Lecture 18 – Normal Forms of Context-Free Grammars

COSE215: Theory of Computation

Jihyeok Park

PLRG

2025 Spring

Recall

- A **context-free grammar (CFG)** is a 4-tuple:

$$G = (V, \Sigma, S, R)$$

  where

  - V: a finite set of **variables** (nonterminals)
  - Σ: a finite set of **symbols** (terminals)
  - $S \in V$: the **start variable**
  - $R \subseteq V \times (V \cup \Sigma)^*$: a set of **production rules**.

- How to **simplify** a CFG?

Recall

- A **context-free grammar (CFG)** is a 4-tuple:

$$G = (V, \Sigma, S, R)$$

  where

  - V: a finite set of **variables** (nonterminals)
  - Σ: a finite set of **symbols** (terminals)
  - $S \in V$: the **start variable**
  - $R \subseteq V \times (V \cup \Sigma)^*$: a set of **production rules**.

- How to **simplify** a CFG?

  Let's put it in **Chomsky normal form (CNF)**!

Contents

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
   - Nullable Variables

3. Eliminating Unit Productions
   - Unit Pairs

4. Eliminating Useless Variables
   - Generating Variables
   - Reachable Variables

5. Putting CFG in CNF

Contents

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
   Nullable Variables

3. Eliminating Unit Productions
   Unit Pairs

4. Eliminating Useless Variables
   Generating Variables
   Reachable Variables

5. Putting CFG in CNF

Chomsky Normal Form (CNF)

Definition (Chomsky Normal Form)

A CFG G is in **Chomsky normal form (CNF)** if all productions are of the form for some $A, B, C \in V$ and $a \in \Sigma$:

$$A \to BC \qquad \text{OR} \qquad A \to a \qquad \text{OR} \qquad S \to \epsilon$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

Chomsky Normal Form (CNF)

Definition (Chomsky Normal Form)

A CFG G is in **Chomsky normal form (CNF)** if all productions are of the form for some $A, B, C \in V$ and $a \in \Sigma$:

$$A \to BC \qquad \text{OR} \qquad A \to a \qquad \text{OR} \qquad S \to \epsilon$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

Consider the following CFG:

$$\begin{array}{lll} S \to 0ABC \mid 1B \mid BB & A \to ABB0 \mid C & C \to CC \mid \epsilon \\ & B \to 0B \mid 1 & D \to 1D \mid AA \end{array}$$

Is it possible to put this CFG in CNF?

Chomsky Normal Form (CNF)

Definition (Chomsky Normal Form)

A CFG G is in **Chomsky normal form (CNF)** if all productions are of the form for some $A, B, C \in V$ and $a \in \Sigma$:

$$A \to BC \qquad \text{OR} \qquad A \to a \qquad \text{OR} \qquad S \to \epsilon$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

Consider the following CFG:

$$\begin{array}{lll} S \to 0ABC \mid 1B \mid BB & A \to ABB0 \mid C & C \to CC \mid \epsilon \\ & B \to 0B \mid 1 & D \to 1D \mid AA \end{array}$$

Is it possible to put this CFG in CNF? **Yes!**

$$\begin{array}{lll} S \to XS_1 \mid XB \mid YB \mid BB & A \to AA_1 \mid BA_2 & B \to XB \mid 1 \\ S_1 \to AB & A_1 \to BA_2 & X \to 0 \\ & A_2 \to BX & Y \to 1 \end{array}$$

Chomsky Normal Form (CNF)

Definition (Chomsky Normal Form)

A CFG G is in **Chomsky normal form (CNF)** if all productions are of the form for some $A, B, C \in V$ and $a \in \Sigma$:

$$A \to BC \qquad \text{OR} \qquad A \to a \qquad \text{OR} \qquad S \to \epsilon$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

Consider the following CFG:

$$\begin{array}{lll} S \to 0ABC \mid 1B \mid BB & A \to ABB0 \mid C & C \to CC \mid \epsilon \\ & B \to 0B \mid 1 & D \to 1D \mid AA \end{array}$$

Is it possible to put this CFG in CNF? **Yes!**

$$\begin{array}{lll} S \to XS_1 \mid XB \mid YB \mid BB & A \to AA_1 \mid BA_2 & B \to XB \mid 1 \\ S_1 \to AB & A_1 \to BA_2 & X \to 0 \\ & A_2 \to BX & Y \to 1 \end{array}$$

Let's learn how to put a CFG in CNF!

