Lecture 12 – Examples of Context-Free Grammars

COSE215: Theory of Computation

Jihyeok Park

PLRG

2023 Spring

Recall

- A **context-free grammar (CFG)**:

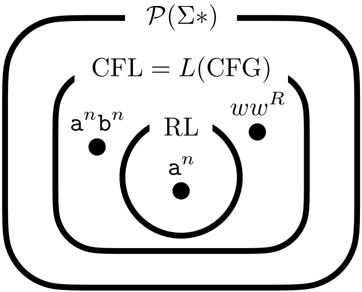
$$G = (V, \Sigma, S, P)$$

- The **language** of a CFG G:

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

- A language L is a **context-free language (CFL)**:

$$\exists \text{ CFG } G.\ L(G) = L$$



Contents

Example 1: Regular Languages

Theorem (RLs are CFLs)

If a language L is a regular language (RL), then L is a CFL.

Proof) For a given RE R, construct a CFG G such that $L(G) = L(R)$.

RE R	CFG G
$\varnothing$	$S \to S$
ϵ	$S \to \epsilon$
$a \in \Sigma$	$S \to a$
$R_1 \mid R_2$	$S \to S_1 \mid S_2$
$R_1 \cdot R_2$	$S \to S_1 S_2$
R_1^*	$S \to \epsilon \mid S_1 S$
(R_1)	$S \to S_1$

where S_1 and S_2 are start variables of CFGs G_1 and G_2 such that $L(G_1) = L(R_1)$ and $L(G_2) = L(R_2)$.

Example 1: Regular Languages

For a given RE R, construct a CFG G such that $L(G) = L(R)$.

- $R = \epsilon \,|\, \mathtt{ab} \,|\, \mathtt{ba}$

$$
\begin{array}{llll}
S \to FD & A \to \mathtt{a} & C \to AB & E \to \epsilon \\
 & B \to \mathtt{b} & D \to BA & F \to EC
\end{array}
$$

Example 1: Regular Languages

For a given RE R, construct a CFG G such that $L(G) = L(R)$.

- $R = \epsilon | \texttt{ab} | \texttt{ba}$

$$
\begin{array}{llll}
S \to FD & A \to \texttt{a} & C \to AB & E \to \epsilon \\
 & B \to \texttt{b} & D \to BA & F \to EC
\end{array}
$$

Its simplified version:

$$S \to \epsilon \mid \texttt{ab} \mid \texttt{ba}$$

Example 1: Regular Languages

For a given RE R, construct a CFG G such that $L(G) = L(R)$.

- $R = \epsilon|\texttt{ab}|\texttt{ba}$

$$
\begin{array}{llll}
S \to FD & A \to \texttt{a} & C \to AB & E \to \epsilon \\
 & B \to \texttt{b} & D \to BA & F \to EC
\end{array}
$$

  Its simplified version:

$$S \to \epsilon \mid \texttt{ab} \mid \texttt{ba}$$

- $R = (\epsilon|\texttt{a})^*$

$$S \to \epsilon \mid AS \qquad A \to \epsilon \mid \texttt{a}$$

Example 1: Regular Languages

For a given RE R, construct a CFG G such that $L(G) = L(R)$.

- $R = \epsilon \,|\, \texttt{ab} \,|\, \texttt{ba}$

$$S \to FD \qquad A \to \texttt{a} \qquad C \to AB \qquad E \to \epsilon$$
$$B \to \texttt{b} \qquad D \to BA \qquad F \to EC$$

  Its simplified version:

$$S \to \epsilon \mid \texttt{ab} \mid \texttt{ba}$$

- $R = (\epsilon \,|\, \texttt{a})^*$

$$S \to \epsilon \mid AS \qquad A \to \epsilon \mid \texttt{a}$$

- $R = (\texttt{0} \,|\, \texttt{1}(\texttt{0}\texttt{1}^*\texttt{0})^*\texttt{1})^*$

$$S \to \epsilon \mid AE \qquad A \to \texttt{0} \mid \texttt{1}B\texttt{1} \qquad C \to \texttt{0}D\texttt{0}$$
$$B \to \epsilon \mid CB \qquad D \to \epsilon \mid \texttt{1}D$$

