Introduction to Context-Free Grammars

Deepak D'Souza

Department of Computer Science and Automation
Indian Institute of Science, Bangalore.

29 September 2016

Outline

Why study Context-Free Grammars?

- Arise naturally in syntax of programming languages, parsing, compiling.
- Characterize languages accepted by Pushdown automata.
- Pushdown automata are important class of system models:
    - They can model programs with procedure calls
    - Can model other infinite-state systems.
- Easier to prove properties of Pushdown languages using CFG's:
    - Pumping lemma
    - Ultimate periodicity
    - PDA = PDA without ϵ-transitions.
- Parsing algo leads to solution to "CFL reachability" problem: Given a finite A-labelled graph, a CFG G, are two given vertices u and v connected by a path whose label is in $L(G)$.

Context-Free Grammars: Example 1

CFG G_1

$$
\begin{aligned}
S &\rightarrow aX \\
X &\rightarrow aX \\
X &\rightarrow bX \\
X &\rightarrow b
\end{aligned}
$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.

Example derivation:

$$S$$

Context-Free Grammars: Example 1

CFG G_1

$$
\begin{array}{lcl}
S & \rightarrow & aX \\
X & \rightarrow & aX \\
X & \rightarrow & bX \\
X & \rightarrow & b
\end{array}
$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.

Example derivation:

$$S \quad \Rightarrow aX$$

Context-Free Grammars: Example 1

CFG G_1

$$
\begin{aligned}
S &\rightarrow aX \\
X &\rightarrow aX \\
X &\rightarrow bX \\
X &\rightarrow b
\end{aligned}
$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.

Example derivation:

$$S \Rightarrow aX \Rightarrow abX$$

Context-Free Grammars: Example 1

CFG G_1

$$\begin{array}{rcl} S & \rightarrow & aX \\ X & \rightarrow & aX \\ X & \rightarrow & bX \\ X & \rightarrow & b \end{array}$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.

Example derivation:

$$S \quad \Rightarrow aX \quad \Rightarrow abX \quad \Rightarrow abb.$$

Context-Free Grammars: Example 1

CFG G_1

$$
\begin{array}{rcl}
S & \rightarrow & aX \\
X & \rightarrow & aX \\
X & \rightarrow & bX \\
X & \rightarrow & b
\end{array}
$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.
Example derivation:

$$S \quad \Rightarrow aX \quad \Rightarrow abX \quad \Rightarrow abb.$$

Language defined by G, written $L(G)$, is the set of all terminal strings that can be generated by G.

Context-Free Grammars: Example 1

CFG G_1

$$
\begin{aligned}
S &\rightarrow aX \\
X &\rightarrow aX \\
X &\rightarrow bX \\
X &\rightarrow b
\end{aligned}
$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.
Example derivation:

$$S \quad \Rightarrow aX \quad \Rightarrow abX \quad \Rightarrow abb.$$

Language defined by G, written $L(G)$, is the set of all terminal strings that can be generated by G.
What is the language defined by G_1 above?

Context-Free Grammars: Example 1

CFG G_1

$$\begin{aligned} S &\rightarrow aX \\ X &\rightarrow aX \\ X &\rightarrow bX \\ X &\rightarrow b \end{aligned}$$

Derivation of a string: Begin with S and keep rewriting the current string by replacing a non-terminal by its RHS in a production of the grammar.

Example derivation:

$$S \quad \Rightarrow aX \quad \Rightarrow abX \quad \Rightarrow abb.$$

Language defined by G, written $L(G)$, is the set of all terminal strings that can be generated by G.

What is the language defined by G_1 above? $a(a+b)^*b$.

