



# CMP 670

Statistical Natural Language Processing (Spring 2019)  
Homework 2

EROL ÖZKAN  
N18245609  
[erolozkan@outlook.com](mailto:erolozkan@outlook.com)

## 1. CONTEXT-FREE GRAMMAR (CFG)

A context-free grammar(CFG) consists of a set of rules or productions.

CFGs are defined by four parameters N,  $\Sigma$ , R, S:

- S: start symbol,
- N: a set of non-terminal symbols (or variables),
- $\Sigma$ : a set of terminal symbols - Words like (“the”, “car”, ”class”),
- R: a set of rules or productions, each of the form  $A \rightarrow \beta$ .

## 2. LANGUAGE GENERATION WITH CFG

CFGs are used to derive Context-free languages. Table 1 shows an example for language generation using grammar rules defined in the assignment. The words shown in bold are selected.

*Table 1: Depth First Search Language Generation*

Current	Status (Selection)	Next Status (Selection)
ROOT	Non-terminal	S -> NP VP
S	Non-terminal	NP -> Det Noun NP PP
NP	Non-terminal	Det -> the a every
Det	Terminal	<b>every</b>
NP	Non-terminal	Noun -> Adj Noun president sandwich pickle mouse floor
Noun	Terminal	<b>sandwich</b>
S	Non-terminal	VP -> Verb NP
VP	Non-terminal	Verb -> ate wanted kissed washed pickled
Verb	Terminal	<b>pickled</b>
VP	Non-terminal	NP -> Det Noun NP PP
NP	Non-terminal	Det -> the a every
Det	Terminal	<b>the</b>
NP	Non-terminal	Noun -> Adj Noun president sandwich pickle mouse floor
Noun	Terminal	<b>mouse</b>
ROOT	Terminal	!

**Generated Sentence: “every sandwich pickled the mouse !”**

Table 2 shows some generated sentences using grammar rules in the assignment.

*Table 2: Some Generated Sentences*

the floor pickled a floor .
is it true that the delicious president pickled the president ?
every president kissed every pickle with every pickle with a president .
is it true that every sandwich washed every mouse under the mouse with a mouse ?
is it true that the mouse on every mouse pickled every sandwich in every old president in a president ?
the floor pickled the mouse !
a pickle in a mouse wanted a pickle on every beautiful president .
is it true that every pickle kissed every mouse ?

**NOTE:** Generated sentences are grammatically correct. But, given that, words are selected randomly, most of them have no sense. Also, PCFGs could be better. Rules & word are selected using more information.

### 3. EQUIVALENCE OF CFGS

A formal language is defined as a (possibly infinite) set of strings of words. It is possible that two distinct CFGs to generate the same language. Two grammars are equivalent if they generate the same set of strings. Two kinds of grammar equivalence exists: weak equivalence and strong equivalence.

**Strong Equivalence:** Two grammars are strongly equivalent if they generate the same set of strings and if they assign the same phrase structure to each sentence.

**Weak Equivalence:** Two grammars are weakly equivalent if they generate the same set of strings but do not assign the same phrase structure to each sentence.

### 4. CHOMSKY NORMAL FORM (CNF)

A CFG is in Chomsky Normal Form (CNF), if it has no Empty rule and each production is either in the form of " $A \rightarrow B C$ " or " $A \rightarrow a$ ". Any grammar can be converted into a weakly-equivalent CNF form. Table 3 shows an example CNG conversion.

Table 3: Example CNF Conversion

$A \rightarrow B C D$	$A \rightarrow B X$ $X \rightarrow C D$
-----------------------	--

### 5. PARSING SENTENCES WITH CYK PARSER

CYK Parser is used to decide whether a given string belongs to a language of grammar or not.

The grammars used with CKY algorithm must be in Chomsky Normal Form (CNF). The right-hand side of each rule must expand to either two non-terminals or to a single terminal. In the assignment, in order to apply CKY algorithm, CNF conversion must be applied. Table 4 shows applied conversion. Following rules are updated.

