Languages that are and are not context free

CENG 280

Course outline

- Preliminaries: Alphabets and languages
- Regular languages PL,
- Context-free languages
  - Context-free grammars
  - Parse trees
  - Push-down automaton
  - Push-down automaton - context-free languages
  - Languages that are and that are not context-free, Pumping lemma
- Turing-machines

Pumping theorem for CFL

Given a context free grammar $G = (V, \Sigma, R, S)$.

- The **fanout** of G, denoted by $\phi(G)$ is the largest number of symbols on the right hand side of any rule in R.
- A **path** in a parse tree is a sequence of distinct nodes connected with line segments, where the first one is the root and the last one is a leaf.
- The **length** of a path is the number of line segments in it.
- The **height** of a parse tree is the length of the longest path in it.

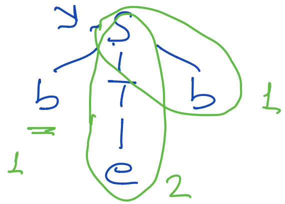
Pumping theorem for CFL

Example

Consider $G = (V, \Sigma, R, S)$ with $|V - \Sigma| = 2$ $\phi(G) = 3$

$R : S \rightarrow \underbrace{aSa}_{3}, S \rightarrow \underbrace{bTb}_{3}, T \rightarrow \underbrace{bTb}_{3}, T \rightarrow \underbrace{e}_{1}$. S, T

$S \Rightarrow bTb \Rightarrow bb$

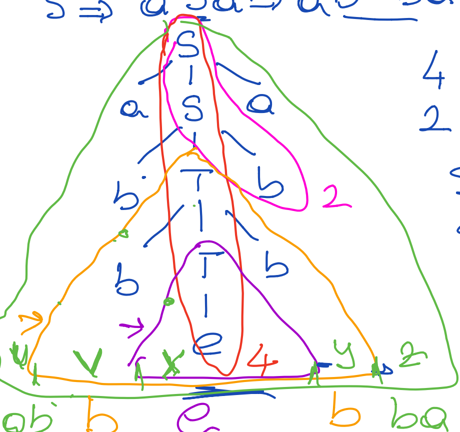


length = 2

uv^2xy^2z

abbbbbba

$S \Rightarrow aSa \Rightarrow abTba \Rightarrow abbTbba \Rightarrow e$



4 non-terminals in the path

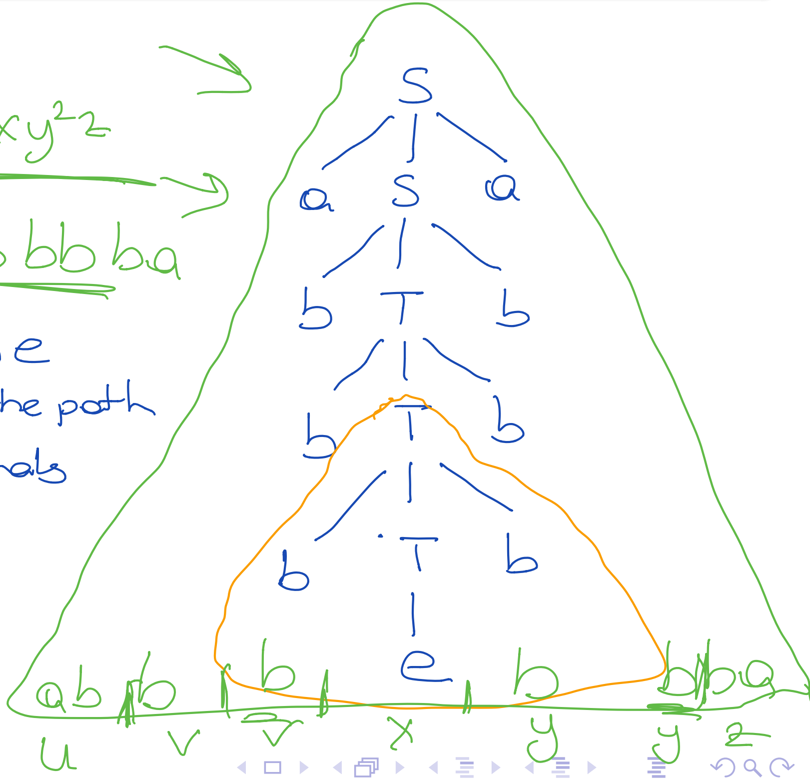
2 distinct nonterminals

$S \Rightarrow^* abbbba$

$S \Rightarrow^* bbbb$

$T \Rightarrow^* bb$

$T \Rightarrow^* e$



Pumping theorem for CFL

Lemma

The yield of a parse tree of G of height h has length at most $\phi(G)^h$.

