Lecture 20 – The Pumping Lemma for Context-Free Languages

COSE215: Theory of Computation

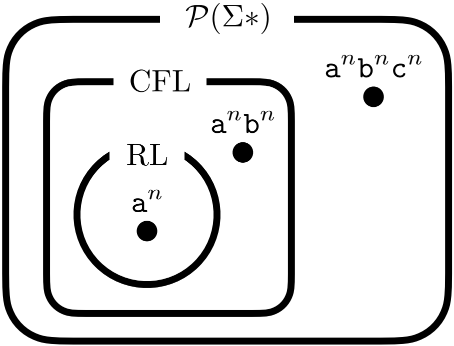
Jihyeok Park

2025 Spring

Recall

- We have learned about the **Pumping Lemma** for **Regular Languages (RLs)**.
- We could use it to **prove** that some languages are **NOT** regular.
- Is there a similar lemma for **Context-Free Languages (CFLs)**?
- Yes. We can use it to prove the following language is **NOT** a CFL:

$$L = \{a^n b^n c^n \mid n \geq 0\}$$



Contents

Contents

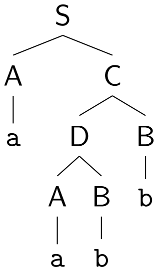
Size of Parse Trees in Chomsky Normal Form

Theorem (Size of Parse Trees in Chomsky Normal Form)

For a given CFG G in Chomsky Normal Form, for all $w \in L(G)$, if the length of the longest path in the parse tree of w is n, then $|w| \leq 2^{n-1}$. Note that the length of a path is the number of edges in the path.

For example, consider the following CFG in CNF, and the parse tree of $w = \texttt{aabb}$. The length of the longest path in the parse tree is 4, and the length of the word is 4. Thus, $|w| = 4 \leq 2^3 = 2^{n-1}$.

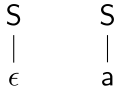
$$
\begin{array}{rcl}
S & \rightarrow & \epsilon \mid AC \mid AB \\
D & \rightarrow & AC \mid AB \\
C & \rightarrow & DB \\
A & \rightarrow & a \\
B & \rightarrow & b
\end{array}
$$



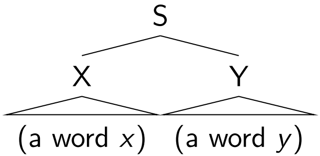
Size of Parse Trees of Chomsky Normal Form – Proof

Proof) Let's perform induction on the length of the longest path n.

- **(Basis Case)** $n = 1$. Then, $|\epsilon| = 0 \leq 2^{1-1}$ and $|a| = 1 \leq 2^{1-1}$.



- **(Induction Case)** The first rule of S is in the form of $S \to XY$. The length of the longest path in the parse tree of X (or Y) is less than or equal to $n - 1$. If $X \Rightarrow^* x \in \Sigma^*$ and $Y \Rightarrow^* y \in \Sigma^*$, then $|x| \leq 2^{n-2}$ and $|y| \leq 2^{n-2}$ ($\because$ I.H.). Thus, $|w| = |x| + |y| \leq 2^{n-2} + 2^{n-2} = 2^{n-1}$.



Pumping Lemma for Context-Free Languages

Lemma (Pumping Lemma for Context-Free Languages)

*For a given CFL L, **there exists** a positive integer n such that **for all** $z \in L$, if $|z| \geq n$, **there exists** a split $z = uvwxy$ such that*

1. $|vx| > 0$
2. $|vwx| \leq n$
3. $\forall i \geq 0.\ uv^iwx^iy \in L$

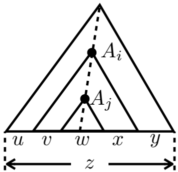
$$A \quad = \quad L \text{ is context-free}$$

$$\Downarrow$$

$$B \quad = \quad \exists n > 0.\ \forall z \in L.\ |z| \geq n \Rightarrow \exists z = uvwxy.\ 1 \land 2 \land 3$$

Proof of Pumping Lemma

- Let L be a context-free language.
- Then, $\exists$ CFG G in Chomsky Normal Form. s.t. $L(G) = L$.
- Let $m \geq 0$ be the number of variables in G, and n be $2^m > 0$.
- Take any $z = \mathtt{a}_1\mathtt{a}_2 \cdots \mathtt{a}_k \in L$ s.t. $|z| = k \geq n$.
- Consider the longest path $A_1(= S), A_2, \cdots, A_p$ in the parse tree of z. Then, $k = |z| \leq 2^{p-1} by$ Theorem of Size of Parse Trees in CNF. It means that $p \geq m + 1$ ($\because 2^{p-1} \geq k \geq n = 2^m$).
- Pick $m + 1$ variables from the bottom of the path: $A_{p-m}, \cdots, A_p$.
- Then, $\exists i, j.\ (p - m \leq i < j \leq p) \land (A_i = A_j)$ by Pigeonhole Principle.
- Split the word $z = uvwxy$ as follows:

