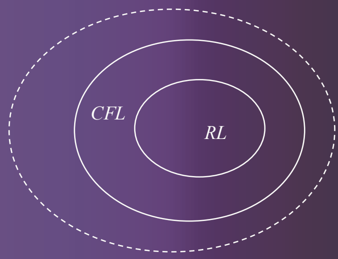
Session 15

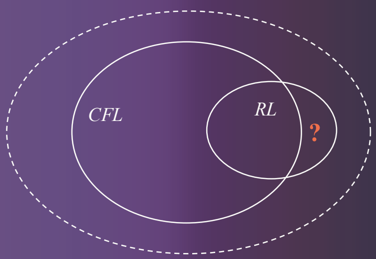
Context Free Grammars
and
Regular Expressions

CFG and FA



- ✓ Not all *CFL* are *RL* (the pumping lemma)
- If *L* is a regular language, is it a *CFL*?

CFG and FA



- If *L* is a regular language is it a *CFL*?
- If it is, what is the form of its grammar?

All *RL* are *CFL*

- Every *RE* has a corresponding CFG
- Every FA has a corresponding CFG:
  - Regular Grammar (RG)
- Every RG has a corresponding FA
- Every *RE* has a corresponding RG

All *RL* are *CFL*

- Every regular language is a Context-free language
- There is a CFG for the basic regular languages in Σ:
  - $L(\Phi) = L(G_\Phi)$ where $G = (V, \Sigma, S, \Phi)$
  - $L(\Lambda) = L(G_\Lambda)$ where $G_\Lambda = (V, \Sigma, S, \{S \rightarrow \Lambda\})$
  - If $\boldsymbol{a} \in \Sigma$ then $L(\boldsymbol{a}) = \{a\} = L(G_a)$ and $G_a = (V, \Sigma, S, \{S \rightarrow a\})$

Operations of CFG

- The inductive part:
- Given:
  - $L_1 = L(G_1)$ and $G_1 = (V_1, \Sigma, S_1, P_1)$
  - $L_2 = L(G_2)$ and $G_2 = (V_2, \Sigma, S_2, P_2)$
- $L_u = L_1 \cup L_2 = L(G_1)$ where
  - $G_u = (V_u, \Sigma, S_u, P_u)$ and $P_u = P_1 \cup P_2 \cup \{S_u \rightarrow S_1 \mid S_2\}$
- $L_c = L_1 L_2 = L(G_c)$ where
  - $G_c = (V_c, \Sigma, S_c, P_c)$ and $P_c = P_1 \cup P_2 \cup \{S_c \rightarrow S_1 S_2\}$
- $L_* = L_1^* = L(G_*)$ where
  - $G_* = (V, \Sigma, S, P)$ where $P = P_1 \cup \{S \rightarrow S_1 S \mid \Lambda\}$

Renaming variables

- Names of variables should be unique:
  – $V_1 \cap V_2 \cap V_u \cap V_c \cap V = \Phi$
- Consider G_1 and G_2 such that $V_1 \cap V_2 = X \neq \Phi$
  – $G_1 = (V_1, \Sigma, S_1, \{S_1 \to XA, X \to c, A \to a\})$
  – $G_2 = (V_2, \Sigma, S_2, \{S_2 \to XB, X \to d, B \to b\})$
- L_1 and L_2:
  – $S_1 \Rightarrow XA \Rightarrow cA \Rightarrow ca$ and $L(G_1) = \{ca\}$
  – $S_2 \Rightarrow XB \Rightarrow dB \Rightarrow db$ and $L(G_2) = \{db\}$
- Consider the union grammar:
  – $P_u = P_1 \cup P_2 \cup \{S_u \to S_1 \mid S_2\}$
  – $S_u \Rightarrow S_1 \mid S_2 \Rightarrow XA \Rightarrow dA \Rightarrow da \notin L_1 \cup L_2$

RE are CFG

- The CFG corresponding to $(011 + 1)^*(01)^*$
- Productions of basic grammars: $P_1 = \{A \to 1\}$ and $P_2 = \{B \to 1\}$
  $A \Rightarrow 1$ and $B \Rightarrow 1$ $\{1\}$ & $\{1\}$
- Concatenation: $P_c = \{A \to 1\} \cup \{B \to 1\} \cup \{C \to AB\}$
  $C \Rightarrow AB \Rightarrow 1B \Rightarrow 11$ $\{11\}$
- Concatenation again: $P_c = \{D \to 0\} \cup \{C \to AB\} \cup \{E \to DC\}$
  $E \Rightarrow DC \Rightarrow 0C \Rightarrow 011$ $\{011\}$
- Union: $P_u = \{E \to DC\} \cup \{F \to 1\} \cup \{G \to E \mid F\}$
  $G \Rightarrow E \mid F \Rightarrow 011 \mid F \Rightarrow 011 \mid 1$ $\{011 + 1\}$

