Formal Languages
Context-Free Languages

$$\{a^n b^n : n \geq 0\} \qquad \{ww^R\}$$

Regular Languages

$$a*b* \qquad (a+b)*$$
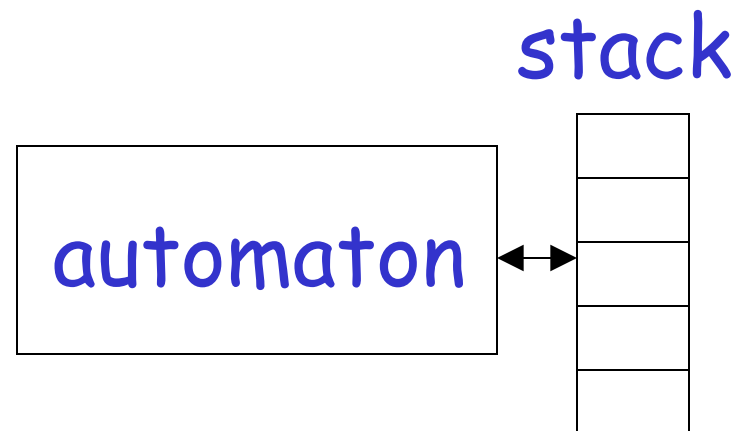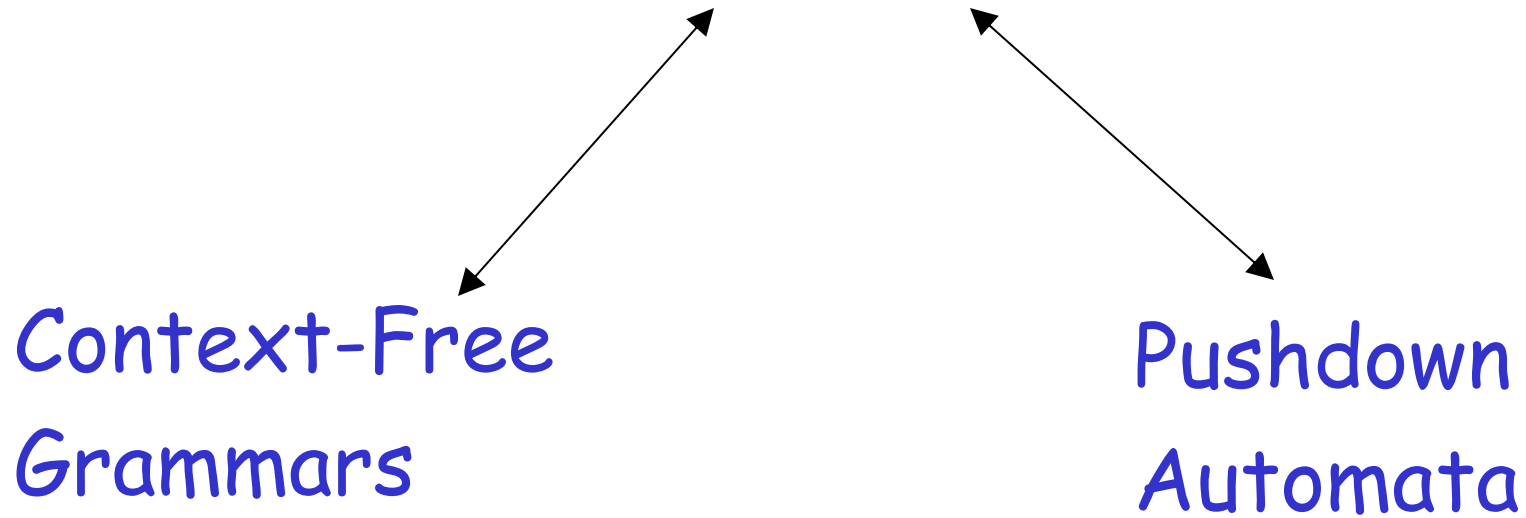
Context-Free Languages

$\{a^n b^n\}$

$\{ww^R\}$

Regular Languages

Context-Free Languages

Context-Free
Grammars

Pushdown
Automata

stack

automaton

Context-Free Grammars

Grammars

Grammars express languages

Example: the English language

$$\langle sentence \rangle \rightarrow \langle noun\_phrase \rangle \ \langle predicate \rangle$$

$$\langle noun\_phrase \rangle \rightarrow \langle article \rangle \ \langle noun \rangle$$

$$\langle predicate \rangle \rightarrow \langle verb \rangle$$

$$\langle article \rangle \rightarrow a$$

$$\langle article \rangle \rightarrow the$$

$$\langle noun \rangle \rightarrow cat$$

$$\langle noun \rangle \rightarrow dog$$

$$\langle verb \rangle \rightarrow runs$$

$$\langle verb \rangle \rightarrow walks$$

A derivation of "the dog walks":

$$\langle sentence \rangle \Rightarrow \langle noun\_phrase \rangle \ \langle predicate \rangle$$

$$\Rightarrow \langle noun\_phrase \rangle \ \langle verb \rangle$$

$$\Rightarrow \langle article \rangle \ \langle noun \rangle \ \langle verb \rangle$$

$$\Rightarrow the \ \langle noun \rangle \ \langle verb \rangle$$

$$\Rightarrow the \ dog \ \langle verb \rangle$$

$$\Rightarrow the \ dog \ walks$$

A derivation of "a cat runs":

$$\langle sentence \rangle \Rightarrow \langle noun\_phrase \rangle \; \langle predicate \rangle$$

$$\Rightarrow \langle noun\_phrase \rangle \; \langle verb \rangle$$

$$\Rightarrow \langle article \rangle \; \langle noun \rangle \; \langle verb \rangle$$

$$\Rightarrow a \; \langle noun \rangle \; \langle verb \rangle$$

$$\Rightarrow a \; cat \; \langle verb \rangle$$

$$\Rightarrow a \; cat \; runs$$

Language of the grammar:

L = { "a cat runs",
"a cat walks",
"the cat runs",
"the cat walks",
"a dog runs",
"a dog walks",
"the dog runs",
"the dog walks" }

Notation

Production Rules

$$\langle noun \rangle \rightarrow cat$$
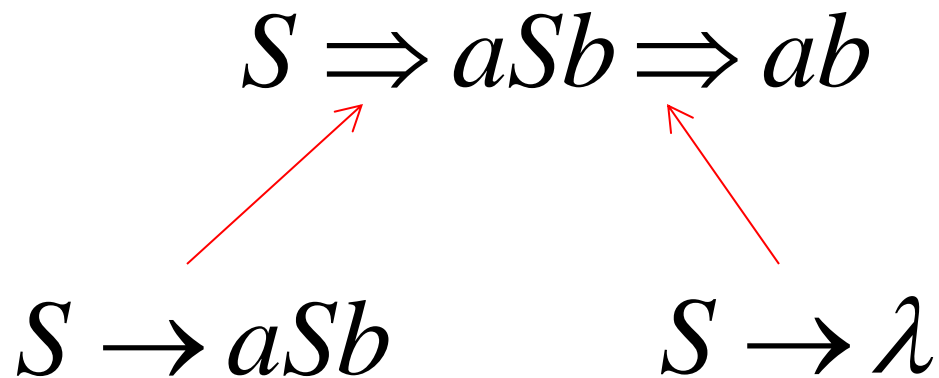
$$\langle noun \rangle \rightarrow dog$$

Variable

Terminal

Another Example

Grammar:
$$S \rightarrow aSb$$
$$S \rightarrow \lambda$$

Derivation of sentence ab :

$$S \Rightarrow aSb \Rightarrow ab$$

$$S \rightarrow aSb \qquad S \rightarrow \lambda$$

Language?

Grammar:

$$S \rightarrow aSb$$

$$S \rightarrow \lambda$$

Derivation of sentence $aabb$:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aabb$$

$$S \rightarrow aSb \qquad S \rightarrow \lambda$$

Other derivations:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb$$

$$\Rightarrow aaaaSbbbb \Rightarrow aaaabbbb$$

Language of the grammar

$$S \rightarrow aSb$$

$$S \rightarrow \lambda$$

$$L = \{a^n b^n : n \geq 0\}$$

More Notation

Grammar $\quad G = \left(V, T, S, P \right)$

$V:$ Set of variables

$T:$ Set of terminal symbols

$S:$ Start variable

$P:$ Set of Production rules

Example

Grammar G: $S \rightarrow aSb$

$S \rightarrow \lambda$

$$G = (V, T, S, P)$$

$V = \{S\}$ $T = \{a, b\}$

$P = \{S \rightarrow aSb, \ S \rightarrow \lambda\}$

More Notation

Sentential Form:

A sentence that contains variables and terminals

Example:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

Sentential Forms   sentence

We write: $$S \overset{*}{\Rightarrow} aaabbb$$

Instead of:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

In general we write: $$w_1 \overset{*}{\Longrightarrow} w_n$$

If: $$w_1 \Rightarrow w_2 \Rightarrow w_3 \Rightarrow \cdots \Rightarrow w_n$$

By default: $\qquad w \overset{*}{\Longrightarrow} w$

Example

Grammar

$$S \to aSb$$

$$S \to \lambda$$

Derivations

$$S \overset{*}{\Rightarrow} \lambda$$

$$S \overset{*}{\Rightarrow} ab$$

$$S \overset{*}{\Rightarrow} aabb$$

$$S \overset{*}{\Rightarrow} aaabbb$$

Example

Grammar

$$S \rightarrow aSb$$

$$S \rightarrow \lambda$$

Derivations

$$S \overset{*}{\Rightarrow} aaSbb$$

$$aaSbb \overset{*}{\Rightarrow} aaaaaSbbbbb$$

Another Grammar Example

Grammar G: $S \rightarrow Ab$

$A \rightarrow aAb$

$A \rightarrow \lambda$

Derivations:

$S \Rightarrow Ab \Rightarrow b$

$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow abb$

$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow aaAbbb \Rightarrow aabbb$

Language?

More Derivations

$$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow aaAbbb \Rightarrow aaaAbbbb$$

$$\Rightarrow aaaaAbbbbb \Rightarrow aaaabbbbb$$

$$S \overset{*}{\Rightarrow} aaaabbbbb$$

$$S \overset{*}{\Rightarrow} aaaaaabbbbbb$$

$$S \overset{*}{\Rightarrow} a^n b^n b$$

Language of a Grammar

For a grammar G
with start variable S :

$$L(G) = \{w : \quad S \overset{*}{\Rightarrow} w\}$$

String of terminals

Example

For grammar $G:$ $\quad S \rightarrow Ab$

$$A \rightarrow aAb$$

$$A \rightarrow \lambda$$

$$L(G) = \{a^n b^n b : \quad n \geq 0\}$$

Since: $\quad S \overset{*}{\Rightarrow} a^n b^n b$

A Convenient Notation

$$A \rightarrow aAb$$

$$A \rightarrow \lambda$$

$\Longrightarrow$

$$A \rightarrow aAb \mid \lambda$$

$$\langle article \rangle \rightarrow a$$

$$\langle article \rangle \rightarrow the$$

$\Longrightarrow$

$$\langle article \rangle \rightarrow a \mid the$$

Example

A context-free grammar G:

$$S \to aSb$$
$$S \to \lambda$$

A derivation:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aabb$$

A context-free grammar G:

$$S \rightarrow aSb$$
$$S \rightarrow \lambda$$

Another derivation:

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaabbb$$

$$S \rightarrow aSb$$

$$S \rightarrow \lambda$$

$$L(G) = \{a^n b^n : n \geq 0\}$$

Describes parentheses: (((( ))))

Example

A context-free grammar G:

$$S \rightarrow aSa$$

$$S \rightarrow bSb$$

$$S \rightarrow \lambda$$

A derivation:

$$S \Rightarrow aSa \Rightarrow abSba \Rightarrow abba$$

Language?

