Assignment 3 (Take Home)

17000912

# Grammar

S -> S\_SG\_1 | S\_PL\_1 | S\_SG\_2 | S\_PL\_2 | S\_SG\_M | S\_SG\_F | S\_SG\_N | S\_PL  
  
S\_SG\_1 -> NP\_SG\_1 VP\_SG\_1  
S\_PL\_1 -> NP\_PL\_1 VP\_PL\_1  
  
S\_SG\_2 -> NP\_SG\_2 VP\_SG\_2  
S\_PL\_2 -> NP\_PL\_2 VP\_PL\_2  
  
S\_SG\_M -> NP\_SG\_M VP\_PRES\_SG | NP\_SG\_M VP\_PAST\_SG\_M  
S\_SG\_F -> NP\_SG\_F VP\_PRES\_SG | NP\_SG\_F VP\_PAST\_SG\_F  
S\_SG\_N -> NP\_SG\_N VP\_PRES\_SG  
S\_PL -> NP\_PL\_M VP\_PRES\_PL | NP\_PL\_M VP\_PAST\_PL | NP\_PL\_F VP\_PRES\_PL | NP\_PL\_F VP\_PAST\_PL  
  
NP -> NP\_SG\_M | NP\_SG\_F | NP\_SG\_N | NP\_PL\_M | NP\_PL\_F  
  
NP\_SG\_1 -> PRE\_N N\_SG\_1 | N\_SG\_1  
NP\_PL\_1 -> PRE\_N N\_PL\_1 | N\_PL\_1  
  
NP\_SG\_2 -> PRE\_N N\_SG\_2 | N\_SG\_2  
NP\_PL\_2 -> PRE\_N N\_PL\_2 | N\_PL\_2  
  
NP\_SG\_M -> PRE\_N N\_SG\_M | N\_SG\_M  
NP\_PL\_M -> PRE\_N N\_PL\_M | N\_PL\_M  
NP\_SG\_N -> PRE\_N N\_SG\_N | N\_SG\_N  
NP\_SG\_F -> PRE\_N N\_SG\_F | N\_SG\_F  
NP\_PL\_F -> PRE\_N N\_PL\_F | N\_PL\_F  
  
PRE\_N -> Det | ADJP | Det ADJP  
ADJP -> Deg Adj | Adj  
  
VP\_SG\_1 -> PRE\_V V\_SG\_1 | V\_SG\_1  
VP\_PL\_1 -> PRE\_V V\_PL\_1 | V\_PL\_1  
  
VP\_SG\_2 -> PRE\_V V\_SG\_2 | V\_SG\_2  
VP\_PL\_2 -> PRE\_V V\_PL\_2 | V\_PL\_2  
  
VP\_PRES\_SG -> PRE\_V V\_PRES\_SG | V\_PRES\_SG  
VP\_PAST\_SG\_M -> PRE\_V V\_PAST\_SG\_M | V\_PAST\_SG\_M  
VP\_PAST\_SG\_F -> PRE\_V V\_PAST\_SG\_F | V\_PAST\_SG\_F  
VP\_PRES\_PL -> PRE\_V V\_PRES\_PL | V\_PRES\_PL  
VP\_PAST\_PL -> PRE\_V V\_PAST\_PL | V\_PAST\_PL  
  
PRE\_V -> NP | ADVP | NP ADVP  
ADVP -> Deg Adv | Adv  
  
Deg -> 'ඉතා'  
Det -> 'ඒ' | 'ෙම්' | 'අර' | 'ඔය' | 'සමහර' | 'ඇතැම්'  
Adj -> 'ලස්සන' | 'කැත' | 'මහත' | 'සුදු' | 'කලු' | 'ෙලොකු' | 'ෙපොඩි' | 'පුංචි' | 'උස'  
Adv -> 'ෙව්ගෙයන්' | 'ලස්සනට' | 'ෙහොඳට' | 'ඉක්මනින්' | 'ෙසෙමන්'  
  
N\_SG\_1 -> 'මම'  
N\_PL\_1 -> 'අපි'  
  
N\_SG\_2 -> 'නුබ' | 'ඔබ'  
N\_PL\_2 -> 'ඔබලා' | 'නුබලා'  
  
