

# 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' | 'මේ' | 'අර' | 'මය' | 'සමහර' | 'ඇතැම්'
Adj -> 'ලස්සන' | 'කැන' | 'මනන' | 'සුදු' | 'කලු' | 'ලොකු' | 'හොඳ' | 'පුංචි' | 'උස'
Adv -> 'වේගයෙන්' | 'ලස්සනට' | 'හොඳට' | 'ඉක්මනින්' | 'සෙමෙන්'

N_SG_1 -> 'මම'
N_PL_1 -> 'අපි'

N_SG_2 -> 'නුබ' | 'ඔබ'
N_PL_2 -> 'ඔබලා' | 'නුබලා'

N_SG_M -> 'මිනු' | 'බල්ලා' | 'මිනිසා' | 'ප්‍රමයා' | 'කොල්ලා' | 'තාත්තා' | 'සර්පයා'
N_PL_M -> 'මවුනු' | 'බල්ලෙයෝ' | 'මිනිස්සු' | 'කොල්ලෙයෝ'
N_SG_N -> 'ගස' | 'මල' | 'අත්ත' | 'අඹ' | 'ගෙදර' | 'පොත' | 'වතුර' | 'සඳ'
```

N\_SG\_F -> 'ඇය' | 'බැල්ලි' | 'කෙල්ල' | 'අම්මා'  
N\_PL\_F -> 'ඔවුහු' | 'හැතැණු' | 'කෙල්ලෙෝ'

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 -> 'දෙමි' | 'කමි' | 'බොමි'  
V\_PRES\_PL\_1 -> 'දෙමු' | 'කමු' | 'බොමු'  
V\_PAST\_SG\_1 -> 'දුන්නෙමි' | 'කෑවෙමි' | 'බිවෙමි'  
V\_PAST\_PL\_1 -> 'දුන්නෙමු' | 'කෑවෙමු' | 'බිවෙමු'

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' | 'බිවෙහි'  
V\_PRES\_PL\_2 -> 'ගියෙහු' | 'කවෙහු' | 'බිවෙහු'  
V\_PAST\_SG\_2 -> 'කයි' | 'කියයි' | 'ගසයි'  
V\_PAST\_PL\_2 -> 'ගියෙෝය' | 'කෑවෙෝය'

V\_PAST\_SG\_M -> 'ගියේය'  
V\_PAST\_SG\_F -> 'ඉල්ලීය'  
V\_PRES\_SG -> 'දෙයි' | 'කයි' | 'කඩයි' | 'කියයි' | 'ගසයි' | 'සිටියි' | 'නිබි'  
V\_PRES\_PL -> 'යති' | 'කති'  
V\_PAST\_PL -> 'ගියෙෝය' | 'කෑවෙෝය'

## 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("\nParse Success\n")
        else:
            print("\nParse 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 ටොන ටොමි

මම බිවෙමි

අරි අඹ කමු

අරි අර අඹ කමු

අරි අර අඹ ඉක්මනින් කමු

අරි අඹ ඉතා ටෙසමන් කමු

අරි ටොන දුන්නෙමු

නුබ ටේගෙයන් කවෙති

නුබලා ටෙදර ගියෙනු

ඔබ ටෙදර ගියෙහි

ඔබලා වතුර බිවෙනු

තාත්තා මල කඩයි

අම්මා ටෙදර සිටියි

ඔහු ටෙදර ගියේය

ඇය ටොන ඉල්ලීය

කොල්ලෙට් ටෙදර යති

කෙල්ලෙට් අඹ කති

කොල්ලෙට් ටෙදර ගියෙව්ය

කෙල්ලෙට් අඹ කෑවෙව්ය

### Incorrect Sentences (10)

මම කමු

අරි අඹ කමි

නුබ ටේගෙයන් ගියෙනු

ඔබලා වතුර බිවෙති

තාත්තා මල ඉල්ලීය

අම්මා ටෙදර ගියේය

කොල්ලෙට් ටෙදර ගියේය

කෙල්ලෙට් අඹ ඉල්ලීය

මම ටෙදර ගියේය

අරි අඹ කෑවෙව්ය

## 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 ඌරෙකු දෙමි

(S

(S\_SG\_1  
(NP\_SG\_1 (N\_SG\_1 මම))  
(VP\_SG\_1  
(PRE\_V (NP (NP\_SG\_N (N\_SG\_N ඌරෙකු))))  
(V\_SG\_1 (V\_PRES\_SG\_1 දෙමි))))))

Parse Success

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

මම බිවෙමි

(S

(S\_SG\_1  
(NP\_SG\_1 (N\_SG\_1 මම))  
(VP\_SG\_1 (V\_SG\_1 (V\_PAST\_SG\_1 බිවෙමි))))))

Parse Success

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

අපි අපි කමු

(S

(S\_PL\_1  
(NP\_PL\_1 (N\_PL\_1 අපි))  
(VP\_PL\_1  
(PRE\_V (NP (NP\_SG\_N (N\_SG\_N අපි))))  
(V\_PL\_1 (V\_PRES\_PL\_1 කමු))))))

Parse Success

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

අරි අර අඹ කමු

```
(S
  (S_PL_1
    (NP_PL_1 (N_PL_1 අරි))
    (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 -----

