

Lecture 20 – The Pumping Lemma for Context-Free Languages

COSE215: Theory of Computation

Jihyeok Park



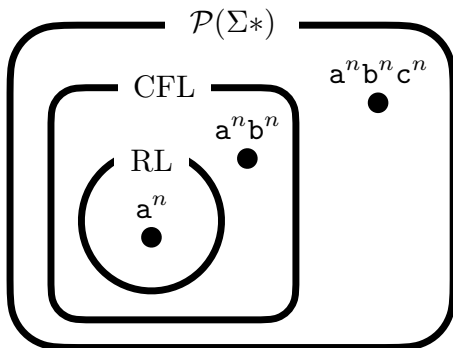
2023 Spring

- We have learned about the **Pumping Lemma** for **Regular Languages (RLs)**.
- We could use it to **prove** that some languages are **NOT** regular.

- 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)**?

- 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)**?
- For example, is it possible to prove that the following language?

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



1. Pumping Lemma for Context-Free Languages

Size of Parse Trees in Chomsky Normal Form

Pumping Lemma

Proof of Pumping Lemma

2. Proving Languages are Not Context-Free

Example 1: $L = \{a^n b^n c^n \mid n \geq 0\}$

Example 2: $L = \{0^n 10^n 10^n \mid n \geq 0\}$

Example 3: $L = \{ww \mid w \in \{a, b\}^*\}$

Example 4: $L = \{a^n b^m \mid m = n^2\}$

Example 5: $L = \{w \in \{a, b, c\}^* \mid N_a(w) = N_b(w) = N_c(w)\}$

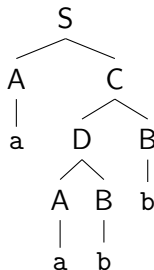
Theorem (Size of Parse Trees in Chomsky Normal Form)

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.

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

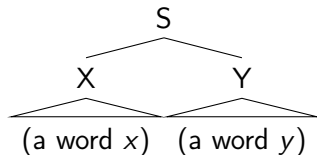
$$\begin{aligned} S &\rightarrow \epsilon \mid AC \mid AB \\ D &\rightarrow AC \mid AB \\ C &\rightarrow DB \\ A &\rightarrow a \\ B &\rightarrow b \end{aligned}$$


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 $S \rightarrow XY$. 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, $|w| = |x| + |y| \leq 2^{n-2} + 2^{n-2} = 2^{n-1}$.



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 $\forall i \geq 0. uv^iwx^iy \in L$

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 $\forall i \geq 0. uv^iwx^iy \in L$

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



$B = \exists n > 0. \forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}$

- Let L be a context-free language.

- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.

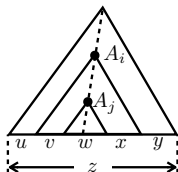
- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.
Let $m \geq 0$ be the number of variables in G and $n = 2^m \geq 1$.

- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.
Let $m \geq 0$ be the number of variables in G and $n = 2^m \geq 1$.
- Take any $z = a_1 a_2 \cdots a_k \in L$ s.t. $|z| = k \geq n$.

- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.
Let $m \geq 0$ be the number of variables in G and $n = 2^m \geq 1$.
- 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, $p \geq m + 1$ by Theorem of Size of Parse Trees in CNF.

- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.
Let $m \geq 0$ be the number of variables in G and $n = 2^m \geq 1$.
- 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, $p \geq m + 1$ by Theorem of Size of Parse Trees in CNF.
- By Pigeonhole Principle, $\exists i, j. p - m \leq i < j \leq p$ and $A_i = A_j$.
- Split the word $z = uvwxy$ as follows:

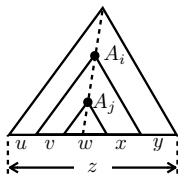
- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.
Let $m \geq 0$ be the number of variables in G and $n = 2^m \geq 1$.
- Take any $z = a_1 a_2 \cdots a_k \in L$ s.t. $|z| = k \geq n$.
- Consider the longest path $(A_1(= S), A_2, \dots, A_p)$ in the parse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



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

and

$$A_i = A_j$$



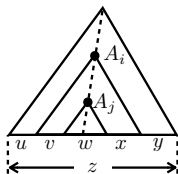
$$p - m \leq i < j \leq p$$

and

$$A_i = A_j$$

- ① $|vx| > 0$

 - Since $i < j$, the word $vw x$ derived from A_i is not equal to the word $vw x$ derived from A_j .
 - Thus, vx is not an empty word, and $|vx| > 0$.