N\_SG\_M -> 'ඔහු' | 'බල්ලා' | 'මිනිසා' | 'ළමයා' | 'ෙකොල්ලා' | 'තාත්තා' | 'සර්පයා'  
N\_PL\_M -> 'ඔවුහු' | 'බල්ෙලෝ' | 'මිනිස්සු' | 'කොල්ලෝ'  
N\_SG\_N -> 'ගස' | 'මල' | 'අත්ත' | 'අඹ' | 'ෙගදර' | 'ෙපොත' | 'වතුර' | 'සඳ'  
N\_SG\_F -> 'ඇය' | 'බැල්ලි' | 'කෙල්ල' | 'අම්මා'  
N\_PL\_F -> 'ඔවුහු' | 'ගැහැණු' | 'කෙල්ලෝ'  
  
V\_SG\_1 -> V\_PRES\_SG\_1 | V\_PAST\_SG\_1  
V\_PL\_1 -> V\_PRES\_PL\_1 | V\_PAST\_PL\_1  
V\_PRES\_SG\_1 -> 'ෙදමි' | 'කමි' | 'ෙබොමි'  
V\_PRES\_PL\_1 -> 'ෙදමු' | 'කමු' | 'ෙබොමු'  
V\_PAST\_SG\_1 -> 'දුන්නෙමි' | 'කෑවෙමි' | 'බීවෙමි'  
V\_PAST\_PL\_1 -> 'දුන්නෙමු' | 'කෑවෙමු' | 'බීවෙමු'  
  
V\_SG\_2 -> V\_PRES\_SG\_2 | V\_PAST\_SG\_2  
V\_PL\_2 -> V\_PRES\_PL\_2 | V\_PAST\_PL\_2  
V\_PRES\_SG\_2 -> 'ගියෙහි' | 'කවෙහි' | 'බීවෙහි'  
V\_PRES\_PL\_2 -> 'ගියෙහු' | 'කවෙහු' | 'බිවෙහු'  
V\_PAST\_SG\_2 -> 'කයි' | 'කියයි' | 'ගසයි'  
V\_PAST\_PL\_2 -> 'ගියෝය' | 'කෑවෝය'  
  
V\_PAST\_SG\_M -> 'ගියේය'  
V\_PAST\_SG\_F -> 'ඉල්ලීය'  
V\_PRES\_SG -> 'ෙදයි' | 'කයි' | 'කඩයි' | 'කියයි' | 'ගසයි' | 'සිටියි' | 'තිෙබ්'  
V\_PRES\_PL -> 'යති' | 'කති'  
V\_PAST\_PL -> 'ගියෝය' | 'කෑවෝය'

# Code

import sys  
import nltk  
import argparse  
import codecs  
  
  
def run(grammar\_file, input\_file):  
 grammar = nltk.data.load(**'file:'** + grammar\_file)  
 parser = nltk.RecursiveDescentParser(grammar)  
  
 file\_input = codecs.open(input\_file, **'r'**, **'utf-8-sig'**)  
  
 count = 1  
 for sent in file\_input:  
 sent\_split = sent.split()  
 print(**"----- Sentence:"**, count, **"-----"**)  
 print(sent)  
 count += 1  
  
 parsed = False;  
 for tree in parser.parse(sent\_split):  
 print(tree)  
  
 if not parsed:  
 parsed = True  
  
 if parsed:  
 print(**"**\n**Parse Success**\n**"**)  
 else:  
 print(**"**\n**Parse Failed**\n**"**)  
  
  
def main(arguments):  
 parser = argparse.ArgumentParser()  
  
 parser.add\_argument(**"-g"**, **"--grammar"**, type=str, default=**"grammar.cfg"**)  
 parser.add\_argument(**"-i"**, **"--input\_file"**, type=str, default=**"input.txt"**)  
  
 args = parser.parse\_args(arguments)  
  
 grammar = args.grammar  
 input\_file = args.input\_file  
  
 run(grammar, input\_file)  
  
  
if \_\_name\_\_ == **'\_\_main\_\_'**:  
 sys.exit(main(sys.argv[1:]))

## How to run

Above script takes two files **grammar.cfg** containing the grammar and **input.txt** containing the sentences as inputs. (Also, alternative file paths can be given by using arguments **-g** and **-i** for grammar file and input file respectively)

# Evaluation

### Correct sentences (20)

මම කමි  
මම ෙපොත ෙදමි  
මම බීවෙමි  
අපි අඹ කමු  
අපි අර අඹ කමු  
අපි අර අඹ ඉක්මනින් කමු  
අපි අඹ ඉතා ෙසෙමන් කමු  
අපි ෙපොත දුන්නෙමු  
නුබ ෙව්ගෙයන් කවෙහි  
නුබලා ෙගදර ගියෙහු  
ඔබ ෙගදර ගියෙහි  
ඔබලා වතුර බිවෙහු  
තාත්තා මල කඩයි  
අම්මා ෙගදර සිටියි  
ඔහු ෙගදර ගියේය  
ඇය ෙපොත ඉල්ලීය  
කොල්ලෝ ෙගදර යති  
කෙල්ලෝ අඹ කති  
කොල්ලෝ ෙගදර ගියෝය  
කෙල්ලෝ අඹ කෑවෝය

