Theory of Computation
Chapter 2

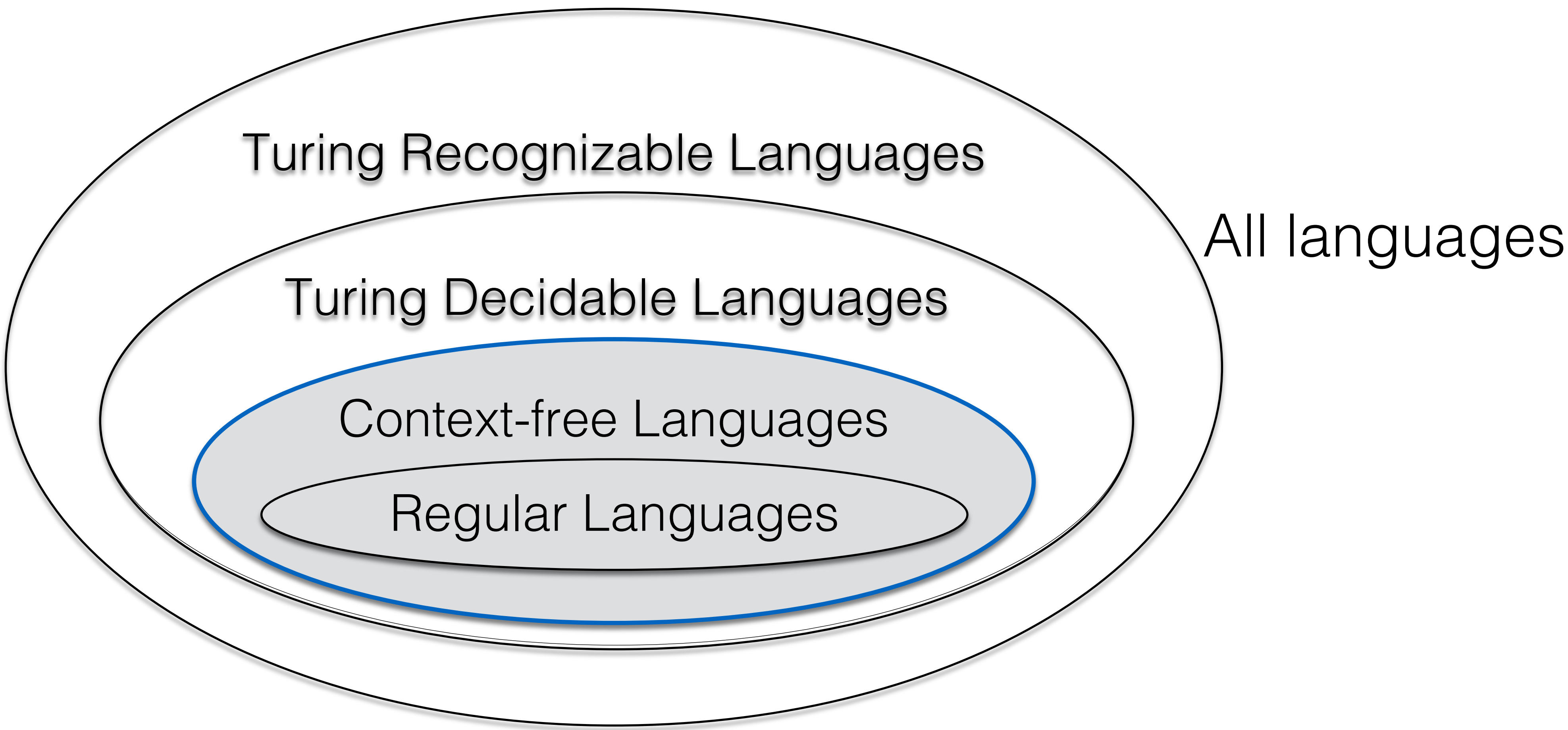
Properties of Context-Free Languages

School of Engineering | Computer Science

Noam Chomsky

- Sometimes called "the father of modern linguistics”
- Major figure in analytic philosophy and one of the founders of the field of cognitive science
- Wrote *Syntactic Structures* that revolutionized the scientific study of language and played a major role in remodeling the study of language
- Political activist and defender of free speech