Example 2: $\mathtt{b}^n\mathtt{a}^m\mathtt{b}^{2n}$

Construct a CFG for the language:

$$L = \{\mathtt{b}^n\mathtt{a}^m\mathtt{b}^{2n} \mid n, m \geq 0\}$$

Example 2: $\texttt{b}^n\texttt{a}^m\texttt{b}^{2n}$

Construct a CFG for the language:

$$L = \{\texttt{b}^n\texttt{a}^m\texttt{b}^{2n} \mid n, m \geq 0\}$$

$$\begin{array}{l} S \to A \mid \texttt{b}S\texttt{bb} \\ A \to \epsilon \mid \texttt{a}A \end{array}$$

Example 2: $\mathtt{b}^n\mathtt{a}^m\mathtt{b}^{2n}$

Construct a CFG for the language:

$$L = \{\mathtt{b}^n\mathtt{a}^m\mathtt{b}^{2n} \mid n, m \geq 0\}$$

$$\begin{array}{l} S \to A \mid \mathtt{b}S\mathtt{bb} \\ A \to \epsilon \mid \mathtt{a}A \end{array}$$

A derivation for `bbaaabbbb`:

$$\begin{array}{lllllll} S & \Rightarrow & \mathtt{b}S\mathtt{bb} & \Rightarrow & \mathtt{bb}S\mathtt{bbbb} & \Rightarrow & \mathtt{bb}A\mathtt{bbbb} \\ & \Rightarrow & \mathtt{bba}A\mathtt{bbbb} & \Rightarrow & \mathtt{bbaa}A\mathtt{bbbb} & \Rightarrow & \mathtt{bbaaa}A\mathtt{bbbb} \\ & \Rightarrow & \mathtt{bbaaabbbb} & & & & \end{array}$$

Example 3: Well-Formed Brackets

Construct a CFG for the language:

$$L = \{w \in \{(,),\{,\},[,]\}^* \mid w \text{ is well-formed}\}$$

Example 3: Well-Formed Brackets

Construct a CFG for the language:

$$L = \{w \in \{(,),\{,\},[,]\}^* \mid w \text{ is well-formed}\}$$

$$S \to \epsilon \mid (S) \mid \{S\} \mid [S] \mid SS$$

Example 3: Well-Formed Brackets

Construct a CFG for the language:

$$L = \{w \in \{(,),\{,\},[,]\}^* \mid w \text{ is well-formed}\}$$

$$S \rightarrow \epsilon \mid (S) \mid \{S\} \mid [S] \mid SS$$

A derivation for $(\{\})\{\}[()[]]$:

$$\begin{array}{llllll} S & \Rightarrow SS & \Rightarrow & SSS & \Rightarrow & (S)SS \\ & \Rightarrow (\{S\})SS & \Rightarrow & (\{\})SS & \Rightarrow & (\{\})\{S\}S \\ & \Rightarrow (\{\})\{\}S & \Rightarrow & (\{\})\{\}[S] & \Rightarrow & (\{\})\{\}[SS] \\ & \Rightarrow (\{\})\{\}[(S)S] & \Rightarrow & (\{\})\{\}[()S] & \Rightarrow & (\{\})\{\}[()[S]] \\ & \Rightarrow (\{\})\{\}[()[]] & & & & \end{array}$$

Example 4: Equal Number of a and b

Construct a CFG for the language:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) = N_{\mathtt{b}}(w)\}$$

where $N_{\mathtt{a}}(w)$ and $N_{\mathtt{b}}(w)$ are the number of a's and b's in w, respectively.

Example 4: Equal Number of a and b

Construct a CFG for the language:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) = N_{\mathtt{b}}(w)\}$$

where $N_{\mathtt{a}}(w)$ and $N_{\mathtt{b}}(w)$ are the number of a's and b's in w, respectively.

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

Example 4: Equal Number of a and b

Construct a CFG for the language:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) = N_{\mathtt{b}}(w)\}$$

where $N_{\mathtt{a}}(w)$ and $N_{\mathtt{b}}(w)$ are the number of a's and b's in w, respectively.