Context-Free Grammars: Example 2

CFG G_2

$$
\begin{aligned}
S &\rightarrow aSb \\
S &\rightarrow \epsilon .
\end{aligned}
$$

Example derivation:

$$S$$

Context-Free Grammars: Example 2

CFG G_2

$$\begin{aligned} S &\rightarrow aSb \\ S &\rightarrow \epsilon. \end{aligned}$$

Example derivation:

$$S \Rightarrow aSb$$

Context-Free Grammars: Example 2

CFG G_2

$$\begin{aligned} S &\rightarrow aSb \\ S &\rightarrow \epsilon. \end{aligned}$$

Example derivation:

$$S \Rightarrow aSb \Rightarrow aaSbb$$

Context-Free Grammars: Example 2

CFG G_2

$$\begin{aligned} S &\rightarrow aSb \\ S &\rightarrow \epsilon. \end{aligned}$$

Example derivation:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb.$$

Context-Free Grammars: Example 2

CFG G_2

$$\begin{array}{lcl} S & \rightarrow & aSb \\ S & \rightarrow & \epsilon. \end{array}$$

Example derivation:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb.$$

What is the language defined by G_2 above?

Context-Free Grammars: Example 2

CFG G_2

$$\begin{aligned} S &\rightarrow aSb \\ S &\rightarrow \epsilon. \end{aligned}$$

Example derivation:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb.$$

What is the language defined by G_2 above? $\{a^n b^n \mid n \geq 0\}$.

Context-Free Grammars: Example 3

CFG G_3

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon.$$

Example derivation:

$$S$$

Context-Free Grammars: Example 3

CFG G_3

$$S \quad\rightarrow\quad aSa \mid bSb \mid a \mid b \mid \epsilon.$$

Example derivation:

$$S \quad\Rightarrow aSa$$

Context-Free Grammars: Example 3

CFG G_3

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon.$$

Example derivation:

$$S \Rightarrow aSa \Rightarrow abSba$$

Context-Free Grammars: Example 3

CFG G_3

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon.$$

Example derivation:

$$S \Rightarrow aSa \Rightarrow abSba \Rightarrow abbSbba \Rightarrow abbbba.$$

Context-Free Grammars: Example 3

CFG G_3

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon.$$

Example derivation:

$$S \Rightarrow aSa \Rightarrow abSba \Rightarrow abbSbba \Rightarrow abbbba.$$

What is the language defined by G_3 above?

Context-Free Grammars: Example 3

CFG G_3

$$S \quad \to \quad aSa \mid bSb \mid a \mid b \mid \epsilon.$$

Example derivation:

$$S \quad \Rightarrow aSa \quad \Rightarrow abSba \quad \Rightarrow abbSbba \quad \Rightarrow abbbba.$$

What is the language defined by G_3 above? Palindromes: $\{w \in \{a, b\}^* \mid w = w^R\}$.

Context-Free Grammars: Example 4

CFG G_4

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

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive “((()())()())”.

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

S

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$S \Rightarrow (S)$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow (SSS) \\
&\Rightarrow ((S)SS) \\
&\Rightarrow ((SS)SS) \\
&\Rightarrow (((S)S)SS) \\
&\Rightarrow ((()S)SS)
\end{aligned}
$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \\ &\Rightarrow ((()S)SS) \\ &\Rightarrow ((()(S))SS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \\ &\Rightarrow ((()S)SS) \\ &\Rightarrow ((()(S))SS) \\ &\Rightarrow ((()())SS) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \\ &\Rightarrow ((()S)SS) \\ &\Rightarrow ((()(S))SS) \\ &\Rightarrow ((()())SS) \\ &\Rightarrow ((()())(S)S) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow (SSS) \\
&\Rightarrow ((S)SS) \\
&\Rightarrow ((SS)SS) \\
&\Rightarrow (((S)S)SS) \\
&\Rightarrow ((()S)SS) \\
&\Rightarrow ((()(S))SS) \\
&\Rightarrow ((()())SS) \\
&\Rightarrow ((()())(S)S) \\
&\Rightarrow ((()())()S)
\end{aligned}
$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \\ &\Rightarrow ((()S)SS) \\ &\Rightarrow ((()(S))SS) \\ &\Rightarrow ((()())SS) \\ &\Rightarrow ((()())(S)S) \\ &\Rightarrow ((()())()S) \\ &\Rightarrow ((()())()(S)) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \\ &\Rightarrow ((()S)SS) \\ &\Rightarrow ((()(S))SS) \\ &\Rightarrow ((()())SS) \\ &\Rightarrow ((()())(S)S) \\ &\Rightarrow ((()())()S) \\ &\Rightarrow ((()())()(S)) \\ &\Rightarrow ((()())()()) \end{aligned}$$

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow (SSS) \\
&\Rightarrow ((S)SS) \\
&\Rightarrow ((SS)SS) \\
&\Rightarrow (((S)S)SS) \\
&\Rightarrow ((()S)SS) \\
&\Rightarrow ((()(S))SS) \\
&\Rightarrow ((()())SS) \\
&\Rightarrow ((()())(S)S) \\
&\Rightarrow ((()())()S) \\
&\Rightarrow ((()())()(S)) \\
&\Rightarrow ((()())()())
\end{aligned}$$

What is the language defined by G_4 above?

