Lecture 5
Context-Free Languages

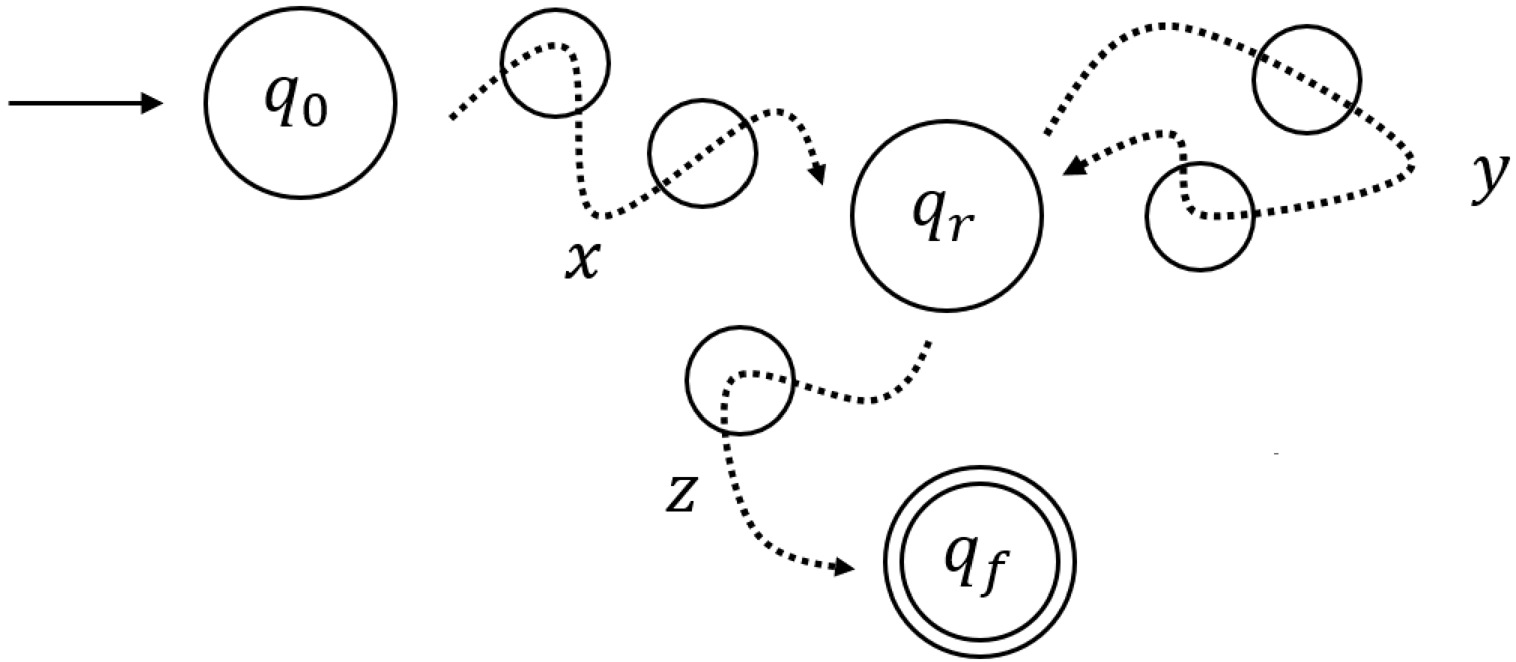
COSE215: Theory of Computation

Seunghoon Woo

Fall 2023

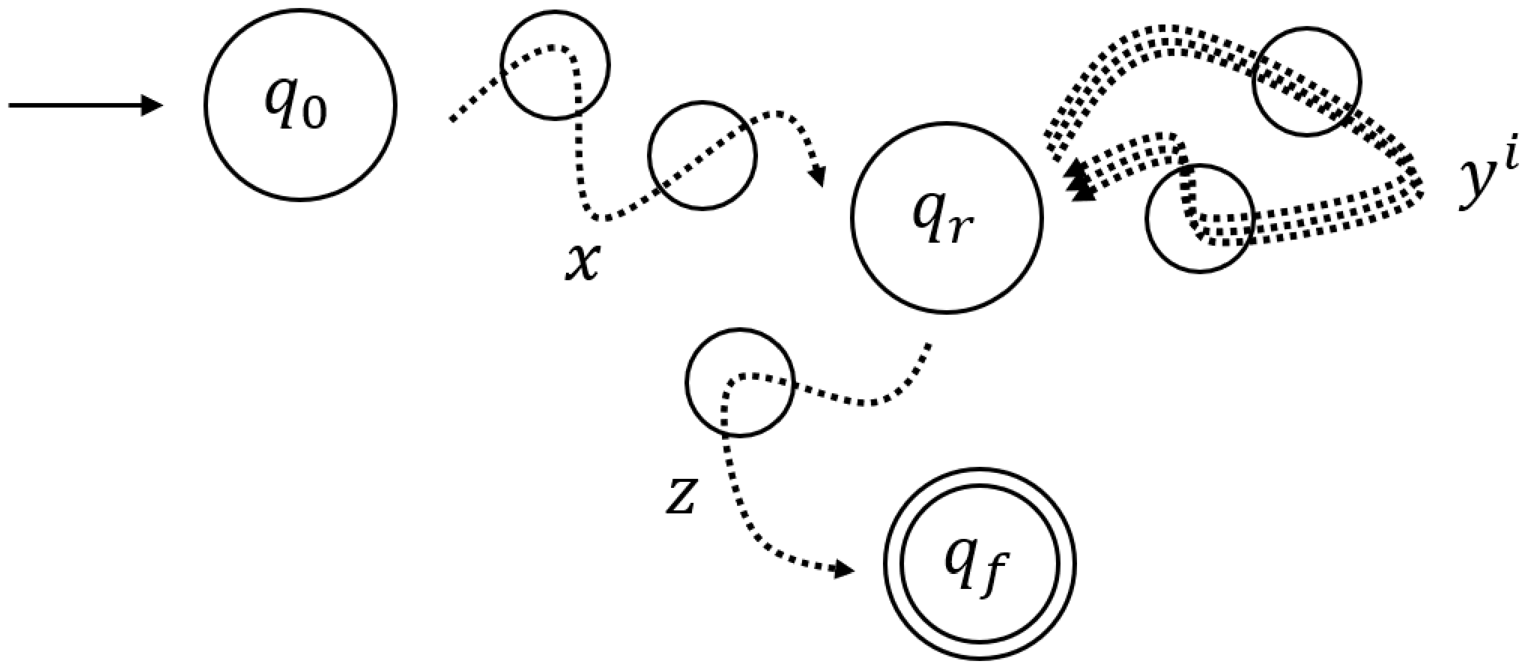
Review: Identifying non-regular languages

- **Pumping lemma**



Review: Identifying non-regular languages

- **Pumping lemma**



Contents

- **Context-Free Languages**

Context-Free Languages

- **There are languages that are not regular**
  - E.g., $L(G) = \{a^n b^n \mid n \geq 0\}$
- **What are the other categories that define such languages?**
  - Context-Free Languages
    - More generative than regular languages
    - Widely used in source code parsing, programming languages, etc.

