Chomsky NF & Decision Problems of CFL

Chomsky Normal Form

A **context-free grammar** is said to be in **Chomsky Normal Form** if each **production rule** is in **one** of the following **two forms**.

$$A \rightarrow BC,\ A \rightarrow \sigma,$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are non-terminals and σ is a terminal. If ε is in the language, then the **start symbol** will produce ε.

A Theorem

Every context-free grammar can be transformed to CNF.

An Example

Consider the following grammar.

$$
\begin{aligned}
E &\rightarrow E + T \mid T \\
T &\rightarrow T * F \mid F \\
F &\rightarrow (E) \mid c
\end{aligned}
$$

Removal of Unit Productions

$$
\begin{aligned}
E &\rightarrow E + T \mid T * F \mid (E) \mid c \\
T &\rightarrow T * F \mid (E) \mid c \\
F &\rightarrow (E) \mid c
\end{aligned}
$$

Replacing Terminals

$$
\begin{aligned}
E &\rightarrow EPT \mid TMF \mid LER \mid c \\
T &\rightarrow TMF \mid LER \mid c \\
F &\rightarrow LER \mid c \\
P &\rightarrow + \\
M &\rightarrow * \\
L &\rightarrow (\\
R &\rightarrow)
\end{aligned}
$$

Replacing Longer Right Side

$$
\begin{aligned}
E &\rightarrow EA \mid TB \mid LC \mid c \\
T &\rightarrow TB \mid LC \mid c \\
F &\rightarrow LC \mid c \\
P &\rightarrow + \\
M &\rightarrow * \\
L &\rightarrow (\\
R &\rightarrow) \\
A &\rightarrow PT\ ,\ B \rightarrow MF\ ,\ \ C \rightarrow ER
\end{aligned}
$$

The 1st Decision Problem

Given a **context-free** grammar $\mathbf{G}$ and a string $\mathbf{x}$ over $\mathbf{\Sigma}^*$, can we **decide** whether $\mathbf{x} \in \mathbf{L(G)}$?

CKY Algorithm : An Example

Consider the previous **expression grammar** in **Chomsky Normal Form** and the input $\mathbf{c} + \mathbf{c} * \mathbf{c}$. This can be generated as

$\{E, T, F\}$	P	$\{E, T, F\}$	M	$\{E, T, F\}$
c	$+$	c	$*$	c

CKY Algorithm : An Example

2	$\emptyset$	**A**	$\emptyset$	**B**	
1	$\{E, T, F\}$	**P**	$\{E, \mathbf{T}, F\}$	**M**	$\{E, T, \mathbf{F}\}$
	c	$+$	c	$*$	c

CKY Algorithm : An Example

3	$\mathbf{E}$	$\emptyset$	$\{\mathbf{E}, \mathbf{T}\}$		
2	$\emptyset$	$\mathbf{A}$	$\emptyset$	$\mathbf{B}$	
1	$\{\mathbf{E}, T, F\}$	P	$\{E, \mathbf{T}, F\}$	M	$\{E, T, F\}$
	c	$+$	c	$*$	c

CKY Algorithm : An Example

4	$\emptyset$	**A**			
3	E	$\emptyset$	$\{E, \mathbf{T}\}$		
2	$\emptyset$	A	$\emptyset$	B	
1	$\{E, T, F\}$	**P**	$\{E, T, F\}$	M	$\{E, T, F\}$
	c	$+$	c	$*$	c

CKY Algorithm : An Example

5	$\mathbf{E}$				
4	$\emptyset$	$\mathbf{A}$			
3	E	$\emptyset$	$\{E, T\}$		
2	$\emptyset$	A	$\emptyset$	B	
1	$\{\mathbf{E}, T, F\}$	P	$\{E, T, F\}$	M	$\{E, T, F\}$
	c	$+$	c	$*$	c

The string $\mathbf{c} + \mathbf{c} * \mathbf{c}$ **is in** $L(G)$.