A context-free grammar $G:$ $S \rightarrow aSa$

$$S \rightarrow bSb$$

$$S \rightarrow \lambda$$

Another derivation:

$$S \Rightarrow aSa \Rightarrow abSba \Rightarrow abaSaba \Rightarrow abaaba$$

$$S \to aSa$$

$$S \to bSb$$

$$S \to \lambda$$

$$L(G) = \{ww^R : \quad w \in \{a,b\}^*\}$$

Example

A context-free grammar G: $\quad S \rightarrow aSb$

$$S \rightarrow SS$$

$$S \rightarrow \lambda$$

A derivation:

$$S \Rightarrow SS \Rightarrow aSbS \Rightarrow abS \Rightarrow ab$$

Language?

A context-free grammar $G:$ $S \rightarrow aSb$

$$S \rightarrow SS$$

$$S \rightarrow \lambda$$

A derivation:

$$S \Rightarrow SS \Rightarrow aSbS \Rightarrow abS \Rightarrow abaSb \Rightarrow abab$$

$$S \rightarrow aSb$$

$$S \rightarrow SS$$

$$S \rightarrow \lambda$$

$$L(G) = \{w \quad : n_a(w) = n_b(w),$$
$$\text{and } n_a(v) \geq n_b(v)$$
$$\text{in any prefix } v\}$$

Interpretation?

$$S \rightarrow aSb$$

$$S \rightarrow SS$$

$$S \rightarrow \lambda$$

$$L(G) = \{w \quad : n_a(w) = n_b(w),$$
$$\text{and } n_a(v) \geq n_b(v)$$
$$\text{in any prefix } v\}$$

Describes matched parentheses: () ((( ))) (( ))

Definition: Context-Free Grammars

Grammar $\quad G = (V, T, S, P)$

Variables   Terminal   Start

             symbols   variable

Productions of the form:
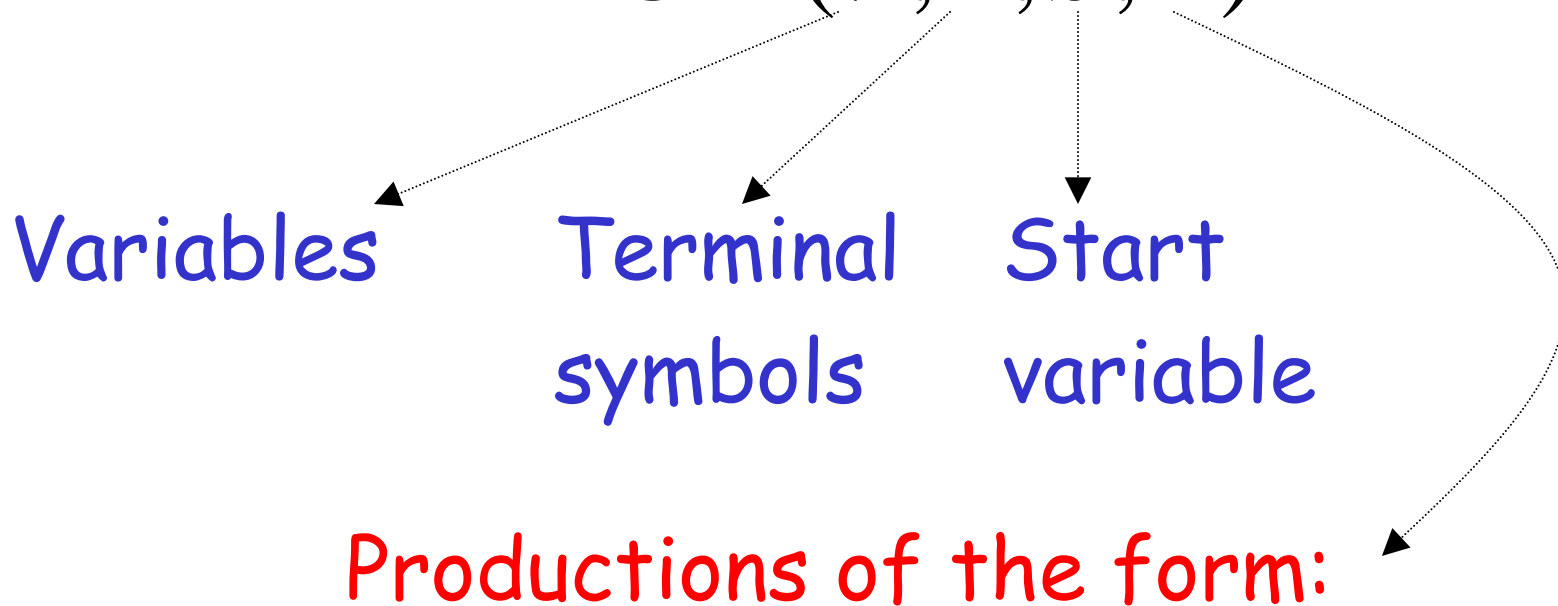
$$A \rightarrow x$$

Variable   String of variables
                 and terminals

$$G = (V, T, S, P)$$

$$L(G) = \{w: \ S \overset{*}{\Rightarrow} w, \ w \in T^*\}$$

Definition: Context-Free Languages

A language L is context-free

if and only if

there is a context-free grammar G
with $L = L(G)$

Derivation Order

1. $S \rightarrow AB$ 2. $A \rightarrow aaA$ 4. $B \rightarrow Bb$

3. $A \rightarrow \lambda$ 5. $B \rightarrow \lambda$

Leftmost derivation:

$$\overset{1}{S} \Rightarrow \overset{2}{AB} \Rightarrow \overset{3}{aaAB} \Rightarrow \overset{4}{aaB} \Rightarrow \overset{5}{aaBb} \Rightarrow aab$$

Rightmost derivation:

$$\overset{1}{S} \Rightarrow \overset{4}{AB} \Rightarrow \overset{5}{ABb} \Rightarrow \overset{2}{Ab} \Rightarrow \overset{3}{aaAb} \Rightarrow aab$$

Language?

$$S \rightarrow aAB$$

$$A \rightarrow bBb$$

$$B \rightarrow A \mid \lambda$$

Leftmost derivation:

$$S \Rightarrow aAB \Rightarrow abBbB \Rightarrow abAbB \Rightarrow abbBbbB$$

$$\Rightarrow abbbbB \Rightarrow abbbb$$

Rightmost derivation:

$$S \Rightarrow aAB \Rightarrow aA \Rightarrow abBb \Rightarrow abAb$$

$$\Rightarrow abbBbb \Rightarrow abbbb$$

Language?

Derivation Trees

$$S \rightarrow AB \qquad A \rightarrow aaA \mid \lambda \qquad B \rightarrow Bb \mid \lambda$$

$$S \Rightarrow AB$$
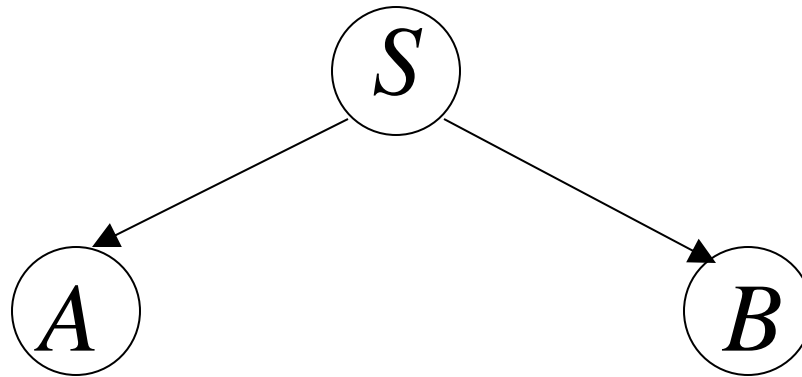
$$S \to AB \qquad A \to aaA \mid \lambda \qquad B \to Bb \mid \lambda$$

$$S \Rightarrow AB \Rightarrow aaAB$$
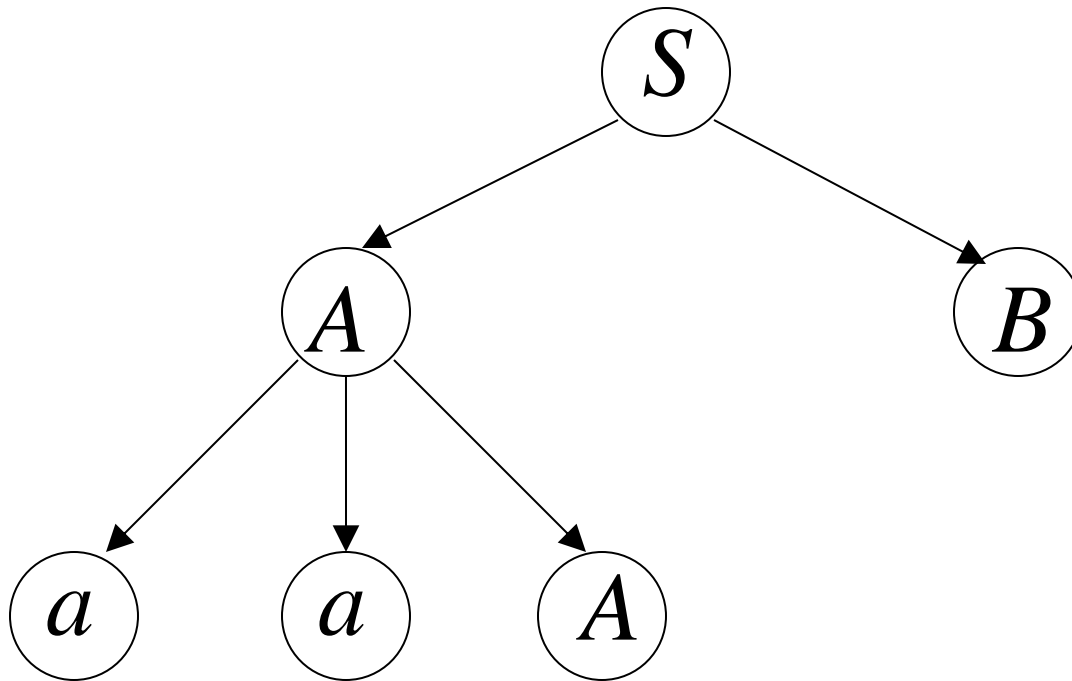
$$S \to AB \qquad A \to aaA \mid \lambda \qquad B \to Bb \mid \lambda$$

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaABb$$
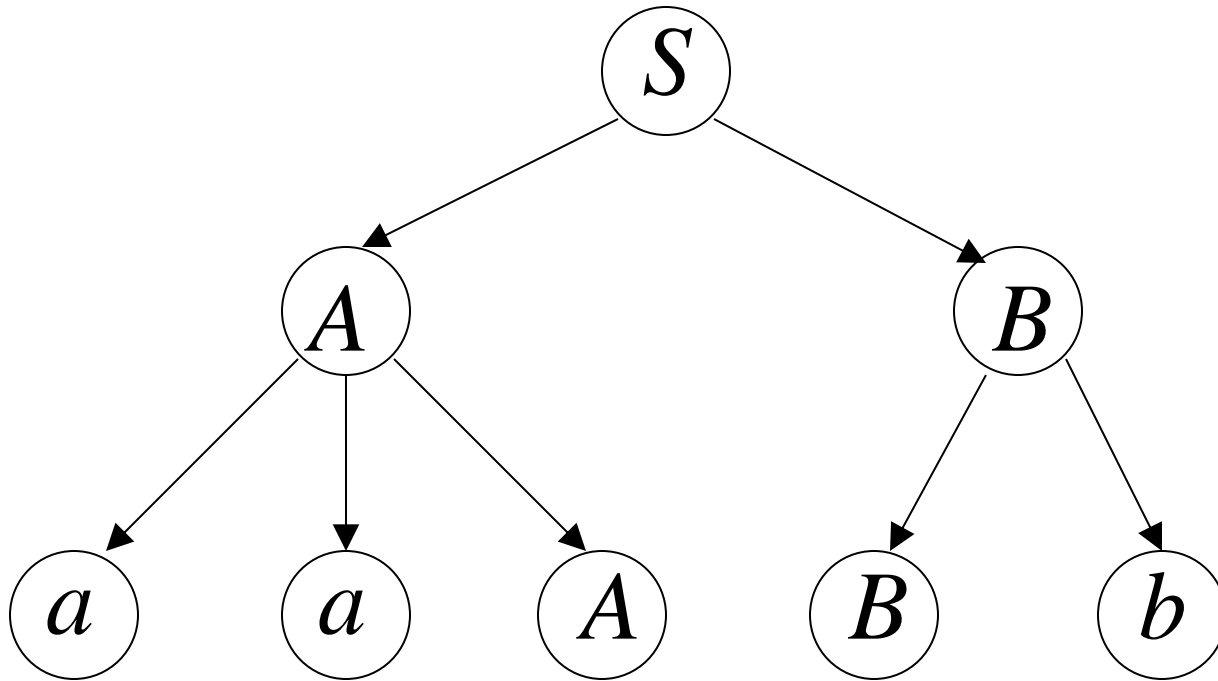
$$S \rightarrow AB \qquad A \rightarrow aaA \,|\, \lambda \qquad B \rightarrow Bb \,|\, \lambda$$