RE are CFG

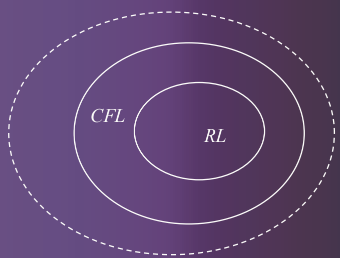
- Provide a CFG for $(011 + 1)^*(01)^*$
- Closure: $P = P_1 \cup \{S \to S_1 S \mid \Lambda\}$
  – $P_1 = \{A \to 011 \mid 1\}$ (Reusing the names for clarity)
  – $P_2 = \{A \to 011 \mid 1\} \cup \{B \to AB \mid \Lambda\}$ $\{011 + 1\}^*$
  $B \Rightarrow (011 \mid 1)B$
  $\Rightarrow (011 \mid 1)(011 \mid 1)B$ (AAB)
  ...
  $\Rightarrow (011 \mid 1)(011 \mid 1)\ldots(011 \mid 1)\Lambda$ $(AA\ldots A\Lambda)$
- Closure: $P_3 = \{C \to 01\} \cup \{D \to CD \mid \Lambda\}$ $\{01\}^*$
- Concatenation: $P_c = P_2 \cup P_3 \cup \{S \to BD\}$ $\{011 + 1\}^*\{01\}^*$

RE are CFG

- $RE = (011 + 1)^*(01)^*$
- The corresponding CFG is
  $G = (\{A, B, C, D, S\}, \{0, 1\}, S, P)$
  and P has the productions:

$A \to 011 \mid 1$	$\{011 + 1\}$
$B \to AB \mid \Lambda$	$\{011 + 1\}^*$
$C \to 01$	$\{01\}$
$D \to CD \mid \Lambda$	$\{01\}^*$
$S \to BD$	$\{011 + 1\}^*\{01\}^*$

CFG and FA

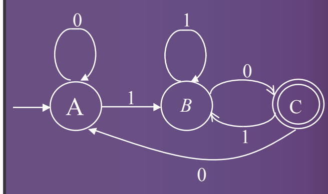


✓ Not all *CFL* are *RL* (the pumping lemma)
✓ If *L* is regular it is a *CFL*

What is the form of a grammar for a *RL*?

✓ Every *RE* has a corresponding CFG
- Every FA has a corresponding CFG:
  – Regular Grammar (RG)
- Every RG has a corresponding FA
- Every *RE* has a corresponding RG

CFG corresponding to a FA

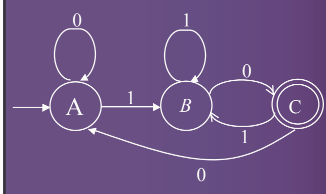


An old friend!

$\delta^*(A, \Lambda) = A$
$\delta^*(A, \Lambda 1) = B$
$\delta^*(A, \Lambda 11) = B$
$\delta^*(A, \Lambda 110) = C$
$\delta^*(A, \Lambda 1100) = A$
$\delta^*(A, \Lambda 11000) = A$
$\delta^*(A, \Lambda 110001) = B$
$\delta^*(A, \Lambda 1100010) = C$
$\delta^*(A, \Lambda 11000101) = B$
$\delta^*(A, \Lambda 110001010) = C$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA



$\delta^*(A, \Lambda) \Rightarrow A$
$\delta^*(A, \Lambda 1) \Rightarrow B$
$\delta^*(A, \Lambda 11) \Rightarrow B$
$\delta^*(A, \Lambda 110) \Rightarrow C$
$\delta^*(A, \Lambda 1100) \Rightarrow A$
$\delta^*(A, \Lambda 11000) \Rightarrow A$
$\delta^*(A, \Lambda 110001) \Rightarrow B$
$\delta^*(A, \Lambda 1100010) \Rightarrow C$
$\delta^*(A, \Lambda 11000101) \Rightarrow B$
$\delta^*(A, \Lambda 110001010) \Rightarrow C$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