Context-Free Languages



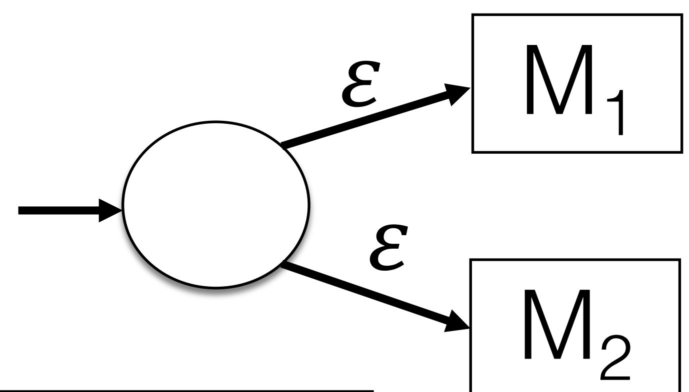
Context-Free Languages

PDA = CFG = CNF

Closed under union, ∪, concatenation, ∘, and star, ∗.

Properties of Context-Free Languages

- <u>Lemma</u>: Context-free languages are closed under the <u>union</u> operation
- <u>Proof</u>: Let G_1 and G_2 be CFG's generating CFL's: $L(G_1)$ and $L(G_2)$ respectively. Let S_1 and S_2 be start variables for G_1 and G_2 respectively. Then the CFG for $L(G_1) \cup L(G_2)$ is:
  - $S \rightarrow S_1 \mid S_2$
    $S_1 \rightarrow \ldots$
    $S_2 \rightarrow \ldots$



M_1 and M_2 are the machines that represent G_1 and G_2 respectively

Create a new start state & variable

Properties of Context-Free Languages

- Lemma: Context-free languages are closed under the concatenation operation

- Proof: Let G_1 and G_2 be CFG's generating CFL's: $L(G_1)$ and $L(G_2)$ respectively. Let S_1 and S_2 be start variables for G_1 and G_2 respectively. Then the CFG for $L(G_1) \circ L(G_2)$ is:

  - $S \rightarrow S_1S_2$
    $S_1 \rightarrow \ldots$
    $S_2 \rightarrow \ldots$

$$\boxed{M_1} \xrightarrow{\varepsilon} \boxed{M_2}$$

M_1's accept states become regular states

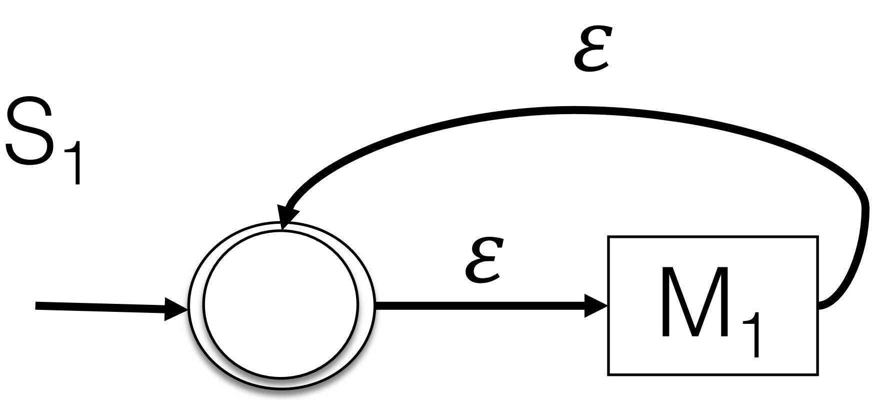
M_1 and M_2 are the machines that represent G_1 and G_2 respectively

Properties of Context-Free Languages

- Lemma: Context-free languages are closed under the star operation

- Proof: Let G_1 be a CFG generating a CFL: $L(G_1)$. Let S_1 be the start variable for G_1. Then the CFG for $L(G_1)^*$ is:

  - $S \rightarrow \varepsilon \mid SS_1$
    $S_1 \rightarrow \ldots$



M_1 is the machine that represents G_1

Create a new start state that is an accept state

Comparison of Regular & Context-Free Languages

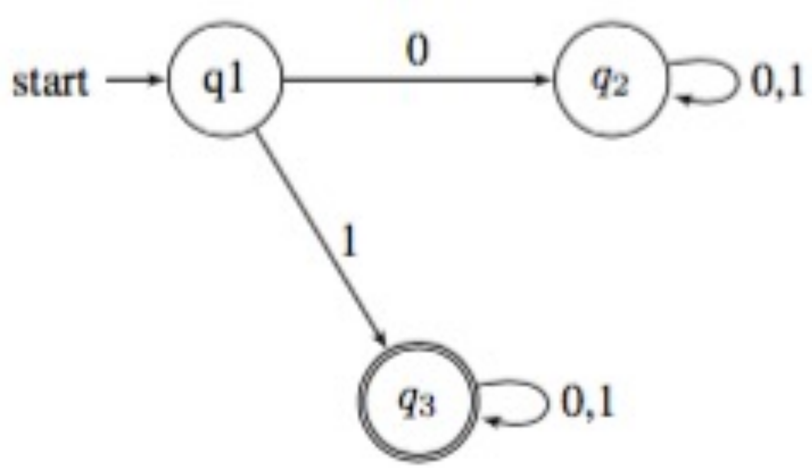
- Take a non-regular language L, can we make it a context-free language?
  - Ex: $L = \{0^n1^n \mid n \geq 0\}$ We learned in Chapter 1 that L is not a regular language
    - We can describe it with productions: $S \rightarrow 0S1 \mid \varepsilon$
    - Since we can define the grammar, the language is context-free
- <u>Claim</u>: All regular languages are context-free languages

Comparison of Regular & Context-Free Languages

- Claim: All regular languages are context-free languages
- Proof: Let L be a regular language with a DFA M = (Q, Σ, δ, q_0, F)
  - We can create a grammar G by replacing the states in M with the variables in G
  - We construct the grammar as follows:
    1. For each state $q_i \in Q$, add variable R_i in G
    2. For each transition $\delta(q_i, a) = q_j$ for the DFA, add a substitution rule to G: $R_i \rightarrow aR_j$
    3. For any $q_i \in F$, add $R_i \rightarrow \varepsilon$
    4. Set R_0 in G to be the start variable q_0

Comparison of Regular & Context-Free Languages

- Claim: All regular languages are context-free languages
- Ex:

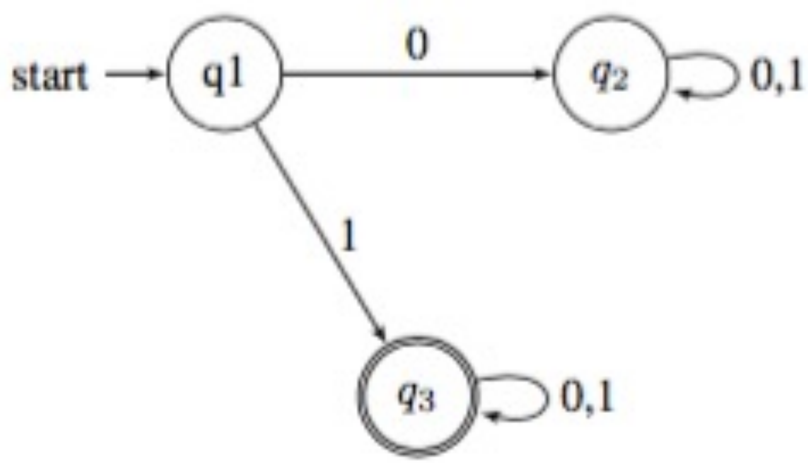


1. For each state $q_i \in Q$, add variable R_i in G
2. For each transition $\delta(q_i, a) = q_j$ for the DFA, add a substitution rule to G: $R_i \rightarrow aR_j$
3. For any $q_i \in F$, add $R_i \rightarrow \varepsilon$
4. Set R_0 in G to be the start variable q_0

- Create the CFG

Comparison of Regular & Context-Free Languages

- Claim: All regular languages are context-free languages
- Ex:



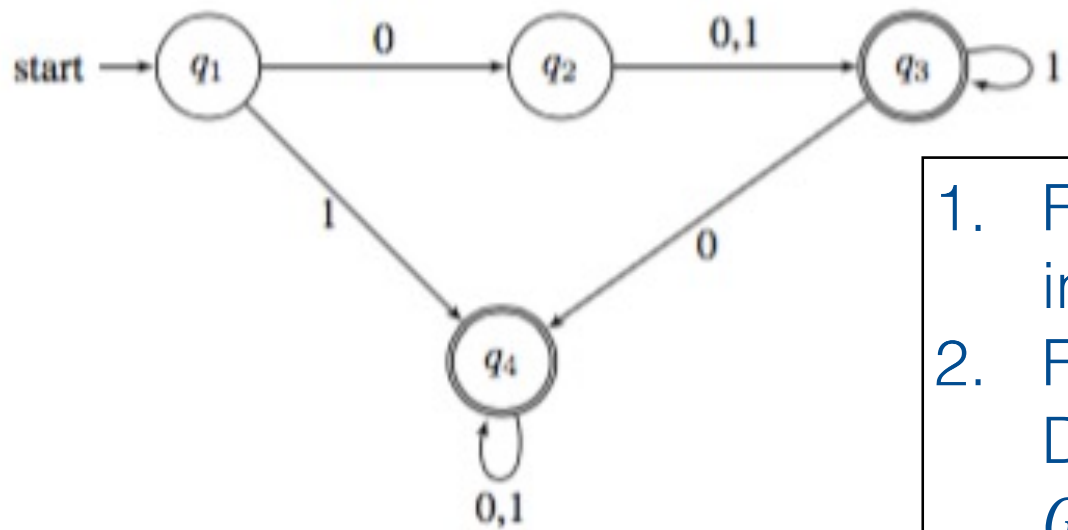
1. Variables: R_1, R_2, R_3
2. $R_1 \rightarrow 0R_2 \mid 1R_3$
   $R_2 \rightarrow 0R_2 \mid 1R_2$
   $R_3 \rightarrow 0R_3 \mid 1R_3 \mid \varepsilon$
4. R_1 is the start variable

The rules match the transitions, such as: on q_1 with a $\mathbf{0}$ move to q_2, so get $R_1 \rightarrow \mathbf{0}R_2$
Step 3 adds the ε rule to the variable that represents the accept state

Comparison of Regular & Context-Free Languages

- <u>Claim</u>: All regular languages are context-free languages

- Ex:



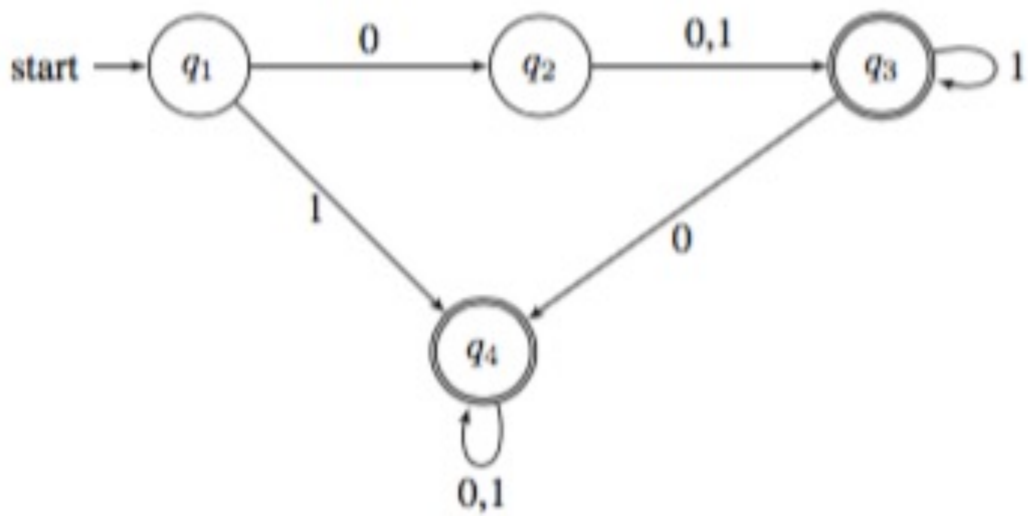
- Create the CFG

1. For each state $q_i \in Q$, add variable R_i in G
2. For each transition $\delta(q_i, a) = q_j$ for the DFA, add a substitution rule to G: $R_i \rightarrow aR_j$
3. For any $q_i \in F$, add $R_i \rightarrow \varepsilon$
4. Set R_0 in G to be the start variable q_0

School of Engineering | Computer Science

Comparison of Regular & Context-Free Languages

- Claim: All regular languages are context-free languages
- Ex:

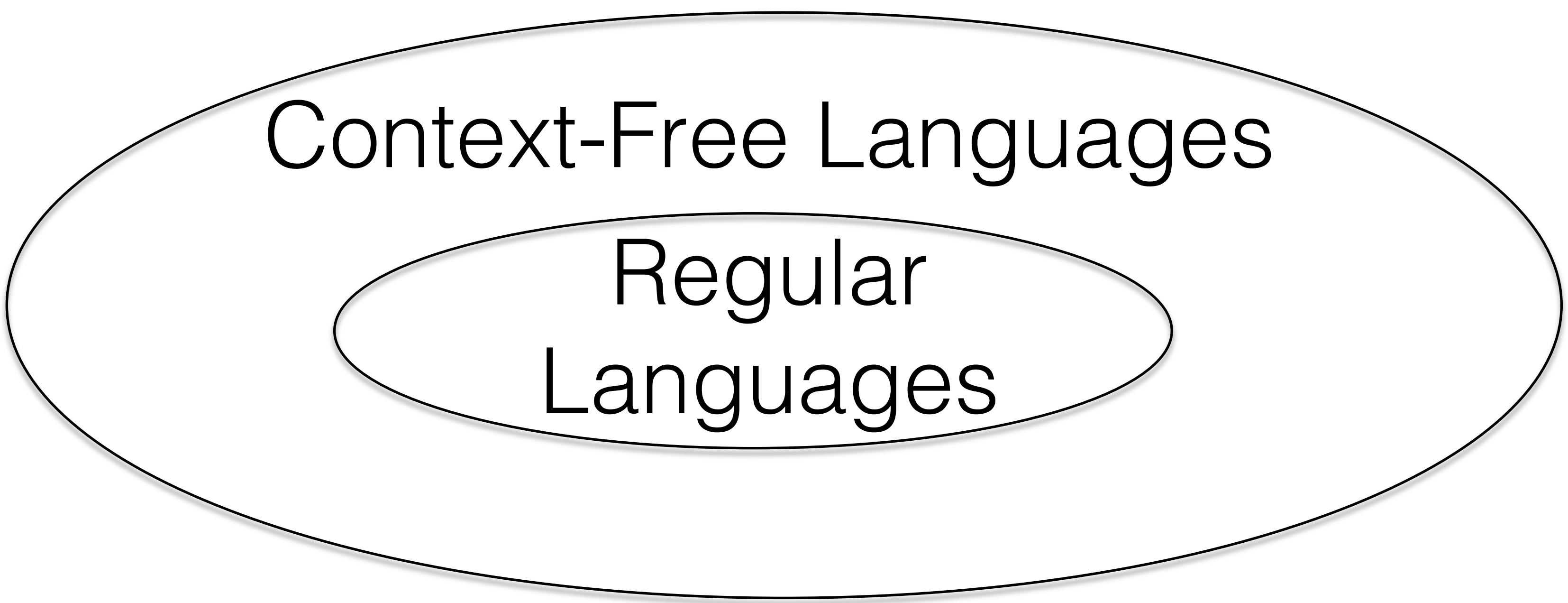


1. Variables: R_1, R_2, R_3, R_4
2. $R_1 \rightarrow 0R_2 \mid 1R_4$
   $R_2 \rightarrow 0R_3 \mid 1R_3$
   $R_3 \rightarrow 0R_4 \mid 1R_3 \mid \varepsilon$
   $R_4 \rightarrow 0R_4 \mid 1R_4 \mid \varepsilon$
4. R_1 is the start variable

School of Engineering | Computer Science

Comparison of Regular & Context-Free Languages

- All regular languages are context-free languages



Chomsky Normal Form

- When working with CFG's, it is convenient to have them in simplified form
- A CFG is in Chomsky Normal form if every rule is of the form:
  - A → BC   or   A → a
  - Where a is a terminal and A, B, C are variables and B and C cannot be the start variable
  - The rule: S → ε is also allowed for the start variable S

Chomsky Normal Form

- Theorem: Any CFL is generated by a CFG in Chomsky Normal Form
  - Ex: $S \rightarrow ASA \mid aB$
    $A \rightarrow B \mid S$
    $B \rightarrow b \mid \varepsilon$
  - Need new start variable that is not also on the right-hand side, S_0

$S_0 \rightarrow S$
$S \rightarrow ASA \mid aB$
$A \rightarrow B \mid S$
$B \rightarrow b \mid \varepsilon$

All rules are in the form:
$A \rightarrow BC$
$A \rightarrow a$

Chomsky Normal Form

- <u>Theorem</u>: Any CFL is generated by a CFG in Chomsky Normal Form

  Ex cont: $S_0 \rightarrow S$
  $S \rightarrow ASA \mid aB$
  $A \rightarrow B \mid S$
  $B \rightarrow b \mid \varepsilon$

  - Get rid of the rule $B \rightarrow \varepsilon$ (must add a rule for every B with the result of the rule $B \rightarrow \varepsilon$)