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaABb \Rightarrow aaBb$$
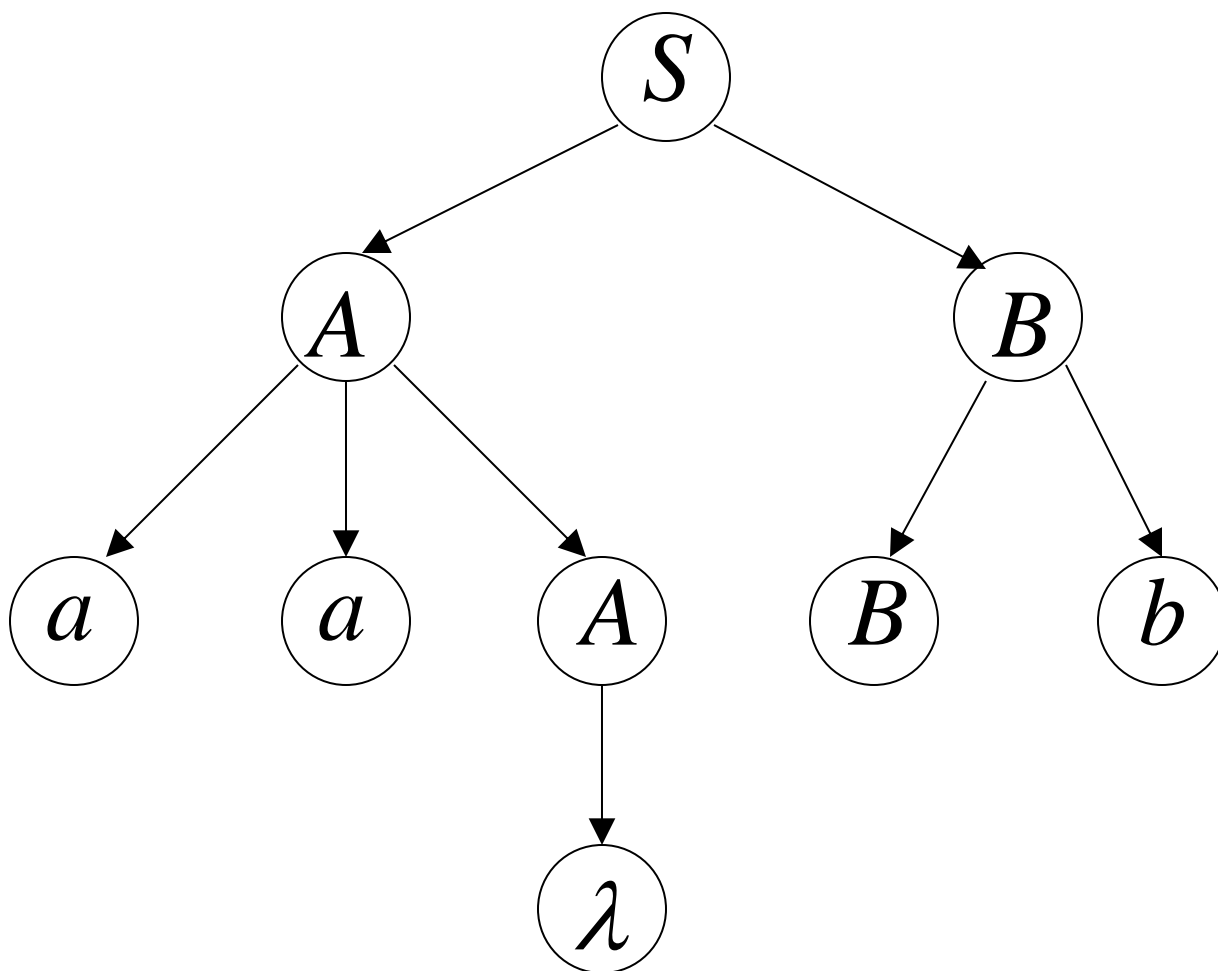
$$S \rightarrow AB \qquad A \rightarrow aaA \mid \lambda \qquad B \rightarrow Bb \mid \lambda$$

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaABb \Rightarrow aaBb \Rightarrow aab$$
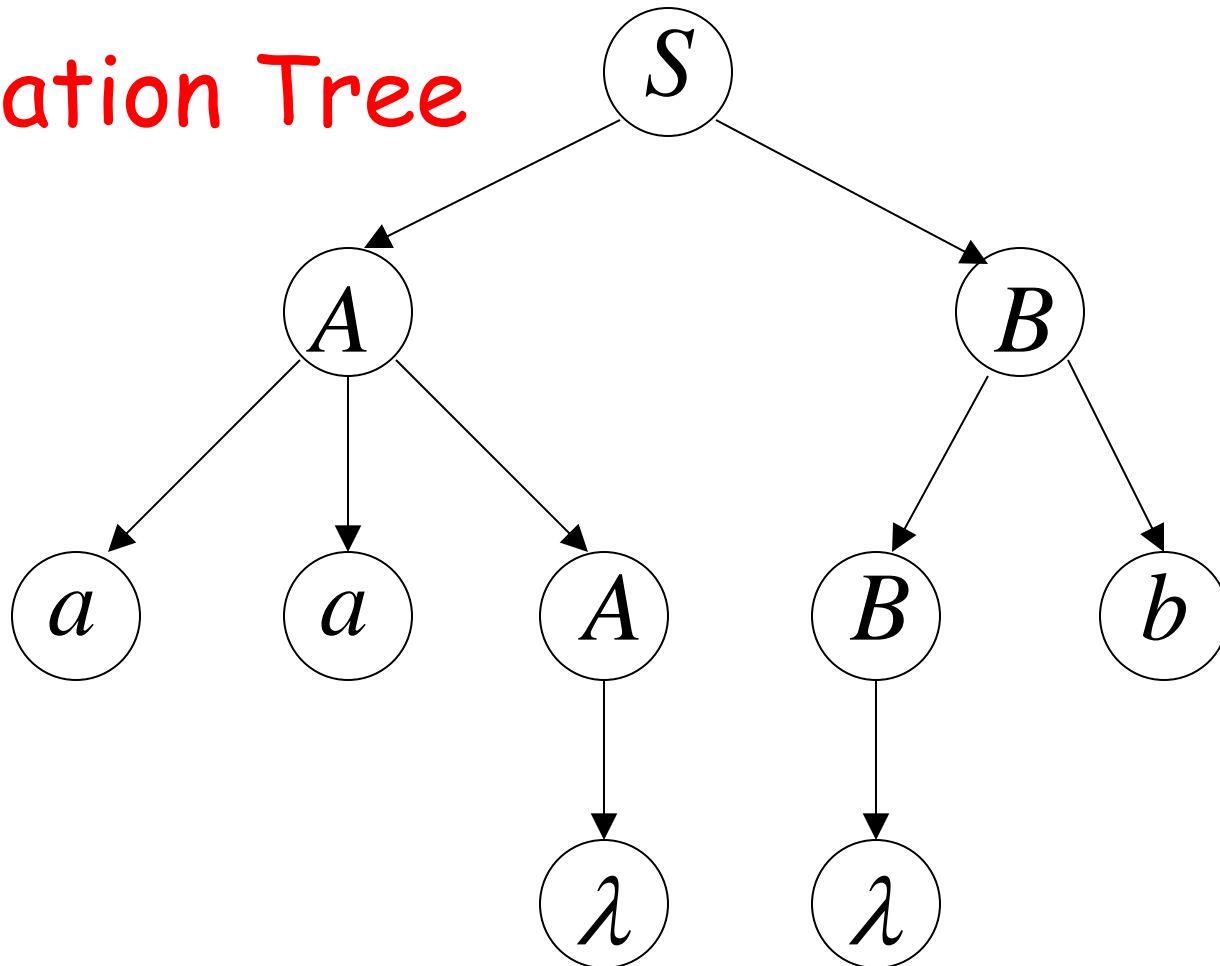
Derivation Tree

$$S \rightarrow AB \qquad A \rightarrow aaA \,|\, \lambda \qquad B \rightarrow Bb \,|\, \lambda$$

$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaABb \Rightarrow aaBb \Rightarrow aab$$

Derivation Tree

yield

$$aa\lambda\lambda b$$

$$= aab$$
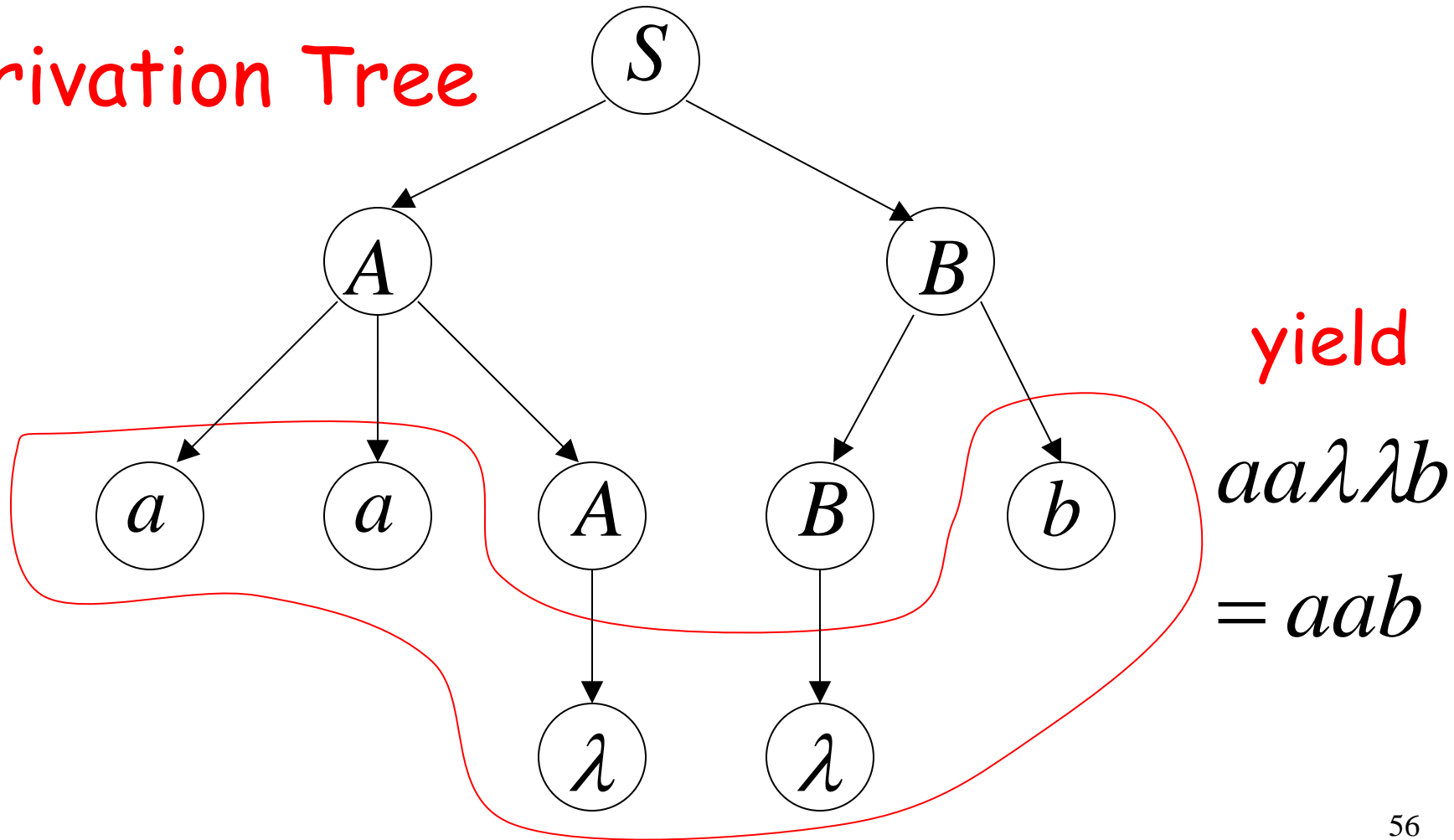
Partial Derivation Trees

$$S \rightarrow AB \qquad A \rightarrow aaA \mid \lambda \qquad B \rightarrow Bb \mid \lambda$$

