# Lecture 19 – Closure Properties of Context-Free Languages

## COSE215: Theory of Computation

Jihyeok Park

PLRG

2023 Spring

# Recall

- A **context-free language (CFL)** is defined in three different ways:
  - A **context free grammar (CFG)**
  - A **pushdown automaton (PDA)** with **final states**
  - A **pushdown automaton (PDA)** with **empty stacks**
- We have learned that the class of **regular languages** is **closed** under various operations. (**Closure Properties**)
- For which operations is the class of **CFLs closed**?

# Contents

# Closure Properties of CFLs

**Definition (Closure Properties)**

The class of CFLs is **closed** under an $n$-ary operator op if and only if $\text{op}(L_1, \cdots, L_n)$ is context-free for any CFLs $L_1, \cdots, L_n$. We say that such properties are **closure properties** of CFLs.

The class of CFLs is closed under the following operations:

- Union
- Concatenation
- Kleene Star
- Homomorphism
- Reverse

# Closure under Union

**Theorem (Closure under Union)**

*If $L_1$ and $L_2$ are context-free languages, then so is $L_1 \cup L_2$.*

# Closure under Union

PLRG

**Theorem (Closure under Union)**

*If $L_1$ and $L_2$ are context-free languages, then so is $L_1 \cup L_2$.*

**Proof)** For given two CFLs $L_1$ and $L_2$, we can always construct two CFGs:

$$\begin{aligned} G_1 &= (V_1, \Sigma, S_1, R_1) \\ G_2 &= (V_2, \Sigma, S_2, R_2) \end{aligned}$$

such that $L_1 = L(G_1)$ and $L_2 = L(G_2)$.
Note that the variables of $G_1$ and $G_2$ should be disjoint. (i.e., $V_1 \cap V_2 = \varnothing$)
Then, $L_1 \cup L_2$ is accepted by the CFG $G = (V, \Sigma, S, R)$ where:

- $V = V_1 \cup V_2 \cup \{S\}$
- $S$ is a new start variable (i.e., $S \notin V_1 \cup V_2$)
- $R = R_1 \cup R_2 \cup \{S \to S_1, S \to S_2\}$

□

# Closure under Union – Example

For example, consider the following two CFLs:

$$L_1 = \{\mathtt{ab}^n \mid n \geq 0\} \qquad L_2 = \{\mathtt{ac}^n \mid n \geq 0\}$$

Then, $L_1$ is accepted by:

$$S_1 \to \mathtt{a}X \qquad X \to \mathtt{b}X \mid \epsilon$$

and $L_2$ is accepted by:

$$S_2 \to \mathtt{a}X \qquad X \to \mathtt{c}X \mid \epsilon$$

But, the same variable $X$ is used in both grammars.
So, we need to rename it to different variables, such as $B$ and $C$.

# Closure under Union – Example

For example, consider the following two CFLs:

$$L_1 = \{\mathtt{ab}^n \mid n \geq 0\} \qquad L_2 = \{\mathtt{ac}^n \mid n \geq 0\}$$

Then, $L_1$ is accepted by:

$$S_1 \rightarrow \mathtt{a}B \qquad B \rightarrow \mathtt{b}B \mid \epsilon$$

and $L_2$ is accepted by:

$$S_2 \rightarrow \mathtt{a}C \qquad C \rightarrow \mathtt{c}C \mid \epsilon$$

# Closure under Union – Example

For example, consider the following two CFLs:

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

Then, $L_1$ is accepted by:

$$S_1 \rightarrow \mathtt{a}B \qquad B \rightarrow \mathtt{b}B \mid \epsilon$$

and $L_2$ is accepted by:

$$S_2 \rightarrow \mathtt{a}C \qquad C \rightarrow \mathtt{c}C \mid \epsilon$$

Then, $L_1 \cup L_2$ is accepted by the following CFG:

$$\begin{array}{llll} S & \rightarrow S_1 \mid S_2 & & \\ S_1 & \rightarrow \mathtt{a}B & B & \rightarrow \mathtt{b}B \mid \epsilon \\ S_2 & \rightarrow \mathtt{a}C & C & \rightarrow \mathtt{c}C \mid \epsilon \end{array}$$

