Quick Review

- (CFG) 4-tuple \((V, \sum, R, S)\):
  - \(V\) finite set of variables
  - \(\sum\) finite set of terminals
  - \(R\) set of rules of form:
    - variable \(\rightarrow\) (string of variables and terminals)
    - \(S \in V\), start variable
    - \(L(G) = \{ w \in \sum^* | S \rightarrow^* w \}\)
  - \(w\) is in \(\sum^*\) and can be derived from \(S\)

Example

\[
A \rightarrow 0A1 \\
A \rightarrow B \\
B \rightarrow #
\]
Chomsky Normal Form

• CNF presents a grammar in a standard, simplified form:
  
  \[ A \rightarrow BC \]
  
  \[ A \rightarrow a \]

1. where A, B, C are variables and B and C are not the start variable AND

2. \( a \) is a terminal

NOTE: The rule \( S \rightarrow \varepsilon \) (S Start Variable) is also allowed so the language can generate the empty string (optional)

Chomsky Normal Form Examples

Which CFG’s are in Chomsky Normal Form?

• \( S \rightarrow XM \)
  
  \( M \rightarrow SY \)
  
  \( X \rightarrow x \)
  
  \( Y \rightarrow y \)

• \( S \rightarrow xSy \)
  
  \( S \rightarrow xy \)
CNF Benefits

• Easier to prove statements about CFG’s when in CNF
• Any CFG can be converted to CNF
• Remove productions:
  1. A → ε to empty
  2. A → B Unit rule
  3. A → s, s contains a terminal and |s| > 1
  4. A → s, |s| > 2, s ∈ { V U ∑ }*

Step 1: Removing A → ε

Example:

S → UAV
A → ε

In CNF? If not, how to fix?
Nullable variables

- A variable A is **nullable** if $A \rightarrow^* \varepsilon$

**So we must:**
- Find all nullable variables
- Remove all $\varepsilon$ transitions
- If $T \rightarrow X_1AX_2$ and A is nullable then add $T \rightarrow X_1X_2$

---

**Step 1 Example**

$S \rightarrow TU$
$T \rightarrow AB$
$A \rightarrow aA \mid \varepsilon$
$B \rightarrow bB \mid \varepsilon$
$U \rightarrow ccA \mid B$

Nullable variables?
Productions removed?
Productions added?
Step 2: Removing A → B (Unit Productions)

A → B  Must add A → s to compensate!
B → s
S ∈ \{ V U \sum \}^*

• A variable B is **A-derivable** if A → *B
• Must:
  • Find all A-derivable variables for each A
  • Remove all unit transitions
  • If B → s and B is A-derivable then add A → s

---

Step 2 Example

S → TU | T | U  B → bB | b
T → AB | A | B  U → ccA | B | cc
A → aA | a

S-derivable:
T-derivable:
U-derivable:

NOTE: Not concerned about A, B since not involved in unit production
Step 2 Example, cont.

S → TU | T | U B → bB | b
T → AB | A | B U → ccA | B | cc
A → aA | a

Productions removed:

Productions added:

---

Step 2 Example, cont.

S → TU | T | U | ε B → bB | b
T → AB | A | B U → ccA | B | cc
A → aA | a

Productions that still violate CNF?:

---

CS 310 – Fall 2016
Pacific University
Steps 3, 4 Example, cont.

Grammar should now have NO unit productions or null productions $A \rightarrow \varepsilon$ (except $A$ may be start variable in null production)

Still need to fix productions with terminals on the RHS

Steps 3, 4: Remove $A \rightarrow S_1aS_2$

$A \rightarrow S_1aS_2$

$a \in \Sigma$, $S_1$ and $S_2$ are strings, at least one is not empty

Create

$X_a \rightarrow a$

$A \rightarrow S_1X_aS_2$

Then fix up $A \rightarrow S_1X_aS_2$

–why? what rule is violated?
–how?
Remove $A \rightarrow S_1 X_a S_2$

$A \rightarrow S_1 X_a S_2$

$A \rightarrow$

Note: In general, chain productions:

$A \rightarrow A_1 A_2 \ldots A_k$, then decompose as:

$A \rightarrow A_1 W_1, W_1 \rightarrow A_2 W_2 \ldots W_{k-1} \rightarrow A_{k-1} A_k$

---

**Step 3,4 Example, cont.**

$S \rightarrow TU | aA | A | bB | AB | ccA | cc | \epsilon$

$T \rightarrow AB | aA | a | bB | b$

$A \rightarrow aA | a$

$B \rightarrow bB | b$

$U \rightarrow ccA | cc \ bB \ b$

$A \rightarrow aA | a$

Finish!
S → ASA | aB  
A → B | S  
B → b | ε  

Put in to CNF  

Should get a grammar equivalent to:  

S₀ → AA₁ | XₐB | a | SA | AS  
S → AA₁ | XₐB | a | SA | AS  
A → b | a A₁ | XₐB | a | SA | AS  
A₁ → SA  
Xₐ → a  
B → b