$$S \Rightarrow AB$$

Partial derivation tree
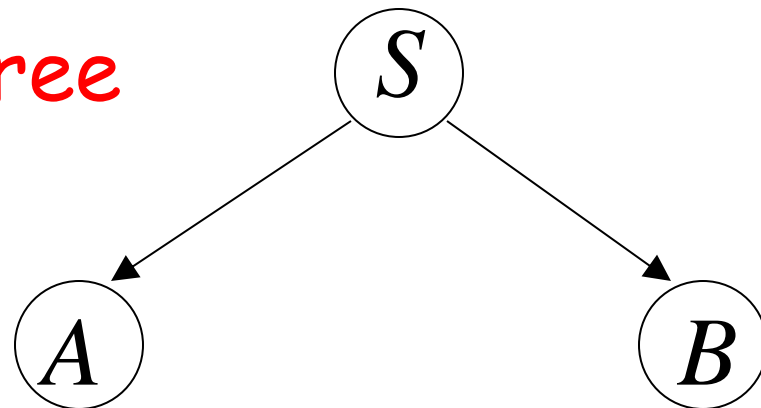
$$S \Rightarrow AB \Rightarrow aaAB$$

Partial derivation tree
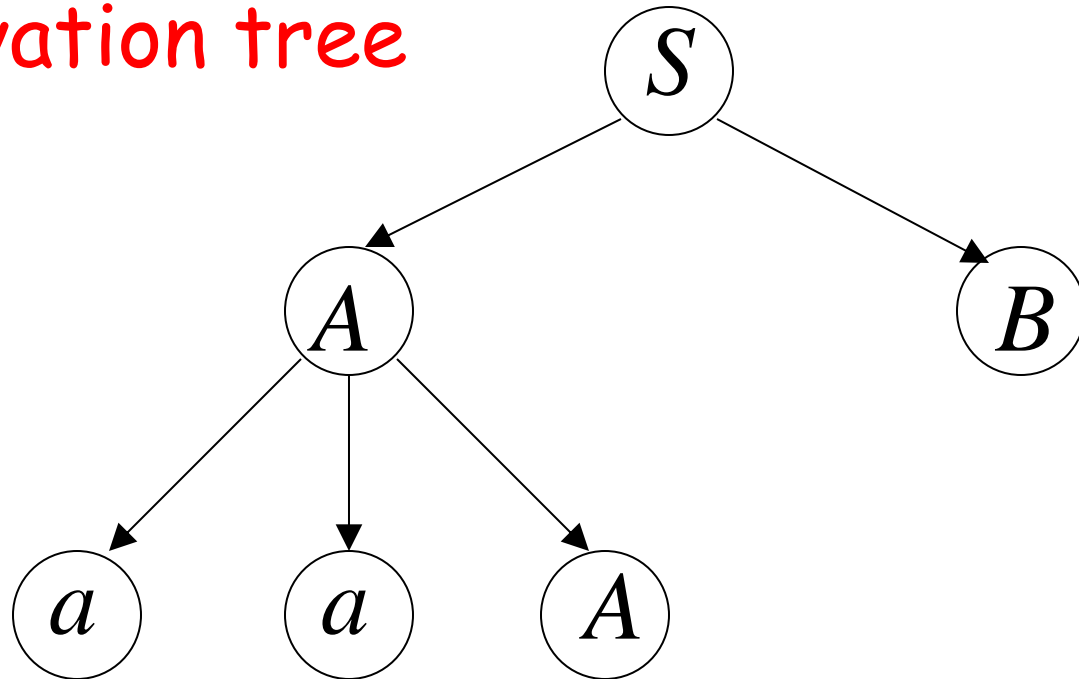
$$S \Rightarrow AB \Rightarrow aaAB$$

sentential form

Partial derivation tree

yield
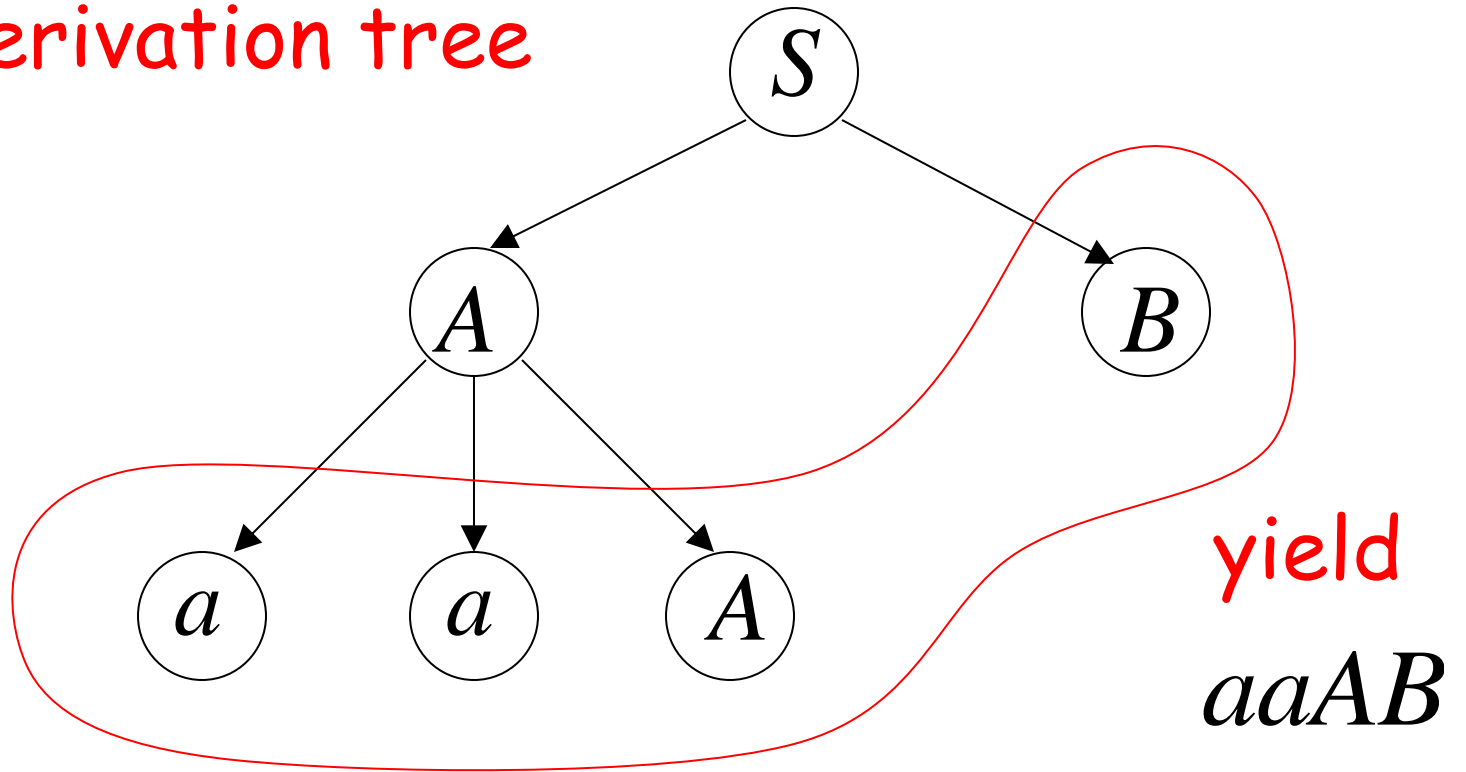
$aaAB$

Sometimes, derivation order doesn't matter

Leftmost:
$$S \Rightarrow AB \Rightarrow aaAB \Rightarrow aaB \Rightarrow aaBb \Rightarrow aab$$

Rightmost:
$$S \Rightarrow AB \Rightarrow ABb \Rightarrow Ab \Rightarrow aaAb \Rightarrow aab$$