Contents

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
   Nullable Variables

3. Eliminating Unit Productions
   Unit Pairs

4. Eliminating Useless Variables
   Generating Variables
   Reachable Variables

5. Putting CFG in CNF

Eliminating ϵ-Productions

The following productions are called ϵ-**productions**:

$$A \to \epsilon$$

Is it possible to eliminate all ϵ-**productions** from a CFG?

Eliminating ϵ-Productions

The following productions are called ϵ-**productions**:

$$A \to \epsilon$$

Is it possible to eliminate all ϵ-**productions** from a CFG?

No, if an empty word ϵ is in the language of the CFG (i.e., $\epsilon \in L(G)$), then we cannot generate the empty word without ϵ-productions.

Eliminating ϵ-Productions

The following productions are called ϵ-**productions**:

$$A \to \epsilon$$

Is it possible to eliminate all ϵ-**productions** from a CFG?

No, if an empty word ϵ is in the language of the CFG (i.e., $\epsilon \in L(G)$), then we cannot generate the empty word without ϵ-productions.

However, we can eliminate all ϵ-productions from a CFG G to construct a new CFG G' such that:

$$L(G') = L(G) \setminus \{\epsilon\}$$

Eliminating ϵ-Productions

The following productions are called ϵ-**productions**:

$$A \to \epsilon$$

Is it possible to eliminate all ϵ-**productions** from a CFG?

No, if an empty word ϵ is in the language of the CFG (i.e., $\epsilon \in L(G)$), then we cannot generate the empty word without ϵ-productions.

However, we can eliminate all ϵ-productions from a CFG G to construct a new CFG G' such that:

$$L(G') = L(G) \setminus \{\epsilon\}$$

We can do it by following the steps below:

1. Find all **nullable variables**.
2. Construct a new CFG by **replacing** nullable variables with ϵ in **all combinations** and **removing** all ϵ-productions in production rules.

Nullable Variables

Definition (Nullable Variables)

For a given CFG $G = (V, \Sigma, S, R)$, a variable $A \in V$ is **nullable** if

$$A \Rightarrow^* \epsilon$$

Nullable Variables

Definition (Nullable Variables)

For a given CFG $G = (V, \Sigma, S, R)$, a variable $A \in V$ is **nullable** if

$$A \Rightarrow^* \epsilon$$

We can inductively define the set of **nullable variables**:

- **(Basis Case)** If $A \to \epsilon \in R$, then A is nullable.
- **(Induction Case)** If $A \to X_1 X_2 \cdots X_n \in R$ and $X_1, X_2, \ldots, X_n$ are all nullable, then A is nullable.

Eliminating ϵ-Productions – Example

Consider the following CFG:

$$
\begin{aligned}
S &\rightarrow 0ABC \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid C \\
B &\rightarrow 0B \mid 1 \\
C &\rightarrow CC \mid \epsilon \\
D &\rightarrow 1D \mid AA
\end{aligned}
$$

1. Find all **nullable variables**:

Eliminating ϵ-Productions – Example

Consider the following CFG:

$$
\begin{aligned}
S &\rightarrow 0ABC \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid C \\
B &\rightarrow 0B \mid 1 \\
C &\rightarrow CC \mid \epsilon \\
D &\rightarrow 1D \mid AA
\end{aligned}
$$

1. Find all **nullable variables**: $\{A, C, D\}$

Eliminating ϵ-Productions – Example

Consider the following CFG:

$$
\begin{aligned}
S &\to 0ABC \mid 1B \mid BB \\
A &\to ABB0 \mid C \\
B &\to 0B \mid 1 \\
C &\to CC \mid \epsilon \\
D &\to 1D \mid AA
\end{aligned}
$$

1. Find all **nullable variables**: $\{A, C, D\}$
2. Construct a new CFG by **replacing** nullable variables with ϵ in **all combinations** and **removing** all ϵ-productions in production rules:

Eliminating ϵ-Productions – Example

Consider the following CFG:

$$
\begin{aligned}
S &\to 0ABC \mid 1B \mid BB \\
A &\to ABB0 \mid C \\
B &\to 0B \mid 1 \\
C &\to CC \mid \epsilon \\
D &\to 1D \mid AA
\end{aligned}
$$

1. Find all **nullable variables**: $\{A, C, D\}$
2. Construct a new CFG by **replacing** nullable variables with ϵ in **all combinations** and **removing** all ϵ-productions in production rules:

$$
\begin{aligned}
S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \mid C \\
B &\to 0B \mid 1 \\
C &\to CC \mid C \\
D &\to 1D \mid 1 \mid AA \mid A
\end{aligned}
$$

Contents

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
   - Nullable Variables

3. Eliminating Unit Productions
   - Unit Pairs

4. Eliminating Useless Variables
   - Generating Variables
   - Reachable Variables

5. Putting CFG in CNF

Eliminating Unit Productions

The following productions are called **unit productions**:

$$A \to B$$

Eliminating Unit Productions

The following productions are called **unit productions**:

$$A \to B$$

Is it possible to eliminate **unit productions**?

Eliminating Unit Productions

The following productions are called **unit productions**:

$$A \to B$$

Is it possible to eliminate **unit productions**?

Yes, we can do it by following the steps below:

1. Find all **unit pairs**.
2. Construct a new CFG by **adding** all possible non-unit productions of B to A for each unit pair (A, B).

Unit Pairs

Definition (Unit Pairs)

For a given CFG $G = (V, \Sigma, S, R)$, a pair of variables $(A, B) \in V \times V$ is a **unit pair** if

$$A \Rightarrow^* B$$

Unit Pairs

Definition (Unit Pairs)

For a given CFG $G = (V, \Sigma, S, R)$, a pair of variables $(A, B) \in V \times V$ is a **unit pair** if

$$A \Rightarrow^* B$$

We can inductively define the set of **unit pairs**:

- **(Basis Case)** (A, A) is a unit pair for all $A \in V$.
- **(Induction Case)** If (A, B) is a unit pair and $B \to C \in R$, then (A, C) is a unit pair.

Eliminating Unit Productions – Example

After eliminating ϵ-productions:

$$
\begin{aligned}
S &\rightarrow 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid BB0 \mid C \\
B &\rightarrow 0B \mid 1 \\
C &\rightarrow CC \mid C \\
D &\rightarrow 1D \mid 1 \mid AA \mid A
\end{aligned}
$$

1. Find all **unit pairs**:

Eliminating Unit Productions – Example

After eliminating ϵ-productions:

$$\begin{aligned} S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\ A &\to ABB0 \mid BB0 \mid C \\ B &\to 0B \mid 1 \\ C &\to CC \mid C \\ D &\to 1D \mid 1 \mid AA \mid A \end{aligned}$$

1. Find all **unit pairs**:

$$\{(S,S),(A,A),(A,C),(B,B),(C,C),(D,D),(D,A),(D,C)\}$$

Eliminating Unit Productions – Example

After eliminating ϵ-productions:

$$\begin{aligned} S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\ A &\to ABB0 \mid BB0 \mid C \\ B &\to 0B \mid 1 \\ C &\to CC \mid C \\ D &\to 1D \mid 1 \mid AA \mid A \end{aligned}$$

1. Find all **unit pairs**:

$$\{(S,S),(A,A),(A,C),(B,B),(C,C),(D,D),(D,A),(D,C)\}$$

2. Construct a new CFG by **adding** all possible non-unit productions of B to A for each unit pair (A, B).

Eliminating Unit Productions – Example

After eliminating ϵ-productions:

$$\begin{aligned} S &\rightarrow 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\ A &\rightarrow ABB0 \mid BB0 \mid C \\ B &\rightarrow 0B \mid 1 \\ C &\rightarrow CC \mid C \\ D &\rightarrow 1D \mid 1 \mid AA \mid A \end{aligned}$$