Context-Free Languages

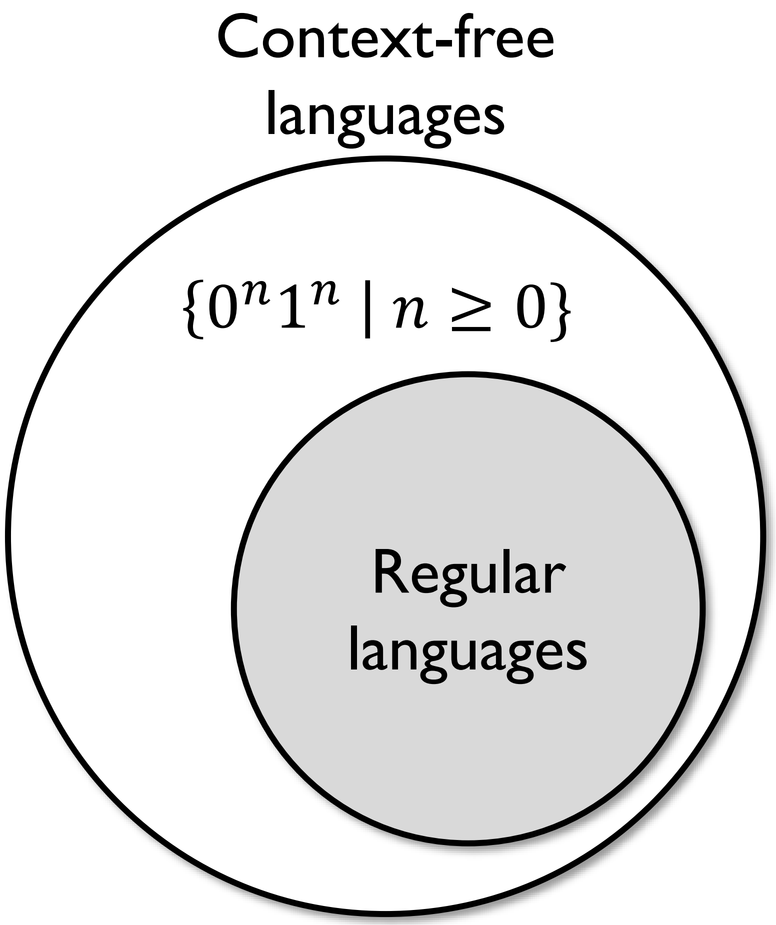
- **Definition**
  - A context-free grammar is a 4-tuple: $G = (V, T, S, P)$
    - ❖V: a finite set of variables (non-terminals)
    - ❖T: a finite set of symbols (terminals)
    - ❖S: the start variable ($S \in V$)
    - ❖P: a finite set of production rules ($P \subseteq V \times (V \cup T)^*$)
      - All productions in P have the form $\boldsymbol{A \rightarrow x}$**, where** $\boldsymbol{A \in V}$ **and** $\boldsymbol{x \in (V \cup T)^*}$
      - **(Regular language: right linear):** P have the form $A \rightarrow xB \mid x$, where $A, B \in V$ and $x \in T^*$

Context-Free Languages

- **Definition**
  - A context-free grammar is a 4-tuple: $G = (V, T, S, P)$
    - V: a finite set of variables (non-terminals)
    - T: a finite set of symbols (terminals)
    - S: the start variable ($S \in V$)
    - P: a finite set of production rules ($P \subseteq V \times (V \cup T)^*$)
      - All productions in P have the form $\boldsymbol{A \rightarrow x}$**, where** $\boldsymbol{A \in V}$ **and** $\boldsymbol{x \in (V \cup T)^*}$
  - A language L is said to be context-free if and only if there is a context-free grammar G such that $L = L(G)$

Context-Free Languages

- **vs. Regular languages**



Context-Free Languages

- **Example**
  - Consider the grammar $G = (\{S\}, \{a, b\}, S, P)$, with P given by $S \rightarrow aSb \mid \lambda$
    - $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aabb$ $(S \overset{*}{\Rightarrow} aabb)$

Note that

$S \rightarrow aSb$
$S \rightarrow \lambda$

can be written as

$S \rightarrow aSb \mid \lambda$

Context-Free Languages

- **Example**
  - Consider the grammar $G = (\{S\}, \{a, b\}, S, P)$, with P given by $S \rightarrow aSb \mid \lambda$
    - $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aabb$ ($S \overset{*}{\Rightarrow} aabb$)
    - $L(G) = \{a^n b^n \mid n \geq 0\}$