$S_0 \rightarrow S$
$S \rightarrow ASA \mid aB \mid \mathbf{a}$
$A \rightarrow B \mid S \mid \boldsymbol{\varepsilon}$
$B \rightarrow b$

All rules are in the form:
$A \rightarrow BC$
$A \rightarrow a$

Chomsky Normal Form

- <u>Theorem</u>: Any CFL is generated by a CFG in Chomsky Normal Form

Ex cont: $S_0 \rightarrow S$
$S \rightarrow ASA \mid aB \mid a$
$A \rightarrow B \mid S \mid \varepsilon$
$B \rightarrow b$

- Get rid of the rule $A \rightarrow \varepsilon$ (must add a rule for every A with the result of the rule $A \rightarrow \varepsilon$)

$S_0 \rightarrow S$
$S \rightarrow ASA \mid aB \mid a \mid \mathbf{SA} \mid \mathbf{AS} \mid \mathbf{S}$
$A \rightarrow B \mid S$
$B \rightarrow b$

All rules are in the form:
$A \rightarrow BC$
$A \rightarrow a$

Chomsky Normal Form

- Theorem: Any CFL is generated by a CFG in Chomsky Normal Form

  Ex cont: $S_0 \rightarrow S$
  $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$
  $A \rightarrow B \mid S$
  $B \rightarrow b$

  - Get rid of the rules $S \rightarrow S$ and $S_0 \rightarrow S$

    $S_0 \rightarrow$ **ASA | aB | a | SA | AS**
    $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
    $A \rightarrow B \mid S$
    $B \rightarrow b$

All rules are in the form:
$A \rightarrow BC$
$A \rightarrow a$

Chomsky Normal Form

- Theorem: Any CFL is generated by a CFG in Chomsky Normal Form

Ex cont:

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
$A \rightarrow B \mid S$
$B \rightarrow b$

- Get rid of the rules $A \rightarrow B$ and $A \rightarrow S$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
$A \rightarrow \mathbf{b \mid ASA \mid aB \mid a \mid SA \mid AS}$
$B \rightarrow b$

All rules are in the form:
$A \rightarrow BC$
$A \rightarrow a$

Chomsky Normal Form

- <u>Theorem</u>: Any CFL is generated by a CFG in Chomsky Normal Form

  Ex cont: $S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
  $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
  $A \rightarrow b \mid ASA \mid aB \mid a \mid SA \mid AS$
  $B \rightarrow b$

  - Convert the rules to the form $A \rightarrow BC$ and $A \rightarrow a$ (must add new variables, T and U)

    - $S_0 \rightarrow \mathbf{AT} \mid \mathbf{UB} \mid a \mid SA \mid AS$
      $S \rightarrow \mathbf{AT} \mid \mathbf{UB} \mid a \mid SA \mid AS$
      $A \rightarrow b \mid \mathbf{AT} \mid \mathbf{UB} \mid a \mid SA \mid AS$
      $\mathbf{T \rightarrow SA}$
      $\mathbf{U \rightarrow a}$
      $B \rightarrow b$

All rules are in the form:
$A \rightarrow BC$
$A \rightarrow a$

Chomsky Normal Form

- Second Example to convert to Chomsky Normal Form:
  - $R \rightarrow XRX \mid S$
    $S \rightarrow aTb \mid bTa$
    $T \rightarrow XTX \mid X \mid \varepsilon$
    $X \rightarrow a \mid b$

  Make all rules in the form:
  $A \rightarrow BC$
  $A \rightarrow a$

  - What steps do we need to take?
    - New start variable that does not go to itself
    - Remove the empty string transitions (the new start state can go to the empty string)
    - Remove transitions of the form $A \rightarrow B$
    - Add variables and rules to convert the rules to the correct form of $A \rightarrow BC$ or $A \rightarrow a$

Chomsky Normal Form

- Second Example to convert to Chomsky Normal Form:
  - $R \rightarrow XRX \mid S$
    $S \rightarrow aTb \mid bTa$
    $T \rightarrow XTX \mid X \mid \varepsilon$
    $X \rightarrow a \mid b$

Make all rules in the form:
$A \rightarrow BC$
$A \rightarrow a$

- Need new start variable that is not also on the right-hand side, S_0
  - **$S_0 \rightarrow R$**
    $R \rightarrow XRX \mid S$
    $S \rightarrow aTb \mid bTa$
    $T \rightarrow XTX \mid X \mid \varepsilon$
    $X \rightarrow a \mid b$