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

The left-most derivation for abbaaabb:

$$\begin{array}{llllll} S & \xRightarrow{\text{lm}} \mathtt{a}S\mathtt{b} & \xRightarrow{\text{lm}} & \mathtt{a}SS\mathtt{b} & \xRightarrow{\text{lm}} & \mathtt{ab}S\mathtt{a}S\mathtt{b} \\ & \xRightarrow{\text{lm}} \mathtt{abb}S\mathtt{aa}S\mathtt{b} & \xRightarrow{\text{lm}} & \mathtt{abbaa}S\mathtt{b} & \xRightarrow{\text{lm}} & \mathtt{abbaaa}S\mathtt{bb} \\ & \xRightarrow{\text{lm}} \mathtt{abbaaabb} & & & & \end{array}$$

Example 5: Unequal Number of a and b

Construct a CFG for the language:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) \neq N_{\mathtt{b}}(w)\}$$

where $N_{\mathtt{a}}(w)$ and $N_{\mathtt{b}}(w)$ are the number of a's and b's in w, respectively.

Example 5: Unequal Number of a and b

Construct a CFG for the language:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) \neq N_{\mathtt{b}}(w)\}$$

where $N_{\mathtt{a}}(w)$ and $N_{\mathtt{b}}(w)$ are the number of a's and b's in w, respectively.

$$\begin{aligned} S &\rightarrow P \mid N \\ P &\rightarrow ZP \mid \mathtt{a}P \mid \mathtt{a}Z \\ N &\rightarrow ZN \mid \mathtt{b}N \mid \mathtt{b}Z \\ Z &\rightarrow \epsilon \mid \mathtt{a}Z\mathtt{b} \mid \mathtt{b}Z\mathtt{a} \mid ZZ \end{aligned}$$

Example 5: Unequal Number of a and b

Construct a CFG for the language:

$$L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) \neq N_{\mathtt{b}}(w)\}$$

where $N_{\mathtt{a}}(w)$ and $N_{\mathtt{b}}(w)$ are the number of a's and b's in w, respectively.

$$\begin{aligned}
S &\rightarrow P \mid N \\
P &\rightarrow ZP \mid \mathtt{a}P \mid \mathtt{a}Z \\
N &\rightarrow ZN \mid \mathtt{b}N \mid \mathtt{b}Z \\
Z &\rightarrow \epsilon \mid \mathtt{a}Z\mathtt{b} \mid \mathtt{b}Z\mathtt{a} \mid ZZ
\end{aligned}$$

The right-most derivation for aabbbaaab:

$$\begin{array}{llllll}
S \overset{\text{rm}}{\Longrightarrow} & P & \overset{\text{rm}}{\Longrightarrow} & ZP & \overset{\text{rm}}{\Longrightarrow} & Z\mathtt{a}Z \\
\overset{\text{rm}}{\Longrightarrow} & Z\mathtt{aa}Z\mathtt{b} & \overset{\text{rm}}{\Longrightarrow} & Z\mathtt{aab} & \overset{\text{rm}}{\Longrightarrow} & ZZ\mathtt{aab} \\
\overset{\text{rm}}{\Longrightarrow} & Z\mathtt{b}Z\mathtt{aaab} & \overset{\text{rm}}{\Longrightarrow} & Z\mathtt{baaab} & \overset{\text{rm}}{\Longrightarrow} & \mathtt{a}Z\mathtt{bbaaab} \\
\overset{\text{rm}}{\Longrightarrow} & \mathtt{aa}Z\mathtt{bbbaaab} & \overset{\text{rm}}{\Longrightarrow} & \mathtt{aabbbaaab} & &
\end{array}$$

Example 6: Arithmetic Expressions

An **arithmetic expression** is defined with the following CFG:

$$
\begin{aligned}
S &\rightarrow N \mid X \mid S\texttt{+}S \mid S\texttt{*}S \mid \texttt{(}S\texttt{)} \\
N &\rightarrow \texttt{1}D \mid \cdots \mid \texttt{9}D \\
D &\rightarrow \epsilon \mid \texttt{0}D \mid \cdots \mid \texttt{9}D \\
X &\rightarrow \texttt{a} \mid \cdots \mid \texttt{z}
\end{aligned}
$$

Example 6: Arithmetic Expressions

An **arithmetic expression** is defined with the following CFG:

$$
\begin{array}{lcl}
S & \rightarrow & N \mid X \mid S\texttt{+}S \mid S\texttt{*}S \mid \texttt{(}S\texttt{)} \\
N & \rightarrow & \texttt{1}D \mid \cdots \mid \texttt{9}D \\
D & \rightarrow & \epsilon \mid \texttt{0}D \mid \cdots \mid \texttt{9}D \\
X & \rightarrow & \texttt{a} \mid \cdots \mid \texttt{z}
\end{array}
$$