Same derivation tree
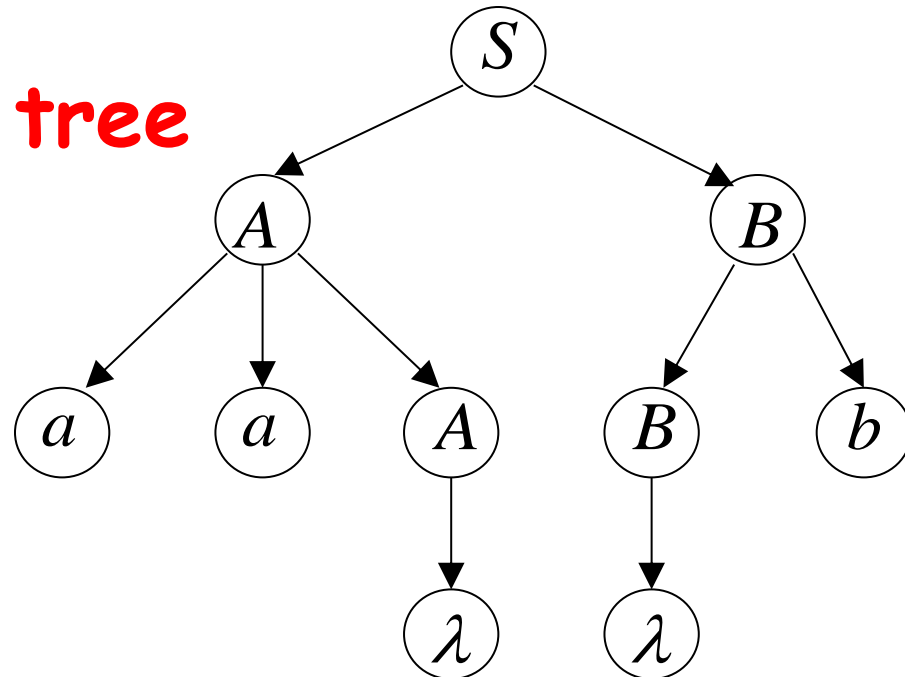
Ambiguity

$$E \rightarrow E + E \mid E * E \mid (E) \mid a$$

$$a + a * a$$

$$E \Rightarrow E + E \Rightarrow a + E \Rightarrow a + E * E$$

$$\Rightarrow a + a * E \Rightarrow a + a * a$$

leftmost derivation

$$E \rightarrow E + E \mid E * E \mid (E) \mid a$$

$$a + a * a$$
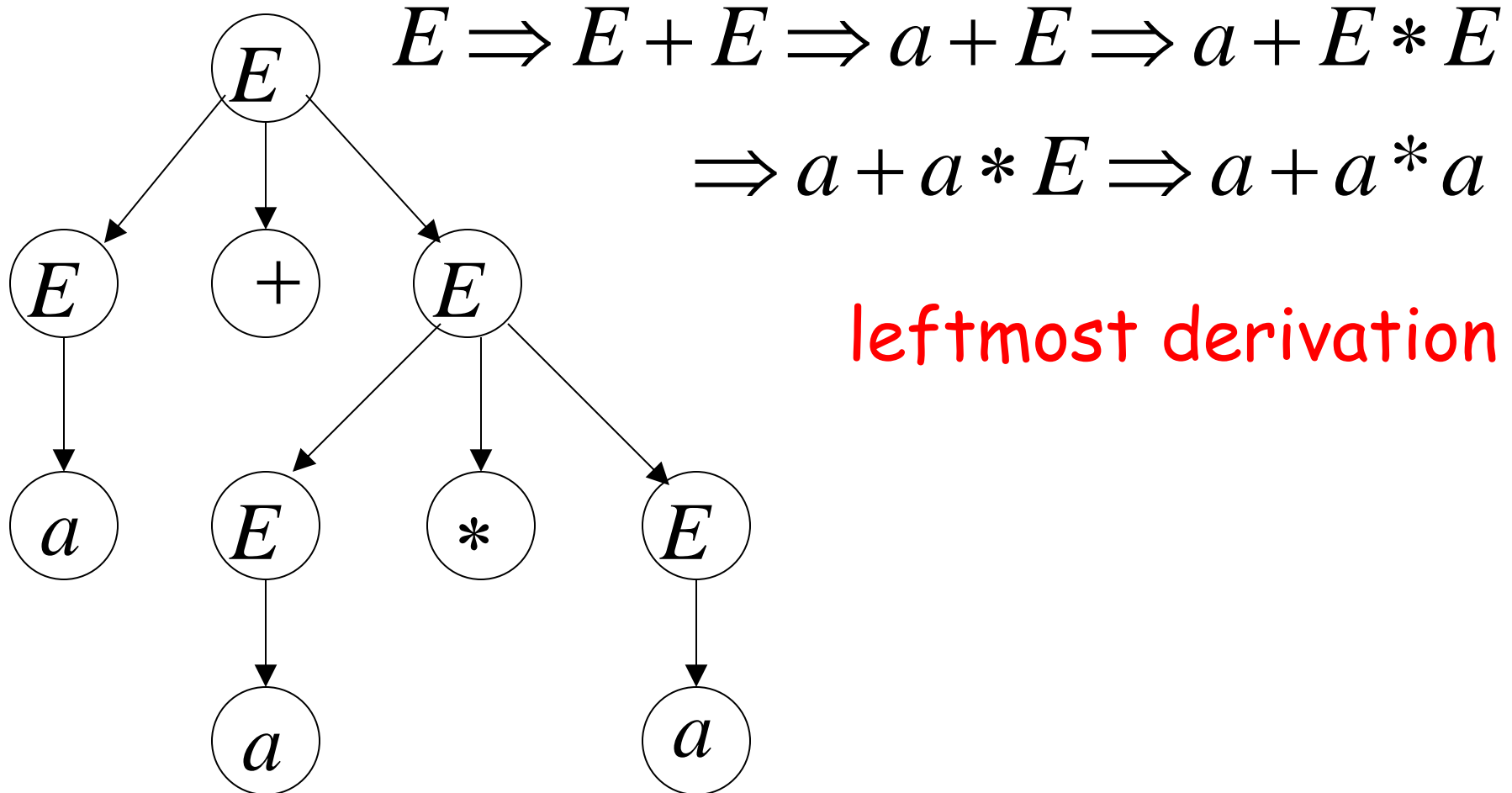
$$E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow a + E * E$$

$$\Rightarrow a + a * E \Rightarrow a + a * a$$

leftmost derivation
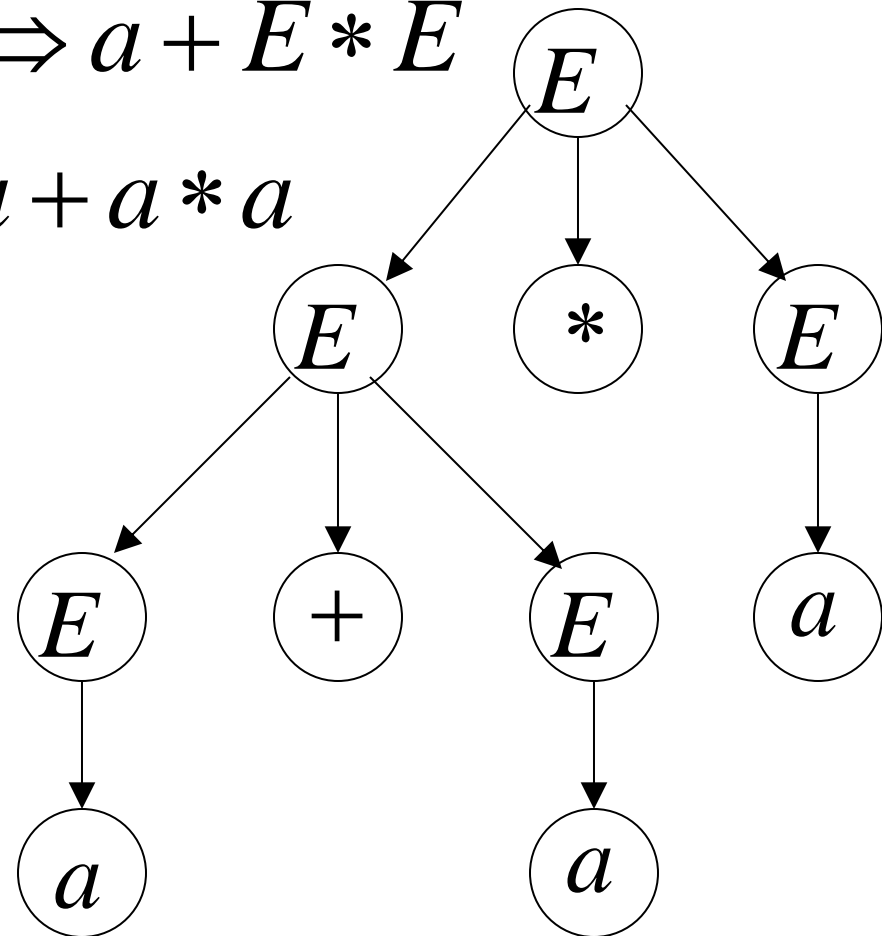
$$E \rightarrow E + E \mid E * E \mid (E) \mid a$$

$$a + a * a$$

Two derivation trees
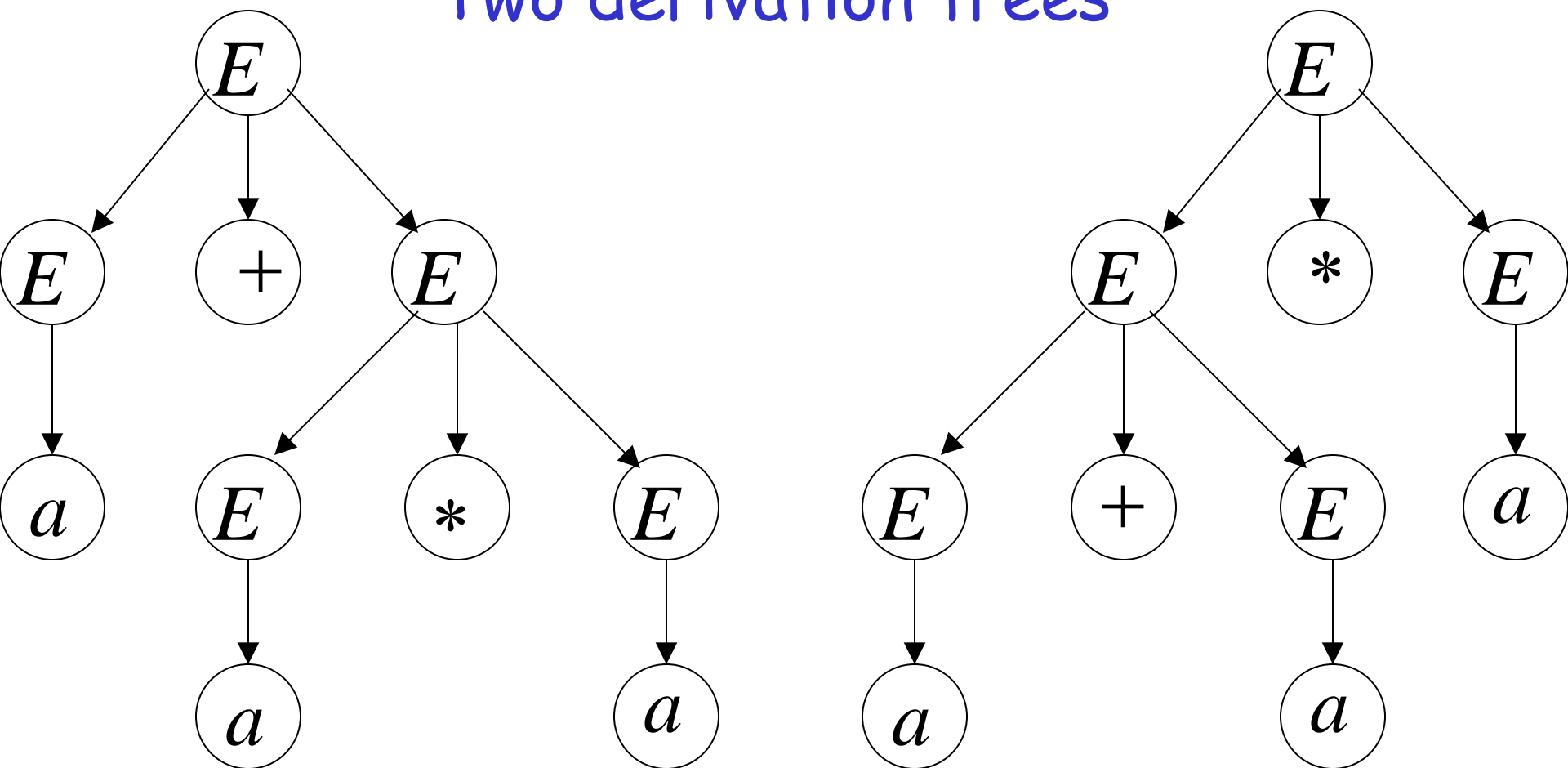
The grammar $E \rightarrow E + E \mid E * E \mid (E) \mid a$

is ambiguous:

string $a + a * a$ has two derivation trees
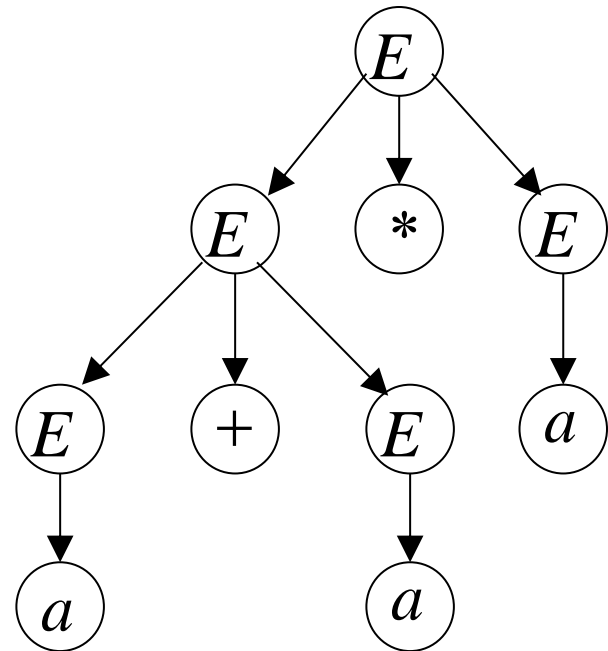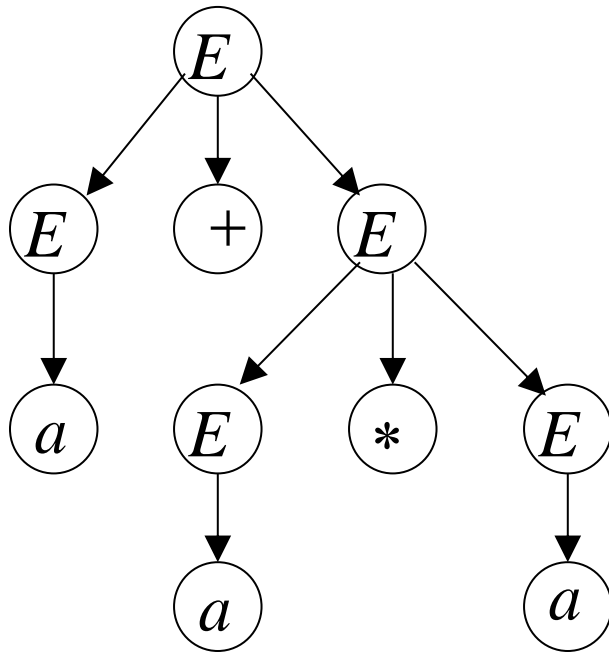
The grammar $E \rightarrow E + E \mid E * E \mid (E) \mid a$

is ambiguous:

string $a + a * a$ has two leftmost derivations

$$E \Rightarrow E + E \Rightarrow a + E \Rightarrow a + E * E$$

$$\Rightarrow a + a * E \Rightarrow a + a * a$$

$$E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow a + E * E$$

$$\Rightarrow a + a * E \Rightarrow a + a * a$$

Definition:

A context-free grammar G is **ambiguous**

if some string $w \in L(G)$ has:

two or more derivation trees

In other words:

A context-free grammar G is **ambiguous**

if some string $w \in L(G)$ has:

two or more leftmost derivations

(or rightmost)

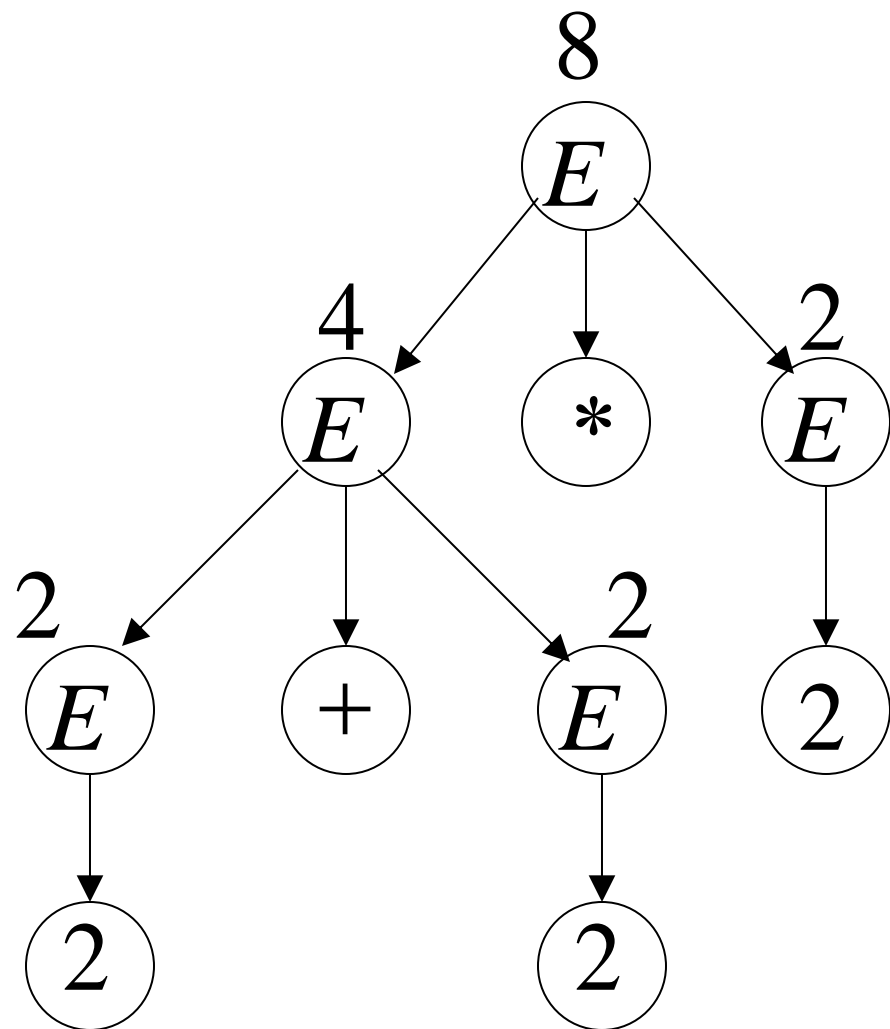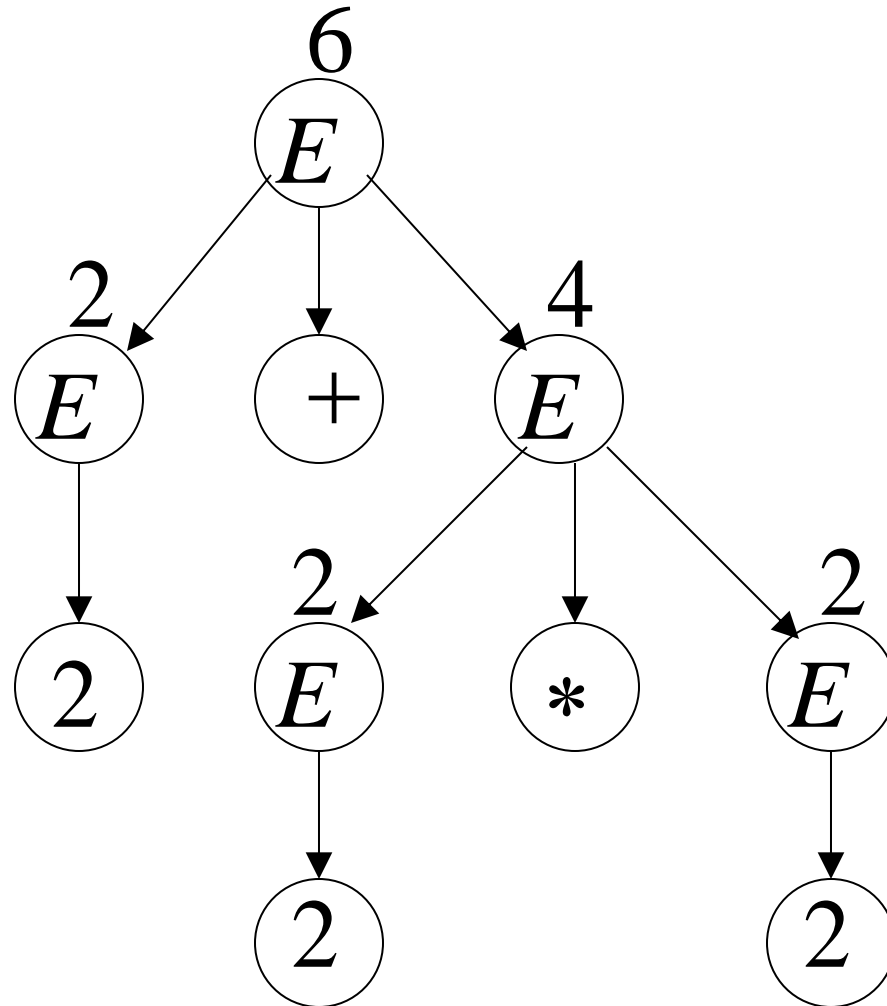# Why do we care about ambiguity?

