

Lecture 11 – Context-Free Grammars (CFGs) and Languages (CFLs)

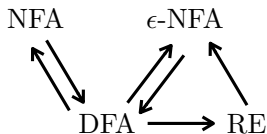
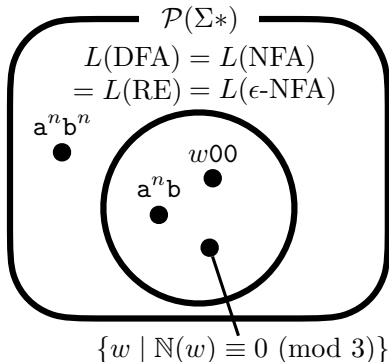
COSE215: Theory of Computation

Jihyeok Park



2023 Spring

- Regular Languages
 - Finite Automata - DFA, NFA, ϵ -NFA
 - Regular Expressions



- Is there a way to describe languages that are not regular?

1. Context-Free Grammars (CFGs)

- Definition

- Derivation Relations

- Leftmost and Rightmost Derivations

- Sentential Forms

- Context-Free Languages (CFLs)

- Examples

- Consider the following language:

$$L = \{w \in \{(,)\}^* \mid w \text{ is balanced}\}$$

For example, the following words are in (or not in) L :

$L \ni \epsilon, (), (()), ()(), (())(), (())(), ((())), \dots$

$L \not\ni (,),)(), ((), ()), (())(), \dots$

- Consider the following language:

$$L = \{w \in \{(,)\}^* \mid w \text{ is balanced}\}$$

For example, the following words are in (or not in) L :

$L \ni \epsilon, (), (()), ()(), (())(), (())(), ((())), \dots$

$L \not\ni (,),)(), ((), ()), (())(), \dots$

- This language is **NOT regular**.
(Do it yourself using the pumping lemma).

- Consider the following language:

$$L = \{w \in \{(,)\}^* \mid w \text{ is balanced}\}$$

For example, the following words are in (or not in) L :

$L \ni \epsilon, (), (()), ()(), (())(), (())(), ((())), \dots$

$L \not\ni (,),)(), ((), ()), (())(), (())(), \dots$

- This language is **NOT regular**.
(Do it yourself using the pumping lemma).
- Then, how can we describe this language?

- Consider the following language:

$$L = \{w \in \{(,)\}^* \mid w \text{ is balanced}\}$$

For example, the following words are in (or not in) L :

$L \ni \epsilon, (), (()), ()(), (())(), (())(), ((())), \dots$

$L \not\ni (,),)(), ((), (()), (()), (())(), \dots$

- This language is **NOT regular**.
(Do it yourself using the pumping lemma).
- Then, how can we describe this language?
Context-Free Grammars (CFGs)

Definition (Context-Free Grammar (CFG))

A **context-free grammar** is a 4-tuple:

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

where

- V : a finite set of **variables** (nonterminals)
- Σ : a finite set of **symbols** (terminals)
- $S \in V$: the **start variable**
- $P \subseteq V \times (V \cup \Sigma)^*$: a set of **production rules**.

Definition (Context-Free Grammar (CFG))

A **context-free grammar** is a 4-tuple:

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

where

- V : a finite set of **variables** (nonterminals)
- Σ : a finite set of **symbols** (terminals)
- $S \in V$: the **start variable**
- $P \subseteq V \times (V \cup \Sigma)^*$: a set of **production rules**.

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

where P is defined as:

$$\begin{array}{lll}
 S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\
 A \rightarrow (S) & B \rightarrow SS &
 \end{array}$$

```
// The type definitions of symbols and variables
type Symbol = Char
type Variable = String
// The definition of context-free grammars
case class CFG(
  variables: Set[Variable],
  symbols: Set[Symbol],
  start: Variable,
  productions: Set[(Variable, List[Variable | Symbol])],
)
// An example of CFG
val cfg1: CFG = CFG(
  variables = Set("S", "A", "B"),
  symbols = Set('(', ')'),
  start = "S",
  productions = Set(
    "S" -> Nil,
    "S" -> List("A"),
    "S" -> List("B"),
    "A" -> List('(', "S", ')'),
    "B" -> List("S", "S"),
  )
)
```

Definition (Derivation Relation (\Rightarrow))

Consider a CFG $G = (V, \Sigma, S, P)$. If a production rule $A \rightarrow \gamma \in P$ exists, the **derivation relation** $\Rightarrow \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$ is defined as:

$$\alpha A \beta \Rightarrow \alpha \gamma \beta$$

for all $\alpha, \beta \in (V \cup \Sigma)^*$. We say that $\alpha A \beta$ **derives** $\alpha \gamma \beta$.