The left-most derivation for `13*(2+x)`:

$$
\begin{array}{llllll}
S & \overset{\text{lm}}{\Longrightarrow} & S\texttt{*}S & \overset{\text{lm}}{\Longrightarrow} N\texttt{*}S & \overset{\text{lm}}{\Longrightarrow} & \texttt{1}D\texttt{*}S \\
 & \overset{\text{lm}}{\Longrightarrow} & \texttt{13}D\texttt{*}S & \overset{\text{lm}}{\Longrightarrow} \texttt{13*}S & \overset{\text{lm}}{\Longrightarrow} & \texttt{13*(}S\texttt{)} \\
 & \overset{\text{lm}}{\Longrightarrow} & \texttt{13*(}S\texttt{+}S\texttt{)} & \overset{\text{lm}}{\Longrightarrow} \texttt{13*(}N\texttt{+}S\texttt{)} & \overset{\text{lm}}{\Longrightarrow} & \texttt{13*(2}D\texttt{+}S\texttt{)} \\
 & \overset{\text{lm}}{\Longrightarrow} & \texttt{13*(2+}S\texttt{)} & \overset{\text{lm}}{\Longrightarrow} \texttt{13*(2+}X\texttt{)} & \overset{\text{lm}}{\Longrightarrow} & \texttt{13*(2+x)}
\end{array}
$$

Example 7: Regular Expressions

Is the following language regular? or context-free?

$$L = \{w \in \{\varnothing, \epsilon, \mathtt{a}, \mathtt{b}, \mathsf{|}, \cdot, {}^*, \mathtt{(}, \mathtt{)}\}^* \mid w \text{ is a regular expression over } \{\mathtt{a}, \mathtt{b}\}\}$$

Example 7: Regular Expressions

Is the following language regular? or context-free?

$$L = \{w \in \{\varnothing, \epsilon, \mathtt{a}, \mathtt{b}, \mathtt{|}, \cdot, {}^{*}, \mathtt{(}, \mathtt{)}\}^* \mid w \text{ is a regular expression over } \{\mathtt{a}, \mathtt{b}\}\}$$

We can prove that L is not regular using the pumping lemma.
(Hint: consider a word $\mathtt{(}^n\epsilon\mathtt{)}^n$ for a given $n > 0$)

Is the following language regular? or context-free?

$$L = \{w \in \{\varnothing, \epsilon, \mathtt{a}, \mathtt{b}, \mathtt{|}, \cdot, {}^*, \mathtt{(}, \mathtt{)}\}^* \mid w \text{ is a regular expression over } \{\mathtt{a}, \mathtt{b}\}\}$$

We can prove that L is not regular using the pumping lemma.
(Hint: consider a word $\mathtt{(}^n\epsilon\mathtt{)}^n$ for a given $n > 0$)
The language L is context-free:

$$S \to \varnothing \mid \epsilon \mid \mathtt{a} \mid \mathtt{b} \mid S \,\mathtt{|}\, S \mid S \cdot S \mid S^* \mid \mathtt{(}S\mathtt{)}$$

Example 7: Regular Expressions

Is the following language regular? or context-free?

$$L = \{w \in \{\varnothing, \epsilon, \texttt{a}, \texttt{b}, \texttt{|}, \cdot, {}^*, \texttt{(}, \texttt{)}\}^* \mid w \text{ is a regular expression over } \{\texttt{a}, \texttt{b}\}\}$$

We can prove that L is not regular using the pumping lemma.
(Hint: consider a word $\texttt{(}^n\epsilon\texttt{)}^n$ for a given $n > 0$)
The language L is context-free:

$$S \rightarrow \varnothing \mid \epsilon \mid \texttt{a} \mid \texttt{b} \mid S \,\texttt{|}\, S \mid S \cdot S \mid S^* \mid \texttt{(}S\texttt{)}$$