$$a + a * a$$

take $a = 2$
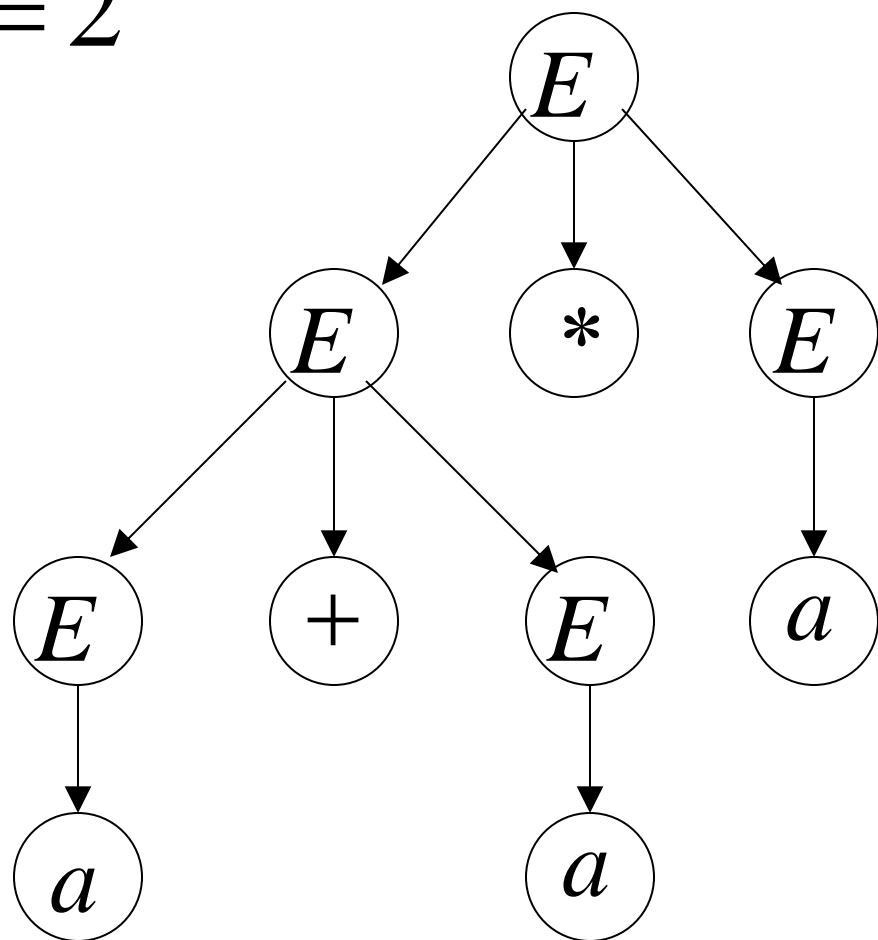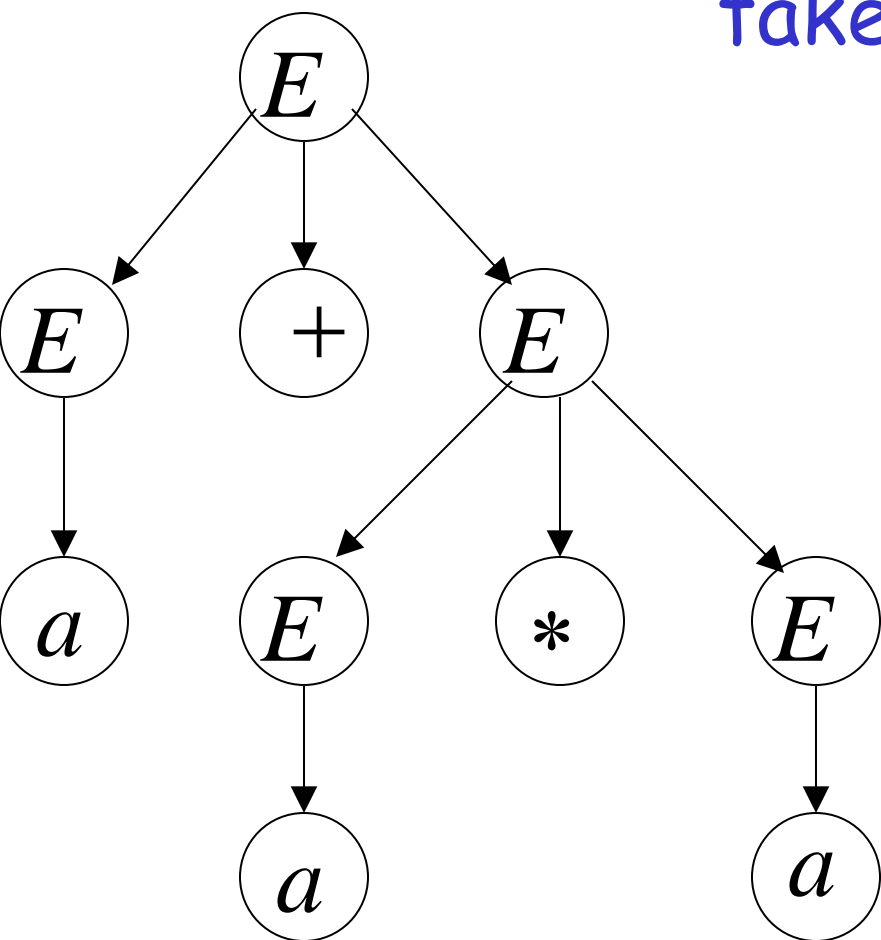
$$2+2*2$$
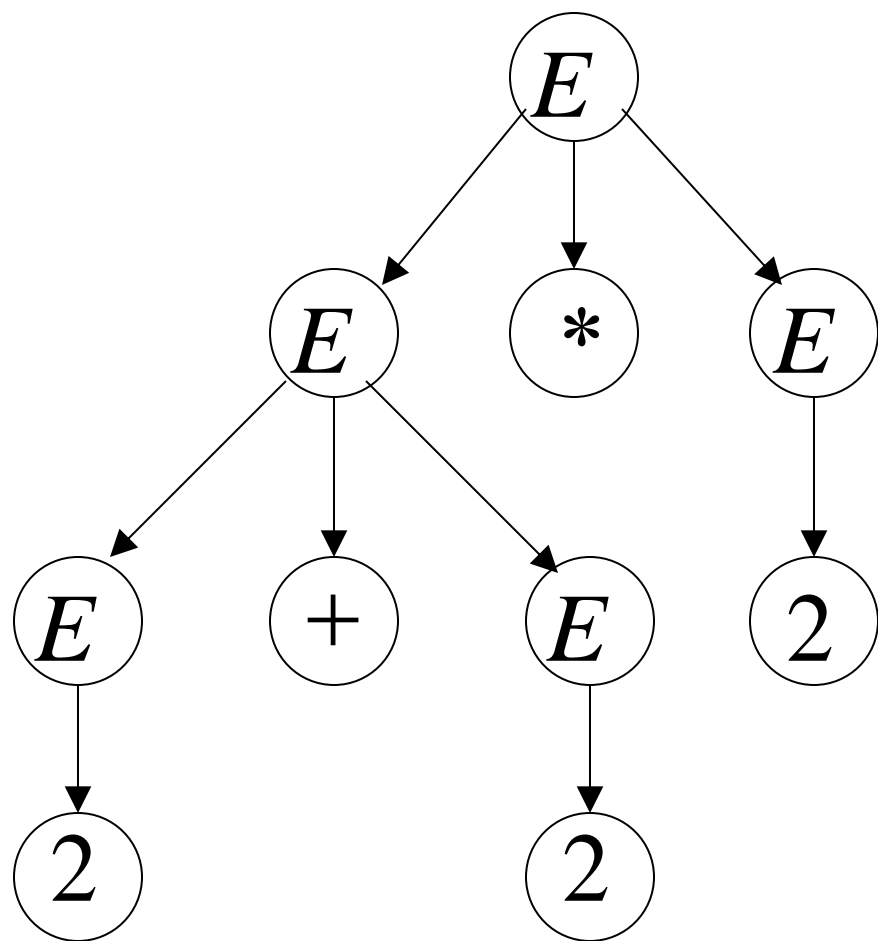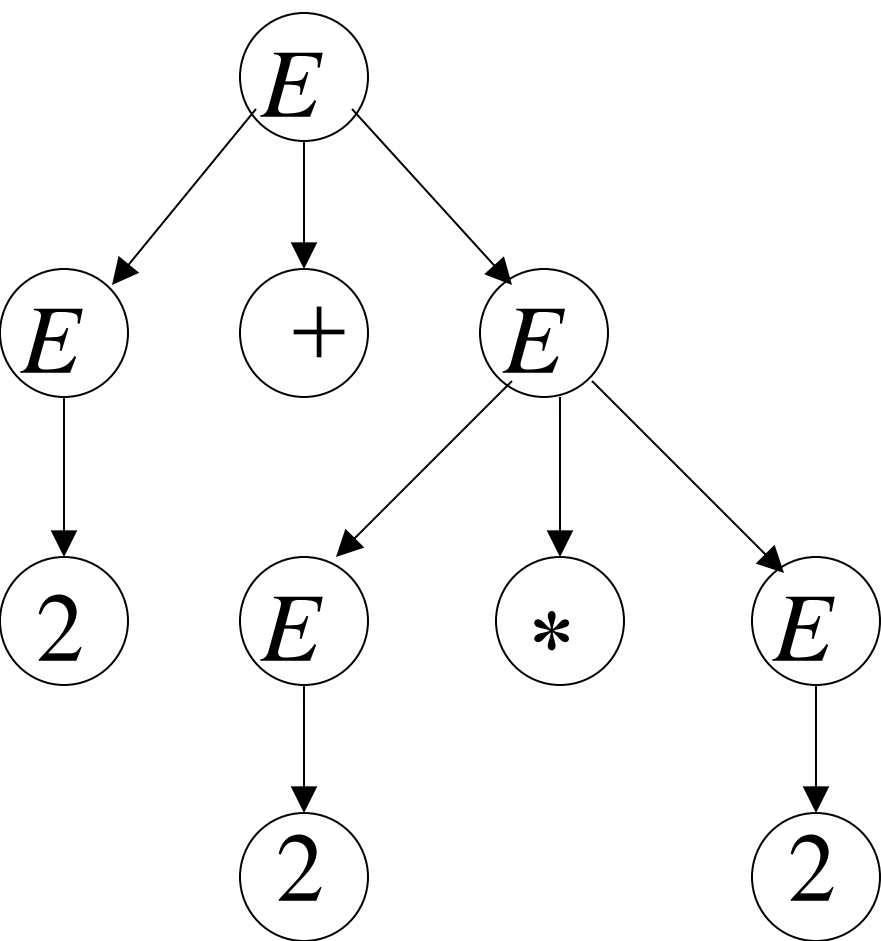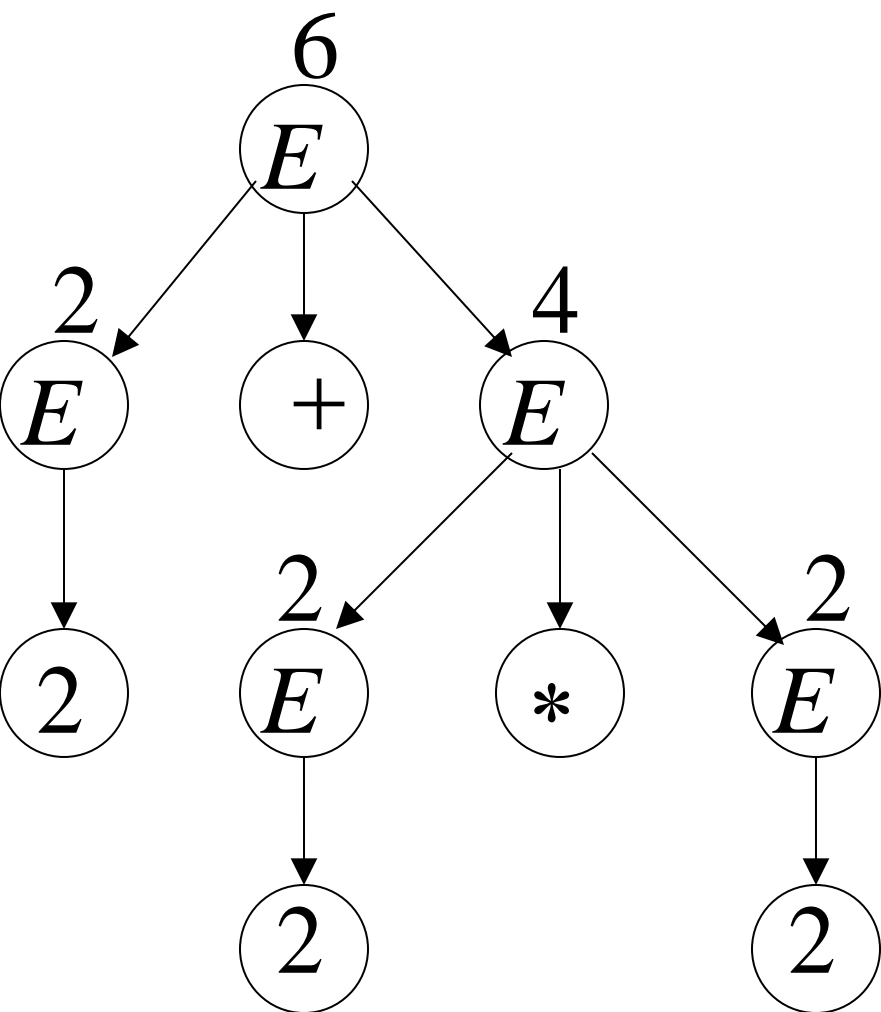
$$2 + 2 * 2 = 6 \qquad 2 + 2 * 2 = 8$$

Correct result: $2+2*2=6$

6
E

2
E $+$ 4
E

2 2
E $*$ 2
E

2 2

- Ambiguity is **bad** for programming languages

- We want to remove ambiguity

We fix the ambiguous grammar:

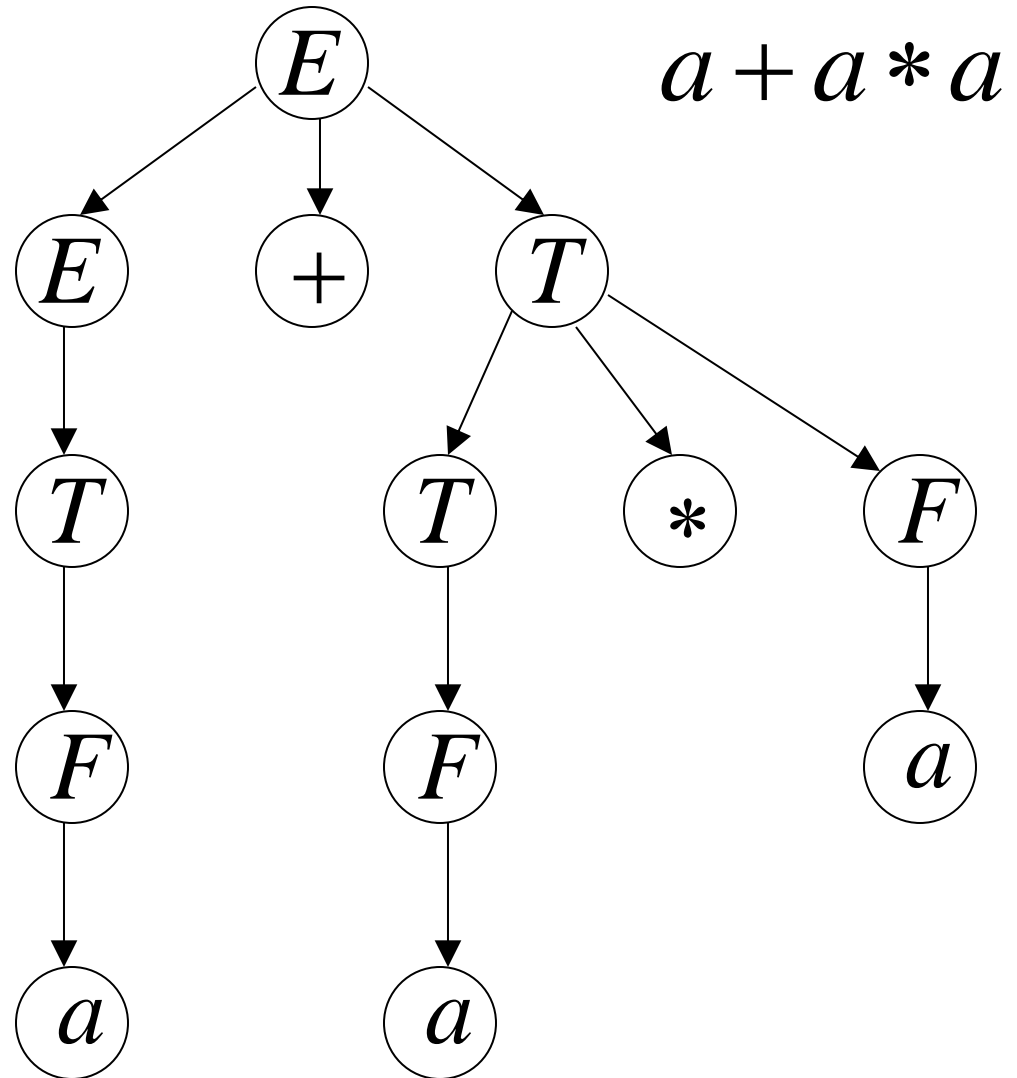$$E \rightarrow E + E \mid E * E \mid (E) \mid a$$

New non-ambiguous grammar:

$$E \rightarrow E + T$$
$$E \rightarrow T$$
$$T \rightarrow T * F$$
$$T \rightarrow F$$
$$F \rightarrow (E)$$
$$F \rightarrow a$$

