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
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 $L_1 \cup L_2$.

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

$$G_1 = (V_1, \Sigma, S_1, R_1)$$
$$G_2 = (V_2, \Sigma, S_2, R_2)$$

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., $S \notin V_1 \cup V_2$)
- $R = R_1 \cup R_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}$

$\square$

Closure under Union – Example

For example, consider the following two CFLs:

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

Then, L_1 is accepted by:

$$S_1 \rightarrow aX \qquad X \rightarrow bX \mid \epsilon$$

and L_2 is accepted by:

$$S_2 \rightarrow aX \qquad X \rightarrow cX \mid \epsilon$$

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:

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

and L_2 is accepted by:

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

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

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

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:

$$G_1 = (V_1, \Sigma, S_1, R_1)$$
$$G_2 = (V_2, \Sigma, S_2, R_2)$$

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., $S \notin V_1 \cup V_2$)
- $R = R_1 \cup R_2 \cup \{S \rightarrow S_1 S_2\}$

$\square$

Closure under Union – Example

For example, consider the following two CFLs:

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

Then, L_1 is accepted by:

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

and L_2 is accepted by:

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

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

$$
\begin{aligned}
S &\to S_1 S_2 \\
S_1 &\to \text{a}B \qquad & B &\to \text{b}B \mid \epsilon \\
S_2 &\to \text{a}C \qquad & C &\to \text{c}C \mid \epsilon
\end{aligned}
$$

Closure under Kleene Star

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., $S' \notin V$)
- $R' = R \cup \{S' \to \epsilon, S' \to SS'\}$

$\square$

For example, consider the following CFL:

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

Then, L is accepted by:

$$S \rightarrow \epsilon \mid aSb$$

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

$$S' \rightarrow \epsilon \mid SS'$$
$$S \rightarrow \epsilon \mid aSb$$

Closure under Homomorphism

Definition (Homomorphism)

Suppose Σ and Γ 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, $h : \{0, 1\} \to \{a, b\}^*$ be a homomorphism such that:

$$h(0) = ab \qquad h(1) = a$$

Then,

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

Closure under Homomorphism

Theorem (Closure under Homomorphism)

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

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 $\forall 1 \leq i \leq n.\ Y_i' = \begin{cases} Y_i & \text{if } Y_i \in V \\ X_a & \text{if } Y_i = a \in \Sigma \end{cases}$

$\square$

PLRG

For example, consider the following CFL:

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

Then, L is accepted by:

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

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

$$h(0) = ab \qquad h(1) = a$$

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

$$
\begin{aligned}
S &\to \epsilon \mid X_0SX_0 \mid X_1SX_1 \\
X_0 &\to ab \\
X_1 &\to a
\end{aligned}
$$

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:

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

$\square$

PLRG

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:

$$S \rightarrow X \mid S\mathtt{d}$$
$$X \rightarrow \epsilon \mid \mathtt{ab}X\mathtt{c}$$

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

$$S \rightarrow X \mid \mathtt{d}S$$
$$X \rightarrow \epsilon \mid \mathtt{c}X\mathtt{ba}$$

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:

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

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

Theorem (Non-Closure under Intersection)

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

Proof) Consider the following two languages:

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

Then, L_1 is accepted by:

$$S_1 \to X \mid S_1 c \qquad X \to \epsilon \mid a X b$$

and L_2 is accepted by:

$$S_2 \to Y \mid a S_2 \qquad Y \to \epsilon \mid b Y c$$

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

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

$\square$

OPLRG

Definition (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. $\quad\square$

Definition (Non-Closure under Difference)

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

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

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

$\square$

Closure Properties of CFLs with Regular Languages ◆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.
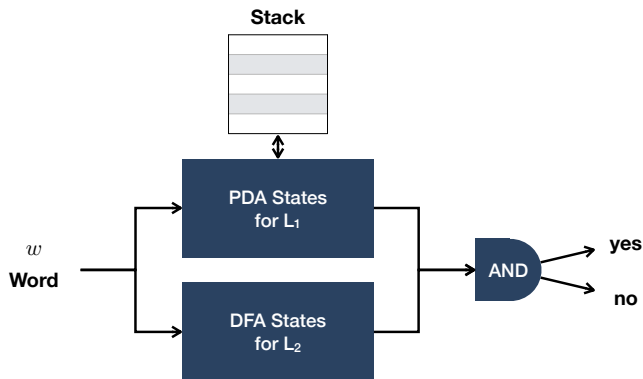
The class of CFLs is closed under the following operations with RLs:

- 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.

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:

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:

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

$\square$

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

$$P = \quad \text{start } [Z] \longrightarrow p_0 \xrightarrow{\ \epsilon\,[Z \to Z]\ }_{\epsilon\,[X \to X]} p_1 \xrightarrow{\ \epsilon\,[Z \to Z]\ } p_2$$

with transitions $a\,[Z \to XZ]$, $a\,[X \to XX]$ on p_0, and $b\,[X \to \epsilon]$ on p_1.

$$D = \quad \text{start} \longrightarrow q_0 \ (a) \xrightarrow[\ b\ ]{\ b\ } q_1 \ (a)$$

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

$$P' = \quad \text{start } [Z] \longrightarrow$$

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.

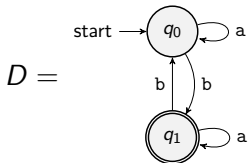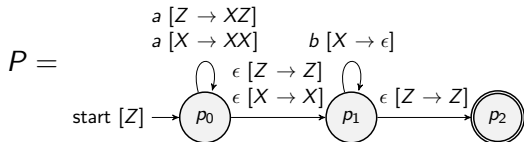**Proof)** We know the following fact:

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

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. □
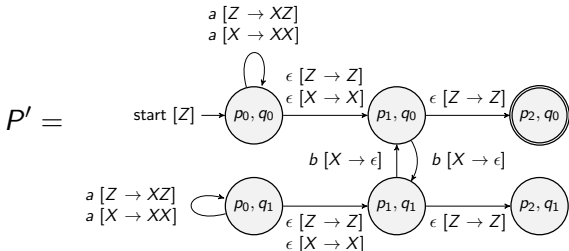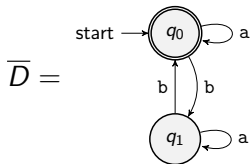
◆PLRG

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

$$P = \quad \quad \quad \quad \quad \quad \quad \quad \quad D =$$

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:

$$\overline{D} = \quad \quad \quad \quad P' =$$

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

- The Pumping Lemma for Context-Free Languages

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