### Incorrect Sentences (10)

මම කමු  
අපි අඹ කමි  
නුබ ෙව්ගෙයන් ගියෙහු  
ඔබලා වතුර බීවෙහි  
තාත්තා මල ඉල්ලීය  
අම්මා ෙගදර ගියේය  
කොල්ලෝ ෙගදර ගියේය  
කෙල්ලෝ අඹ ඉල්ලීය  
මම ෙගදර ගියේය  
අපි අඹ කෑවෝය

# Results

### Output of the script

----- Sentence: 1 -----

මම කමි

(S

(S\_SG\_1 (NP\_SG\_1 (N\_SG\_1 මම)) (VP\_SG\_1 (V\_SG\_1 (V\_PRES\_SG\_1 කමි)))))

Parse Success

----- Sentence: 2 -----

මම ෙපොත ෙදමි

(S

(S\_SG\_1

(NP\_SG\_1 (N\_SG\_1 මම))

(VP\_SG\_1

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙපොත))))

(V\_SG\_1 (V\_PRES\_SG\_1 ෙදමි)))))

Parse Success

----- Sentence: 3 -----

මම බීවෙමි

(S

(S\_SG\_1

(NP\_SG\_1 (N\_SG\_1 මම))

(VP\_SG\_1 (V\_SG\_1 (V\_PAST\_SG\_1 බීවෙමි)))))

Parse Success

----- Sentence: 4 -----

අපි අඹ කමු

(S

(S\_PL\_1

(NP\_PL\_1 (N\_PL\_1 අපි))

(VP\_PL\_1

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N අඹ))))

(V\_PL\_1 (V\_PRES\_PL\_1 කමු)))))

Parse Success

----- Sentence: 5 -----

අපි අර අඹ කමු

(S

(S\_PL\_1

(NP\_PL\_1 (N\_PL\_1 අපි))

(VP\_PL\_1

(PRE\_V (NP (NP\_SG\_N (PRE\_N (Det අර)) (N\_SG\_N අඹ))))

(V\_PL\_1 (V\_PRES\_PL\_1 කමු)))))

Parse Success

----- Sentence: 6 -----

අපි අර අඹ ඉක්මනින් කමු

(S

(S\_PL\_1

(NP\_PL\_1 (N\_PL\_1 අපි))

(VP\_PL\_1

(PRE\_V

(NP (NP\_SG\_N (PRE\_N (Det අර)) (N\_SG\_N අඹ)))

(ADVP (Adv ඉක්මනින්)))

(V\_PL\_1 (V\_PRES\_PL\_1 කමු)))))

Parse Success

----- Sentence: 7 -----

අපි අඹ ඉතා ෙසෙමන් කමු

(S

(S\_PL\_1

(NP\_PL\_1 (N\_PL\_1 අපි))

(VP\_PL\_1

(PRE\_V

(NP (NP\_SG\_N (N\_SG\_N අඹ)))

(ADVP (Deg ඉතා) (Adv ෙසෙමන්)))

(V\_PL\_1 (V\_PRES\_PL\_1 කමු)))))

Parse Success

----- Sentence: 8 -----

අපි ෙපොත දුන්නෙමු

(S

(S\_PL\_1

(NP\_PL\_1 (N\_PL\_1 අපි))

(VP\_PL\_1

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙපොත))))

(V\_PL\_1 (V\_PAST\_PL\_1 දුන්නෙමු)))))

Parse Success

----- Sentence: 9 -----

නුබ ෙව්ගෙයන් කවෙහි

(S

(S\_SG\_2

(NP\_SG\_2 (N\_SG\_2 නුබ))

(VP\_SG\_2

(PRE\_V (ADVP (Adv ෙව්ගෙයන්)))

(V\_SG\_2 (V\_PRES\_SG\_2 කවෙහි)))))

Parse Success

----- Sentence: 10 -----

නුබලා ෙගදර ගියෙහු

(S

(S\_PL\_2

(NP\_PL\_2 (N\_PL\_2 නුබලා))

(VP\_PL\_2

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙගදර))))

(V\_PL\_2 (V\_PRES\_PL\_2 ගියෙහු)))))