$\delta^*(q_0, ya) = \delta(\delta^*(q_0, y), a)$

$\delta^*(q_0, x)$ correspond to $S \Rightarrow^* x$

$\delta^*(A, \Lambda) = A$	$S \Rightarrow \Lambda A$
$\delta^*(A, \Lambda 1) = B$	$\Rightarrow \Lambda 1B$
$\delta^*(A, \Lambda 11) = B$	$\Rightarrow \Lambda 11B$
$\delta^*(A, \Lambda 110) = C$	$\Rightarrow \Lambda 110C$
$\delta^*(A, \Lambda 1100) = A$	$\Rightarrow \Lambda 1100A$
$\delta^*(A, \Lambda 11000) = A$	$\Rightarrow \Lambda 11000A$
$\delta^*(A, \Lambda 110001) = B$	$\Rightarrow \Lambda 110001B$
$\delta^*(A, \Lambda 1100010) = C$	$\Rightarrow \Lambda 1100010C$
$\delta^*(A, \Lambda 11000101) = B$	$\Rightarrow \Lambda 11000101B$
$\delta^*(A, \Lambda 110001010) = C$	$\Rightarrow \Lambda 110001010C$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

The productions:

$A \Rightarrow \Lambda A$	$A \rightarrow \Lambda A$
$\Rightarrow \Lambda 1B$	$A \rightarrow 1B$
$\Rightarrow \Lambda 11B$	$B \rightarrow 1B$
$\Rightarrow \Lambda 110C$	$B \rightarrow 0C$
$\Rightarrow \Lambda 1100A$	$C \rightarrow 0A$
$\Rightarrow \Lambda 11000A$	$A \rightarrow 0A$
$\Rightarrow \Lambda 110001B$	$A \rightarrow 1B$
$\Rightarrow \Lambda 1100010C$	$B \rightarrow 0C$
$\Rightarrow \Lambda 11000101B$	$C \rightarrow 1B$
$\Rightarrow \Lambda 110001010C$	$B \rightarrow 0C$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

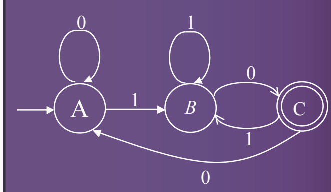
getting rid of the Λ:

$\Rightarrow 1B$	$A \rightarrow 1B$
$\Rightarrow 11B$	$B \rightarrow 1B$
$\Rightarrow 110C$	$B \rightarrow 0C$
$\Rightarrow 1100A$	$C \rightarrow 0A$
$\Rightarrow 11000A$	$A \rightarrow 0A$
$\Rightarrow 110001B$	$A \rightarrow 1B$
$\Rightarrow 1100010C$	$B \rightarrow 0C$
$\Rightarrow 11000101B$	$C \rightarrow 1B$
$\Rightarrow 110001010C$	$B \rightarrow 0C$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

Eliminating duplications:

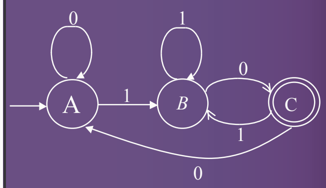


$A \rightarrow 0A$
$A \rightarrow 1B$
$B \rightarrow 0C$
$B \rightarrow 1B$
$C \rightarrow 0A$
$C \rightarrow 1B$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

The transition function

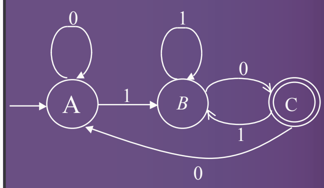


$\delta(A, 0) = A$
$\delta(A, 1) = B$
$\delta(B, 0) = C$
$\delta(B, 1) = B$
$\delta(C, 0) = A$
$\delta(C, 1) = B$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

The corresponding grammar:



$\delta(A, 0) = A$	$A \rightarrow 0A$
$\delta(A, 1) = B$	$A \rightarrow 1B$
$\delta(B, 0) = C$	$B \rightarrow 0C$
$\delta(B, 1) = B$	$B \rightarrow 1B$
$\delta(C, 0) = A$	$C \rightarrow 0A$
$\delta(C, 1) = B$	$C \rightarrow 1B$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

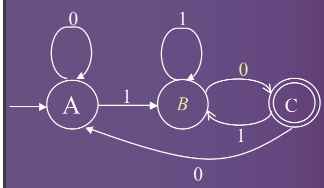
Accepting a string

$\Rightarrow 1B$	$A \rightarrow 1B$
$\Rightarrow 11B$	$B \rightarrow 1B$
$\Rightarrow 110C$	$B \rightarrow 0C$
$\Rightarrow 1100A$	$C \rightarrow 0A$
$\Rightarrow 11000A$	$A \rightarrow 0A$
$\Rightarrow 110001B$	$A \rightarrow 1B$
$\Rightarrow 1100010C$	$B \rightarrow 0C$
$\Rightarrow 11000101B$	$C \rightarrow 1B$
$\Rightarrow 110001010$	$B \rightarrow 0$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

