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

COSE215: Theory of Computation

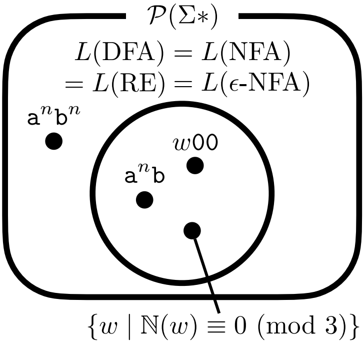
Jihyeok Park

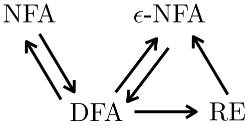
PLRG

2025 Spring

Recall

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





- The minimized DFA is **unique** up to isomorphism by the **Myhill-Nerode Theorem**.[1]

[1] https://en.wikipedia.org/wiki/Myhill-Nerode_theorem

Recall

	Automata	Grammars	Languages
(Part 3) Turing Machines	(Lecture 23) ETM ⇄ (Lecture 21/22) TM ⇄	(Lecture 24) LC	(Lecture 21) REL $\cup$ DL $\supset$ (Lecture 25); (Lecture 26) **NP** $\overset{?}{=}$ **P** $\supset$
(Part 2) Pushdown Automata	(Lecture 14/15) PDA_{FS} ⇄ PDA_{ES}; (Lecture 16) ⇄; $\text{PDA}_{\text{FS}} \cup \text{DPDA}_{\text{FS}}$ $\supset$ DPDA_{ES} $\cup$ (Lecture 17) $\not\cup$	(Lecture 11/12) CFG; **Chomsky Normal Form** (Lecture 18)	(Lecture 11) CFL; (Lecture 13) **Parse Trees & Ambiguity**; **Closure Properties** (Lecture 19); **Pumping Lemma** (Lecture 20)
(Part 1) Finite Automata	(Lecture 4) NFA ⇄ (Lecture 3) DFA ⇄ (Lecture 5) ϵ-NFA; (Lecture 7); **Equivalence & Minimization** (Lecture 10)	(Lecture 6) RE	(Lecture 3) RL; **Closure Properties** (Lecture 8); **Pumping Lemma** (Lecture 9)
(Part 0) Basic Concepts	(Lecture 1) **Mathematical Preliminaries**	(Lecture 2) **Scala**	

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:

$$L \ni \epsilon,\ (),\ (()),\ ()(),\ (()()),\ (()(())),\ ((())),\ \cdots$$
$$L \not\ni (,\ ),\ )(,\ ((),\ ()),\ (())),\ (()((),\ \cdots$$

- Is this language regular? **No**, we can prove that this language is **not regular** using the **Pumping Lemma** (Do it yourself!).

- Is there a way to describe this language?

- Yes, let's learn **Context-Free Grammars (CFGs)**!

Contents

Basic Idea

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

How to **inductively generate** (or produce) words in the language L?

- **Base Case:** $\epsilon \in L$
- **Inductive Case:** There are two inductive rules:
  - If $w \in L$, then $\texttt{(}w\texttt{)} \in L$
  - If $w_1, w_2 \in L$, then $w_1 w_2 \in L$

ϵ `(()())` `(())()` `()(())()` $\cdots$

Context-Free Grammars (CFGs) provide a way to describe languages with such **inductive rules** to generate words in the language.

Context-Free Grammars (CFGs)

PLRG

Definition (Context-Free Grammar (CFG))

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

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

where

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

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

where R 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)

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

where R is defined as:

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

We often call the sequence of variables and symbols in the production rule a **right-hand side** (RHS) of the production rule.

$$S \to \epsilon \mid A \mid B \qquad A \to (S) \qquad B \to SS$$

We can simplify the notation using the bar (|) notation by **combining** multiple production rules for the **same variable**.