# Closure under Concatenation

PLRG

**Theorem (Closure under Concatenation)**

*If $L_1$ and $L_2$ are context-free languages, then so is $L_1 \cdot L_2$.*

# Closure under Concatenation

**Theorem (Closure under Concatenation)**

*If $L_1$ and $L_2$ are context-free languages, then so is $L_1 \cdot L_2$.*

**Proof)** For given two CFLs $L_1$ and $L_2$, we can always construct two CFGs:

$$\begin{aligned} G_1 &= (V_1, \Sigma, S_1, R_1) \\ G_2 &= (V_2, \Sigma, S_2, R_2) \end{aligned}$$

such that $L_1 = L(G_1)$ and $L_2 = L(G_2)$.
Note that the variables of $G_1$ and $G_2$ should be disjoint. (i.e., $V_1 \cap V_2 = \varnothing$)
Then, $L_1 \cdot L_2$ is accepted by the CFG $G = (V, \Sigma, S, R)$ where:

- $V = V_1 \cup V_2 \cup \{S\}$
- $S$ is a new start variable (i.e., $S \notin V_1 \cup V_2$)
- $R = R_1 \cup R_2 \cup \{S \to S_1 S_2\}$

□

# Closure under Union – Example

For example, consider the following two CFLs:

$$L_1 = \{\mathtt{ab}^n \mid n \geq 0\} \qquad L_2 = \{\mathtt{ac}^n \mid n \geq 0\}$$

Then, $L_1$ is accepted by:

$$S_1 \to \mathtt{a}B \qquad B \to \mathtt{b}B \mid \epsilon$$

and $L_2$ is accepted by:

$$S_2 \to \mathtt{a}C \qquad C \to \mathtt{c}C \mid \epsilon$$

Then, $L_1 \cdot L_2$ is accepted by the following CFG:

$$\begin{array}{llll} S & \to S_1 S_2 & & \\ S_1 & \to \mathtt{a}B & B & \to \mathtt{b}B \mid \epsilon \\ S_2 & \to \mathtt{a}C & C & \to \mathtt{c}C \mid \epsilon \end{array}$$

# Closure under Kleene Star

**Theorem (Closure under Kleene Star)**

*If $L$ is a context-free language, then so is $L^*$.*

**Theorem (Closure under Kleene Star)**

*If $L$ is a context-free language, then so is $L^*$.*

**Proof)** For a given CFL $L$, we can always construct a CFG:

$$G = (V, \Sigma, S, R)$$

such that $L = L(G)$.

Then, $L^*$ is accepted by the CFG $G' = (V', \Sigma, S', R')$ where:

- $V' = V \cup \{S'\}$
- $S'$ is a new start variable (i.e., $S' \notin V$)
- $R' = R \cup \{S' \to \epsilon, S' \to SS'\}$

□

# Closure under Kleene Star – Example

For example, consider the following CFL:

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

Then, $L$ is accepted by:

$$S \to \epsilon \mid \mathtt{a}S\mathtt{b}$$

Then, $L^*$ is accepted by the following CFG:

$$\begin{aligned} S' &\to \epsilon \mid SS' \\ S &\to \epsilon \mid \mathtt{a}S\mathtt{b} \end{aligned}$$

# Closure under Homomorphism

PLRG

## Definition (Homomorphism)

Suppose $\Sigma$ and $\Gamma$ are two finite sets of symbols. Then, a function $h : \Sigma \to \Gamma^*$ is called a **homomorphism**. For a given word $w = a_1 a_2 \cdots a_n$,

$$h(w) = h(a_1)h(a_2) \cdots h(a_n)$$

For a language $L$, $h(L) = \{h(w) \mid w \in L\}$.

# Closure under Homomorphism

PLRG

## Definition (Homomorphism)

Suppose $\Sigma$ and $\Gamma$ are two finite sets of symbols. Then, a function $h : \Sigma \to \Gamma^*$ is called a **homomorphism**. For a given word $w = a_1 a_2 \cdots a_n$,

$$h(w) = h(a_1)h(a_2) \cdots h(a_n)$$

For a language $L$, $h(L) = \{h(w) \mid w \in L\}$.