1. Find all **unit pairs**:

$$\{(S,S),(A,A),(A,C),(B,B),(C,C),(D,D),(D,A),(D,C)\}$$

2. Construct a new CFG by **adding** all possible non-unit productions of B to A for each unit pair (A,B).

$$\begin{aligned} S &\rightarrow 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\ A &\rightarrow ABB0 \mid BB0 \mid CC \\ B &\rightarrow 0B \mid 1 \\ C &\rightarrow CC \\ D &\rightarrow 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC \end{aligned}$$

Contents

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
    Nullable Variables

3. Eliminating Unit Productions
    Unit Pairs

4. Eliminating Useless Variables
    Generating Variables
    Reachable Variables

5. Putting CFG in CNF

Eliminating Useless Variables

What are useless variables?

- **Non-generating variables**: Variables that cannot derive any word.
- **Unreachable variables**: Variables unreachable from the start variable.

Is it possible to eliminate **useless variables**?

Eliminating Useless Variables

What are useless variables?

- **Non-generating variables**: Variables that cannot derive any word.
- **Unreachable variables**: Variables unreachable from the start variable.

Is it possible to eliminate **useless variables**?

Yes, we can do it by following the steps below:

1. Find all **generating variables**.
2. Find all **reachable variables**.
3. Construct a new CFG by **removing** all productions that contain non-generating variables or come from unreachable variables.

Generating Variables

Definition (Generating Variables)

For a given CFG $G = (V, \Sigma, S, R)$, a variable $A \in V$ is a **generating variable** if for some $w \in \Sigma^*$,

$$A \Rightarrow^* w$$

Generating Variables

Definition (Generating Variables)

For a given CFG $G = (V, \Sigma, S, R)$, a variable $A \in V$ is a **generating variable** if for some $w \in \Sigma^*$,

$$A \Rightarrow^* w$$

We can inductively define the set of **generating variables**:

- **(Basis Case)** There is no basis case.
- **(Induction Case)** If $A \to \alpha \in R$ and α contains only symbols or generating variables, then A is a generating variable.

Reachable Variables

Definition (Reachable Variables)

For a given CFG $G = (V, \Sigma, S, R)$, a variable $A \in V$ is a **reachable variable** if there exists a derivation:

$$S \Rightarrow^* \alpha A \beta$$

Reachable Variables

Definition (Reachable Variables)

For a given CFG $G = (V, \Sigma, S, R)$, a variable $A \in V$ is a **reachable variable** if there exists a derivation:

$$S \Rightarrow^* \alpha A \beta$$

We can inductively define the set of **reachable variables**:

- **(Basis Case)** The start variable S is reachable variable.
- **(Induction Case)** If $A \in V$ is a reachable variable and $A \rightarrow \alpha \in R$, then all variables in α are reachable variables.

Eliminating Useless Variables – Example

After eliminating ϵ-productions and unit productions:

$$
\begin{aligned}
S &\rightarrow 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid BB0 \mid CC \\
B &\rightarrow 0B \mid 1 \\
C &\rightarrow CC \\
D &\rightarrow 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC
\end{aligned}
$$

1. Find all **generating variables**:

Eliminating Useless Variables – Example

After eliminating ϵ-productions and unit productions:

$$
\begin{aligned}
S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \mid CC \\
B &\to 0B \mid 1 \\
C &\to CC \\
D &\to 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC
\end{aligned}
$$

1. Find all **generating variables**: $\{S, A, B, D\}$ – C is non-generating.

Eliminating Useless Variables – Example

After eliminating ϵ-productions and unit productions:

$$
\begin{aligned}
S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \mid CC \\
B &\to 0B \mid 1 \\
C &\to CC \\
D &\to 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC
\end{aligned}
$$

1. Find all **generating variables**: $\{S, A, B, D\}$ – C is non-generating.
2. Find all **reachable variables**:

Eliminating Useless Variables – Example

After eliminating ϵ-productions and unit productions:

$$\begin{aligned}
S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \mid CC \\
B &\to 0B \mid 1 \\
C &\to CC \\
D &\to 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC
\end{aligned}$$

1. Find all **generating variables**: $\{S, A, B, D\}$ – C is non-generating.
2. Find all **reachable variables**: $\{S, A, B, C\}$ – D is unreachable.