The corresponding grammar



$A \rightarrow 0A$
$A \rightarrow 1B$
$B \rightarrow 0C$
$B \rightarrow 1B$
$C \rightarrow 0A$
$C \rightarrow 1B$

A transition to the final state: $B \rightarrow 0$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Regular Grammars

- A grammar G = (V, Σ, S, P) is regular if every production takes one of two forms:
  - $B \rightarrow aC$
  - $B \rightarrow a$
- Let L be a RL and $M = (Q, \Sigma, q_0, A, \delta)$ such that $L(M) = L$; there is a Regular Grammar $G = (V, \Sigma, S, P)$ accepting L, defined as follows:
  - $V = Q$ The variables of G are the states of M
  - $S = q_0$ The start symbol of G is the initial state of M
  - $P = \{B \rightarrow aC \mid \delta(B, a) = C\} \cup$
    $\{B \rightarrow a \mid \delta(B, a) = F \text{ and } F \in A\}$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

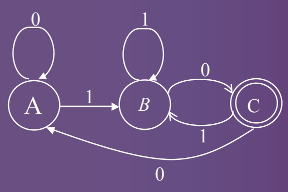
$M = (Q, \Sigma, A, \{C\}, \delta)$ $\qquad$ $G = (\{A, B, C\}, \Sigma, A, P)$

δ:	P:
$\delta(A, 0) = A$	$A \rightarrow 0A$
$\delta(A, 1) = B$	$A \rightarrow 1B$
$\delta(B, 0) = C$	$B \rightarrow 0C \mid 0$
$\delta(B, 1) = B$	$B \rightarrow 1B$
$\delta(C, 0) = A$	$C \rightarrow 0A$
$\delta(C, 1) = B$	$C \rightarrow 1B$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

CFG corresponding to a FA

$M = (Q, \Sigma, A, \{C\}, \delta)$ $G = (\{A, B, C\}, \Sigma, A, P)$



$A \rightarrow 0A$
$A \rightarrow 1B$
$B \rightarrow 0C \mid 0$
$B \rightarrow 1B$
$C \rightarrow 0A$
$C \rightarrow 1B$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

All *RL* are *CFL*

- ✓ Every *RE* has a corresponding CFG
- ✓ Every FA has a corresponding CFG:
  - – Regular Grammar (RG)
- Every RG has a corresponding FA
- Every *RE* has a corresponding RG

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

A FA corresponding to CFG

- We can go in the other direction!

$A \rightarrow 0A$		$\delta(A, 0) = A$
$A \rightarrow 1B$		$\delta(A, 1) = B$
$B \rightarrow 0C$	➡	$\delta(B, 0) = C$
$B \rightarrow 1B$		$\delta(B, 1) = B$
$C \rightarrow 0A$		$\delta(C, 0) = A$
$C \rightarrow 1B$		$\delta(C, 1) = B$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

A FA corresponding to CFG

- What is the state corresponding to the final condition?

$A \rightarrow 0A$		$\delta(A, 0) = A$
$A \rightarrow 1B$		$\delta(A, 1) = B$
$B \rightarrow 0C$	➡	$\delta(B, 0) = C$
$B \rightarrow 1B$		$\delta(B, 1) = B$
$C \rightarrow 0A$		$\delta(C, 0) = A$
$C \rightarrow 1\text{B}$		$\delta(C, 1) = B$
$B \rightarrow 0$		$\delta(B, 0) = ?$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

A FA corresponding to CFG

- We can use the productions in reverse, but…

$A \rightarrow 1B$	$\Rightarrow 1B$
$B \rightarrow 1B$	$\Rightarrow 11B$
$B \rightarrow 0C$	$\Rightarrow 110C$
$C \rightarrow 0A$	$\Rightarrow 1100A$
$A \rightarrow 0A$	$\Rightarrow 11000A$
$A \rightarrow 1B$	$\Rightarrow 110001B$
$B \rightarrow 0C$	$\Rightarrow 1100010C$
$C \rightarrow 1B$	$\Rightarrow 11000101B$
$B \rightarrow 0$	$\Rightarrow 110001010$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

A FA corresponding to CFG

- We need an additional accepting state: F

$A \rightarrow 1B$	$\Rightarrow 1B$
$B \rightarrow 1B$	$\Rightarrow 11B$
$B \rightarrow 0C$	$\Rightarrow 110C$
$C \rightarrow 0A$	$\Rightarrow 1100A$
$A \rightarrow 0A$	$\Rightarrow 11000A$
$A \rightarrow 1B$	$\Rightarrow 110001B$
$B \rightarrow 0C$	$\Rightarrow 1100010C$
$C \rightarrow 1B$	$\Rightarrow 11000101B$
$B \rightarrow 0$	$\Rightarrow 110001010\text{F}$ ⬅

- Only one will do!