For example, $h : \{\mathtt{0}, \mathtt{1}\} \to \{\mathtt{a}, \mathtt{b}\}^*$ be a homomorphism such that:

$$h(\mathtt{0}) = \mathtt{ab} \qquad h(\mathtt{1}) = \mathtt{a}$$

Then,

$$h(\mathtt{10}) = \mathtt{aab} \qquad h(\mathtt{010}) = \mathtt{abaab} \qquad h(\mathtt{1100}) = \mathtt{aaabab}$$

# Closure under Homomorphism

**Theorem (Closure under Homomorphism)**

*If $h$ is a homomorphism and $L$ is a context-free language, then so is $h(L)$.*

PLRG

## Theorem (Closure under Homomorphism)

*If $h$ is a homomorphism and $L$ is a context-free language, then so is $h(L)$.*

**Proof)** For a given CFL $L$, we can always construct a CFG:

$$G = (V, \Sigma, S, R)$$

such that $L = L(G)$.
Then, for a given homomorphism $h : \Sigma \to \Gamma^*$, $h(L)$ is accepted by the CFG $G' = (V', \Gamma, S, R')$ where:

- $V' = V \cup \{X_a \mid a \in \Sigma\}$
- $R' = \{Y \to Y'_1 \cdots Y'_n \mid Y \to Y_1 \cdots Y_n \in R\} \cup \{X_a \to h(a) \mid a \in \Sigma\}$
  where $\forall 1 \leq i \leq n.\ Y'_i = \begin{cases} Y_i & \text{if } Y_i \in V \\ X_a & \text{if } Y_i = a \in \Sigma \end{cases}$

□

# Closure under Homomorphism – Example

For example, consider the following CFL:

$$L = \{ww^R \mid w \in \{0, 1\}^*\}$$

Then, $L$ is accepted by:

$$S \rightarrow \epsilon \mid 0S0 \mid 1S1$$

If a homomorphism $h : \{0, 1\} \rightarrow \{\texttt{a}, \texttt{b}\}^*$ is defined as follows:

$$h(0) = \texttt{ab} \qquad h(1) = \texttt{a}$$

Then, $h(L)$ is accepted by the following CFG:

$$\begin{aligned} S &\rightarrow \epsilon \mid X_0 S X_0 \mid X_1 S X_1 \\ X_0 &\rightarrow \texttt{ab} \\ X_1 &\rightarrow \texttt{a} \end{aligned}$$

# Closure under Reverse

**Theorem (Closure under Reverse)**

*If $L$ is a context-free language, then so is $L^R$.*

# Closure under Reverse

**Theorem (Closure under Reverse)**

*If $L$ is a context-free language, then so is $L^R$.*

**Proof)** For a given CFL $L$, we can always construct a CFG:

$$G = (V, \Sigma, S, R)$$

such that $L = L(G)$.
Then, $L^R$ is accepted by the CFG $G' = (V, \Sigma, S, R')$ where:

- $R' = \{X \to \alpha^R \mid X \to \alpha \in R\}$

□

# Closure under Reverse – Example

For example, consider the following CFL:

$$L = \{(\mathtt{ab})^n \mathtt{c}^n \mathtt{d}^m \mid n, m \geq 0\}$$

Then, $L$ is accepted by:

$$\begin{aligned} S &\rightarrow X \mid S\mathtt{d} \\ X &\rightarrow \epsilon \mid \mathtt{ab}X\mathtt{c} \end{aligned}$$

Then, $L^R$ is accepted by the following CFG:

$$\begin{aligned} S &\rightarrow X \mid \mathtt{d}S \\ X &\rightarrow \epsilon \mid \mathtt{c}X\mathtt{ba} \end{aligned}$$

# Non-Closure Properties of CFLs

PLRG

## Definition (Closure Properties)

The class of CFLs is **closed** under an $n$-ary operator op if and only if op$(L_1, \cdots, L_n)$ is context-free for any CFLs $L_1, \cdots, L_n$. We say that such properties are **closure properties** of CFLs.