Definition (Closure of Derivation Relation (\Rightarrow^*))

The **closure of derivation relation** \Rightarrow^* is defined as:

- **(Basis Case)** $\forall \alpha \in (V \cup \Sigma)^*. \alpha \Rightarrow^* \alpha$
- **(Induction Case)** $\forall \alpha, \beta, \gamma \in (V \cup \Sigma)^*. (\alpha \Rightarrow^* \gamma)$ if

$$(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow^* \gamma)$$

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

A derivation for $((())())$:

$$\begin{array}{l} S \Rightarrow B \Rightarrow SS \Rightarrow AS \Rightarrow (S)S \\ \Rightarrow (A)S \Rightarrow ((S))S \Rightarrow (())S \Rightarrow (())A \\ \Rightarrow (())(S) \Rightarrow (())() \end{array}$$

Thus,

$$\begin{array}{llll} S \Rightarrow^* S & S \Rightarrow^* B & S \Rightarrow^* SS & \dots \\ \dots & S \Rightarrow^* (())A & S \Rightarrow^* (())(S) & S \Rightarrow^* (())() \end{array}$$

- **Leftmost Derivation** ($\xRightarrow{\text{lm}}$): always derive the *leftmost* variable.
- **Rightmost Derivation** ($\xRightarrow{\text{rm}}$): always derive the *rightmost* variable.

- **Leftmost Derivation** ($\xRightarrow{\text{lm}}$): always derive the *leftmost* variable.
- **Rightmost Derivation** ($\xRightarrow{\text{rm}}$): always derive the *rightmost* variable.

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

- **Leftmost Derivation** ($\xRightarrow{\text{lm}}$): always derive the *leftmost* variable.
- **Rightmost Derivation** ($\xRightarrow{\text{rm}}$): always derive the *rightmost* variable.

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

For example, the **leftmost derivation** for $((()))()$:

$$\begin{array}{ccccccccc} S & \xRightarrow{\text{lm}} & B & \xRightarrow{\text{lm}} & SS & \xRightarrow{\text{lm}} & AS & \xRightarrow{\text{lm}} & (S)S & \xRightarrow{\text{lm}} & (A)S \\ & \xRightarrow{\text{lm}} & ((S))S & \xRightarrow{\text{lm}} & ((()))S & \xRightarrow{\text{lm}} & ((()))A & \xRightarrow{\text{lm}} & ((()))(S) & \xRightarrow{\text{lm}} & ((()))() \end{array}$$

- **Leftmost Derivation** ($\xRightarrow{\text{lm}}$): always derive the *leftmost* variable.
- **Rightmost Derivation** ($\xRightarrow{\text{rm}}$): always derive the *rightmost* variable.

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

For example, the **leftmost derivation** for $((()))()$:

$$\begin{array}{ccccccccc} S & \xRightarrow{\text{lm}} & B & \xRightarrow{\text{lm}} & SS & \xRightarrow{\text{lm}} & AS & \xRightarrow{\text{lm}} & (S)S & \xRightarrow{\text{lm}} & (A)S \\ & \xRightarrow{\text{lm}} & ((S))S & \xRightarrow{\text{lm}} & ((()))S & \xRightarrow{\text{lm}} & ((()))A & \xRightarrow{\text{lm}} & ((()))(S) & \xRightarrow{\text{lm}} & ((()))() \end{array}$$

and, the **rightmost derivation** for $((()))()$:

$$\begin{array}{ccccccccc} S & \xRightarrow{\text{rm}} & B & \xRightarrow{\text{rm}} & SS & \xRightarrow{\text{rm}} & SA & \xRightarrow{\text{rm}} & S(S) & \xRightarrow{\text{rm}} & S() \\ & \xRightarrow{\text{rm}} & A() & \xRightarrow{\text{rm}} & (S)() & \xRightarrow{\text{rm}} & (A)() & \xRightarrow{\text{rm}} & ((S))() & \xRightarrow{\text{rm}} & ((()))() \end{array}$$

Definition (Sentential Form)

For a given CFG $G = (V, \Sigma, S, P)$, a sequence of variables or symbols $\alpha \in (V \cup \Sigma)^*$ is a **sentential form** if and only if $S \Rightarrow^* \alpha$.

Definition (Sentential Form)

For a given CFG $G = (V, \Sigma, S, P)$, a sequence of variables or symbols $\alpha \in (V \cup \Sigma)^*$ is a **sentential form** if and only if $S \Rightarrow^* \alpha$.

