Simplification of CFG

BITS Pilani
Pilani Campus

Shashank Gupta
Assistant Professor
Department of Computer Science and Information Systems

Simplification of CFG

CFG is often required in its simplest form.

One of the simplest and most useful forms is called as the Chomsky Normal Form

Chomsky Normal Form (CNF)

A context-free grammar is in Chomsky normal form if every rule is of the form

- $A \rightarrow BC$
- $A \rightarrow a$

where a is any terminal and A, B, and C are any variables

- Except that B and C may not be the start variable. In addition, we permit the rule $S \rightarrow \in$, where S is the start variable.

Properties of CNF Grammar

CNF produces the same language as generated by CFG.

For generating string w of length ‘x’ requires ‘2x-1’ derivation steps in CNF.

Chomsky Normal Form

Every CFL has a Context-Free Grammar G.

We can convert any grammar G into Chomsky normal form.

- The conversion has several stages wherein rules that violate the conditions are replaced with equivalent ones that are satisfactory.

Conversion: **Step 1**

First, we add a new start variable S_0 and the rule $S_0 \rightarrow S$, where S was the original start variable.

This change guarantees that the start variable doesn't occur on the right-hand side of a rule.

Conversion: **Step 2**

Remove an $\in$-rule $A \rightarrow \in$, where A is not the start variable.

- Then, for every occurrence of an A on the right-hand side of a rule, add a new rule with that occurrence deleted.
- Consider the following Grammar
- $A \rightarrow \in$
- $B \rightarrow uAv$
- $C \rightarrow u_1Au_2Au_3$

$\mathbf{B \rightarrow uAv \mid uv}$

$\mathbf{C \rightarrow u_1Au_2Au_3 \mid u_1u_2Au_3 \mid u_1Au_2u_3 \mid u_1u_2u_3}$

Conversion: **Step 3**

Remove a unit rule $A \rightarrow B$. Then, whenever a rule $B \rightarrow u$ appears, we add the rule $A \rightarrow u$.

We repeat these steps until we eliminate all the unit rules.

Conversion: **Step 3** (Continued….)

Suppose a grammar had the following rules:

$A \rightarrow B$

$B \rightarrow u$

Then the grammar formed by removing the rule $A \rightarrow B$ will have the corresponding set of rules

$B \rightarrow u$

$A \rightarrow u$ (new rule added)

Conversion: **Step 4**

We replace each rule $A \rightarrow u_1u_2 \cdot \cdot \cdot u_k$, (where $k \geq 3$ and each u_i is a variable or terminal symbol) with the following rules

- $A \rightarrow u_1A_1$,
- $A_1 \rightarrow u_2A_2$,
- $A_2 \rightarrow u_3A_3$,
- .
- .
- . ,
- and $A_{k-2} \rightarrow u_{k-1}u_k$.

The A_i's are new variables added to the Grammar We replace any terminal u_i in the preceding rule(s) with the new variable u_i and add the rule $U_i \rightarrow u_i$.

Conversion: **Step 5**

Replacing few terminal symbols on the RHS.

If there is a rule of the form A → uv where at least one of either u or v is a terminal symbol (let's assume u), then replace the rule A → uv with

- **A → Uv**
- **U → u (new rule added)**
- where U is a new variable added to the grammar. Repeat until no such rules remain.

Example

CFG

$S \rightarrow ASA \mid aB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \in$

Step 1: Add a new start symbol

$S_0 \rightarrow S$ (New Rule Added)

$S \rightarrow ASA \mid aB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \in$

Example (Continued……)

Step 2: Remove $\in$ Productions

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \in$

Step 2a: Removing $B \rightarrow \in$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid \mathbf{a}$

$A \rightarrow B \mid S \mid \in$

$B \rightarrow b$

Example (Continued……)

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a$

$A \rightarrow B \mid S \mid \in$

$B \rightarrow b$

Step 2b: Removing $A \rightarrow \in$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid \mathbf{SA} \mid \mathbf{AS}$

$A \rightarrow B \mid S$

$B \rightarrow b$

Example (Continued……)