Note that

$S \rightarrow aSb$
$S \rightarrow \lambda$

can be written as

$S \rightarrow aSb \mid \lambda$

Context-Free Languages

- **Example**
  - Consider $G = (\{S\}, \{(,)\}, S, P)$ and $L(G) = \{w \in \{(,)\}^* \mid w \text{ is balanced}\}$
    - $\lambda, (\,), ((\,)), (\,)(\,), ((\,)(\,)), \ldots \in L$
    - $(,), ((\,))), (\,)))(\,), (((\,)), \ldots \notin L$

Context-Free Languages

- **Example**
  - Consider $G = (\{S\}, \{(,)\}, S, P)$ and $L(G) = \{w \in \{(,)\}^* \mid w \text{ is balanced}\}$
    - ❖ $\lambda, (\,), \big((\,)\big), (\,)(\,), \big((\,)(\,)\big), \ldots \in L$
    - ❖ $(,), \big((\,))\big), (\,)))(\,), \big(((\,)\big), \ldots \notin L$
    - ❖ $S \rightarrow (S) \mid SS \mid \lambda$

Context-Free Languages

- **Example**
  - Find a grammar for $L = \{a^n b^{2n} \mid n \geq 1\}$
    - $abb, aabbbb, aaabbbbbb \ldots \in L$
    - $ab, aab, \ ababa \ldots \notin L$

Context-Free Languages

- **Example**
  - Find a grammar for $L = \{a^n b^{2n} \mid n \geq 1\}$
    - $abb, aabbbb, aaabbbbbb \ldots \in L$
    - $ab, aab,\ ababa \ldots \notin L$
    - $S \rightarrow aSbb \mid abb$

Context-Free Languages

- **Practice**
  - Consider $L(G) = \{ww^R \mid w \in \{a, b\}^*\}$
    - ❖ Find P in the context-free grammar $G = (\{S\}, \{a, b\}, S, P)$
  - Consider $L(G) = \{a^n b^m \mid n \neq m\}$
    - ❖ Find P in the context-free grammar $G = (\{S, A, B, C\}, \{a, b\}, S, P)$

Context-Free Languages

- **Practice**
  - Show that the following language is not regular and find the CFG
    - ❖ $L = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w)\}$
      - $abba, aaabbb, ababab, abbaaabb \ldots \in L$
      - $abb, aabbbaa, ababa, \ldots \notin L$

Context-Free Languages

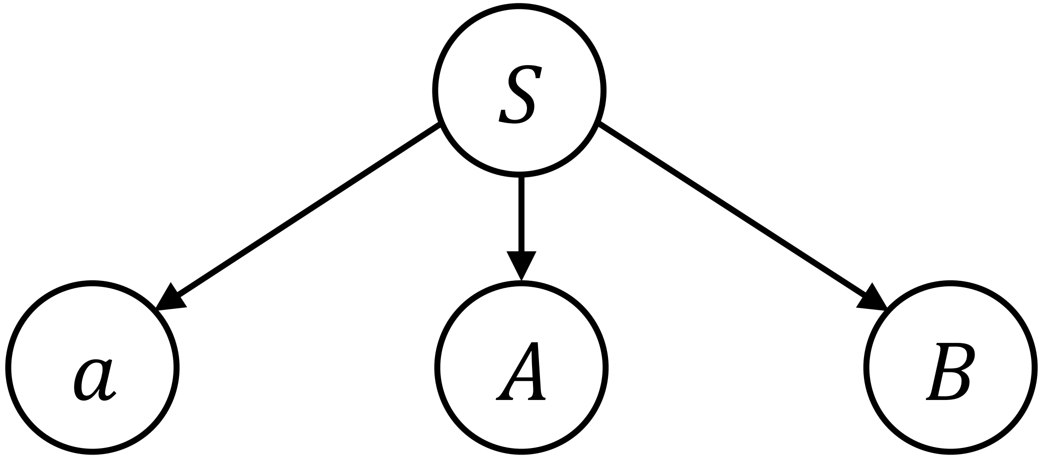
- **Leftmost and Rightmost derivation**
  - A derivation is said to be leftmost (rightmost) if in each step the leftmost (rightmost) variable is replaced
  - E.g., Consider the grammar G with productions
    - $S \rightarrow aAB, \quad A \rightarrow bBb, \quad B \rightarrow A \mid \lambda$