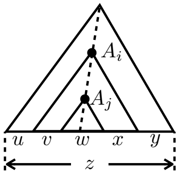


$p - m \leq i < j \leq p$

and

$A_i = A_j$

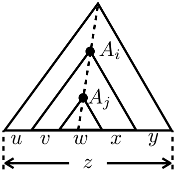
Proof of Pumping Lemma - ① and ②



$$p - m \leq i < j \leq p$$
and
$$A_i = A_j$$

- $\boxed{① \; |vx| > 0}$
  - Since $i < j$, the word vwx derived from A_i is not equal to the word w derived from A_j. ($\because S \to \epsilon$ never occurs in the middle of the parse tree.)
  - Thus, vx is not an empty word, and $|vx| > 0$.
- $\boxed{② \; |vwx| \leq n}$
  - Since $p - m \leq i$, the length of the longest path from A_i in the parse tree of z is $p - i + 1$ is less than or equal to $m + 1$.
  - By Theorem of Size of Parse Trees in CNF, the length of the word vwx is less than or equal to $2^m = n$.

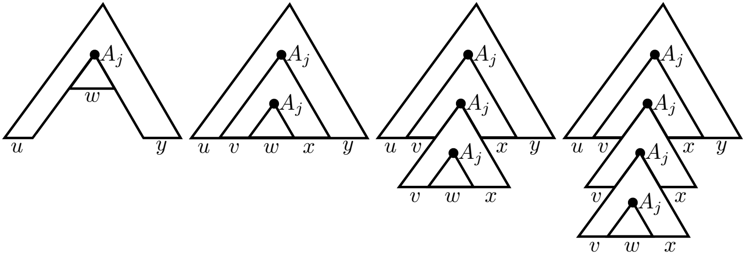
Proof of Pumping Lemma - ③



$$p - m \leq i < j \leq p$$
and
$$A_i = A_j$$

- $\boxed{③\ \forall i \geq 0.\ uv^iwx^iy \in L}$

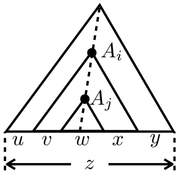
uwy	$uvwxy$	$uvvwxxy$	$uvvvwxxxy$	$\cdots$
(uv^0wx^0y)	(uv^1wx^1y)	(uv^2wx^2y)	(uv^3wx^3y)	



$\cdots$

Proof of Pumping Lemma

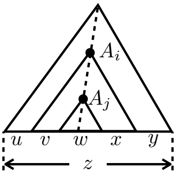
- Let L be a context-free language.
- Then, $\exists$ CFG G in Chomsky Normal Form. s.t. $L(G) = L$.
- Let $m \geq 0$ be the number of variables in G, and n be $2^m > 0$.
- Take any $z = \mathtt{a}_1\mathtt{a}_2 \cdots \mathtt{a}_k \in L$ s.t. $|z| = k \geq n$.
- Consider the longest path $A_1(= S), A_2, \cdots, A_p$ in the parse tree of z. Then, $k = |z| \leq 2^{p-1}$*by* Theorem of Size of Parse Trees in CNF. It means that $p \geq m + 1$ ($\because 2^{p-1} \geq k \geq n = 2^m$).
- Pick $m + 1$ variables from the bottom of the path: $A_{p-m}, \cdots, A_p$.
- Then, $\exists i, j.\ (p - m \leq i < j \leq p) \wedge (A_i = A_j)$ by Pigeonhole Principle.
- Split the word $z = uvwxy$ as follows. Then, it satisfies ①, ②, and ③.