Chomsky Normal Form

- Second Example to convert to Chomsky Normal Form:
  - $S_0 \rightarrow R$
    $R \rightarrow XRX \mid S$
    $S \rightarrow aTb \mid bTa$
    $T \rightarrow XTX \mid X \mid \varepsilon$
    $X \rightarrow a \mid b$

Make all rules in the form:
$A \rightarrow BC$
$A \rightarrow a$

- Remove the rule $T \rightarrow \varepsilon$ (must add a rule for every T with the result of the rule $T \rightarrow \varepsilon$)
  - $S_0 \rightarrow R$
    $R \rightarrow XRX \mid S$
    $S \rightarrow aTb \mid bTa \mid \mathbf{ab} \mid \mathbf{ba}$
    $T \rightarrow XTX \mid X \mid \mathbf{XX}$
    $X \rightarrow a \mid b$

Chomsky Normal Form

- Second Example to convert to Chomsky Normal Form:
  - $S_0 \rightarrow R$
    $R \rightarrow XRX \mid S$
    $S \rightarrow aTb \mid bTa \mid ab \mid ba$
    $T \rightarrow XTX \mid X \mid XX$
    $X \rightarrow a \mid b$

Make all rules in the form:
$A \rightarrow BC$
$A \rightarrow a$

- Remove the rules $S_0 \rightarrow R$ and $R \rightarrow S$
  - $S_0 \rightarrow \mathbf{XRX \mid S}$
    $R \rightarrow XRX \mid \mathbf{aTb \mid bTa \mid ab \mid ba}$
    $S \rightarrow aTb \mid bTa \mid ab \mid ba$
    $T \rightarrow XTX \mid X \mid XX$
    $X \rightarrow a \mid b$

Chomsky Normal Form

- Second Example to convert to Chomsky Normal Form:
  - $S_0 \rightarrow XRX \mid S$
    $R \rightarrow XRX \mid aTb \mid bTa \mid ab \mid ba$
    $S \rightarrow aTb \mid bTa \mid ab \mid ba$
    $T \rightarrow XTX \mid X \mid XX$
    $X \rightarrow a \mid b$

Make all rules in the form:
$A \rightarrow BC$
$A \rightarrow a$

- Remove the rules $S_0 \rightarrow S$ and $T \rightarrow X$
  - $S_0 \rightarrow XRX \mid \mathbf{aTb \mid bTa \mid ab \mid ba}$
    $R \rightarrow XRX \mid aTb \mid bTa \mid ab \mid ba$
    $S \rightarrow aTb \mid bTa \mid ab \mid ba$
    $T \rightarrow XTX \mid \mathbf{a \mid b} \mid XX$
    $X \rightarrow a \mid b$

Chomsky Normal Form

- Second Example to convert to Chomsky Normal Form:
  - $S_0 \rightarrow XRX \mid aTb \mid bTa \mid ab \mid ba$
    $R \rightarrow XRX \mid aTb \mid bTa \mid ab \mid ba$
    $S \rightarrow aTb \mid bTa \mid ab \mid ba$
    $T \rightarrow XTX \mid a \mid b \mid XX$
    $X \rightarrow a \mid b$

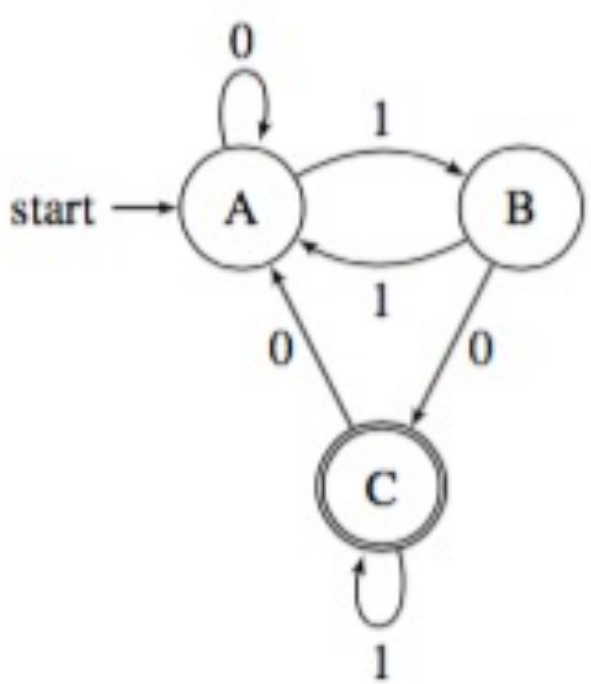
Make all rules in the form:
$A \rightarrow BC$
$A \rightarrow a$