Step 3: Remove Unit Rules

$S_0 \rightarrow S$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

Step 3a: Removing $S_0 \rightarrow S$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$A \rightarrow B \mid S$

$B \rightarrow b$

Example (Continued……)

Remove other Unit Rules

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$\mathbf{A \rightarrow B} \mid S$

$B \rightarrow b$

Step 3b: Removing $\mathbf{A \rightarrow B}$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$\mathbf{A} \rightarrow S \mid \mathbf{b}$

$B \rightarrow b$

Example (Continued……)

Remove other Unit Rules

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$\mathbf{A \rightarrow S} \mid b$

$B \rightarrow b$

Step 3c: Removing $\mathbf{A \rightarrow S}$

$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

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

$B \rightarrow b$

Example (Continued……)

Step 4: Shortening the RHS

$S_0 \rightarrow$ **ASA** | aB | a | SA | AS

$S \rightarrow$ **ASA** | aB | a | SA | AS

$A \rightarrow b$ | **ASA** | aB | a | SA | AS

$B \rightarrow b$

Step 4: Shortening the RHS

$S_0 \rightarrow$ AC | aB | a | SA | AS

$S \rightarrow$ AC | aB | a | SA | AS

$A \rightarrow b$ | AC | aB | a | SA | AS

$B \rightarrow b$

$C \rightarrow SA$

Example (Continued……)

Step 5: Replacing few Terminals on RHS

$S_0 \rightarrow AC \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AC \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AC \mid aB \mid a \mid SA \mid AS$

$B \rightarrow b$

$C \rightarrow SA$

Step 5: Replacing few Terminals on RHS

$S_0 \rightarrow AC \mid DB \mid a \mid SA \mid AS$

$S \rightarrow AC \mid DB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AC \mid DB \mid a \mid SA \mid AS$

$B \rightarrow b$

$C \rightarrow SA$

$D \rightarrow a$

Resultant CFG is in CNF

More Examples

Consider the following CFG and convert it into CNF CFG.

$S \rightarrow aXbX$

$X \rightarrow aY \mid bY \mid \in$

$Y \rightarrow X \mid c$

Example (Continued….)

Given CFG

$S \rightarrow aXbX$

$X \rightarrow aY \mid bY \mid \in$

$Y \rightarrow X \mid c$

Step 1 can be skipped

$S \rightarrow aXbX$

$X \rightarrow aY \mid bY \mid \in$

$Y \rightarrow X \mid c$

Example (Continued….)

Step 2: Remove ∈ Productions

S → aXbX

X → aY | bY | ∈

Y → X | c

Step:2a Removing X →∈

S → aXbX | **abX | aXb | ab**

X → aY | bY

Y → X | c | ∈

Example (Continued….)

$S \rightarrow aXbX \mid abX \mid aXb \mid ab$

$X \rightarrow aY \mid bY$

$Y \rightarrow X \mid c \mid \in$

Step:2b Removing $Y \rightarrow \in$

$S \rightarrow aXbX \mid abX \mid aXb \mid ab$

$X \rightarrow aY \mid bY \mid \mathbf{a \mid b}$

$Y \rightarrow X \mid c$

Example (Continued….)

Step 3: Remove Unit Rules

S → aXbX | abX | aXb | ab

X → aY | bY | a | b

Y → X | c

Step3: Removing Y → X

S → aXbX | abX | aXb | ab

X → aY | bY | a | b

Y → c | **aY | bY | a | b**

Example (Continued….)

Step 4: Replacing certain terminals

S → aXbX | abX | aXb | ab

X → aY | bY | a | b

Y → c | aY | bY | a | b

Step 4: Replacing certain terminals

S → AXBX | ABX | AXB | AB

X → AY | BY | a | b

Y → c | AY | BY | a | b

A → a

B → a

Example (Continued….)

Step 5: Shortening RHS

S → AXBX | ABX | AXB | AB

X → AY | BY | a | b

Y → c | AY | BY | a | b

A → a

B → a

Step 5: Shortening RHS

S → PQ | AQ | PB | AB

X → AY | BY | a | b

Y → c | AY | BY | a | b

A → a

B → a

P → AX

Q → BX