The right-most derivation for $\texttt{(}\epsilon\,\texttt{|}\,\texttt{a} \cdot \texttt{b}\texttt{)}^*$:

$$\begin{array}{llllll} S & \stackrel{\text{rm}}{\Longrightarrow} & S^* & \stackrel{\text{rm}}{\Longrightarrow} & \texttt{(}S\texttt{)}^* & \stackrel{\text{rm}}{\Longrightarrow} & \texttt{(}S\,\texttt{|}\,S\texttt{)}^* \\ & \stackrel{\text{rm}}{\Longrightarrow} & \texttt{(}S\,\texttt{|}\,S \cdot S\texttt{)}^* & \stackrel{\text{rm}}{\Longrightarrow} & \texttt{(}S\,\texttt{|}\,S \cdot \texttt{b}\texttt{)}^* & \stackrel{\text{rm}}{\Longrightarrow} & \texttt{(}S\,\texttt{|}\,\texttt{a} \cdot \texttt{b}\texttt{)}^* \\ & \stackrel{\text{rm}}{\Longrightarrow} & \texttt{(}\epsilon\,\texttt{|}\,\texttt{a} \cdot \texttt{b}\texttt{)}^* & & & & \end{array}$$

Example 8: Simplified Scala

We can define a CFG for a simplified version of Scala[1]:

$$
\begin{array}{lrl}
\text{(Scala Program)} & S \rightarrow & E \mid E \texttt{ ; } S \\
\text{(Expressions)} & E \rightarrow & N \mid X \mid E \texttt{ + } E \mid E \texttt{ - } E \mid E \texttt{ * } E \mid E \texttt{ / } E \\
& \mid & \texttt{val } X \texttt{ : } T \texttt{ = } E \\
& \mid & \texttt{def } X \texttt{ (} P \texttt{): } T \texttt{ = } E \\
& \mid & E \texttt{ (} A \texttt{)} \\
& \mid & \texttt{if (} E \texttt{) } E \texttt{ else } E \\
& \mid & \texttt{trait } T \texttt{ (} P \texttt{)} \\
& \mid & \texttt{case class } T \texttt{ (} P \texttt{)} \\
& \mid & E \texttt{ match \{ } C \texttt{ \}} \\
\text{(Numbers)} & N \rightarrow & \texttt{1}D \mid \cdots \mid \texttt{9}D \\
& D \rightarrow & \epsilon \mid \texttt{0}D \mid \cdots \mid \texttt{9}D \\
\text{(Variables)} & X \rightarrow & A \mid AX \\
& A \rightarrow & \texttt{a} \mid \cdots \mid \texttt{z} \mid \texttt{A} \mid \cdots \mid \texttt{Z} \\
\text{(Types)} & T \rightarrow & X \mid T \texttt{ [} T \texttt{]} \mid T \texttt{ => } T \\
\text{(Parameters)} & P \rightarrow & \epsilon \mid X \texttt{ : } T \mid P \texttt{ , } X \texttt{ : } T \\
\text{(Arguments)} & A \rightarrow & \epsilon \mid E \mid A \texttt{ , } E \\
\text{(Cases)} & C \rightarrow & \texttt{case } E \texttt{ => } E \mid C \texttt{ ; case } E \texttt{ => } E
\end{array}
$$

[1] https://docs.scala-lang.org/scala3/reference/syntax.html

Example 8: Simplified Scala

```
def sum(n: Int): Int = n match { case 0 => 0; case n => n + sum(n - 1) }
```

The left-most derivation for this program:

$S \overset{\text{lm}}{\Longrightarrow}$ `def` X (P) `:` T = E $\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(` P `):` T = E

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(` X`:` T `):` T = E $\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int =` E

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int =` E match { C }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match {` C }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case` E => E ; C }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case 0 => 0;` C }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case 0 => 0; case` E => E }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case 0 => 0; case n =>` E }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case 0 => 0; case n =>` E + E }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case 0 => 0; case n => n +` E }

$\overset{\text{lm}}{\Longrightarrow}{}^{*}$ `def sum(n: Int): Int = n match { case 0 => 0; case n => n + sum(n - 1) }`

Summary

1. Examples of Context-Free Grammars
    - Example 1: Regular Languages
    - Example 2: $\mathtt{b}^n\mathtt{a}^m\mathtt{b}^{2n}$
    - Example 3: Well-Formed Brackets
    - Example 4: Equal Number of `a` and `b`
    - Example 5: Unequal Number of `a` and `b`
    - Example 6: Arithmetic Expressions
    - Example 7: Regular Expressions
    - Example 8: Simplified Scala

Next Lecture

- Parse Trees and Ambiguity

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