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

COSE215: Theory of Computation

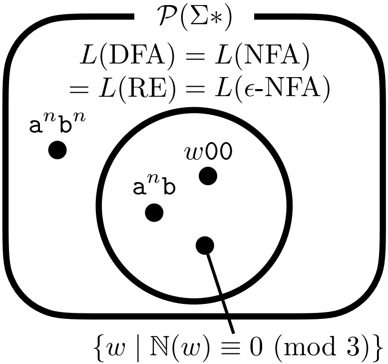
Jihyeok Park

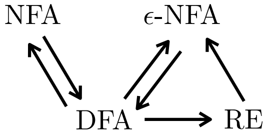
PLRG

2023 Spring

Recall

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





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

Contents

Context-Free Grammars (CFGs)

- Consider the following language:

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

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

$$\begin{array}{rcl} L & \ni & \epsilon,\ (),\ (()),\ ()(),\ (()()),\ (()(())),\ ((())),\ \cdots \\ L & \not\ni & (,\ ),\ )(,\ ((),\ ()),\ (())),\ (()((),\ \cdots \end{array}$$

Context-Free Grammars (CFGs)

- Consider the following language:

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

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

$$\begin{array}{lcl} L & \ni & \epsilon,\ (),\ (()),\ ()(),\ (()()),\ (()(())),\ ((())),\ \cdots \\ L & \not\ni & (,\ ),\ )(,\ ((),\ ()),\ (())),\ ()((),\ \cdots \end{array}$$

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

Context-Free Grammars (CFGs)

- Consider the following language:

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

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

$$\begin{array}{lcl} L & \ni & \epsilon,\ (),\ (()),\ ()(),\ (()()),\ (()(())),\ ((())),\ \cdots \\ L & \not\ni & (,\ ),\ )(,\ ((),\ ()),\ (())),\ ()((),\ \cdots \end{array}$$

- 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:

$$\begin{array}{lcl} L & \ni & \epsilon, \; (), \; (()), \; ()(), \; (()()), \; (()(())), \; ((())), \; \cdots \\ L & \not\ni & (, \;), \;)(, \; ((), \; ()), \; (())), \; ()((), \; \cdots \end{array}$$

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

- Then, how can we describe this language?
  Context-Free Grammars (CFGs)

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**.

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**.

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

where P is defined as:

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

Context-Free Grammars (CFGs)

```scala
// 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", 'b'),
  )
)
```

Derivation Relations

Definition (Derivation Relation ($\Rightarrow$))

Consider a CFG $G = (V, \Sigma, S, P)$. If a production rule $A \to \gamma \in P$ 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)** $\forall \alpha, \beta, \gamma \in (V \cup \Sigma)^*.$

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

Derivation Relations

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

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

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

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

Thus,

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

Leftmost and Rightmost Derivations

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

Leftmost and Rightmost Derivations

- **Leftmost Derivation** ($\overset{\text{lm}}{\Longrightarrow}$): always derive the *leftmost* variable.
- **Rightmost Derivation** ($\overset{\text{rm}}{\Longrightarrow}$): 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 and Rightmost Derivations

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

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

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

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

$$\begin{array}{llllllllll} S & \overset{\text{lm}}{\Longrightarrow} & B & \overset{\text{lm}}{\Longrightarrow} & SS & \overset{\text{lm}}{\Longrightarrow} & AS & \overset{\text{lm}}{\Longrightarrow} & \mathtt{(}S\mathtt{)}S & \overset{\text{lm}}{\Longrightarrow} & \mathtt{(}A\mathtt{)}S \\ & \overset{\text{lm}}{\Longrightarrow} & \mathtt{((}S\mathtt{))}S & \overset{\text{lm}}{\Longrightarrow} & \mathtt{(())}S & \overset{\text{lm}}{\Longrightarrow} & \mathtt{(())}A & \overset{\text{lm}}{\Longrightarrow} & \mathtt{(())(}S\mathtt{)} & \overset{\text{lm}}{\Longrightarrow} & \mathtt{(())()} \end{array}$$

Leftmost and Rightmost Derivations

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

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

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

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

$$\begin{array}{lllllllllll} S & \overset{\text{lm}}{\Longrightarrow} & B & \overset{\text{lm}}{\Longrightarrow} & SS & \overset{\text{lm}}{\Longrightarrow} & AS & \overset{\text{lm}}{\Longrightarrow} & (S)S & \overset{\text{lm}}{\Longrightarrow} & (A)S \\ & \overset{\text{lm}}{\Longrightarrow} & ((S))S & \overset{\text{lm}}{\Longrightarrow} & (())S & \overset{\text{lm}}{\Longrightarrow} & (())A & \overset{\text{lm}}{\Longrightarrow} & (())(S) & \overset{\text{lm}}{\Longrightarrow} & (())() \end{array}$$

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

$$\begin{array}{lllllllllll} S & \overset{\text{rm}}{\Longrightarrow} & B & \overset{\text{rm}}{\Longrightarrow} & SS & \overset{\text{rm}}{\Longrightarrow} & SA & \overset{\text{rm}}{\Longrightarrow} & S(S) & \overset{\text{rm}}{\Longrightarrow} & S() \\ & \overset{\text{rm}}{\Longrightarrow} & A() & \overset{\text{rm}}{\Longrightarrow} & (S)() & \overset{\text{rm}}{\Longrightarrow} & (A)() & \overset{\text{rm}}{\Longrightarrow} & ((S))() & \overset{\text{rm}}{\Longrightarrow} & (())() \end{array}$$

Sentential Forms

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 $S \Rightarrow^* \alpha$.

Sentential Forms

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 $S \xRightarrow{\text{rm}}^* \alpha$.

Sentential Forms

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

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

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

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

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

Context-Free Languages (CFLs)

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** if there exists a CFG G such that $LG = L$.

Context-Free Languages (CFLs)

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** if there exists a CFG G such that $LG = L$.

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

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

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

Example 1

What is the language of the following CFG?

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

$$S \rightarrow \epsilon \qquad S \rightarrow A \qquad S \rightarrow B$$
$$A \rightarrow (S) \qquad B \rightarrow SS$$

Example 1

What is the language of the following CFG?

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

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

The language of G is:

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

Example 1

What is the language of the following CFG?

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

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

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:

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

Example 1

What is the language of the following CFG?

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

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

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:

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

In addition, it is equivalent to the following CFG:

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

Example 2

Define a CFG whose language is:

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

Example 2

Define a CFG whose language is:

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

The answer is:

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

Example 3

Define a CFG whose language is:

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

Example 3

Define a CFG whose language is:

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

The answer is:

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

Summary

1. Context-Free Grammars (CFGs)
    - Definition
    - Derivation Relations
    - Leftmost and Rightmost Derivations
    - Sentential Forms
    - Context-Free Languages (CFLs)
    - Examples

Next Lecture

- Examples of Context-Free Grammars

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