pickles the dictionary.

The dictionary contains the right-hand side of the productions as keys and the value as a list of all corresponding Left hand side of the productions for the key.

In case of terminals the dictionary key is a string and in case of no.-terminals the dictionary key is a tuple. I have created the pickle file of the dictionary and then used it further.

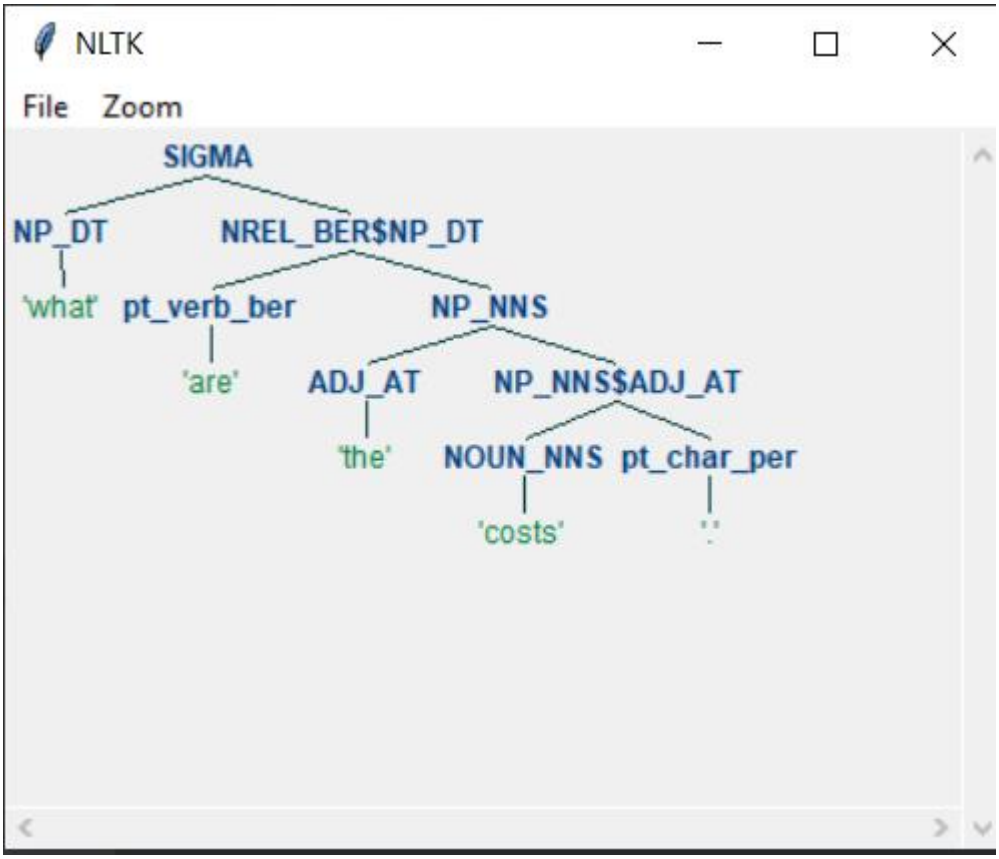
(1.) Creating a file containing the list of ATIS sentences with tab separated no. of parse trees.

