# CS 301 - Lecture 10 Chomsky and Greibach **Normal Forms** Fall 2008

#### Review

- Languages and Grammars
- Alphabets, strings, languages
- Regular Languages
  - Deterministic Finite and Nondeterministic Automata
     Equivalence of NFA and DFA
     Regular Expressions

  - Regular Grammars
  - Properties of Regular Languages
- Languages that are not regular and the pumping lemma
  Context Free Languages
- - Context Free Grammars - Derivations: leftmost, rightmost and derivation trees

  - Parsing and ambiguity
    Simplifying Context Free Grammars
- Today:
   More Simplifications
  - Normal Forms

### **Nullable Variables**

 $\lambda$  – production :  $A \rightarrow \lambda$ 

 $A \Rightarrow ... \Rightarrow \lambda$ Nullable Variable:

#### Which Variables are Nullable?

0) Nullable Variables =  $V_n = \emptyset$ 

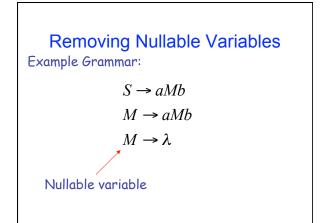
For every production  $A \to \lambda$  Add A to  $V_n$ 

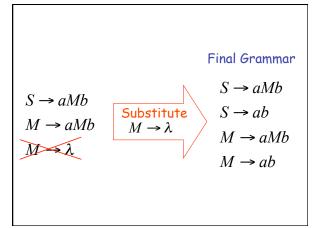
1) For every variable  $B \not\in V_n$ 

check each production  $B \to A_1 A_2 ... A_n$ 

Add B to  $V_n$  if all  $A_i \in V_n$ 

2) If step 1 added any B to  $V_n$ repeat step 1





# **Unit-Productions**

Unit Production:  $A \rightarrow B$ 

(a single variable in both sides)

# Removing Unit Productions

Observation:

$$A \rightarrow A$$

Is removed immediately

### Example Grammar:

$$S \to aA$$

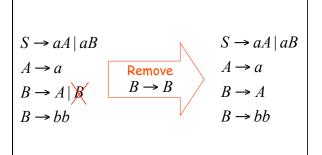
$$A \to a$$

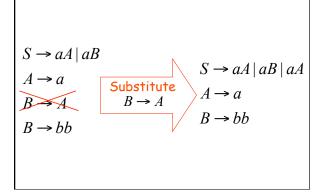
$$A \to B$$

$$B \to A$$

$$B \to bb$$

$$S \rightarrow aA$$
 $A \rightarrow a$ 
 $B \rightarrow A$ 
 $B \rightarrow bb$ 
 $S \rightarrow aA \mid aB$ 
 $A \rightarrow a$ 
 $A \rightarrow B$ 
 $B \rightarrow A \mid B$ 
 $B \rightarrow bb$ 





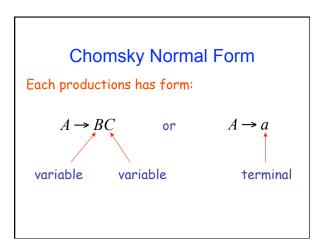
### Remove repeated productions

Final grammar 
$$S \rightarrow aA \mid aB \mid aA$$
  $S \rightarrow aA \mid aB$   $A \rightarrow a$   $A \rightarrow a$   $B \rightarrow bb$   $B \rightarrow bb$ 

# Removing All

- Step 1: Remove Nullable Variables
- Step 2: Remove Unit-Productions
- Step 3: Remove Useless Variables