- α is a **left-sentential form** if $S \xRightarrow{lm}^* \alpha$.
- α is a **right-sentential form** if $S \xRightarrow{rm}^* \alpha$.

Definition (Sentential Form)

For a given CFG $G = (V, \Sigma, S, P)$, a sequence of variables or symbols $\alpha \in (V \cup \Sigma)^*$ is a **sentential form** if and only if $S \Rightarrow^* \alpha$.

- α is a **left-sentential form** if $S \xRightarrow{\text{lm}}^* \alpha$.
- α is a **right-sentential form** if $S \xRightarrow{\text{rm}}^* \alpha$.

For example, $(A)S$ is a left-sentential form:

$$S \xRightarrow{\text{lm}} B \xRightarrow{\text{lm}} SS \xRightarrow{\text{lm}} AS \xRightarrow{\text{lm}} (S)S \xRightarrow{\text{lm}} (A)S$$

and, $S(S)$ is a right-sentential form:

$$S \xRightarrow{\text{rm}} B \xRightarrow{\text{rm}} SS \xRightarrow{\text{rm}} SA \xRightarrow{\text{rm}} S(S)$$

Definition (Language of CFG)

For a given CFG $G = (V, \Sigma, S, P)$, the **language** of G is defined as:

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

Definition (Context-Free Language)

A language L is **context-free language (CFL)** if and only if there exists a CFG G such that $L(G) = L$.

Definition (Language of CFG)

For a given CFG $G = (V, \Sigma, S, P)$, the **language** of G is defined as:

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

Definition (Context-Free Language)

A language L is **context-free language (CFL)** if and only if there exists a CFG G such that $L(G) = L$.

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

Then, $((())) \in L(G)$ because $S \Rightarrow^* ((()))$.

Example 1

What is the language of the following CFG?

$$G = (\{S, A, B\}, \{(,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

Example 1

What is the language of the following CFG?

$$G = (\{S, A, B\}, \{(\,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

The language of G is:

$$L(G) = \{w \in \{(\,)\}^* \mid w \text{ is balanced}\}$$

Example 1

What is the language of the following CFG?

$$G = (\{S, A, B\}, \{(\, , \,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

The language of G is:

$$L(G) = \{w \in \{(\, , \,)\}^* \mid w \text{ is balanced}\}$$

We can define this CFG in a more compact way using the bar ($|$) notation:

$$\begin{array}{l} S \rightarrow \epsilon \mid A \mid B \\ A \rightarrow (S) \\ B \rightarrow SS \end{array}$$

Example 1

What is the language of the following CFG?

$$G = (\{S, A, B\}, \{(\, , \,)\}, S, P)$$

$$\begin{array}{lll} S \rightarrow \epsilon & S \rightarrow A & S \rightarrow B \\ A \rightarrow (S) & B \rightarrow SS & \end{array}$$

The language of G is:

$$L(G) = \{w \in \{(\, , \,)\}^* \mid w \text{ is balanced}\}$$

We can define this CFG in a more compact way using the bar ($|$) notation:

$$\begin{array}{l} S \rightarrow \epsilon \mid A \mid B \\ A \rightarrow (S) \\ B \rightarrow SS \end{array}$$

In addition, it is equivalent to the following CFG:

$$S \rightarrow \epsilon \mid (S) \mid SS$$

Example 2

Define a CFG whose language is:

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

Example 2

Define a CFG whose language is:

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

The answer is:

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

Example 3

Define a CFG whose language is:

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

Example 3

Define a CFG whose language is:

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

The answer is:

$$S \rightarrow \epsilon \mid aSa \mid bSb$$

- Midterm exam will be given in class.
- **Date:** 14:00-15:15 (1 hour 15 minutes), April 24 (Mon.).
- **Location:** 302, Aegineung (애기능생활관)
- **Coverage:** Lectures 1 – 13
- **Format:** short- or long-answer questions, including proofs
 - Closed book, closed notes
 - Some questions may require you to fill in the blanks in Scala code. But, it is about the concepts we have learned so far, not Scala itself.

1. Context-Free Grammars (CFGs)

Definition

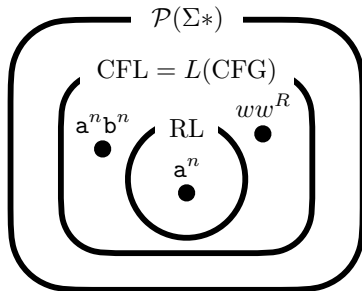
Derivation Relations

Leftmost and Rightmost Derivations

Sentential Forms

Context-Free Languages (CFLs)

Examples



- Examples of Context-Free Grammars

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