Context-Free Grammars: Example 4

CFG G_4

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

Exercise: Derive "((()())()())".

$$\begin{aligned} S &\Rightarrow (S) \\ &\Rightarrow (SS) \\ &\Rightarrow (SSS) \\ &\Rightarrow ((S)SS) \\ &\Rightarrow ((SS)SS) \\ &\Rightarrow (((S)S)SS) \\ &\Rightarrow ((()S)SS) \\ &\Rightarrow ((()(S))SS) \\ &\Rightarrow ((()())SS) \\ &\Rightarrow ((()())(S)S) \\ &\Rightarrow ((()())()S) \\ &\Rightarrow ((()())()(S)) \\ &\Rightarrow ((()())()()) \end{aligned}$$

What is the language defined by G_4 above?

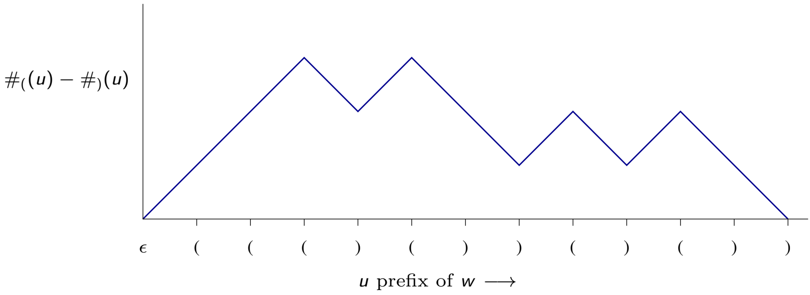
Balanced Parenthesis: $w \in \{(,)\}^*$ such that

- $\#_((w) = \#_)(w)$, and
- for each prefix u of w, $\#_((u) \geq \#_)(u)$.

Visualizing balanced parenthesis

Balanced Parenthesis: $w \in \{(,)\}^*$ such that

- $\#_((w) = \#_)(w)$, and
- for each prefix u of w, $\#_((u) \geq \#_)(u)$.



CFG's more formally

A Context-Free Grammar (CFG) is of the form

$$G = (N, A, S, P)$$

where

- N is a finite set of non-terminal symbols
- A is a finite set of terminal symbols.
- $S \in N$ is the start non-terminal symbol.
- P is a finite subset of $N \times (N \cup A)^*$, called the set of productions or rules. Productions are written as

$$X \to \alpha.$$

Derivations, language etc.

- "α derives β in 0 or more steps": $\alpha \Rightarrow_G^* \beta$.
- First define $\alpha \overset{n}{\Rightarrow} \beta$ inductively:
    - $\alpha \overset{1}{\Rightarrow} \beta$ iff α is of the form $\alpha_1 X \alpha_2$ and $X \rightarrow \gamma$ is a production in P, and $\beta = \alpha_1 \gamma \alpha_2$.
    - $\alpha \overset{n+1}{\Rightarrow} \beta$ iff there exists γ such that $\alpha \overset{n}{\Rightarrow} \gamma$ and $\gamma \overset{1}{\Rightarrow} \beta$.
- Sentential form of G: any $\alpha \in (N \cup A)^*$ such that $S \Rightarrow_G^* \alpha$.
- Language defined by G:

$$L(G) = \{w \in A^* \mid S \Rightarrow_G^* w\}.$$

- $L \subseteq A^*$ is called a Context-Free Language (CFL) if there is a CFG G such that $L = L(G)$.