- Rules that mix terminals with non-terminals on the right-hand side (**Conversion 1**),
- Rules where the right-hand side's length is greater than two. (**Conversion 2**).

Table 4: Applied CNF Conversion

Conversion 1	ROOT $\rightarrow$ S . ROOT $\rightarrow$ S ! ROOT $\rightarrow$ is it true that S ?	ROOT $\rightarrow$ S QQ
		QQ $\rightarrow$ .
		QQ $\rightarrow$ !
		QQ $\rightarrow$ ?
Conversion 2	ROOT $\rightarrow$ is it true that S ?	S $\rightarrow$ AA S
		AA $\rightarrow$ BB CC
		BB $\rightarrow$ A B
		CC $\rightarrow$ C D
		A $\rightarrow$ is B $\rightarrow$ it C $\rightarrow$ true D $\rightarrow$ that

Restricting a grammar to CNF does not lead to any loss in expressiveness since any context-free grammar can be converted into a corresponding CNF grammar that accepts exactly the same set of strings as the original grammar. Note that these conversions are done manually.

Conversion 1 and conversion 2 are shown in Figure 1 and Figure 2 respectively.

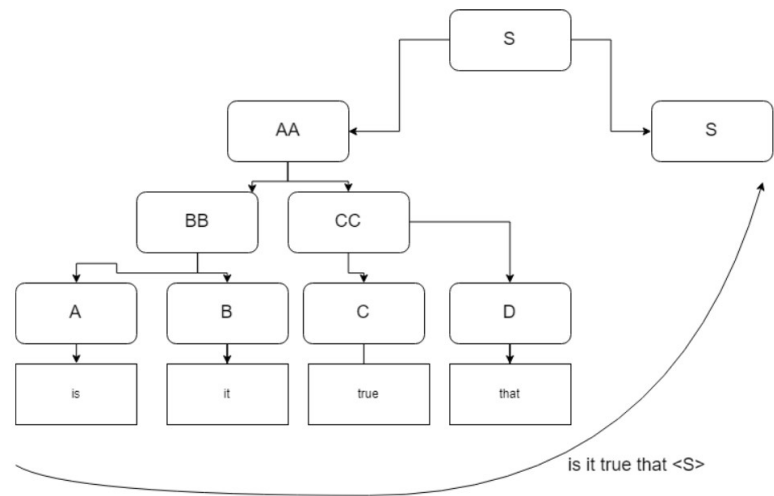


Figure 1: CNF Conversion 1 - “is it true that S”

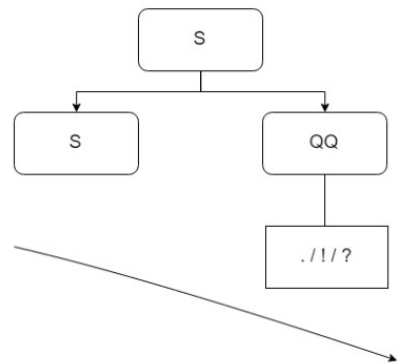


Figure 2: CNF Conversion 2 - “S .!/?”

Sentence	Parse Table	Result
is it true that a floor ate every president ?	<div>['A']['BB'][][]['AA'][][]['S']['ROOT']</div> <div>['B'][][][][]</div> <div>['C']['CC'][][]</div> <div>['D'][][]</div> <div>['Det']['NP'][][]['S']['ROOT']</div> <div>['Noun'][][]</div> <div>['Verb'][][]['VP']</div> <div>['Det']['NP']</div> <div>['Noun']</div> <div>['QQ']</div>	CORRECT
a mouse kissed every beautiful president in the floor !	<div>['Det']['NP'][][]['S'][][]['S']['ROOT']</div> <div>['Noun'][][]</div> <div>['Verb'][][]['VP'][][]['VP']</div> <div>['Det'][][]['NP'][][]['NP']</div> <div>['Adj']['Noun'][][]</div> <div>['Noun'][][]</div> <div>['Prep'][][]['PP']</div> <div>['Det']['NP']</div> <div>['Noun']</div> <div>['QQ']</div>	CORRECT