Context-Free Grammars

1. *Formal Definition*
2. *Construction*
3. *Parse Tree*
4. *Ambiguity*
5. *Simplification of CFG*
6. *CNF & GNF*

English Grammar

$\langle \text{sentence} \rangle \rightarrow \langle \text{noun\_phrase} \rangle \langle \text{predicate} \rangle$

$\langle \text{noun\_phrase} \rangle \rightarrow \langle \text{article} \rangle \langle \text{noun} \rangle$

$\langle \text{predicate} \rangle \rightarrow \langle \text{verb} \rangle$

$\langle \text{article} \rangle \rightarrow \langle \text{a} \rangle \mid \langle \text{an} \rangle \mid \langle \text{the} \rangle$

$\langle \text{noun} \rangle \rightarrow \langle \text{boy} \rangle \mid \langle \text{dog} \rangle$

$\langle \text{verb} \rangle \rightarrow \langle \text{runs} \rangle \mid \langle \text{walks} \rangle$

a boy runs   a dog walks

Context-Free Grammar

A grammar G=(V, T, S, P) is said to be context-free if all productions in P have the form

$$A \rightarrow \alpha \ , \ \text{where} \ A \in V, \alpha \in (V \cup T)^*$$

Palindrome Language

$$L=\{\, w \mid w \in \{0,1\}^* \text{ and } w = w^R \,\}$$

- recursive definition
  - *basis* ε, 0, 1 are palindromes.
  - *induction* If w is a palindrome, so is $0w0$ and $1w1$.

Palindrome Language

$$L=\{ w \mid w \in \{0,1\}^* \text{ and } w = w^R \}$$

- definition with grammars or rules

1. ε is a palindrome.
2. 0 is a palindrome.
3. 1 is a palindrome.
4. If w is a palindrome, so is $0w0$.
5. If w is a palindrome, so is $1w1$.

CFG & Palindrome Language

$L = \{ w \mid w \in \{0,1\}^* \text{ and } w = w^R \}$

1. ε is a P.
2. 0 is a P.
3. 1 is a P.
4. If w is a P, so is $0w0$.
5. If w is a P, so is $1w1$.

1. $P \to \varepsilon$
2. $P \to 0$
3. $P \to 1$
4. $P \to 0P0$
5. $P \to 1P1$

CFG of Palindrome Language

$$L = \{ w \mid w \in \{0,1\}^* \text{ and } w = w^R \}$$

- CFG for palindromes on $\{0,1\}$

  $R = (\{S\}, \{0,1\}, P, S)$, P is defined as follow

  $S \to \varepsilon$, $S \to 0$, $S \to 1$, $S \to 0S0$, $S \to 1S1$

Compact notation

$$S \to \varepsilon \mid 0 \mid 1 \mid 0S0 \mid 1S1$$

Example 7.1

$L = \{ 0^n 1^n \mid n \geq 0 \}$

$R = (\{S\}, \{0,1\}, P, S)$, P is defined as follow

$S \rightarrow \varepsilon \mid 0S1$

Example 7.2

$L=\{ 0^n 1^m \mid n \neq m \}$

$R = (\{S,A,B,C\}, \{0,1\}, P, S)$, P is defined as follow

$S \to AC \mid CB, \; C \to 0C1 \mid \varepsilon$

$A \to A0 \mid 0 , \; B \to 1B \mid 1$

Example 7.3

L={ $w \in \{0,1\}^*$ | w contains same number of 0's and 1's }

R = ({S}, {0,1}, P, S), P is defined as follow

$S \to \varepsilon \mid 0S1 \mid 1S0 \mid SS$

Example 7.4

$L=\{w \in \{0,1\}^* \mid n_0(w)=n_1(w)$ and $n_0(v) \geq n_1(v)$

where v is any prefix of w $\}$

$R = (\{S\}, \{0,1\}, P, S)$, P is defined as follow

$S \rightarrow \varepsilon \mid 0S1 \mid SS$

Derivations and Recursive Inferences

$L=\{a^{2n}b^m \mid n \geq 0, m \geq 0\}$

$R = (\{S, A, B\}, \{a,b\}, P, S)$, P is defined as follow

$S \rightarrow AB$, $A \rightarrow \varepsilon | aaA$, $B \rightarrow \varepsilon | Bb$

for $w = aabb$:

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaABb \Rightarrow aaBb \Rightarrow aaBbb \Rightarrow aabb$$

$S \rightarrow AB$ $\quad$ $A \rightarrow aaA$ $\quad$ $B \rightarrow Bb$ $\quad$ $A \rightarrow \varepsilon$ $\quad$ $B \rightarrow Bb$ $\quad$ $B \rightarrow \varepsilon$

Context-Free Language

Let G=(V, T, S, P) be context-free, then

$$L(G) = \{w \mid w \in T^* \text{ and } S \stackrel{*}{\Rightarrow} w\}$$

Left/Right Most Derivations

$L=\{a^{2n}b^{m} \mid n \geq 0, m \geq 0\}$

$S \rightarrow AB$, $A \rightarrow \varepsilon | aaA$, $B \rightarrow \varepsilon | Bb$

for $w = aabb$:

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaABb \Rightarrow aaBb \Rightarrow aaBbb \Rightarrow aabb$$

Left most :

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaB \Rightarrow aaBb \Rightarrow aaBbb \Rightarrow aabb$$

Right most :

$$S \Rightarrow AB \Rightarrow ABb \Rightarrow ABbb \Rightarrow Abb \Rightarrow aaAbb \Rightarrow aabb$$

Parse Tree

Let $G = (V, T, S, P)$ be a CFG. A tree is a parse tree for G if :

1. Each interior node is labeled by a variable in V

2. Each leaf is labeled by a symbol in $T \cup \{\varepsilon\}$. Any ε-labeled leaf is the only child of its parent.

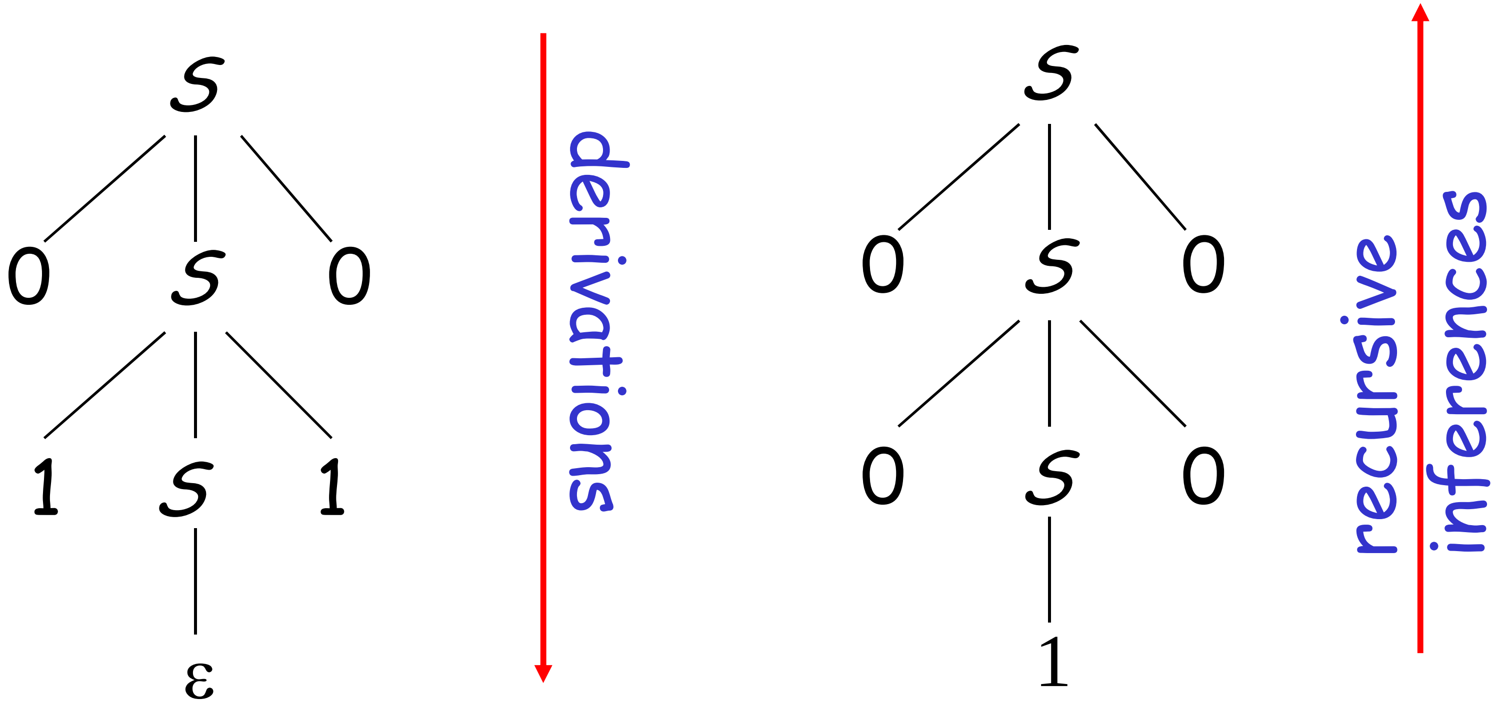
3. If an interior node is labeled A, and its children (from left to right) labeled $x_1, x_2, \ldots, x_k$,