අරි අර අඹ ඉක්මනින් කමු

```
(S
  (S_PL_1
    (NP_PL_1 (N_PL_1 අරි))
    (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 -----

අරි අඹ ඉතා ඌසෙමන් කමු

```
(S
  (S_PL_1
    (NP_PL_1 (N_PL_1 අරි))
    (VP_PL_1
      (PRE_V
        (NP (NP_SG_N (N_SG_N අඹ)))
        (ADVP (Deg ඉතා) (Adv ඌසෙමන්)))
      (V_PL_1 (V_PRES_PL_1 කමු))))))
```

Parse Success

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

අපි ඌරාන දුන්නෙමු

```
(S
  (S_PL_1
    (NP_PL_1 (N_PL_1 අපි))
    (VP_PL_1
      (PRE_V (NP (NP_SG_N (N_SG_N ඌරාන))))
      (V_PL_1 (V_PAST_PL_1 දුන්නෙමු))))
```

Parse Success

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

නුබ ඌවගෙයන් කවෙති

```
(S
  (S_SG_2
    (NP_SG_2 (N_SG_2 නුබ))
    (VP_SG_2
      (PRE_V (ADV (Adv ඌවගෙයන්)))
      (V_SG_2 (V_PRESENT_SG_2 කවෙති))))
```

Parse Success

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

නුබලා ඌගෙර හියෙනු

```
(S
  (S_PL_2
    (NP_PL_2 (N_PL_2 නුබලා))
    (VP_PL_2
      (PRE_V (NP (NP_SG_N (N_SG_N ඌගෙර))))
      (V_PL_2 (V_PRESENT_PL_2 හියෙනු))))
```

Parse Success

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

ඔබ ොගඳර ගියෙහි

```
(S
  (S_SG_2
    (NP_SG_2 (N_SG_2 ඔබ))
    (VP_SG_2
      (PRE_V (NP (NP_SG_N (N_SG_N ොගඳර))))
      (V_SG_2 (V_PRES_SG_2 ගියෙහි))))
```

Parse Success

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

ඔබලා චතුර බිචෙනු

```
(S
  (S_PL_2
    (NP_PL_2 (N_PL_2 ඔබලා))
    (VP_PL_2
      (PRE_V (NP (NP_SG_N (N_SG_N චතුර))))
      (V_PL_2 (V_PRES_PL_2 බිචෙනු))))
```

Parse Success

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

නාන්තා මල කඩයි

```
(S
  (S_SG_M
    (NP_SG_M (N_SG_M නාන්තා))
    (VP_PRES_SG (PRE_V (NP (NP_SG_N (N_SG_N මල)))) (V_PRES_SG කඩයි))))
```

Parse Success



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

අම්මා ගෙදර සිටියි

```
(S
  (S_SG_F
    (NP_SG_F (N_SG_F අම්මා))
    (VP_PRESENT_SG
      (PRE_V (NP (NP_SG_N (N_SG_N ගෙදර))))
      (V_PRESENT_SG සිටියි))))
```

Parse Success

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

ඔහු ගෙදර 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 ගෙදර))))
      (V_PAST_SG_M ගියේය))))
```

Parse Success

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

ඇය තොර 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 තොර))))
      (V_PAST_SG_F ඉල්ලීය))))
```

Parse Success

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

කොල්ලෙෝ ෙගඳර යනි

```
(S
  (S_PL
    (NP_PL_M (N_PL_M කොල්ලෙෝ))
    (VP_PRESENT_PL (PRE_V (NP (NP_SG_N (N_SG_N ෙගඳර)))) (V_PRESENT_PL යනි))))
```

Parse Success

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

කෙල්ලෙෝ අඹ කනි

```
(S
  (S_PL
    (NP_PL_F (N_PL_F කෙල්ලෙෝ))
    (VP_PRESENT_PL (PRE_V (NP (NP_SG_N (N_SG_N අඹ)))) (V_PRESENT_PL කනි))))
```

Parse Success

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

කොල්ලෙෝ ෙගඳර ගියෙෝය

```
(S
  (S_PL
    (NP_PL_M (N_PL_M කොල්ලෙෝ))
    (VP_PAST_PL
      (PRE_V (NP (NP_SG_N (N_SG_N ෙගඳර))))
      (V_PAST_PL ගියෙෝය))))
```

Parse Success

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

කෙල්ලෙෝ අඹ කෑවෙෝය

```
(S
  (S_PL
    (NP_PL_F (N_PL_F කෙල්ලෙෝ))
    (VP_PAST_PL (PRE_V (NP (NP_SG_N (N_SG_N අඹ)))) (V_PAST_PL කෑවෙෝය))))
```

Parse Success

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

මම කමු

Parse Failed

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

අපි අඹ කමි

Parse Failed

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

නුබ ටේගෙයන් ගියෙනු

Parse Failed

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

ඔබලා චතුර බීවෙනි

Parse Failed

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

තාත්තා මල ඉල්ලිය

Parse Failed

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

අම්මා ගෙදර ගියේය

Parse Failed

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

කොල්ලෙට් ගෙදර ගියේය

Parse Failed

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

කෙල්ලෙට් අඹ ඉල්ලිය

Parse Failed

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

මම ගෙදර ගියේය

Parse Failed

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

අපි අඹ කෑවෙට්

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](https://www.researchgate.net/publication/235931895_A_Computational_Grammar_of_Sinhala)