Parse Success

----- Sentence: 11 -----

ඔබ ෙගදර ගියෙහි

(S

(S\_SG\_2

(NP\_SG\_2 (N\_SG\_2 ඔබ))

(VP\_SG\_2

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙගදර))))

(V\_SG\_2 (V\_PRES\_SG\_2 ගියෙහි)))))

Parse Success

----- Sentence: 12 -----

ඔබලා වතුර බිවෙහු

(S

(S\_PL\_2

(NP\_PL\_2 (N\_PL\_2 ඔබලා))

(VP\_PL\_2

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N වතුර))))

(V\_PL\_2 (V\_PRES\_PL\_2 බිවෙහු)))))

Parse Success

----- Sentence: 13 -----

තාත්තා මල කඩයි

(S

(S\_SG\_M

(NP\_SG\_M (N\_SG\_M තාත්තා))

(VP\_PRES\_SG (PRE\_V (NP (NP\_SG\_N (N\_SG\_N මල)))) (V\_PRES\_SG කඩයි))))

Parse Success

----- Sentence: 14 -----

අම්මා ෙගදර සිටියි

(S

(S\_SG\_F

(NP\_SG\_F (N\_SG\_F අම්මා))

(VP\_PRES\_SG

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙගදර))))

(V\_PRES\_SG සිටියි))))

Parse Success

----- Sentence: 15 -----

ඔහු ෙගදර ගියේය

(S

(S\_SG\_M

(NP\_SG\_M (N\_SG\_M ඔහු))

(VP\_PAST\_SG\_M

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙගදර))))

(V\_PAST\_SG\_M ගියේය))))

Parse Success

----- Sentence: 16 -----

ඇය ෙපොත ඉල්ලීය

(S

(S\_SG\_F

(NP\_SG\_F (N\_SG\_F ඇය))

(VP\_PAST\_SG\_F

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙපොත))))

(V\_PAST\_SG\_F ඉල්ලීය))))

Parse Success

----- Sentence: 17 -----

කොල්ලෝ ෙගදර යති

(S

(S\_PL

(NP\_PL\_M (N\_PL\_M කොල්ලෝ))

(VP\_PRES\_PL (PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙගදර)))) (V\_PRES\_PL යති))))

Parse Success

----- Sentence: 18 -----

කෙල්ලෝ අඹ කති

(S

(S\_PL

(NP\_PL\_F (N\_PL\_F කෙල්ලෝ))

(VP\_PRES\_PL (PRE\_V (NP (NP\_SG\_N (N\_SG\_N අඹ)))) (V\_PRES\_PL කති))))

Parse Success

----- Sentence: 19 -----

කොල්ලෝ ෙගදර ගියෝය

(S

(S\_PL

(NP\_PL\_M (N\_PL\_M කොල්ලෝ))

(VP\_PAST\_PL

(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ෙගදර))))

(V\_PAST\_PL ගියෝය))))

Parse Success

----- Sentence: 20 -----

කෙල්ලෝ අඹ කෑවෝය

(S

(S\_PL

(NP\_PL\_F (N\_PL\_F කෙල්ලෝ))

(VP\_PAST\_PL (PRE\_V (NP (NP\_SG\_N (N\_SG\_N අඹ)))) (V\_PAST\_PL කෑවෝය))))

Parse Success

----- Sentence: 21 -----

මම කමු

Parse Failed

----- Sentence: 22 -----

අපි අඹ කමි

Parse Failed

----- Sentence: 23 -----

නුබ ෙව්ගෙයන් ගියෙහු

Parse Failed

----- Sentence: 24 -----

ඔබලා වතුර බීවෙහි

Parse Failed

----- Sentence: 25 -----

තාත්තා මල ඉල්ලීය

Parse Failed

----- Sentence: 26 -----

අම්මා ෙගදර ගියේය

Parse Failed

----- Sentence: 27 -----

කොල්ලෝ ෙගදර ගියේය

Parse Failed

----- Sentence: 28 -----

කෙල්ලෝ අඹ ඉල්ලීය

Parse Failed

----- Sentence: 29 -----

මම ෙගදර ගියේය

Parse Failed

----- Sentence: 30 -----

අපි අඹ කෑවෝය

Parse Failed

# References

1. <https://www.nltk.org/book/ch01.html> (sections 1.1 and 1.2)
2. <https://www.nltk.org/book/ch08.html> (sections 1, 2, 3, 4.1 and 4.2)
3. <https://www.nltk.org/book/ch09.html> (sections 1.1 and 1.2)
4. <https://www.researchgate.net/publication/235931895_A_Computational_Grammar_of_Sinhala>