Leftmost derivations

- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$\underline{S}$$

Leftmost derivations

- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$\underline{S} \quad \Rightarrow (\underline{S})$$

Leftmost derivations

- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$\begin{aligned} \underline{S} &\Rightarrow (\underline{S}) \\ &\Rightarrow (\underline{S}S) \end{aligned}$$

Leftmost derivations

- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$
\begin{aligned}
\underline{S} \quad & \Rightarrow (\underline{S}) \\
& \Rightarrow (\underline{S}S) \\
& \Rightarrow ((\underline{S})S)
\end{aligned}
$$

Leftmost derivations

- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$\begin{aligned} \underline{S} &\Rightarrow (\underline{S}) \\ &\Rightarrow (\underline{S}S) \\ &\Rightarrow ((\underline{S})S) \\ &\Rightarrow (()\underline{S}) \end{aligned}$$

Leftmost derivations

- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$
\begin{aligned}
\underline{S} \quad & \Rightarrow (\underline{S}) \\
& \Rightarrow (\underline{S}S) \\
& \Rightarrow ((\underline{S})S) \\
& \Rightarrow (()\underline{S}) \\
& \Rightarrow (()(\underline{S}))
\end{aligned}
$$

Leftmost derivations

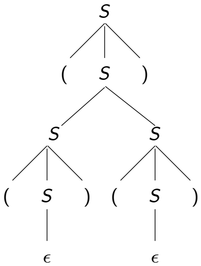
- A leftmost derivation in G is a derivation sequence in which at each step the leftmost non-terminal in the sentential form is re-written.
- Example:

$$\begin{aligned} \underline{S} &\Rightarrow (\underline{S}) \\ &\Rightarrow (\underline{S}S) \\ &\Rightarrow ((\underline{S})S) \\ &\Rightarrow (()\underline{S}) \\ &\Rightarrow (()(\underline{S})) \\ &\Rightarrow (()()) \end{aligned}$$

Parse trees

$$
\begin{aligned}
\underline{S} &\Rightarrow (\underline{S}) \\
&\Rightarrow (\underline{S}S) \\
&\Rightarrow ((\underline{S})S) \\
&\Rightarrow (()\underline{S}) \\
&\Rightarrow (()(\underline{S})) \\
&\Rightarrow (()())
\end{aligned}
$$

Derivation represented as parse tree:



- Sentential form can be read off from the leaves of the parse tree in a left-to-right manner.
- Leftmost derivations and parse trees represent eachother.

Proving that a CFG accepts a certain language

CFG G_1

$$\begin{array}{rcl} S & \rightarrow & aX \\ X & \rightarrow & aX \\ X & \rightarrow & bX \\ X & \rightarrow & b \end{array}$$

Prove that $L(G_1) = a(a+b)^*b$.

Proving that a CFG accepts a certain language

CFG G_1

$$\begin{array}{rcl} S & \rightarrow & aX \\ X & \rightarrow & aX \\ X & \rightarrow & bX \\ X & \rightarrow & b \end{array}$$

Prove that $L(G_1) = a(a+b)^*b$.

- Show that $L(G_1) \subseteq L(a(a+b)^*b)$, and $L(a(a+b)^*b) \subseteq L(G_1)$.
- Use induction statement that talks about sentential forms rather than just terminal strings.
- Eg: "$P(n)$: If $S \overset{n}{\Rightarrow}_{G_1} \alpha$ then α is of the form S, auX, or aub."
- Follows that all terminal sentential forms are of the form "aub" $\in L(a(a+b)^*b)$.
- For $L(a(a+b)^*b) \subseteq L(G_1)$ use induction statement "If $|u| = n$ then $S \Rightarrow^*_{G_1} auX$."

Proving that a CFG accepts a certain language

CFG G_2

$$\begin{array}{rcl} S & \rightarrow & aSb \\ S & \rightarrow & \epsilon. \end{array}$$

Prove that $L(G_2) = \{a^n b^n \mid n \geq 0\}$.

Proving that a CFG accepts a certain language

CFG G_4

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

Prove that $L(G_4) = \mathrm{BP}$.

Visualizing balanced parenthesis

Balanced Parenthesis: $w \in \{(,)\}^*$ such that

- $\#_((w) = \#_)(w)$, and
- for each prefix u of w, $\#_((u) \geq \#_)(u)$.