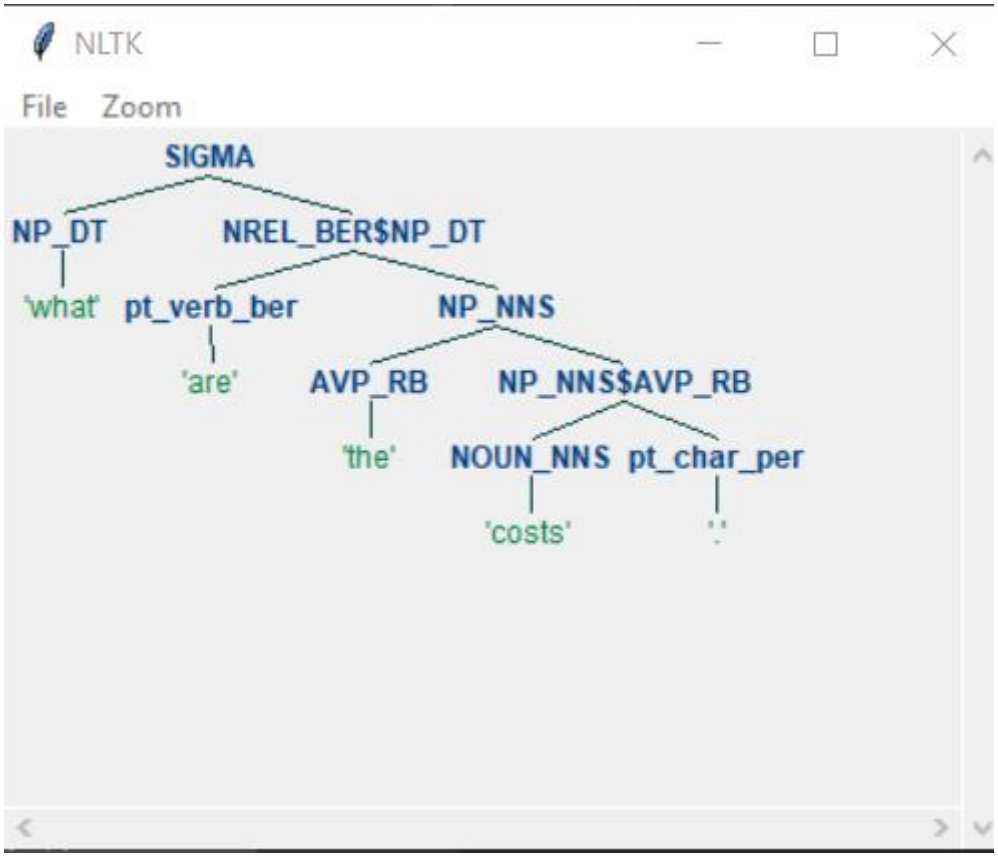
The file ParseTreesCount.py finds the no. of all possible parse trees for each sentence and writes the sentences and tab separated no. of parse trees of the sentence into a file CalculatedParseTrees.txt.
To write into the file just this python file needs to be run.

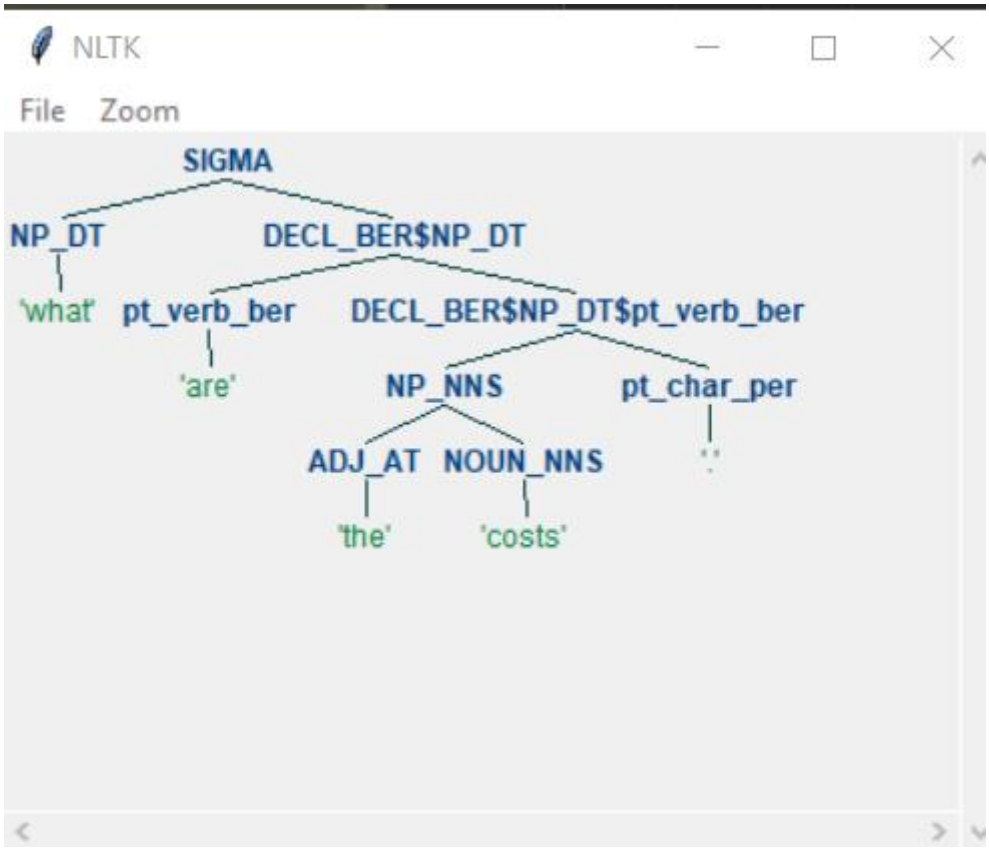
(2.) Pictures of the parse tree for an ATIS sentence with a number of parses p such that 1<p<5.

