Morning

Context-Free Grammars

- Formal definition
- Construction
- Parse tree
- Simplication

English Grammar

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

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

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

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

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

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

a boy runs   a dog walks

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 $1w1$.

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. $P \to 1P1$

Context-Free Grammar

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

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

CFG for Palindrome Language

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

CFG for palindromes on {0,1}

$R = (\{S\}, \{0,1\}, S, P)$, 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 1 CFG for

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

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

$S \rightarrow \varepsilon \mid 0S1$

ε

0 1

0 0 1 1

0 0 0 1 1 1

0 0 0 0 1 1 1 1

Example 2 CFG for

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

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

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

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

$$n \neq m \Rightarrow \begin{cases} n > m \Rightarrow n = (n-m) + m \\ n < m \Rightarrow m = n + (m-n) \end{cases}$$

Example 3 CFG for

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

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

Example 4 CFG for

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

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

Example 5 CFG for

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

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

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

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 $w = aabb$:

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

$S \rightarrow AB$

$A \rightarrow aaA$

$B \rightarrow Bb$

$A \rightarrow \varepsilon$

$B \rightarrow Bb$

$B \rightarrow \varepsilon$

Context-Free Language

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

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

Left Most Derivations

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

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

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.

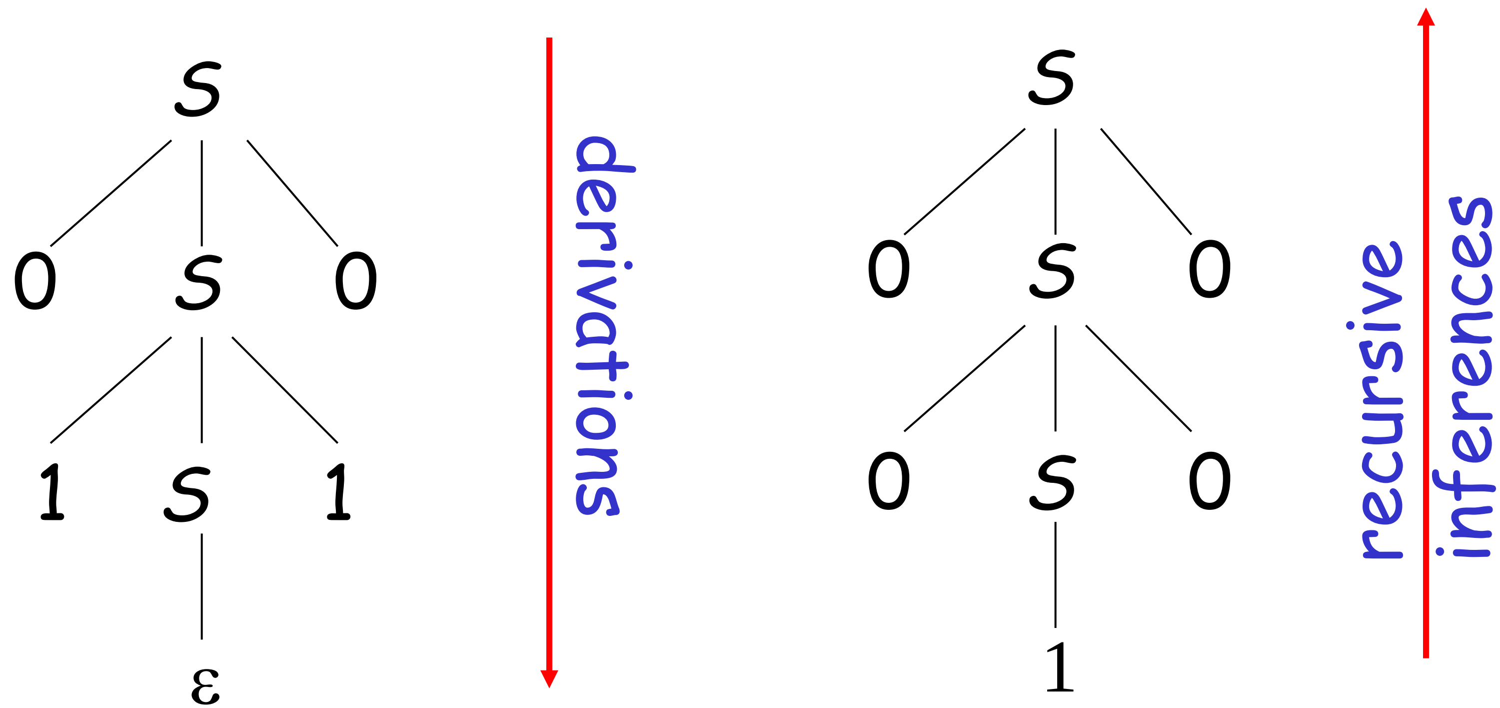
3. If an interior node is labeled A, and its children (from left to right) labeled $x_1, x_2, \ldots, x_k$,