The class of CFLs is **NOT** closed under the following operations:

- Intersection
- Complement
- Difference

# Non-Closure Properties of CFLs

**Definition (Closure Properties)**

The class of CFLs is **closed** under an $n$-ary operator op if and only if $\text{op}(L_1, \cdots, L_n)$ is context-free for any CFLs $L_1, \cdots, L_n$. We say that such properties are **closure properties** of CFLs.

The class of CFLs is **NOT** closed under the following operations:

- Intersection
- Complement
- Difference

We will prove it by using the fact that the following language is not a CFL:

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

# Non-Closure Properties of CFLs

## Definition (Closure Properties)

The class of CFLs is **closed** under an $n$-ary operator op if and only if op$(L_1, \cdots, L_n)$ is context-free for any CFLs $L_1, \cdots, L_n$. We say that such properties are **closure properties** of CFLs.

The class of CFLs is **NOT** closed under the following operations:

- Intersection
- Complement
- Difference

We will prove it by using the fact that the following language is not a CFL:

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

We will learn how to prove that $L$ is not a CFL in the next lecture (Pumping Lemma for CFLs).

# Non-Closure under Intersection

**Theorem (Non-Closure under Intersection)**

*The class of CFLs is* **NOT** *closed under intersection.*

# Non-Closure under Intersection

**Theorem (Non-Closure under Intersection)**

*The class of CFLs is* **NOT** *closed under intersection.*

**Proof)** Consider the following two languages:

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

Then, $L_1$ is accepted by:

$$S_1 \rightarrow X \mid S_1\mathtt{c} \qquad X \rightarrow \epsilon \mid \mathtt{a}X\mathtt{b}$$

and $L_2$ is accepted by:

$$S_2 \rightarrow Y \mid \mathtt{a}S_2 \qquad Y \rightarrow \epsilon \mid \mathtt{b}Y\mathtt{c}$$

Thus, they are both CFLs. However, their intersection is not a CFL:

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

□

# Non-Closure under Complement and Difference

**Theorem (Non-Closure under Complement)**

*The class of CFLs is* **NOT** *closed under complement.*

# Non-Closure under Complement and Difference

**Theorem (Non-Closure under Complement)**

*The class of CFLs is* **NOT** *closed under complement.*

**Proof)** Assume that the class of CFLs is closed under complement. Then, for any two CFLs $L_1$ and $L_2$, $L_1 \cap L_2$ is also a CFL:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

However, we have already proved that the class of CFLs is not closed under intersection. Thus, the class of CFLs is not closed under complement. □

# Non-Closure under Complement and Difference

**Theorem (Non-Closure under Complement)**

*The class of CFLs is* **NOT** *closed under complement.*

**Proof)** Assume that the class of CFLs is closed under complement. Then, for any two CFLs $L_1$ and $L_2$, $L_1 \cap L_2$ is also a CFL:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

However, we have already proved that the class of CFLs is not closed under intersection. Thus, the class of CFLs is not closed under complement. □

**Theorem (Non-Closure under Difference)**

*The class of CFLs is* **NOT** *closed under difference.*

# Non-Closure under Complement and Difference

**Theorem (Non-Closure under Complement)**

*The class of CFLs is* **NOT** *closed under complement.*

**Proof)** Assume that the class of CFLs is closed under complement. Then, for any two CFLs $L_1$ and $L_2$, $L_1 \cap L_2$ is also a CFL:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

However, we have already proved that the class of CFLs is not closed under intersection. Thus, the class of CFLs is not closed under complement. □

**Theorem (Non-Closure under Difference)**

*The class of CFLs is* **NOT** *closed under difference.*

**Proof)** Similarly, we can prove it using the following fact:

$$L_1 \cap L_2 = L_1 \setminus (L_1 \setminus L_2)$$

□

**Definition (Closure Properties)**

The class of CFLs is **closed** under an $n$-ary operator op if and only if op$(L_1, \cdots, L_n)$ is context-free for any CFLs $L_1, \cdots, L_n$. We say that such properties are **closure properties** of CFLs.