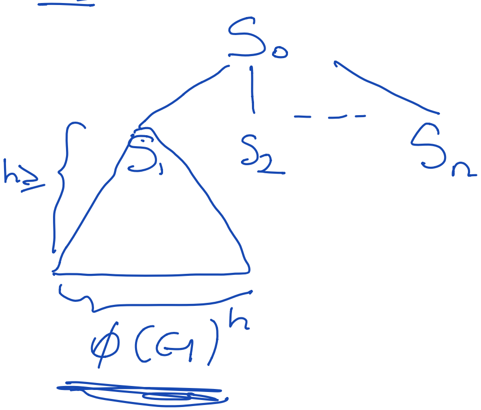
Induction

Base case $h = 1$

$\phi(G)^1$

I.H. Suppose that the result is true for parse trees of length up to $h \geq 1$.

I.S Show that is true for parse trees with length $h+1$.



$n \leq \phi(G)$

$\phi(G)^h \cdot \phi(G) = \phi(G)^{h+1}$

Pumping theorem for CFL

Corollary

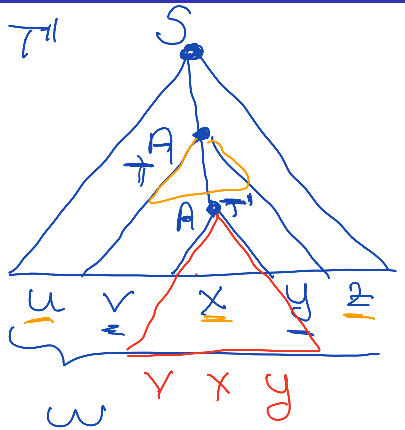
The parse tree of any $w \in L(G)$ *with* $|w| > \phi(G)^h$ *must have a path longer than* h*.*

Theorem

Let $G = (V, \Sigma, R, S)$ *be a context-free grammar. Then any string* $w \in L(G)$ *of length greater than* $\phi(G)^{|V \setminus \Sigma|}$ *can be written as* $w = uvxyz$ *in such a way that*

- v *or* y *is non empty (i.e.* $|vy| > 0$*) and*
- $uv^n xy^n z \in L(G)$ *for every* $n \geq 0$*.*

Proof of pumping theorem for CFL



$|w| > \phi(G)^{|V-\Sigma|}$ pumping length

- $h > |V - \Sigma|$
- There exists a path with h non-terminals

By PHP, there should be a non-terminal repeated along a path

$A \Rightarrow^* x$

$A \Rightarrow^* vxy$

$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$

$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow uvvAyyz \Rightarrow uv^2xy^2z$

$S \Rightarrow^* uAz \Rightarrow^* uxz$

uv^3xy^3z, uv^nxy^nz

$|vy| = 0$

T'' is the parse tree of w with smallest # of leaves.

$|vy| > 0$

Pumping theorem for CFL - Examples

By Pumping theorem L is not CF.

Example

Show that $L = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free. $\phi(G)^{|V-\Sigma|}$

Assume that L is CF, then there exists G, $L(G) = L$

Let $K = \phi(G)^{|V-\Sigma|}$ (pumping length)

w, $|w| > K$, $w \in L$

$a \dots a\, b \dots b\, c \dots c$

vxy vxy vxy

$w = a^K b^K c^K$ $|w| = 3K > K$

Then by Pumping theorem there exists a split $w = uvxyz$, $|vy| > 0$

s.t. $uv^n x y^n z \in L$ for any $n \geq 0$.