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

and

$$A_i = A_j$$

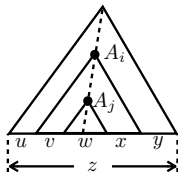
① $|vx| > 0$

- Since $i < j$, the word vwx derived from A_i is not equal to the word w derived from A_j .
- Thus, vx is not an empty word, and $|vx| > 0$.

② $|vwx| \leq n$

- Since $p - m \geq i$, 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



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

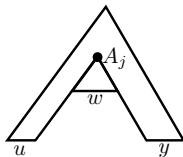
and

$$A_i = A_j$$

- ③ $\forall i \geq 0. uv^iwx^i y \in L$

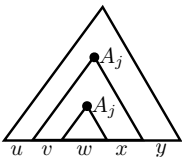
$$uwy$$

$$(uv^0wx^0y)$$



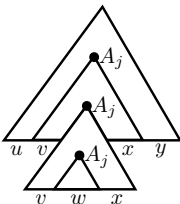
$$uvwxy$$

$$(uv^1wx^1y)$$



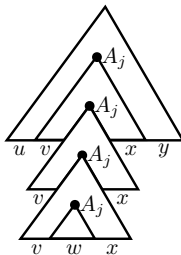
$$uvvwxxy$$

$$(uv^2wx^2y)$$



$$uvvvwxxy$$

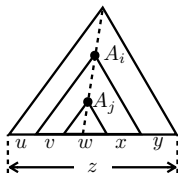
$$(uv^3wx^3y)$$



...

...

- Let L be a context-free language.
- Then, \exists CFG G in Chomsky Normal Form. s.t. $L(D) = L$.
Let $m \geq 0$ be the number of variables in G and $n = 2^m \geq 1$.
- 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, $p \geq m + 1$ by Theorem of Size of Parse Trees in CNF.
- By Pigeonhole Principle, $\exists i, j. p - m \leq i < j \leq p$ and $A_i = A_j$.
- Split the word $z = uvwxy$ as follows. Then, it satisfies ①, ②, and ③.

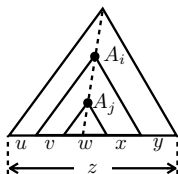


$$p - m \leq i < j \leq p$$

and

$$A_i = A_j$$

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



$$\begin{aligned} p - m \leq i < j \leq p \\ \text{and} \\ A_i = A_j \end{aligned}$$

Lemma (Pumping Lemma for Context-Free Languages)

$A =$ L is context-free



$B = \exists n > 0. \forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}$

Lemma (Pumping Lemma for Context-Free Languages)

$A =$ L is context-free



$B = \exists n > 0. \forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}$

$A \implies B \quad (0)$

Lemma (Pumping Lemma for Context-Free Languages)

$A =$ L is context-free



$B = \exists n > 0. \forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}$

$A \implies B$ (O)

$B \implies A$ (X)

Lemma (Pumping Lemma for Context-Free Languages)

$A =$ L is context-free



$B = \exists n > 0. \forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}$

$A \implies B \quad (0)$

$B \implies A \quad (X)$

$\neg B \implies \neg A \quad (0)$

Lemma (Pumping Lemma for Context-Free Languages)

$A =$ L is context-free



$B = \exists n > 0. \forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}$

$A \Rightarrow B \quad (0)$

$B \Rightarrow A \quad (X)$

$\neg B \Rightarrow \neg A \quad (0)$

$$\begin{aligned}
 \neg B &= \forall n > 0. \neg(\forall z \in L. |z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}) \\
 &= \forall n > 0. \exists z \in L. \neg(|z| \geq n \Rightarrow \exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}) \\
 &= \forall n > 0. \exists z \in L. |z| \geq n \wedge \neg(\exists z = uvwxy. \textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}) \\
 &= \forall n > 0. \exists z \in L. |z| \geq n \wedge \forall z = uvwxy. \neg(\textcircled{1} \wedge \textcircled{2} \wedge \textcircled{3}) \\
 &= \forall n > 0. \exists z \in L. |z| \geq n \wedge \forall z = uvwxy. \neg(\textcircled{1} \wedge \textcircled{2}) \vee \neg\textcircled{3} \\
 &= \forall n > 0. \exists z \in L. |z| \geq n \wedge \forall z = uvwxy. (\textcircled{1} \wedge \textcircled{2}) \Rightarrow \neg\textcircled{3}
 \end{aligned}$$

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

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

$$\textcircled{1} \quad |vx| > 0$$

$$\textcircled{2} \quad |vwx| \leq n$$

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

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

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

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

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

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

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, and $\textcircled{1} |vx| > 0 \wedge \textcircled{2} |vwx| \leq n$.
- ⑤ $\neg \textcircled{3}$ Pick $i \geq 0$, and show that $uv^iwx^iy \notin L$ using $\textcircled{1}$ and $\textcircled{2}$.