Normal Forms for Context-free Grammars



### Examples:

$$S \rightarrow AS$$

$$S \rightarrow a$$

$$A \rightarrow SA$$

$$A \rightarrow b$$

$$S \rightarrow AS$$

$$S \rightarrow \widehat{AAS}$$

$$A \rightarrow SA$$

$$A \rightarrow aa$$

Not Chomsky Normal Form

 $T_c \rightarrow c$ 

# Conversion to Chomsky Normal Form

$$S \rightarrow ABa$$

• Example: 
$$A \rightarrow aab$$

$$B \rightarrow Ac$$

Not Chomsky Normal Form

# Introduce variables for terminals: $T_a, T_b, T_c$

$$S \to ABT_a$$

$$A \to T_a T_a T_b$$

$$A \to aab$$

$$B \to Ac$$

$$T_a \to a$$

$$T_b \to b$$

# Introduce intermediate variable: $V_1$

$$S \rightarrow ABT_{a}$$

$$A \rightarrow T_{a}T_{a}T_{b}$$

$$B \rightarrow AT_{c}$$

$$T_{a} \rightarrow a$$

$$T_{b} \rightarrow b$$

$$T_{c} \rightarrow c$$

$$S \rightarrow AV_{1}$$

$$V_{1} \rightarrow BT_{a}$$

$$A \rightarrow T_{a}T_{a}T_{b}$$

$$B \rightarrow AT_{c}$$

$$T_{a} \rightarrow a$$

$$T_{b} \rightarrow b$$

$$T_{c} \rightarrow c$$

Introduce intermediate variable: 
$$V_2$$

$$S \rightarrow AV_1$$

$$V_1 \rightarrow BT_a$$

$$A \rightarrow T_aT_aT_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$

$$S \rightarrow AV_1$$

$$V_1 \rightarrow BT_a$$

$$A \rightarrow T_aV_2$$

$$V_2 \rightarrow T_aT_b$$

$$B \rightarrow AT_c$$

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$

$$T_c \rightarrow c$$

Final grammar in Chomsky Normal Form: 
$$S \to AV_1$$
 
$$V_1 \to BT_a$$
 
$$A \to T_aV_2$$
 Initial grammar 
$$V_2 \to T_aT_b$$
 
$$S \to ABa$$
 
$$A \to aab$$
 
$$T_a \to a$$
 
$$T_a \to a$$
 
$$T_b \to b$$
 
$$T_c \to c$$

#### In general:

From any context-free grammar (which doesn't produce  $\lambda$ ) not in Chomsky Normal Form

we can obtain:

An equivalent grammar in Chomsky Normal Form

#### The Procedure

First remove:

Nullable variables

Unit productions

Then, for every symbol a:

Add production  $T_a \rightarrow a$ 

In productions: replace a with  $T_a$ 

New variable:  $T_a$ 

Replace any production  $A \rightarrow C_1 C_2 \cdots C_n$ 

with  $A \rightarrow C_1 V_1$   $V_1 \rightarrow C_2 V_2$ ...  $V_{n-2} \rightarrow C_{n-1} C_n$ 

New intermediate variables:  $V_1, V_2, ..., V_{n-2}$ 

**Theorem:** For any context-free grammar (which doesn't produce  $\lambda$  ) there is an equivalent grammar in Chomsky Normal Form

#### Observations

 Chomsky normal forms are good for parsing and proving theorems

• It is very easy to find the Chomsky normal form for any context-free grammar

### **Greibach Normal Form**

All productions have form:

$$A \to a V_1 V_2 \cdots V_k \qquad k \ge 0$$
 symbol variables

### Examples:

$$S \rightarrow cAB$$
  
 $A \rightarrow aA \mid bB \mid b$   
 $B \rightarrow b$   
 $S \rightarrow abSb$   
 $S \rightarrow aa$ 

Greibach
Normal Form
Normal Form

#### Conversion to Greibach Normal Form:

$$S \rightarrow abSb$$
 $S \rightarrow aa$ 

$$S \rightarrow aT_bST_b$$
 $S \rightarrow aT_a$ 

$$T_a \rightarrow a$$

$$T_b \rightarrow b$$
Greibach
Normal Form

Theorem: For any context-free grammar (which doesn't produce  $\lambda$ ) there is an equivalent grammar in Greibach Normal Form

#### Observations

- Greibach normal forms are very good for parsing
- It is hard to find the Greibach normal form of any context-free grammar

### What's Next

- Read
- Linz Chapter 1,2.1, 2.2, 2.3, (skip 2.4), 3, 4, 5, 6.1, 6.2, (skip 6.3), and 7.1
- JFLAP Chapter 1, 2.1, (skip 2.2), 3, 4, 5, 6.1, 7
- Next Lecture Topics from Chapter 7.1
  - Nondeterminstic Pushdown Automata
- Quiz 2 in Recitation on Wednesday 10/1
  - Covers Linz 2, 3, 4 and JFLAP 3, 4
  - Closed book, but you may bring one sheet of 8.5 x 11 inch paper with any notes you like.
  - Quiz will take the full hour
- Homework
  - Homework Due Today
  - New Homework Available Friday Morning
  - New Homework Due Next Thursday