- Convert the rules to the form $A \rightarrow BC$ and $A \rightarrow a$ (must add new variables, Y, Z, W, V, B, and A)
  - $S_0 \rightarrow YX \mid WB \mid VA \mid AB \mid BA$
    $R \rightarrow YX \mid WB \mid VA \mid AB \mid BA$
    $S \rightarrow WB \mid VA \mid AB \mid BA$
    $T \rightarrow ZX \mid a \mid b \mid XX$
    $X \rightarrow a \mid b$
    $Y \rightarrow XR$
    $Z \rightarrow XT$

$W \rightarrow AT$
$V \rightarrow BT$
$B \rightarrow b$
$A \rightarrow a$

Try It

1. Generate the CFG from the DFA given below.



2. Convert the CFG in to Chomsky Normal Form:

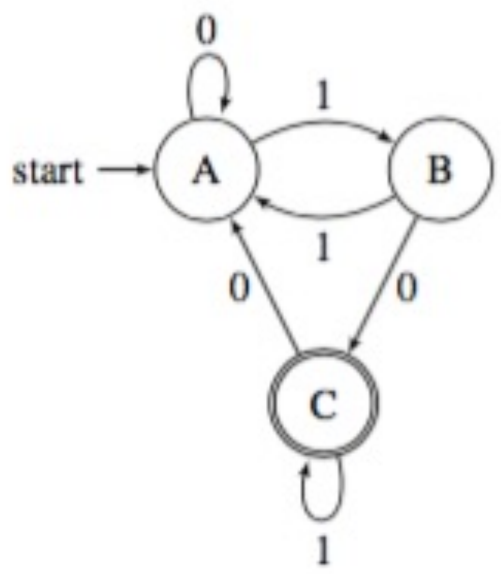
   S → TSV |VST | ε

   T → 0

   V → 1

Try It

1. Generate the CFG from the DFA given below.



- Variables A, B, C, A is the start symbol
- Productions:

  A → 0A | 1B

  B → 0C | 1A

  C → 0A | 1C | ε (ε added since C is an accept state)

Try It

2. Convert the CFG in to Chomsky Normal Form:

   $S \rightarrow TSV \mid VST \mid \varepsilon$

   $T \rightarrow 0$

   $V \rightarrow 1$

- New start state S_0:

  $\mathbf{S_0 \rightarrow S}$

  $S \rightarrow TSV \mid VST \mid \varepsilon$

  $T \rightarrow 0$

  $V \rightarrow 1$

Try It

2. Convert the CFG in to Chomsky Normal Form:

$S_0 \rightarrow S$

$S \rightarrow TSV \mid VST \mid \varepsilon$

$T \rightarrow 0$

$V \rightarrow 1$

- Get rid of ε transition:

$S_0 \rightarrow S \mid \varepsilon$

$S \rightarrow TSV \mid VST \mid \mathbf{TV} \mid \mathbf{VT}$

$T \rightarrow 0$

$V \rightarrow 1$

Note that the rule $S_0 \rightarrow \varepsilon$ is acceptable.

Try It

2. Convert the CFG in to Chomsky Normal Form:

   $S_0 \rightarrow S \mid \varepsilon$

   $S \rightarrow TSV \mid VST \mid TV \mid VT$

   $T \rightarrow 0$

   $V \rightarrow 1$

- Get rid of $S_0 \rightarrow S$ transition:

   $S_0 \rightarrow \mathbf{TSV \mid VST \mid TV \mid VT} \mid \varepsilon$

   $S \rightarrow TSV \mid VST \mid TV \mid VT$

   $T \rightarrow 0$

   $V \rightarrow 1$

Try It

2. Convert the CFG in to Chomsky Normal Form:

   $S_0 \rightarrow TSV \mid VST \mid TV \mid VT \mid \varepsilon$

   $S \rightarrow TSV \mid VST \mid TV \mid VT$

   $T \rightarrow 0$

   $V \rightarrow 1$

- Make all transitions in form of $A \rightarrow BC$ or $A \rightarrow a$:

   $S_0 \rightarrow XV \mid YT \mid TV \mid VT \mid \varepsilon$

   $S \rightarrow XV \mid YT \mid TV \mid VT$

   $T \rightarrow 0$

   $V \rightarrow 1$

   $X \rightarrow TS$

   $Y \rightarrow VS$