1. $vxy = a^i$
2. $vxy = b^i$
3. $vxy = c^i$
4. $vxy = a^i b^j$
5. $vxy = b^i c^j$
6. $vxy = a^i b^K c^j$

(1) $uv^0 x y^0 z = a^{K-k} b^K c^K \notin L$ (2., 3. cases similar)
$|vy| = k > 0$ $+k$

(4) $uv^2 x y^2 z \notin L$ # of a's or b's is strictly greater than the # c's (5. similar)
$|vy| = k > 0$

(6) $uv^2 x y^2 z \notin L$ (1) vy contains all 3 symbols.
$\Rightarrow$ the ordering property is violated
v or y contains 2 symbols, $a^i b^j$ $a^i b^j a^i b^j$
$|vy| = k > 0$ vy doesnot 4

Pumping theorem for CFL - Examples

Example

Show that $L = \{w \in \{a, b, c\}^\star \mid w$ has equal number of $a's, b's$ and $c's\}$ is not context-free.

$L' = \{a^n b^n c^n \mid n \geq 0\}$ L' is not context free

$L'' = \mathcal{L}(a^* b^* c^*)$ L'' is regular

$$\underbrace{L''}_{\text{regular}} \cap L = \underbrace{L'}_{\text{not context free}}$$

if L was context-free $L'' \cap L$ = context free →

re

L is not context-free.

Theorem

The context-free languages are not closed under intersection or complementation.

$$\overline{\left(L_1 \cap L_2 = \overline{(\overline{L_1} \cup \overline{L_2})}\right)}$$

Proof? L_1, L_2 context free

$L_1 \cap L_2$ is not context-free ($a^n b^n c^n$)

- $L_1 = \{a^n b^n c^m \mid n, m \geq 0\}$
- $L_2 = \{a^n b^m c^m \mid n, m \geq 0\}$

$$\left.\begin{aligned} S &\Rightarrow AC \\ A &\Rightarrow aAb \mid e \\ C &\Rightarrow cC \mid e \end{aligned}\right\} L_1$$

$$L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$$

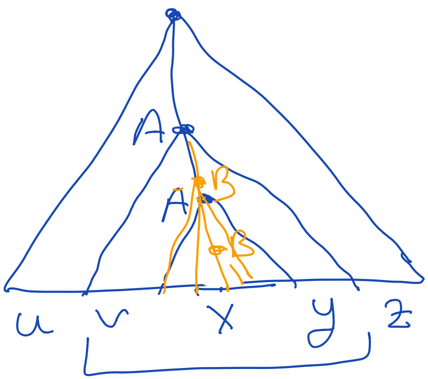
Pumping theorem for CFL

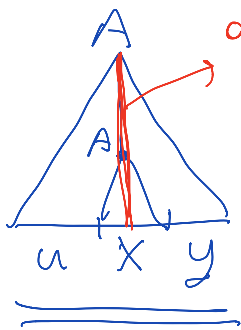
Theorem (Strong version)

Let $G = (V, \Sigma, R, S)$ be a context-free grammar. Then there exists k, K such that any string $w \in L(G)$ with $|w| > K$ can be written as $w = uvxyz$ in such a way that $|vy| > 0$, $|vxy| \leq k$ and $uv^n xy^n z \in L(G)$ for every $n \geq 0$.

$K = k = \phi(G)^{|V-\Sigma|}$

$|w| > K$





at least two non-terminals are repeated

$|uxy| > k = \phi(G)^{|V-\Sigma|}$

$|uxy| \leq K$

Example

Show that $L = \{\underline{a^{n^2}} \mid n \geq 0\}$ is not context free.

Assume that L is context-free. Then by strong PT there exists K, k such that $|w| > K$, there exist a split $w = uvxyz$ $|vxy| \leq k$

$uv^n x y^n z \in L$

$w = a^{K^2}$ $\quad |w| = K^2 > K$

$w = uvxyz$ $\quad uv^2xy^2z = a^{K^2+m} \notin L$

$|vxy| < K$ $\quad K^2 < K^2 + m < (K+1)^2$

$|vy| = m < K$ $\quad K^2 + 2K + 1$

So by PT. L is not context-free