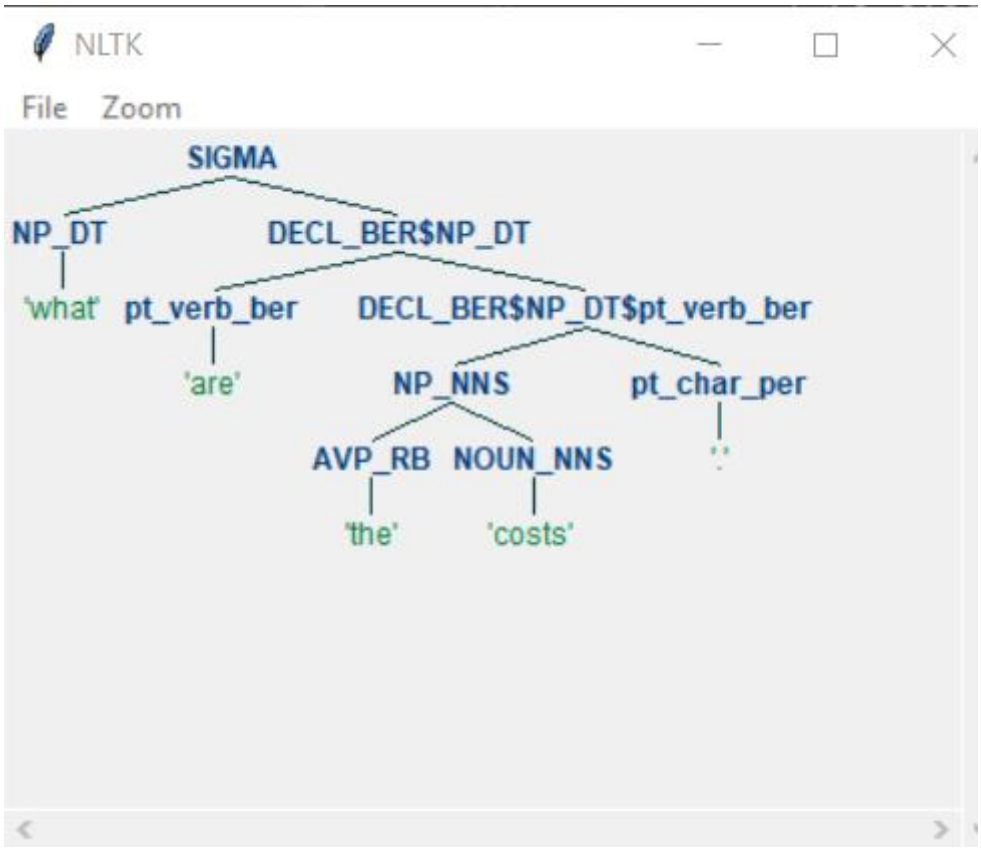
The file parser.py finds the count of parse trees of the sentence whose index is specified and then also draws all possible parse trees for that sentence. The file parser,py needs to be just run to generate the parse trees.

The no. of parse trees for the sentence : what are the costs .
is 4 and the parse trees generated by the parser.py for this sentence are shown below:


NLTK
File Zoom
SIGMA
NP_DT
NREL_BER$NP_DT
'what'
pt_verb_ber
NP_NNS
'are'
ADJ_AT
NP_NNS$ADJ_AT
'the'
NOUN_NNS
pt_char_per
'costs'
'.'


NLTK
File Zoom
SIGMA
NP_DT
NREL_BER$NP_DT
'what'
pt_verb_ber
NP_NNS
'are'
AVP_RB
NP_NNS$AVP_RB
'the'
NOUN_NNS
pt_char_per
'costs'
'.'


NLTK
File Zoom
SIGMA
NP_DT
DECL_BER$NP_DT
'what'
pt_verb_ber
DECL_BERNP_DTpt_verb_ber
'are'
NP_NNS
pt_char_per
ADJ_AT
NOUN_NNS
'.'
'the'
'costs'


NLTK
File Zoom
SIGMA
NP_DT
DECL_BER$NP_DT
'what'
pt_verb_ber
DECL_BERNP_DTpt_verb_ber
'are'
NP_NNS
pt_char_per
AVP_RB
NOUN_NNS
'.'
'the'
'costs'