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

A FA corresponding to CFG

- The resulting M is non-deterministic…

$A \to 1B$	$\Rightarrow 1B$
$B \to 1B$	$\Rightarrow 11B$
$B \to 0C$	$\Rightarrow 110C$
$C \to 0A$	$\Rightarrow 1100A$
$A \to 0A$	$\Rightarrow 11000A$
$A \to 1B$	$\Rightarrow 110001B$
$B \to 0C$	$\Rightarrow 1100010C$ ⇦
$C \to 1B$	$\Rightarrow 11000101B$
$B \to 0$	$\Rightarrow 110001010F$ ⇦

- It is a NFA!

A FA corresponding to CFG

- The resulting NFA:

$A \to 0A$		$\delta(A, 0) = A$
$A \to 1B$		$\delta(A, 1) = B$
$B \to 0C$		$\delta(B, 0) = C$ ⇦
$B \to 1B$	⇨	$\delta(B, 1) = B$
$C \to 0A$		$\delta(C, 0) = A$
$C \to 1B$		$\delta(C, 1) = B$
$B \to 0$		$\delta(B, 0) = F$ ⇦

Regular Grammars

- For any language $L \subseteq \Sigma^*$, L is regular iff there is a RG G so that $L(G) = L - \Lambda$
- Let L be a Regular Grammar $G = (V, \Sigma, S, P)$ such that $L(G) = L$; there is an NFA $M = (Q, \Sigma, q_0, A, \delta)$ defined as follows:
  - $Q = V \cup \{f\}$ The states of M are the variables of G, and an additional final state f
  - $q_0 = S$ G's start symbol is M's initial state
  - δ is defined as follows:

    $\delta(q, a) = \{p\}$ if $q \to ap \in P$ and $q \to a \notin P$

    $\delta(q, a) = \{p\} \cup \{f\}$ if $q \to ap \in P$ and $q \to a \in P$

Regular Grammars

- Let L be a Regular Grammar

  $G = (\{A, B, C\}, \{0, 1\}, A, P)$

  where $P = \{A \to 0A \mid 1B,$

  $B \to 0C \mid 0 \mid 1B,$

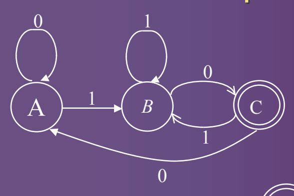
  $C \to 0A \mid 1B\}$

  then $M = (\{A, B, C, F\}, \{0, 1\}, A, \{F\}, \delta)$,

  where δ as follows:

$\delta(A, 0) = \{A\}$	$\delta(A, 1) = \{B\}$
$\delta(B, 0) = \{C, F\}$	$\delta(B, 1) = \{B\}$
$\delta(C, 0) = \{A\}$	$\delta(C, 1) = \{B\}$

CFG corresponding to a FA



$A \to 0A$

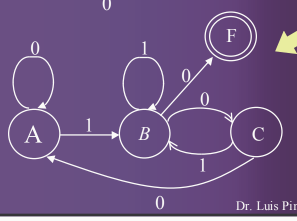
$A \to 1B$

$B \to 0C \mid 0$

$B \to 1B$

$C \to 0A$

$C \to 1B$



All *RL* are *CFL*

- ✓ Every *RE* has a corresponding CFG
- ✓ Every FA has a corresponding CFG:
  - Regular Grammar (RG)
- ✓ Every RG has a corresponding FA
- Every *RE* has a corresponding RG

RG corresponding to a *RE*

- There is a NFA-Λ for every *RE*
  - Kleene's theorem
- There is a NFA for every NFA-Λ
- There is a DFA for every NFA
- There is a RG for every DFA
- Then, there is a RG for every *RE*

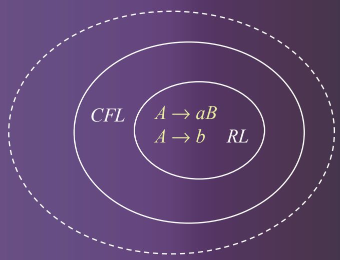
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

RG corresponding to a *RE*

✓ Every *RE* has a corresponding CFG
✓ Every FA has a corresponding CFG:
  – Regular Grammar (RG)
✓ Every *RG* has a corresponding FA
✓ Every *RE* has a corresponding RG

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

A Normal Form for *RL*



✓ If *L* is a regular language, it is a *CFL*
✓ A RG is a CFG in a Normal Form

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003