Context-Free Languages

- **Leftmost and Rightmost derivation**
  - A derivation is said to be leftmost (rightmost) if in each step the leftmost (rightmost) variable is replaced
  - E.g., Consider the grammar G with productions
    - $S \rightarrow aAB, \quad A \rightarrow bBb, \quad B \rightarrow A \mid \lambda$
    - The leftmost derivation of the string $abbbb$ is
      - $S \Rightarrow a\boldsymbol{A}B \Rightarrow ab\boldsymbol{B}bB \Rightarrow abb\boldsymbol{B} \Rightarrow abb\boldsymbol{A} \Rightarrow abbb\boldsymbol{B}b \Rightarrow abbbb$
    - The rightmost derivation of the string $abbbb$ is
      - $S \Rightarrow aA\boldsymbol{B} \Rightarrow a\boldsymbol{A} \Rightarrow ab\boldsymbol{B}b \Rightarrow ab\boldsymbol{A}b \Rightarrow abb\boldsymbol{B}bb \Rightarrow abbbb$

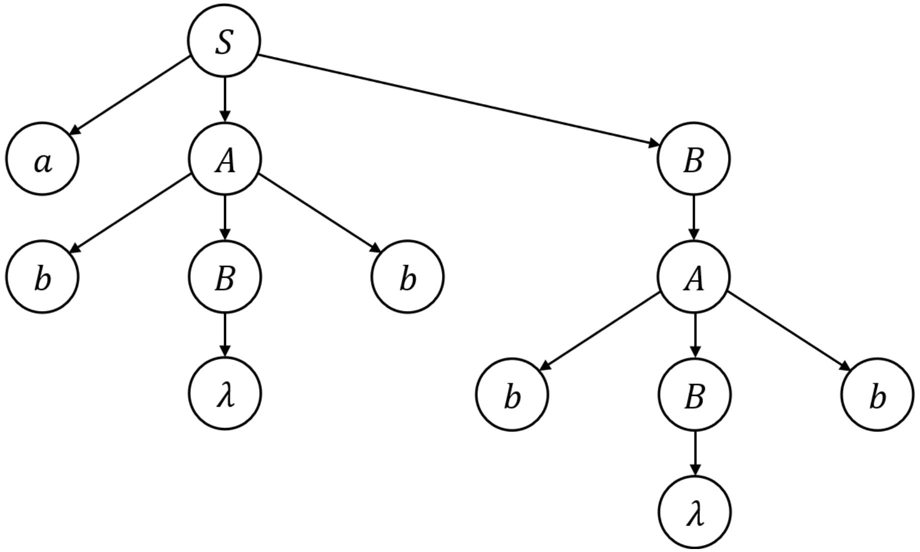
Context-Free Languages

- **Derivation (parse) tree**
  - A derivation is a repeated application of rules
  - An ordered tree
    - ❖Nodes are labeled with the left-hand sides of productions
    - ❖Children of a node represent its corresponding right-hand sides
  - E.g., $S \rightarrow aAB$



Context-Free Languages

- **Example**
  - Consider the grammar G with productions
    - $S \rightarrow aAB, \quad A \rightarrow bBb, \quad B \rightarrow A \mid \lambda$
    - $S \Rightarrow a\boldsymbol{A}B \Rightarrow ab\boldsymbol{B}bB \Rightarrow abb\boldsymbol{B} \Rightarrow abb\boldsymbol{A} \Rightarrow abbb\boldsymbol{B}b \Rightarrow abbbb$



It can indicate derivation, but it cannot show order

Next Lecture

- **Parsing and ambiguity**
- **Context-free grammars and programming languages**