Eliminating Useless Variables – Example

After eliminating ϵ-productions and unit productions:

$$
\begin{aligned}
S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \mid CC \\
B &\to 0B \mid 1 \\
C &\to CC \\
D &\to 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC
\end{aligned}
$$

1. Find all **generating variables**: $\{S, A, B, D\}$ – C is non-generating.
2. Find all **reachable variables**: $\{S, A, B, C\}$ – D is unreachable.
3. Construct a new CFG by **removing** all productions that contain non-generating variables or come from unreachable variables.

Eliminating Useless Variables – Example

After eliminating ϵ-productions and unit productions:

$$\begin{aligned}
S &\to 0ABC \mid 0BC \mid 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \mid CC \\
B &\to 0B \mid 1 \\
C &\to CC \\
D &\to 1D \mid 1 \mid AA \mid ABB0 \mid BB0 \mid CC
\end{aligned}$$

1. Find all **generating variables**: $\{S, A, B, D\}$ – C is non-generating.
2. Find all **reachable variables**: $\{S, A, B, C\}$ – D is unreachable.
3. Construct a new CFG by **removing** all productions that contain non-generating variables or come from unreachable variables.

$$\begin{aligned}
S &\to 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \\
B &\to 0B \mid 1
\end{aligned}$$

Contents

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
   Nullable Variables

3. Eliminating Unit Productions
   Unit Pairs

4. Eliminating Useless Variables
   Generating Variables
   Reachable Variables

5. Putting CFG in CNF

Putting CFG in CNF

Our goal is to put a CFG in **Chomsky normal form (CNF)** consisting of:

$$A \to BC \qquad \text{OR} \qquad A \to a$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

We can put a CFG in CNF by following the steps below:

Putting CFG in CNF

Our goal is to put a CFG in **Chomsky normal form (CNF)** consisting of:

$$A \to BC \qquad \text{OR} \qquad A \to a$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

We can put a CFG in CNF by following the steps below:

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

Putting CFG in CNF

Our goal is to put a CFG in **Chomsky normal form (CNF)** consisting of:

$$A \to BC \qquad \text{OR} \qquad A \to a$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

We can put a CFG in CNF by following the steps below:

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.
2. Eliminate ϵ-productions, unit productions, and useless variables.

Putting CFG in CNF

Our goal is to put a CFG in **Chomsky normal form (CNF)** consisting of:

$$A \to BC \qquad \text{OR} \qquad A \to a$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

We can put a CFG in CNF by following the steps below:

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.
2. Eliminate ϵ-productions, unit productions, and useless variables.
3. Rewrite all RHSs whose length > 1 to contain only variables: if a symbol a a appears in the RHS, replace it with a new variable A and introduce a new production rule $A \to a$.

Putting CFG in CNF

Our goal is to put a CFG in **Chomsky normal form (CNF)** consisting of:

$$A \to BC \qquad \text{OR} \qquad A \to a$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

We can put a CFG in CNF by following the steps below:

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.
2. Eliminate ϵ-productions, unit productions, and useless variables.
3. Rewrite all RHSs whose length > 1 to contain only variables: if a symbol a a appears in the RHS, replace it with a new variable A and introduce a new production rule $A \to a$.
4. Replace all RHSs whose length is greater than 2 with a chain of variables. To do so, if $A \to X_1 X_2 \cdots X_n$ is a production with $n > 2$, then replace it with a sequence of productions:

$$A \to X_1 A_1 \qquad A_1 \to X_2 A_2 \qquad \cdots \qquad A_{n-2} \to X_{n-1} X_n$$

Putting CFG in CNF

Our goal is to put a CFG in **Chomsky normal form (CNF)** consisting of:

$$A \to BC \qquad \text{OR} \qquad A \to a$$

where $B \neq S$ and $C \neq S$. And $S \to \epsilon$ is allowed only if $\epsilon \in L(G)$.