Context-Free Grammars (CFGs)

```scala
// The definition of variables (nonterminals)
type Nt = String
// The type definitions of symbols (terminals)
type Symbol = Char
// The definition of right-hand side of a production rule
case class Rhs(seq: List[Nt | Symbol])
// The definition of context-free grammars
case class CFG(
  nts: Set[Nt],
  symbols: Set[Symbol],
  start: Nt,
  rules: Map[Nt, List[Rhs]],
)
```

```scala
// An example of CFG
val cfg: CFG = CFG(
  nts = Set("S", "A", "B"), symbols = Set('(', ')'), start = "S",
  rules = Map(
    "S" -> List(Rhs(List()), Rhs(List("A")), Rhs(List("B"))),
    "A" -> List(Rhs(List('(', "S", ')'))),
    "B" -> List(Rhs(List("S", "S")))
  ),
)
```

Derivation Relations

Definition (Derivation Relation ($\Rightarrow$))

Consider a CFG $G = (V, \Sigma, S, R)$. If a production rule $A \rightarrow \gamma \in R$ 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

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

Derivation Relations

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

$$S \to \epsilon \mid A \mid B \qquad A \to (S) \qquad B \to SS$$

A derivation for `(())()`:

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

Thus, we can **derive** (or generate/produce) the word `(())()` from S:

$$S \Rightarrow^* (())()$$

Leftmost and Rightmost Derivations

- **Leftmost Derivation** ($\Rightarrow_L$): always derive the *leftmost* variable.
- **Rightmost Derivation** ($\Rightarrow_R$): always derive the *rightmost* variable.

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

$$S \to \epsilon \mid A \mid B \qquad A \to (S) \qquad B \to SS$$

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

$$\begin{array}{llllllll} S & \Rightarrow_L & B & \Rightarrow_L & SS & \Rightarrow_L & AS & \\ & \Rightarrow_L & (S)S & \Rightarrow_L & (A)S & \Rightarrow_L & ((S))S & \\ & \Rightarrow_L & (())S & \Rightarrow_L & (())A & \Rightarrow_L & (())(S) & \Rightarrow_L \quad (())() \end{array}$$

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

$$\begin{array}{llllllll} S & \Rightarrow_R & B & \Rightarrow_R & SS & \Rightarrow_R & SA & \\ & \Rightarrow_R & S(S) & \Rightarrow_R & S() & \Rightarrow_R & A() & \\ & \Rightarrow_R & (S)() & \Rightarrow_R & (A)() & \Rightarrow_R & ((S))() & \Rightarrow_R \quad (())() \end{array}$$

Sentential Forms

Definition (Sentential Form)

For a given CFG $G = (V, \Sigma, S, R)$, 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 $S \Rightarrow_L^* \alpha$.
- α is a **right-sentential form** if $S \Rightarrow_R^* \alpha$.

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

$$S \Rightarrow_L B \Rightarrow_L SS \Rightarrow_L AS \Rightarrow_L (S)S \Rightarrow_L (A)S$$

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

$$S \Rightarrow_R B \Rightarrow_R SS \Rightarrow_R SA \Rightarrow_R S(S)$$

Context-Free Languages (CFLs)

Definition (Language of CFG)

For a given CFG $G = (V, \Sigma, S, R)$, 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 $L(G) = L$.

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

$$S \to \epsilon \mid A \mid B \qquad A \to (S) \qquad B \to SS$$

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

Example 1

What is the language of the following CFG?

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

$$S \to \epsilon \mid A \mid B \qquad A \to (S) \qquad B \to SS$$

The language of G is:

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

In addition, it is equivalent to the following CFG:

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

Example 2

Define a CFG whose language is:

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

The answer is:

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

Example 3

Define a CFG whose language is:

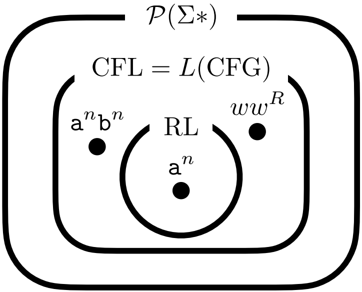
$$L = \{ww^R \mid w \in \{\texttt{a}, \texttt{b}\}^*\}\}$$

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