Recitation

CFG and CKY

11/12/2020

Param P

CFG

- Set of nonterminal symbols N
- Set of terminal symbols Σ
- A start symbol $S \in N$
- Set R of productions of the form $A \rightarrow B$ where $A \in N$ and $B \in (\Sigma \cup N)^*$ ie. B is a string of terminals and non-terminals
- Given G is a CFG, language of G is defined as set of all terminal strings that can be derived from the start symbol
- **Context free:** applicability of a rule depends only on a nonterminal symbol, not on its context

Regular Grammar

- Set of nonterminal symbols N
- Set of terminal symbols Σ
- A start symbol $S \in N$
- Set R of productions of the form $A \rightarrow aB$ or $A \rightarrow a$ where $A, B \in N$ and $a \in \Sigma$
- Can be implemented as finite state automata
- Set of all regular languages is strictly smaller than set of all context-free languages
- Problem: regular languages cannot capture long-distance dependencies

CKY Parsing

- Syntactic parsing is the task of recognizing a sentence and assigning a syntactic structure to it (i.e. adding POS tags).
- The **CKY algorithm** is a bottom-up approach to figuring out if a sentence/sequence of words is in the language/grammar provided, because it starts with the terminals and sees which subtrees it can build, eventually making those subtrees into larger subtrees until you reach S (or don’t, in which case it’s not in the language).
- To use the CKY algorithm, the CFG must be in **CNF**.

CNF

Chomsky Normal Form

- A CFG $G=(N, \Sigma, R, S)$ is in Chomsky Normal Form (CNF) if the rules take one of the following forms:
  - $A \rightarrow B\ C$, where $A \in N, B \in N, C \in N$.
  - $A \rightarrow b$, where $A \in N, b \in \Sigma$.

S $\rightarrow$ NP VP	V $\rightarrow$ saw
VP $\rightarrow$ V NP	P $\rightarrow$ with
VP $\rightarrow$ VP PP	D $\rightarrow$ the
PP $\rightarrow$ P NP	N $\rightarrow$ cat
NP $\rightarrow$ D N	N $\rightarrow$ tail
NP $\rightarrow$ NP PP	N $\rightarrow$ student

Any CFG can be converted to an equivalent grammar in CNF that expresses the same language.

Cannot have null rules and unreachable rules as well

CNF Example

CFG -> CNF

- $S \rightarrow bC$
- $C \rightarrow B$
- $B \rightarrow bAE$
- $A \rightarrow bE$
- $E \rightarrow e$

CFG -> CNF

- $S \rightarrow FC$
- $C \rightarrow FAE$
- $A \rightarrow FE$
- $E \rightarrow e$
- $F \rightarrow b$

Replace ‘b’ by non-terminal ‘F’ by adding a new rule

Merge B into C since C has only 1 non-terminal
On the right side.

CFG -> CNF

- S → FC
- C → FG
- G → AE
- A → FE
- E → e
- F → b

C had 3 non-terminals on the right. Take the last ~~3~~ 2
And make a new rule 'G' for them.

Now each NT has only 2 Non-Terminals on the right
or only 1 Terminal.