We can put a CFG in CNF by following the steps below:

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.
2. Eliminate ϵ-productions, unit productions, and useless variables.
3. Rewrite all RHSs whose length > 1 to contain only variables: if a symbol a a appears in the RHS, replace it with a new variable A and introduce a new production rule $A \to a$.
4. Replace all RHSs whose length is greater than 2 with a chain of variables. To do so, if $A \to X_1 X_2 \cdots X_n$ is a production with $n > 2$, then replace it with a sequence of productions:

$$A \to X_1 A_1 \qquad A_1 \to X_2 A_2 \qquad \cdots \qquad A_{n-2} \to X_{n-1} X_n$$

5. If ϵ is in the original language, add a production $S \to \epsilon$ (or $S' \to \epsilon$).

Putting CFG in CNF – Example 1

Let's put the following CFG in CNF:

$$
\begin{aligned}
S &\rightarrow 0ABC \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid C \\
B &\rightarrow 0B \mid 1 \\
C &\rightarrow CC \mid \epsilon \\
D &\rightarrow 1D \mid AA
\end{aligned}
$$

Putting CFG in CNF – Example 1

Let's put the following CFG in CNF:

$$
\begin{aligned}
S &\rightarrow 0ABC \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid C \\
B &\rightarrow 0B \mid 1 \\
C &\rightarrow CC \mid \epsilon \\
D &\rightarrow 1D \mid AA
\end{aligned}
$$

1. If S on RHSs, add a new start variable S' and a production $S' \rightarrow S$.

Putting CFG in CNF – Example 1

Let's put the following CFG in CNF:

$$\begin{aligned} S &\to 0ABC \mid 1B \mid BB \\ A &\to ABB0 \mid C \\ B &\to 0B \mid 1 \\ C &\to CC \mid \epsilon \\ D &\to 1D \mid AA \end{aligned}$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.
2. Eliminate ϵ-productions, unit productions, and useless variables:

$$\begin{aligned} S &\to 0AB \mid 0B \mid 1B \mid BB \\ A &\to ABB0 \mid BB0 \\ B &\to 0B \mid 1 \end{aligned}$$

Putting CFG in CNF – Example 1

$$
\begin{aligned}
S &\to 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \\
B &\to 0B \mid 1
\end{aligned}
$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

Putting CFG in CNF – Example 1

$$S \to 0AB \mid 0B \mid 1B \mid BB$$
$$A \to ABB0 \mid BB0$$
$$B \to 0B \mid 1$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$\begin{array}{ll} S \to XAB \mid XB \mid YB \mid BB & X \to 0 \\ A \to ABBX \mid BBX & Y \to 1 \\ B \to XB \mid 1 & \end{array}$$

Putting CFG in CNF – Example 1

$$
\begin{aligned}
S &\to 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \\
B &\to 0B \mid 1
\end{aligned}
$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$
\begin{aligned}
S &\to XAB \mid XB \mid YB \mid BB & X &\to 0 \\
A &\to ABBX \mid BBX & Y &\to 1 \\
B &\to XB \mid 1
\end{aligned}
$$

4. Replace all RHSs whose length > 2 with a chain of variables:

Putting CFG in CNF – Example 1

$$
\begin{aligned}
S &\to 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \\
B &\to 0B \mid 1
\end{aligned}
$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$
\begin{aligned}
S &\to XAB \mid XB \mid YB \mid BB & X &\to 0 \\
A &\to ABBX \mid BBX & Y &\to 1 \\
B &\to XB \mid 1
\end{aligned}
$$

4. Replace all RHSs whose length > 2 with a chain of variables:

$$
\begin{aligned}
S &\to XS_1 \mid XB \mid YB \mid BB & A &\to AA_1 \mid BA_2 & B &\to XB \mid 1 \\
S_1 &\to AB & A_1 &\to BA_2 & X &\to 0 \\
& & A_2 &\to BX & Y &\to 1
\end{aligned}
$$

Putting CFG in CNF – Example 1

$$
\begin{aligned}
S &\to 0AB \mid 0B \mid 1B \mid BB \\
A &\to ABB0 \mid BB0 \\
B &\to 0B \mid 1
\end{aligned}
$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$
\begin{aligned}
S &\to XAB \mid XB \mid YB \mid BB & X &\to 0 \\
A &\to ABBX \mid BBX & Y &\to 1 \\
B &\to XB \mid 1
\end{aligned}
$$