Example 1

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

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

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 $|vx| > 0 \wedge$ ② $|vwx| \leq n$.
- ⑤ Let $i = 0$. We need to show that \neg ③ $uv^0wx^0y \notin L$:
 - Since ② $|vwx| \leq n$,

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

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:

$$L = \{0^n 10^n 10^n \mid n \geq 0\}$$

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

$$L = \{0^n 10^n 10^n \mid n \geq 0\}$$

- ① Assume **any** positive integer n is given.
- ② Let $z = 0^n 10^n 10^n \in L$.
- ③ $|z| = n + 1 + n + 1 + n = 3n + 2 \geq n$.
- ④ Assume **any** split $z = uvwxy$ is given, and ① $|vx| > 0 \wedge$ 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 does not contain 0's in the third block (or the first block).
 - 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:

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

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

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

- ① Assume **any** positive integer n is given.
- ② Let $z = a^n b^n a^n b^n \in L$.
- ③ $|z| = n + n + n + n = 4n \geq n$.
- ④ Assume **any** split $z = uvwxy$ is given, and ① $|vx| > 0 \wedge$ 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 does not contain a's (or b's) in both different blocks.
 - Since ① $|vx| > 0$, we can remove at least one a (or b) in one block from z without changing the other one when $i = 0$.
 - It means that $uv^0wx^0y \notin L$. □

Example 4

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

$$L = \{a^n b^m \mid m = n^2\}$$

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

$$L = \{a^n b^m \mid m = n^2\}$$

- ① Assume **any** positive integer n is given.
- ② Let $z = a^n b^{n^2} \in L$.
- ③ $|z| = n + n^2 \geq n$.
- ④ Assume **any** split $z = uvwxy$ is given, and ① $|vx| > 0 \wedge$ ② $|vwx| \leq n$.
- ⑤ Let $i = n + 1$. We need to show that \neg ③ $uv^{n+1}wx^{n+1}y \notin L$:
 - Let's use proof by contradiction. Assume that $uv^{n+1}wx^{n+1}y \in L$.
 - Since ② $|vwx| \leq n$, $v = a^p$ and $u = b^q$ for some $0 \leq p, q \leq n$, and:

$$uv^{n+1}wx^{n+1}y = a^{n+np}b^{n^2+nq} \in L$$

- Then, $(n + np)^2 = n^2 + nq \Rightarrow n^2 p^2 + 2np = nq \Rightarrow np^2 + 2p = q$.
- Since ① $|vx| > 0$, $p > 0$ or $q > 0$. However, $q > n$ if $p > 0$ and $q = 0$ if $p = 0$. Therefore, we have a contradiction. \square

Example 5

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

$$L = \{w \in \{a, b, c\}^* \mid N_a(w) = N_b(w) = N_c(w)\}$$

where $N_a(w)$, $N_b(w)$, and $N_c(w)$ are the number of a's, b's, and c's in w .

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

$$L = \{w \in \{a, b, c\}^* \mid N_a(w) = N_b(w) = N_c(w)\}$$

where $N_a(w)$, $N_b(w)$, and $N_c(w)$ are the number of a's, b's, and c's in w .

- It is much easier to use **closure properties** under **intersection** with regular languages.
- Consider a regular expressions $R = a^*b^*c^*$ and its language:

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

- 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:

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

- Therefore, L is **NOT** context-free. □

1. Pumping Lemma for Context-Free Languages

Size of Parse Trees in Chomsky Normal Form

Pumping Lemma

Proof of Pumping Lemma

2. Proving Languages are Not Context-Free

Example 1: $L = \{a^n b^n c^n \mid n \geq 0\}$

Example 2: $L = \{0^n 10^n 10^n \mid n \geq 0\}$

Example 3: $L = \{ww \mid w \in \{a, b\}^*\}$

Example 4: $L = \{a^n b^m \mid m = n^2\}$

Example 5: $L = \{w \in \{a, b, c\}^* \mid N_a(w) = N_b(w) = N_c(w)\}$

- Please see <https://github.com/ku-plrg-classroom/docs/tree/main/equiv-pda-cfg>.
- The due date is May 30 (Tue.).
- Please only submit `Implementation.scala` file to **Blackboard**.

- Turing Machines (TMs)

Jihyeok Park

jihyeok_park@korea.ac.kr

<https://plrg.korea.ac.kr>