The class of CFLs is closed under the following operations with RLs:

- Intersection
- Difference

# Closure under Intersection with RLs

**Theorem (Closure under Intersection with RLs)**

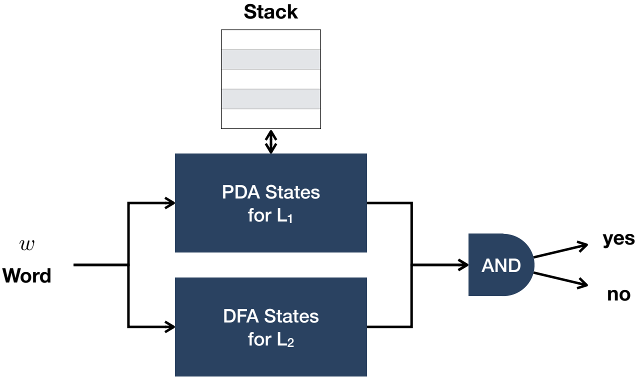
*If $L_1$ Is a CFL and $L_2$ is a RL, then $L_1 \cap L_2$ is a CFL.*

# Closure under Intersection with RLs

**Theorem (Closure under Intersection with RLs)**

*If $L_1$ Is a CFL and $L_2$ is a RL, then $L_1 \cap L_2$ is a CFL.*

There exists a PDA $P$ that accepts $L_1$ by final states and a DFA $D$ that accepts $L_2$. We will construct a PDA $P'$ that accepts $L_1 \cap L_2$ as follows:



# Closure under Intersection with RLs

**Theorem (Closure under Intersection with RLs)**

*If $L_1$ Is a CFL and $L_2$ is a RL, then $L_1 \cap L_2$ is a CFL.*

**Proof)** Consider a PDA $P = (Q_P, \Sigma, \Gamma, \delta_P, q_P, Z, F_P)$ and a DFA $D = (Q_D, \Sigma, \delta_D, q_D, F_D)$ such that:

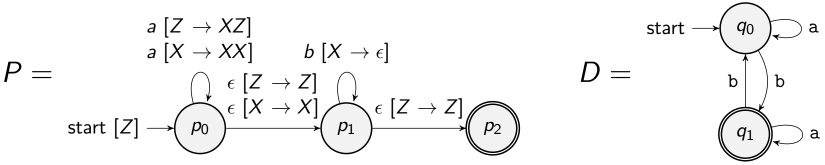
$$L_F(P) = L_1 \qquad L(D) = L_2$$

Then, $L_1 \cap L_2$ is accepted by the PDA $P' = (Q, \Sigma, \Gamma, \delta, q_0, Z, F)$ by final states, where:

- $Q = Q_P \times Q_D$
- $\delta((p, q), \epsilon, X) = \{((p', q), \alpha) \mid (p', \alpha) \in \delta_P(p, \epsilon, X)\}$
- $\delta((p, q), a, X) = \{((p', q'), \alpha) \mid (p', \alpha) \in \delta_P(p, a, X) \wedge q' = \delta_D(q, a)\}$
- $q_0 = (q_P, q_D)$
- $F = F_P \times F_D$

□

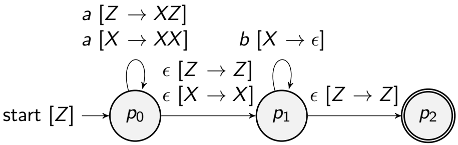
For example, consider the following PDA $P$ and DFA $D$:



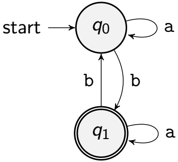
# Closure under Intersection with RLs – Example

For example, consider the following PDA $P$ and DFA $D$:

$P =$

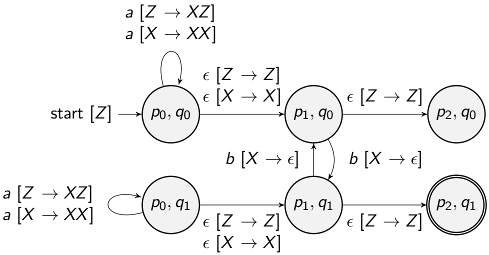



$D =$




Then, a PDA $P'$ that accepts $L_F(P) \cap L(D)$ by the final states can be constructed as follows:

$P' =$



# Closure under Difference with RLs

**Theorem (Closure under Difference with RLs)**

*If $L_1$ Is a CFL and $L_2$ is a RL, then $L_1 \setminus L_2$ is a CFL.*

# Closure under Difference with RLs

**Theorem (Closure under Difference with RLs)**

*If $L_1$ Is a CFL and $L_2$ is a RL, then $L_1 \setminus L_2$ is a CFL.*

**Proof)** We know the following fact:

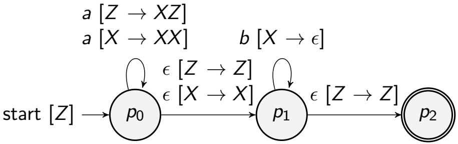
$$L_1 \setminus L_2 = L_1 \cap \overline{L_2}$$

Since the class of RLs is closed under complement, $\overline{L_2}$ is a RL. In addition, we know that the class of CFLs is closed under intersection with RLs. Thus, $L_1 \setminus L_2$ is a CFL. □

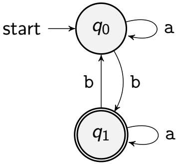
# Closure under Difference with RLs – Example

For example, consider the following PDA $P$ and DFA $D$:

$P =$



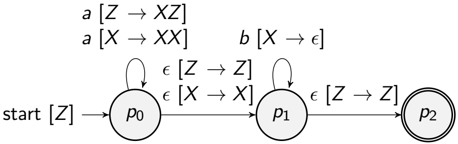

$D =$



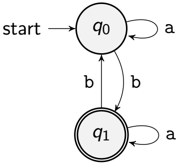
# Closure under Difference with RLs – Example

For example, consider the following PDA $P$ and DFA $D$:

$P =$

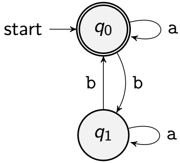



$D =$




Then, a DFA $\overline{D}$ that accepts $\overline{L(D)}$ and a PDA $P'$ that accepts $L_F(P) \setminus L(D) = L_F(P) \cap \overline{L(D)}$ can be constructed as follows:

$\overline{D} =$

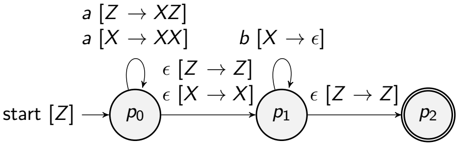


# Closure under Difference with RLs – Example

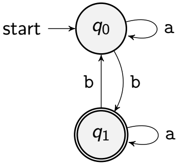
PLRG

For example, consider the following PDA $P$ and DFA $D$:

$P =$

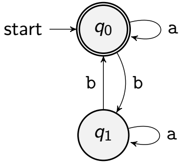



$D =$

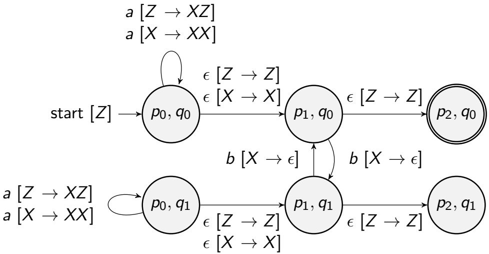



Then, a DFA $\overline{D}$ that accepts $\overline{L(D)}$ and a PDA $P'$ that accepts $L_F(P) \setminus L(D) = L_F(P) \cap \overline{L(D)}$ can be constructed as follows:

$\overline{D} =$




$P' =$



# Summary

1. Closure Properties of Context-Free Languages
    - Union
    - Concatenation
    - Kleene Star
    - Homomorphism
    - Reversal
2. Non-Closure Properties of Context-Free Languages
    - Intersection
    - Complement and Difference
3. Closure Properties of CFLs with Regular Languages
    - Intersection with Regular Languages
    - Difference with Regular Languages

# Next Lecture

- The Pumping Lemma for Context-Free Languages

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