$p - m \leq i < j \leq p$
and
$A_i = A_j$

Proof of Pumping Lemma

- Let L be a context-free language.
- Then, $\exists$ CFG G in Chomsky Normal Form. s.t. $L(G) = L$.
- Let $m \geq 0$ be the number of variables in G, and n be $2^m > 0$.
- Take any $z = a_1 a_2 \cdots a_k \in L$ s.t. $|z| = k \geq n$.
- Consider the longest path $A_1 (= S), A_2, \cdots, A_p$ in the parse tree of z.
  Then, $k = |z| \leq 2^{p-1}$ *by* Theorem of Size of Parse Trees in CNF.
  It means that $p \geq m + 1$ ($\because 2^{p-1} \geq k \geq n = 2^m$).
- Pick $m + 1$ variables from the bottom of the path: $A_{p-m}, \cdots, A_p$.
- Then, $\exists i, j.\ (p - m \leq i < j \leq p) \land (A_i = A_j)$ by Pigeonhole Principle.
- Split the word $z = uvwxy$ as follows. Then, it satisfies 1, 2, and 3.



$$p - m \leq i < j \leq p$$

and

$$A_i = A_j$$

Proving Languages are Not Context-Free

Lemma (Pumping Lemma for Context-Free Languages)

$$
\begin{array}{rcl}
A & = & L \textit{ is context-free} \\
& & \Downarrow \\
B & = & \exists n > 0.\ \forall z \in L.\ |z| \geq n \Rightarrow \exists z = uvwxy.\ ① \land ② \land ③
\end{array}
$$

$$
\begin{array}{rcll}
A & \Longrightarrow & B & \text{(O)} \\
B & \Longrightarrow & A & \text{(X)} \\
\neg B & \Longrightarrow & \neg A & \text{(O)}
\end{array}
$$

$$
\begin{array}{rcl}
\neg B & = & \forall n > 0.\ \neg(\forall z \in L.\ |z| \geq n \Rightarrow \exists z = uvwxy.\ ① \land ② \land ③) \\
& = & \forall n > 0.\ \exists z \in L.\ \neg(|z| \geq n \Rightarrow \exists z = uvwxy.\ ① \land ② \land ③) \\
& = & \forall n > 0.\ \exists z \in L.\ |z| \geq n \land \neg(\exists z = uvwxy.\ ① \land ② \land ③) \\
& = & \forall n > 0.\ \exists z \in L.\ |z| \geq n \land \forall z = uvwxy.\ \neg(① \land ② \land ③) \\
& = & \forall n > 0.\ \exists z \in L.\ |z| \geq n \land \forall z = uvwxy.\ \neg(① \land ②) \lor \neg③ \\
& = & \forall n > 0.\ \exists z \in L.\ |z| \geq n \land \forall z = uvwxy.\ (① \land ②) \Rightarrow \neg③
\end{array}
$$

Proving Languages are Not Context-Free

To prove a language L is **NOT** context-free, we need to show that

$$\forall n > 0.\ \exists z \in L.\ |z| \geq n \land \forall z = uvwxy.\ (① \land ②) \Rightarrow \neg ③$$

1. $|vx| > 0$
2. $|vwx| \leq n$
3. $\forall i \geq 0.\ uv^iwx^iy \in L$

Note that $\neg ③ = \exists i \geq 0.\ uv^iwx^iy \notin L$.

We can prove this by following the steps below:

1. Assume any positive integer n is given.
2. Pick a word $z \in L$.
3. Show that $|z| \geq n$.
4. Assume any split $z = uvwxy$ is given (① $|vx| > 0 \land$ ② $|vwx| \leq n$).
5. $\neg ③$ Pick $i \geq 0$, and show that $uv^iwx^iy \notin L$ using ① and ②.

Contents

Example 1

Let's prove that L is **NOT** context-free using the Pumping Lemma:

$$L = \{a^n b^n c^n \mid n \geq 0\}$$

1. Assume any positive integer n is given.
2. Let $z = a^n b^n c^n \in L$.
3. $|z| = n + n + n = 3n \geq n$.
4. Assume any split $z = uvwxy$ is given, and ① $|vx| > 0 \land$ ② $|vwx| \leq n$.
5. Let $i = 0$. We need to show that $\neg$③ $uv^0wx^0y \notin L$:
   - Since ② $|vwx| \leq n$,

     $$vx = a^p b^q \qquad (\text{or } vx = b^p c^q)$$

     where $0 \leq p, q \leq n$.
   - Since ① $|vx| > 0$, we can remove at least one a or b (or b or c) from z without changing the number of c's (or a's) when $i = 0$.
   - It means that $uv^0wx^0y \notin L$. □

Example 2

Let's prove that L is **NOT** context-free using the Pumping Lemma:

$$L = \{0^n10^n10^n \mid n \geq 0\}$$

1. Assume any positive integer n is given.
2. Let $z = 0^n10^n10^n \in L$.
3. $|z| = n + 1 + n + 1 + n = 3n + 2 \geq n$.
4. Assume any split $z = uvwxy$ is given, and 1 $|vx| > 0 \land$ 2 $|vwx| \leq n$.
5. Let $i = 0$. We need to show that $\neg$3 $uv^0wx^0y \notin L$:
   - Since 2 $|vwx| \leq n$,

     vx cannot cover the third block (or the first block) consisting of 0's.