4. Replace all RHSs whose length > 2 with a chain of variables:

$$
\begin{aligned}
S &\to XS_1 \mid XB \mid YB \mid BB & A &\to AA_1 \mid BA_2 & B &\to XB \mid 1 \\
S_1 &\to AB & A_1 &\to BA_2 & X &\to 0 \\
& & A_2 &\to BX & Y &\to 1
\end{aligned}
$$

5. If ϵ is in the original language, add a production $S \to \epsilon$:

Putting CFG in CNF – Example 1

$$
\begin{aligned}
S &\rightarrow 0AB \mid 0B \mid 1B \mid BB \\
A &\rightarrow ABB0 \mid BB0 \\
B &\rightarrow 0B \mid 1
\end{aligned}
$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$
\begin{aligned}
S &\rightarrow XAB \mid XB \mid YB \mid BB & X &\rightarrow 0 \\
A &\rightarrow ABBX \mid BBX & Y &\rightarrow 1 \\
B &\rightarrow XB \mid 1
\end{aligned}
$$

4. Replace all RHSs whose length > 2 with a chain of variables:

$$
\begin{aligned}
S &\rightarrow XS_1 \mid XB \mid YB \mid BB & A &\rightarrow AA_1 \mid BA_2 & B &\rightarrow XB \mid 1 \\
S_1 &\rightarrow AB & A_1 &\rightarrow BA_2 & X &\rightarrow 0 \\
& & A_2 &\rightarrow BX & Y &\rightarrow 1
\end{aligned}
$$

5. If ϵ is in the original language, add a production $S \rightarrow \epsilon$: **No.**

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$S' \to ASB \mid AB \qquad S \to ASB \mid AB \qquad A \to a \qquad B \to b$$

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$S' \to ASB \mid AB \qquad S \to ASB \mid AB \qquad A \to a \qquad B \to b$$

4. Replace all RHSs whose length > 2 with a chain of variables:

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$S' \to ASB \mid AB \qquad S \to ASB \mid AB \qquad A \to a \qquad B \to b$$

4. Replace all RHSs whose length > 2 with a chain of variables:

$$S' \to AS_1 \mid AB \quad S \to AS_1 \mid AB \quad S_1 \to SB \quad A \to a \quad B \to b$$

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$S' \to ASB \mid AB \qquad S \to ASB \mid AB \qquad A \to a \qquad B \to b$$

4. Replace all RHSs whose length > 2 with a chain of variables:

$$S' \to AS_1 \mid AB \quad S \to AS_1 \mid AB \quad S_1 \to SB \quad A \to a \quad B \to b$$

5. If ϵ is in the original language, add a production $S' \to \epsilon$:

Putting CFG in CNF – Example 2

Let's put the following CFG in CNF:

$$S \to aSb \mid \epsilon$$

1. If S on RHSs, add a new start variable S' and a production $S' \to S$.

$$S' \to S \qquad S \to aSb \mid \epsilon$$

2. Eliminate ϵ-productions, unit productions, and useless variables:

$$S' \to aSb \mid ab \qquad S \to aSb \mid ab$$

3. Rewrite all RHSs whose length > 1 to contain only variables:

$$S' \to ASB \mid AB \qquad S \to ASB \mid AB \qquad A \to a \qquad B \to b$$

4. Replace all RHSs whose length > 2 with a chain of variables:

$$S' \to AS_1 \mid AB \quad S \to AS_1 \mid AB \quad S_1 \to SB \quad A \to a \quad B \to b$$

5. If ϵ is in the original language, add a production $S' \to \epsilon$: **Yes.**

$$S' \to \epsilon \mid AS_1 \mid AB \quad S \to AS_1 \mid AB \quad S_1 \to SB \quad A \to a \quad B \to b$$

Summary

1. Chomsky Normal Form (CNF)

2. Eliminating ϵ-Productions
   - Nullable Variables

3. Eliminating Unit Productions
   - Unit Pairs

4. Eliminating Useless Variables
   - Generating Variables
   - Reachable Variables

5. Putting CFG in CNF

Next Lecture

- Properties of Context-Free Languages

Jihyeok Park
jihyeok_park@korea.ac.kr
https://plrg.korea.ac.kr