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

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

$$E \rightarrow E+T$$

$$E \rightarrow T$$

$$T \rightarrow T*F$$

$$T \rightarrow F$$

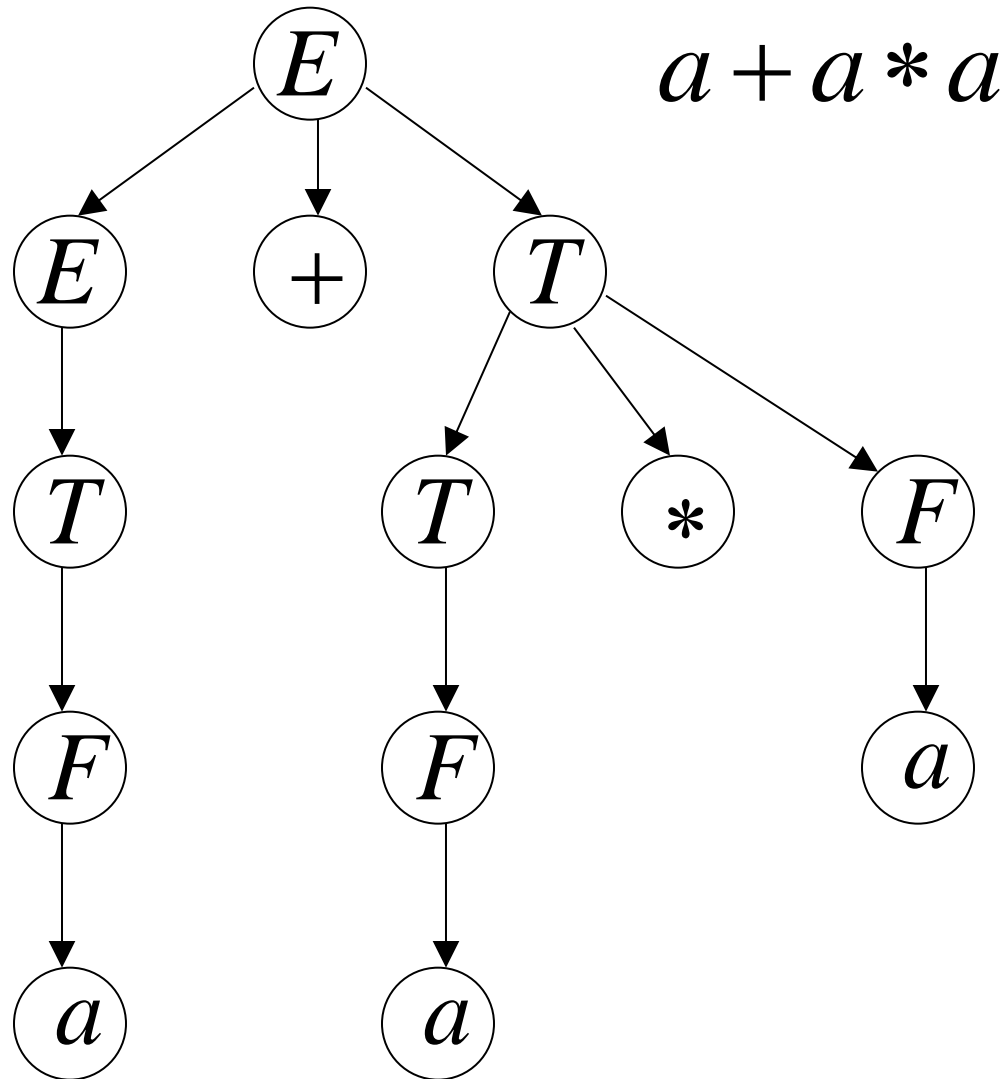$$F \rightarrow (E)$$

$$F \rightarrow a$$

$$a+a*a$$

Unique derivation tree

$$a + a * a$$

The grammar G:

$$E \to E + T$$

$$E \to T$$

$$T \to T * F$$

$$T \to F$$

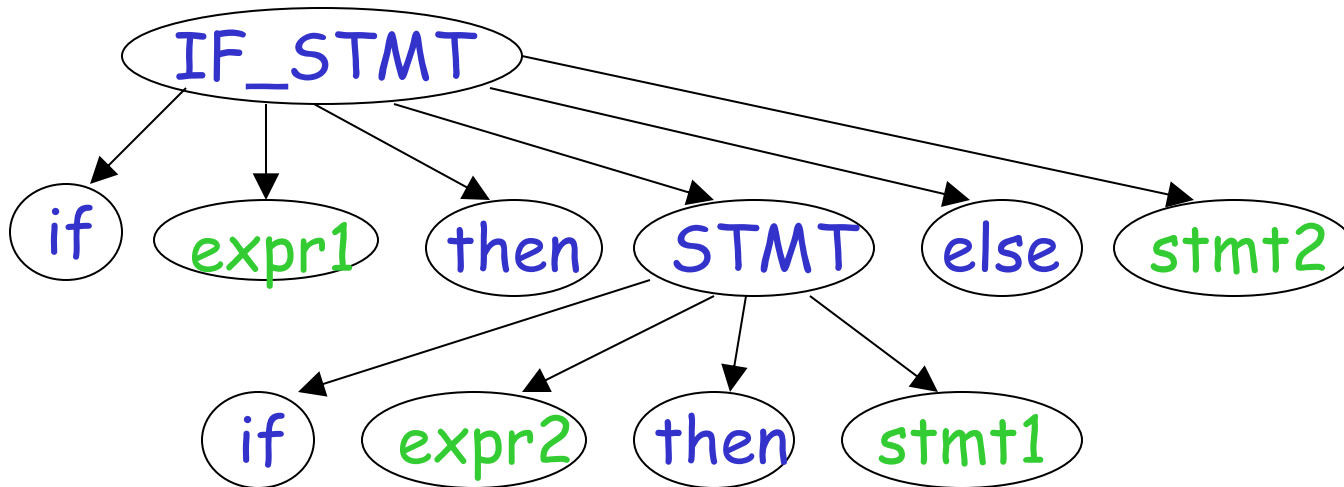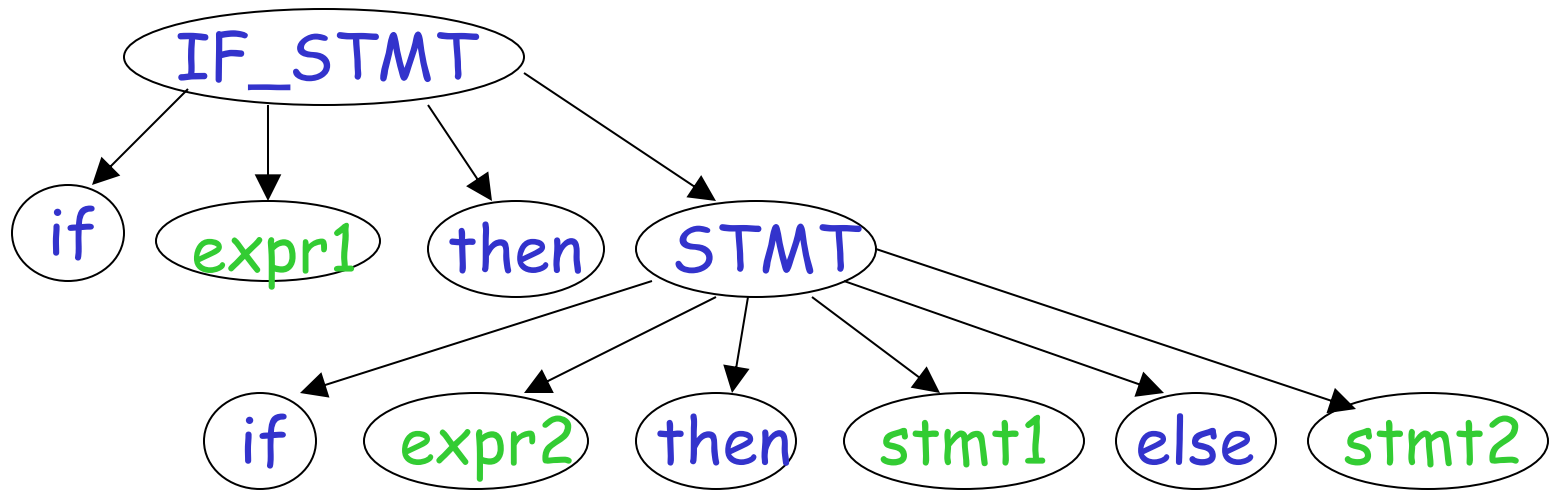$$F \to (E)$$

$$F \to a$$

is non-ambiguous:

Every string $w \in L(G)$ has a unique derivation tree

Another Ambiguous Grammar

IF_STMT $\rightarrow$ if EXPR then STMT

| if EXPR then STMT else STMT

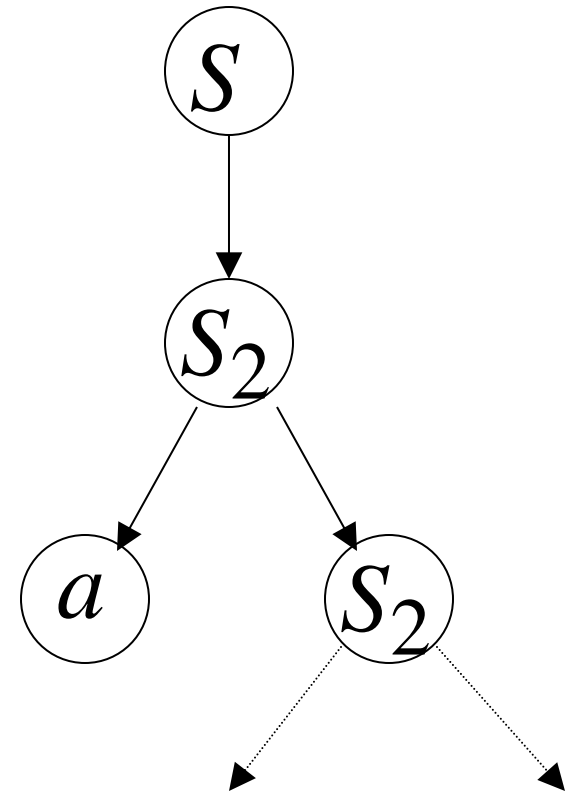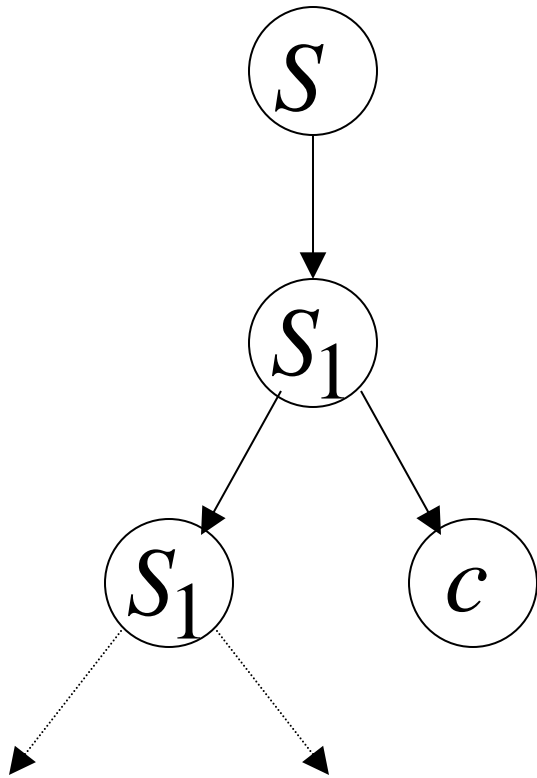If expr1 then if expr2 then stmt1 else stmt2

Inherent Ambiguity

Some context free languages
have only ambiguous grammars

Example: $L = \{a^n b^n c^m\} \cup \{a^n b^m c^m\}$

$S \rightarrow S_1 \mid S_2$

$S_1 \rightarrow S_1 c \mid A$

$A \rightarrow aAb \mid \lambda$

$S_2 \rightarrow aS_2 \mid B$

$B \rightarrow bBc \mid \lambda$

The string $a^n b^n c^n$

has two derivation trees

Ambiguity in natural language?