Then $A \to x_1, x_2, \ldots, x_k \in P$.

Parse Tree

Example 7.5 L={ w | $w \in \{0,1\}^*$ and $w = w^R$ }

$$S \rightarrow \varepsilon \mid 0 \mid 1 \mid 0S0 \mid 1S1$$



w=0110

w=00100

Ambiguity

$G = (\{E, I\}, \{a, b, (,), +, *\}, P, E)$

$E \to I \mid E + E \mid E * E \mid (E), \quad I \to a \mid b$

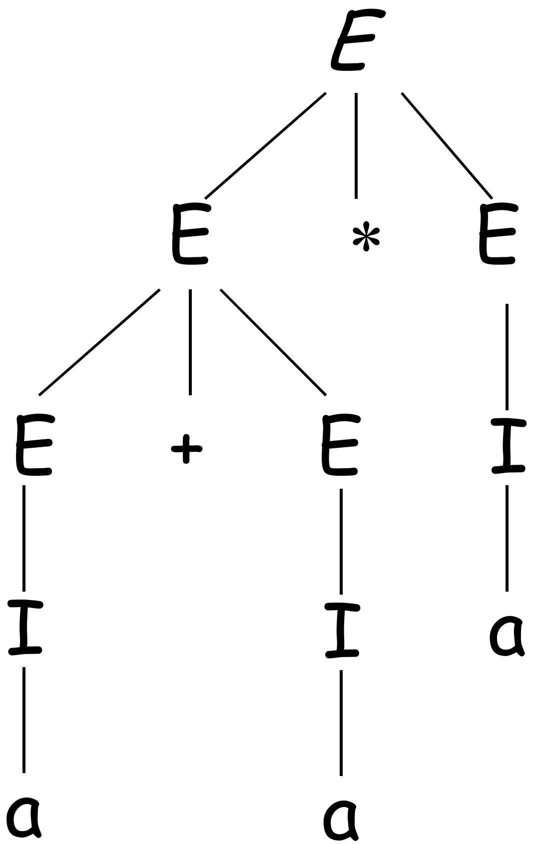
Derivation for $w = a + a * a$:

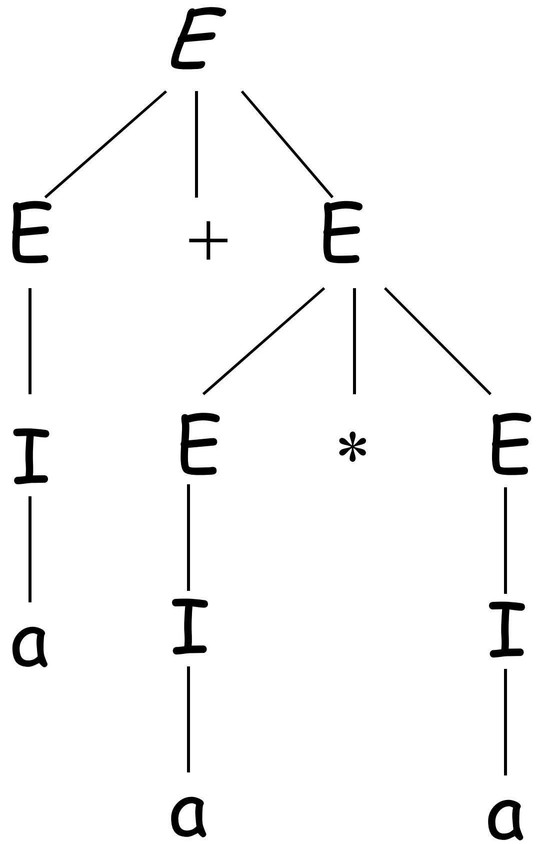
$E \Rightarrow E*E \Rightarrow E+E*E \Rightarrow I+E*E \Rightarrow a+E*E \overset{*}{\Rightarrow} a+a*a$

