# 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: $A \rightarrow \lambda$

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

## Which Variables are Nullable?

0) Nullable Variables = $V_n = \emptyset$

For every production $A \rightarrow \lambda$

Add $A$ to $V_n$

1) For every variable $B \notin V_n$

check each production $B \rightarrow A_1 A_2 \ldots A_n$

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

## Removing Nullable Variables

Example Grammar:

$$S \to aMb$$
$$M \to aMb$$
$$M \to \lambda$$

Nullable variable

$S \to aMb$

$M \to aMb$

$\cancel{M \to \lambda}$

Substitute $M \to \lambda$

Final Grammar

$$S \to aMb$$
$$S \to ab$$
$$M \to aMb$$
$$M \to ab$$

## Unit-Productions

Unit Production: $A \to B$

(a single variable in both sides)

## Removing Unit Productions

Observation:

$$A \to A$$

Is removed immediately

Example Grammar:

$$S \to aA$$
$$A \to a$$
$$A \to B$$
$$B \to A$$
$$B \to bb$$

---

$$S \to aA$$
$$A \to a$$
$$\cancel{A \to B}$$
$$B \to A$$
$$B \to bb$$

Substitute $A \to B$

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

---

$$S \to aA \mid aB$$
$$A \to a$$
$$B \to A \mid \cancel{B}$$
$$B \to bb$$

Remove $B \to B$

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

---

$$S \to aA \mid aB$$
$$A \to a$$
$$\cancel{B \to A}$$
$$B \to bb$$

Substitute $B \to A$

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

Remove repeated productions

Final grammar

$$S \to aA \mid aB \mid \cancel{aA} \qquad \Rightarrow \qquad S \to aA \mid aB$$
$$A \to a \qquad\qquad\qquad A \to a$$
$$B \to bb \qquad\qquad\qquad B \to bb$$

## Removing All

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

# Normal Forms for Context-free Grammars

## Chomsky Normal Form

Each productions has form:

$$A \to \underbrace{B}_{\text{variable}}\underbrace{C}_{\text{variable}} \qquad \text{or} \qquad A \to \underbrace{a}_{\text{terminal}}$$

Examples:

$S \to AS$
$S \to a$
$A \to SA$
$A \to b$

Chomsky
Normal Form

$S \to AS$
$S \to AAS$
$A \to SA$
$A \to aa$

Not Chomsky
Normal Form

## Conversion to Chomsky Normal Form

- Example:

$S \to ABa$
$A \to aab$
$B \to Ac$

Not Chomsky
Normal Form

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

$S \to ABa$
$A \to aab$
$B \to Ac$

$\Longrightarrow$

$S \to ABT_a$
$A \to T_aT_aT_b$
$B \to AT_c$
$T_a \to a$
$T_b \to b$
$T_c \to c$

Introduce intermediate variable: $V_1$

$S \to ABT_a$
$A \to T_aT_aT_b$
$B \to AT_c$
$T_a \to a$
$T_b \to b$
$T_c \to c$

$\Longrightarrow$

$S \to AV_1$
$V_1 \to BT_a$
$A \to T_aT_aT_b$
$B \to AT_c$
$T_a \to a$
$T_b \to b$
$T_c \to c$

Introduce intermediate variable: $V_2$

$$S \to AV_1$$
$$V_1 \to BT_a$$
$$A \to T_aT_aT_b$$
$$B \to AT_c$$
$$T_a \to a$$
$$T_b \to b$$
$$T_c \to c$$

$$S \to AV_1$$
$$V_1 \to BT_a$$
$$A \to T_aV_2$$
$$V_2 \to T_aT_b$$
$$B \to AT_c$$
$$T_a \to a$$
$$T_b \to b$$
$$T_c \to c$$

Initial grammar

$$S \to ABa$$
$$A \to aab$$
$$B \to Ac$$

Final grammar in Chomsky Normal Form:

$$S \to AV_1$$
$$V_1 \to BT_a$$
$$A \to T_aV_2$$
$$V_2 \to T_aT_b$$
$$B \to AT_c$$
$$T_a \to a$$
$$T_b \to b$$
$$T_c \to c$$

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

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

**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:

$$A \to a\, V_1 V_2 \cdots V_k \qquad k \geq 0$$

symbol variables

Examples:

$$S \to cAB$$
$$A \to aA \mid bB \mid b$$
$$B \to b$$

Greibach Normal Form

$$S \to abSb$$
$$S \to aa$$

Not Greibach Normal Form

Conversion to Greibach Normal Form:

$$S \to abSb$$
$$S \to aa$$

$$S \to aT_bST_b$$
$$S \to aT_a$$
$$T_a \to a$$
$$T_b \to b$$

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