Context-free grammars

- a set of terminals or tokens $\Sigma$
- a set of nonterminals or variables $V$
- a starting nonterminal $S$
- a set of productions (rewrite rules) $P$ of the form

$$\text{variable ::= string}$$

Example: $\Sigma = \{a, b\}$, $V = \{S\}$,

$$P = \{S ::= aSb | ab\}$$

Derivation step: $S$ can be replaced with $aSb$ or $ab$
Derivation tree

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

$$P = \{S ::= aSb \mid ab\}$$

![Derivation Tree Diagram]
Ambiguity

A grammar is ambiguous if there is a string with two or more derivation trees.

Example: $\Sigma = \{a, b\}$, $V = \{S\}$,

$$P = \{S ::= SS | a\}$$

String $aaa$ has two distinct derivation trees.
Ambiguity: another example

\[ \Sigma = \{a\}, \ V = \{S\}, \]

\[ P = \{S ::= aSSa \mid a\} \]

String \( a^7 \) (aaaaaaa) has two distinct derivation trees.
Backus-Naur Form (BNF)

a notation used to write context-free grammars of programming languages

- nonterminals are enclosed between ‘<’ and ‘>’
- the empty string is written as <empty>

Example: BNF rules for nonnegative real numbers

<real-number> ::= <integer-part> . <fraction>
<integer-part> ::= <digit>
                | <integer-part> <digit>
<fraction> ::= <digit> | <digit> <fraction>
<digit> ::= 0 | 1 | 2 | 3 | 4 | 5
           | 6 | 7 | 8 | 9
Expressions

<exp> ::= <exp> <op> <exp> | ( <exp> )  
     | <number>
<op>  ::= + | - | * | /
<number> ::= <digit> | <digit> <number>
<digit> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7  
     | 8 | 9

2 * 5 + 3 has two distinct derivation trees.
Expressions, continued

<exp> ::= <exp> <op1> <term> | <term>
<term> ::= <term> <op2> <sterm> | <sterm>
<op1> ::= + | -
<op2> ::= * | /
<sterm> ::= <number> | ( <exp> )
<number> ::= <digit> | <digit> <number>
<digit> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7
| 8 | 9
Dangling “else”

\[
\langle S \rangle ::= \text{if } \langle C \rangle \text{ then } \langle S \rangle
\]

\[
\langle S \rangle ::= \text{if } \langle C \rangle \text{ then } \langle S \rangle \text{ else } \langle S \rangle
\]
Extended BNF (EBNF)

- Brackets, [ and ], enclose an optional construct

\[
<\text{real\_num}> ::= [<\text{integer}>].<\text{fraction}>
\]

- Braces, { and }, represent zero or more repetitions

\[
<\text{ident\_list}> ::= <\text{ident}>\ {,\ <\text{ident}>}
\]

- Parentheses, ( and ), are used for grouping