$E \Rightarrow E+E \Rightarrow I+E \Rightarrow a+E \Rightarrow a+E*E \overset{*}{\Rightarrow} a+a*a$

Ambiguity

parse-tree for $w = a + a * a$:





Removing Ambiguity

$E \to I \mid E+E \mid E*E \mid (E)\ ,\ \ I \to a \mid b$

$E \to T \mid E+T,\ T \to F \mid T*F,\ F \to I \mid (E)\ ,\ I \to a \mid b \mid Ia \mid Ib$

Left most derivation for $w = a + a * a$:

$E \Rightarrow E+T \Rightarrow T+T \Rightarrow F+T \Rightarrow I+T \Rightarrow a+T \Rightarrow a+T*F$

$\Rightarrow a+F*F \Rightarrow a+I*F \Rightarrow a+a*F \Rightarrow a+a*I \Rightarrow a+a*a$

$E \Rightarrow T \Rightarrow T*T \Rightarrow (E)*T \Rightarrow (E+T)*T \overset{*}{\Rightarrow} (a+a)*a$

Inherent Ambiguity

- What is inherent ambiguity

A CFL L is said to be *inherently ambiguous* if every grammar that generates it is ambiguous.

Example 7.6 Let $L=\{ w \mid w \in \{0,1\}^* \text{ and } n_0(w)=n_1(w) \}$

L is not inherently ambiguous ,because there is an unambiguous CFG :

$S \to \varepsilon \mid 0S1 \mid 1S0 \mid 0S11S0 \mid 1S00S1$

Example 7.7

$L=\{a^n b^n c^m d^m \mid n\geq 1, m\geq 1\} \cup \{a^n b^m c^m d^n \mid n\geq 1, m\geq 1\}$

The CFG for L is :

$S \rightarrow AB \mid C$, $A \rightarrow aAb \mid ab$, $B \rightarrow cBd \mid cd$

$C \rightarrow aCd \mid aDd$, $D \rightarrow bDc \mid bc$

Let w= *abcd* , there are two left most derivations

$S \Rightarrow AB \Rightarrow abB \Rightarrow abcd$

$S \Rightarrow C \Rightarrow aDd \Rightarrow abcd$

Simplification of CFG

Why & what :

$S \to A \mid B,\ A \to 1CA \mid 1DE \mid \varepsilon,\ B \to 1CB \mid 1DF,$

$C \to 1CC \mid 1DG \mid 0G,\ D \to 1CD \mid 1DH \mid 0H,$

$E \to 0A,\ F \to 0B,\ G \to \phi,\ H \to 1$

- ε-productions
- unit productions
- useless symbols and productions

Eliminating ε-productions

Variable A is said to be nullable if $A \overset{*}{\Rightarrow} \varepsilon$.

Let $G=(V,T,P,S)$ is a CFG

If $A \rightarrow \varepsilon \in P$, then A is nullable.

If $A \rightarrow A_1 A_2 \ldots A_k \in P$, and $A_i \rightarrow \varepsilon \in P$ for $i=1, \ldots, k$

then A is nullable.

Example 7.8 $G: S \rightarrow AB,\ A \rightarrow aAA|\varepsilon,\ B \rightarrow bBB|\varepsilon$

$$\left.\begin{array}{l} A \rightarrow \varepsilon \Rightarrow A \text{ is nullable.} \\ B \rightarrow \varepsilon \Rightarrow B \text{ is nullable.} \end{array}\right\} S \rightarrow AB \Rightarrow S \text{ is nullable.}$$

Eliminating unit productions

Example 7.9 $G: S \to A|B|0S1,\ A \to 0A|0,\ B \to 1B|1$

$S \to 0A|0|1B|1|0S1$

$A \to 0A|0$

$B \to 1B|1$

Eliminating useless productions

A symbol X is useful for a grammar $G=(V,T,P,S)$,

if there is a derivation for some $w \in T^*$

$$S \stackrel{*}{\Rightarrow} \alpha X \beta \stackrel{*}{\Rightarrow} w$$

A symbol X is generating if $X \stackrel{*}{\Rightarrow} w$ for some $w \in T^*$

A symbol X is reachable if $S \stackrel{*}{\Rightarrow} \alpha X \beta$ for $\{\alpha,\beta\} \subseteq (V \cup T)^*$

Example 7.10 $G: S \to AB|a,\ A \to b$

S and A are generating , B is not.