   - Since 1 $|vx| > 0$, we can remove at least one 0 in the first or second blocks (or second or third blocks) from z without changing the number of 0's in the third block (or first block) when $i = 0$.
   - It means that $uv^0wx^0y \notin L$. □

Example 3

Let's prove that L is **NOT** context-free using the Pumping Lemma:

$$L = \{ww \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$$

1. Assume any positive integer n is given.
2. Let $z = \mathtt{a}^n\mathtt{b}^n\mathtt{a}^n\mathtt{b}^n \in L$.
3. $|z| = n + n + n + n = 4n \geq n$.
4. Assume any split $z = uvwxy$ is given, and ① $|vx| > 0 \land$ ② $|vwx| \leq n$.
5. Let $i = 0$. We need to show that $\neg$③ $uv^0wx^0y \notin L$:
   - Since ② $|vwx| \leq n$,

     vx cannot cover both two different blocks consisting of $\mathtt{a}$'s (or $\mathtt{b}$'s).

   - Since ① $|vx| > 0$, we can remove at least one $\mathtt{a}$ (or $\mathtt{b}$) in one block from z without changing the other one when $i = 0$.
   - It means that $uv^0wx^0y \notin L$. $\square$

Example 4

Let's prove that L is **NOT** context-free using the Pumping Lemma:

$$L = \{\mathtt{a}^i \mathtt{b}^j \mathtt{c}^i \mid i, j \geq 0 \land i \geq 2j\}$$

1. Assume any positive integer n is given.
2. Let $z = \mathtt{a}^{2n}\mathtt{b}^n\mathtt{c}^{2n} \in L$.
3. $|z| = 2n + n + 2n = 5n \geq n$.
4. Assume any split $z = uvwxy$ is given, and ① $|vx| > 0 \land$ ② $|vwx| \leq n$.
5. Let $i = 2$. We need to show that $\neg$③ $uv^{n+1}wx^{n+1}y \notin L$:
   - If vx covers a's (or c's),
     - vx cannot cover both a's and c's at the same time. ($\because$ ② $|vwx| \leq n$)
     - uv^2wx^2y will have more a's (or c's) than c's (or a's) ($\because$ ① $|vx| > 0$).
     - Therefore, $uv^2wx^2y \notin L$.
   - Otherwise,
     - vx covers only b. Thus, $vx = \mathtt{b}^k$ and $k > 0$ ($\because$ ① $k = |vx| > 0$).
     - $v^2wx^2y = \mathtt{a}^{2n}\mathtt{b}^{n+k}\mathtt{c}^{2n} \notin L \quad (\because k > 0 \Rightarrow 2n < 2(n + k))$.

Example 5

Let's prove that L is **NOT** context-free:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid N_\mathtt{a}(w) = N_\mathtt{b}(w) = N_\mathtt{c}(w)\}$$

where $N_\mathtt{a}(w)$, $N_\mathtt{b}(w)$, and $N_\mathtt{c}(w)$ are the number of a's, b's, and c's in w.

- It is much easier to prove that L is **NOT** context-free by using the **closure properties** under **intersection** with regular languages.
- Consider a regular expression $R = \mathtt{a}^*\mathtt{b}^*\mathtt{c}^*$ and its language:

$$L(R) = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i, j, k \geq 0\}$$

- If L is context-free, then $L \cap L(R)$ must be context-free as well because of the closure under intersection with regular languages.
- However, we know that the following language is **NOT** context-free:

$$L \cap L(R) = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \geq 0\}$$

- Since it is a contradiction, L is **NOT** context-free. $\square$

Summary

1. Pumping Lemma for Context-Free Languages
   - Size of Parse Trees in Chomsky Normal Form
   - Pumping Lemma
   - Proof of Pumping Lemma
   - Application: Proving Languages are Not Context-Free

2. Examples
   - Example 1: $L = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \geq 0\}$
   - Example 2: $L = \{0^n 1 0^n 1 0^n \mid n \geq 0\}$
   - Example 3: $L = \{ww \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$
   - Example 4: $L = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^j \mid i, j \geq 0 \land i \geq 2j\}$
   - Example 5: $L = \{w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \mid N_{\mathtt{a}}(w) = N_{\mathtt{b}}(w) = N_{\mathtt{c}}(w)\}$

Next Lecture

- Turing Machines (TMs)

Jihyeok Park
jihyeok_park@korea.ac.kr
https://plrg.korea.ac.kr