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

Example 6

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

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



w=0110

w=00100

Ambiguity

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

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

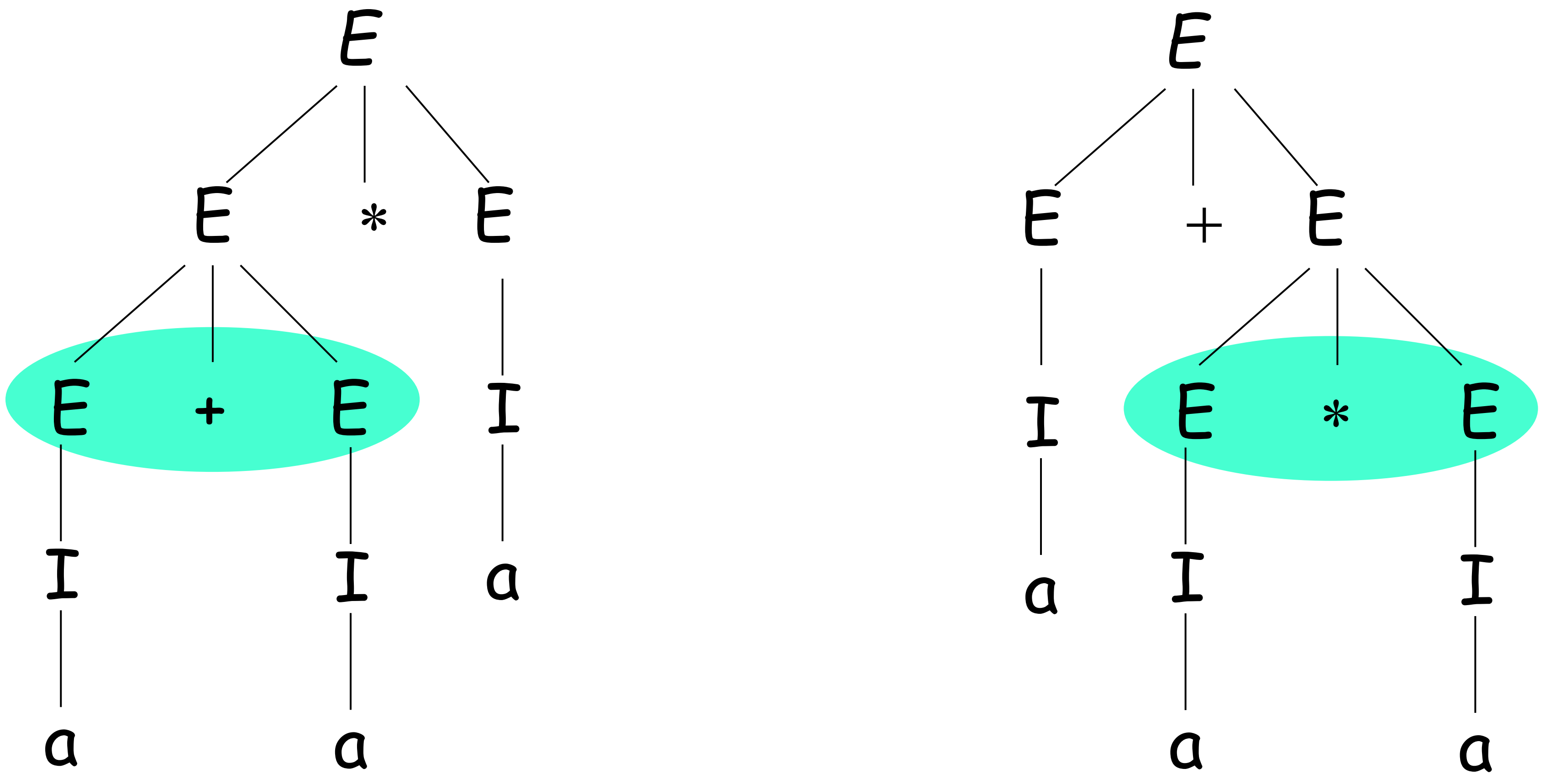
Derivation for $w = a + a * a$:

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

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

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 all grammars that generate it is ambiguous.

Example 7

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 :

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

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

ambiguity

Example 8

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

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

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, ...,k

then A is nullable.

Example 9 ε-production

$G: S \to AB,\ A \to aAA \mid \varepsilon,\ B \to bBB \mid \varepsilon$

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

Example 10 unit productions

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

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

$A \to 0A|0$

$B \to 1B|1$

Useless productions

For a grammar $G=(V,T,P,S)$, a symbol X is

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

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

generating, if $\alpha X \beta \overset{*}{\Rightarrow} w$ for some $w \in T^*$

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

Example 11 Useless productions

$G: S \to AB \mid a,\ A \to b.$

useless

non-genrating

$S \Rightarrow a$ or $S \Rightarrow AB \Rightarrow bB \Rightarrow ?$

non-reachable

$G: S \to a,\ A \to b.$

$G: S \to a$

Example 12 Simplify CFG

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

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

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

- eliminating ε-productions : $A \to \varepsilon$

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

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

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

Example 12 Simplify CFG

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

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

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

- eliminating unit productions : $S \to A,\ S \to B$

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

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

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

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

Example 12 Simplify CFG

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

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

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

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

- eliminating useless productions

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

Chomsky Normal Form

All productions are one of following two forms :

1. $A \rightarrow BC$, $A, B, C \in V$

2. $A \rightarrow a$, $a \in T$

Example 13

Convert following CFG into CNF

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

Greibach Normal Form/GNF

All productions are shown as following form :

$$A \rightarrow ax \text{ , where } a \in T \text{ , } x \in V^*$$

Example 14

Convert following grammar to GNF

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

Example 15

Convert following grammar to GNF

$S \rightarrow 01S1 | 00$

Discussion

- eliminating ε-productions : $\varepsilon \in L$?
- Chomsky normal form

  $A \to a \mid BC$ *advantage ?*
- Greibach normal form

  $A \to a\alpha$ *advantage ?*

Good good study

day day up !