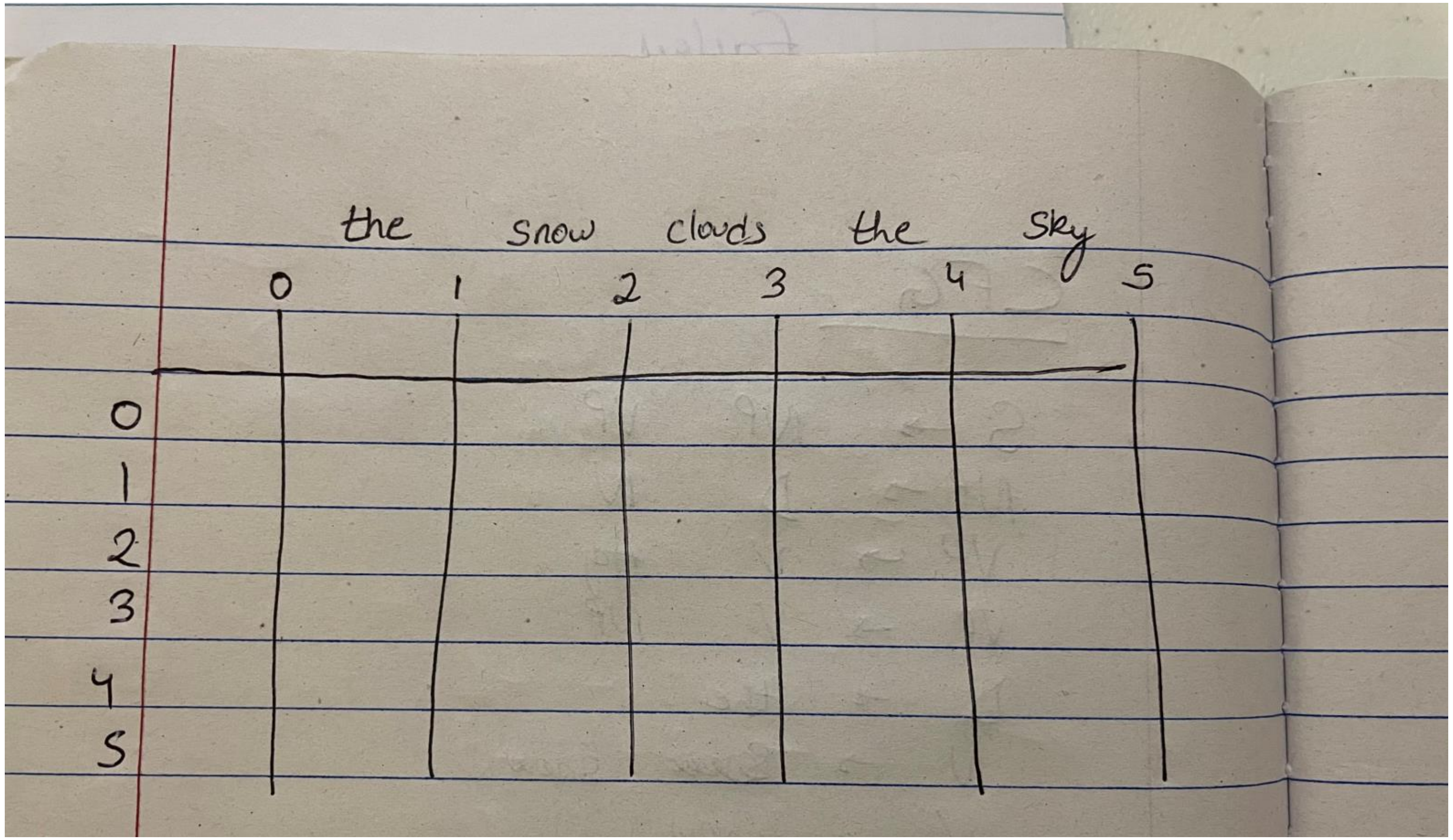
This is final CNF form

CKY Example

CFG

S	$\rightarrow$	NP	VP
NP	$\rightarrow$	D	N
VP	$\rightarrow$	V	Adj
VP	$\rightarrow$	V	NP
D	$\rightarrow$	the	
N	$\rightarrow$	~~Snow~~ snow	
V	$\rightarrow$	snow	
N	$\rightarrow$	clouds	
V	$\rightarrow$	clouds	
N	$\rightarrow$	sky	
Adj	$\rightarrow$	sky	

Sentence: the snow clouds the sky



		the	snow	clouds	the	sky
	0	1	2	3	4	5
0	X					
1	X					
2	X					
3	X					
4	X					
5	X	X	X	X	X	X

CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~Snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky

		the	snow	clouds	the	sky
		0	1	2	3	4
0	X	D				
1	X					
2	X					
3	X					
4	X					
5	X	X	X	X	X	X

CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~Snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky

		the	snow	clouds	the	sky
	0	1	2	3	4	5
0	X	D				
1	X	X	N, V			
2	X	X	X	N, V		
3	X	X	X	X	D	
4	X	X	X	X	X	N, Adj
5	X	X	X	X	X	X

CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky

		the	snow	clouds	the	sky
	0	1	2	3	4	5
0	X	D →	NP			
1	X	X	N, V	X		
2	X	X	X	N, V	X	
3	X	X	X	X	D →	NP
4	X	X	X	X	X	N, Adj
5	X	X	X	X	X	X

CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~Snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky

		the	snow	clouds	the	sky
	0	1	2	3	4	5
0	X	D →	NP	X		
1	X	X	N, V	X	X	
2	X	X	X	N, V	X →	VP
3	X	X	X	X	D →	NP
4	X	X	X	X	X	N, Adj
5	X	X	X	X	X	X

CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~Snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky

		the	snow	clouds	the	sky
		0	1	2	3	4 5
0	X	D →	NP	X	X	
1	X	X	N, V	X	X	X
2	X	X	X	N, V	X	VP
3	X	X	X	X	D →	NP
4	X	X	X	X	X	N, Adj
5	X	X	X	X	X	X

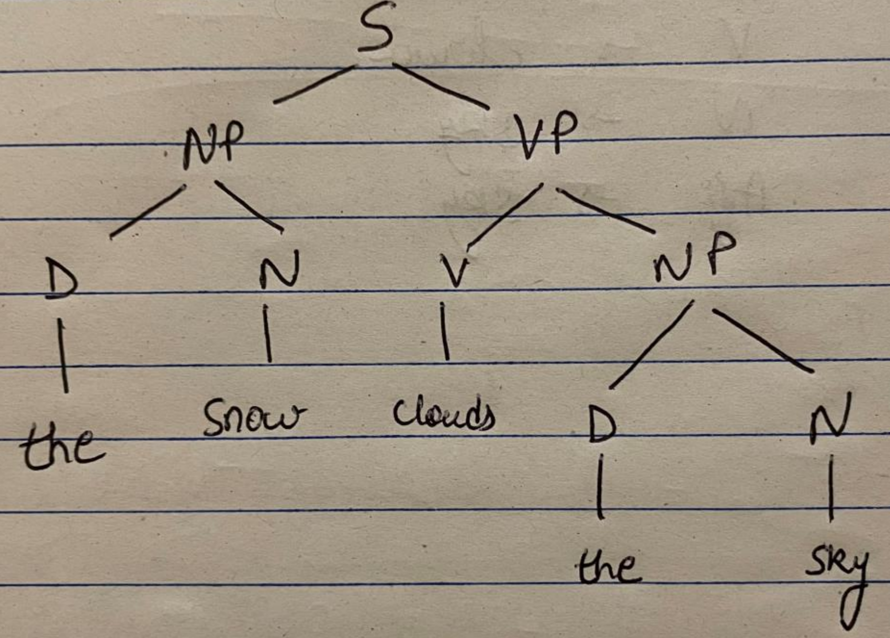
CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~Snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky

		the	snow	clouds	the	sky
	0	1	2	3	4	5
0	X	D	NP	X	X	S
1	X	X	N, V	X	X	X
2	X	X	X	N, V	X	VP
3	X	X	X	X	D	NP
4	X	X	X	X	X	N, Adj
5	X	X	X	X	X	X

CFG

S → NP VP
NP → D N
VP → V Adj
VP → V NP
D → the
N → ~~snow~~ snow
V → snow
N → clouds
V → clouds
N → sky
Adj → sky



CFG

S →	NP	VP
NP →	D	N
VP →	V	Adj
VP →	V	NP
D →	the	
N →	~~Snow~~ snow	
V →	snow	
N →	clouds	
V →	clouds	
N →	sky	
Adj →	sky	