Eliminate B, that eliminate $S \to AB$, leaving

$$S \to a,\ A \to b$$

Now only S is reachable. So there leaves $S \to a$.

If eliminate non-reachable symbol first :

$$S \to AB|a,\ A \to b \Rightarrow S \to AB|a,\ A \to b$$

Then eliminate non-generating symbol :

$$S \to AB|a,\ A \to b \Rightarrow S \to a,\ A \to b$$

Example 7.11 $G: S \to A \mid B,\ A \to 1CA \mid 1DE \mid \varepsilon$

$B \to 1CB \mid 1DF,\ C \to 1CC \mid 1DG \mid 0G,\ D \to 1CD \mid 1DH \mid 0H,$

$E \to 0A,\ F \to 0B,\ G \to \phi,\ H \to 1$

- eliminating ε-productions

  the only one : $A \to \varepsilon$

  $S \to A \mid B,\ A \to 1CA \mid 1C \mid 1DE,\ B \to 1CB \mid 1DF,$

  $C \to 1CC \mid 1DG \mid 0G,\ D \to 1CD \mid 1DH \mid 0H,$

  $E \to 0A \mid 0,\ F \to 0B,\ G \to \phi,\ H \to 1$

$S \to A \mid B,\ A \to 1CA \mid 1C \mid 1DE,\ B \to 1CB \mid 1DF,$

$C \to 1CC \mid 1DG \mid 0G,\ D \to 1CD \mid 1DH \mid 0H,$

$E \to 0A \mid 0,\ F \to 0B,\ G \to \phi,\ H \to 1$

- eliminating unit productions

the only two : $S \to A$ and $S \to B$

$S \to 1CA \mid 1C \mid 1DE \mid 1CB \mid 1DF,$

$A \to 1CA \mid 1C \mid 1DE,\ B \to 1CB \mid 1DF,$

$C \to 1CC \mid 1DG \mid 0G,\ D \to 1CD \mid 1DH \mid 0H,$

$E \to 0A \mid 0,\ F \to 0B,\ G \to \phi,\ H \to 1$

$S \rightarrow 1CA \mid 1C \mid 1DE \mid 1CB \mid 1DF$,

$A \rightarrow 1CA \mid 1C \mid 1DE$, $B \rightarrow 1CB \mid 1DF$,

$C \rightarrow 1CC \mid 1DG \mid 0G$, $D \rightarrow 1CD \mid 1DH \mid 0H$,

$E \rightarrow 0A \mid 0$, $F \rightarrow 0B$, $G \rightarrow \phi$, $H \rightarrow 1$

- eliminating useless symbols and productions

$S \rightarrow 1DE$,$A \rightarrow 1DE$, $D \rightarrow 1DH \mid 0H$, $E \rightarrow 0A \mid 0$, $H \rightarrow 1$

Chomsky Normal Form(CNF)

1. $A \to BC$;
2. $A \to a$.

$S \to 1DE$,$A \to 1DE$, $D \to 1DH \mid 0H$, $E \to 0A \mid 0$, $H \to 1$

Chomsky normal form :

$S \to IE$,$A \to IE$, $D \to IH|EH$, $E \to EA|0$, $I \to HD$, $H \to 1$

$D \to IH|FH$, $E \to FA|0$, $F \to 0$

Chomsky Normal Form(CNF)

Example 7.12 Convert following grammar to CNF

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

Greibach Normal Form(GNF)

$A \to ax$, where $a \in T$, $x \in V^*$

Example 7.13 Convert following grammar to GNF

$S \to AB$, $A \to aA|bB|b$, $B \to b$

Example 7.14 Convert following grammar to GNF

$S \to 01S1|00$

- ? ? ?
  - eliminating ε-productions : $\varepsilon \in L$?
  - Greibach normal form :
    - $A \to a\alpha$ *advantage ?*
  - Chomsky normal form :
    - $A \to a \mid BC$ *advantage ?*
  - left recursiveness
    - $A \to A\alpha$ *shortage ?*