CMP 670 - Statistical Natural Language Processing

Assignment 2 – Context Free Grammar (CFGs) and Parsing
(Due date: 11.04.2019 – 11:59 pm)

Part 1: Language Generation with CFG

In this first part of the assignment, you will implement a random sentence generator. For this, a sample CFG rule set is provided for you in the Chomsky Normal Form (CNF). Using the grammar rules defined in a grammar file, you will generate a sentence randomly.

The name of your Pyhton file will be randsentence.py. The output of the program will be written to an output file called 'random-sentence.txt'.

The format of the CFG rules is as follows:

S	NP VP	*#S → NP VP*
Noun	book	*#Noun → book*

Every sentence begins with a root which is defined as follows:

ROOT	S .	*#ROOT → S .*

Each nonterminal can have multiple expansions. For example, NP either can be expanded as Det Noun or Det Adj Noun. You should choose one of the rules randomly while you generate a sentence.

Run your program multiple times and observe what kind of sentences it tends to generate. Discuss it in your report.

Part 2: Parsing Sentences with CYK Parser

In the second part of the assignment, you will implement CYK parser as a recognizer which tells whether a given sentence is grammatically correct or not according to the same CFG rule set (used in the first part of the assignment).

The name of your Python file will be parse.py. The program will read a sentence from an input file called 'sentence.txt' and then it will print whether this sentence is grammatically correct or not.

Discuss your solution in your report.

Note: No out-of-vocabulary (OOV) word will be included in the given sentence.