

SPECIFYING SYNTAX

Terminology

The **syntax** of a programming language determines the well-formed or grammatically correct programs of the language.

Its **semantics** describes how or whether such programs will execute.

- syntax is how things look
- semantics is how things work (the meaning)

Example: “a = b+c” is correct Java syntax apparently.

Questions:

1) Have b and c been declared as a type that allows the + operation?

Note that Java has five different + operations.

2) Is “a” assignment compatible with the result of the expression “b+c”?

3) Do “b” and “c” have values?

Compile-time or **static** issues: 1) and 2)

Run-time or **dynamic** issues: 3)

Syntax

A **terminal** is a symbol in the language.

Java terminals: **while**, **static**, (, ;, 5, b

A **syntactic category** or **nonterminal** is a set of objects (strings) that will be defined in terms of symbols in the language.

Java nonterminals: <statement>, <expression>

A **metalanguage** is a higher-level language used to discuss, describe, or analyze another language. English is used as a metalanguage for describing programming languages, but because of the ambiguities in English, more formal metalanguages have been proposed.

Backus-Naur Form (BNF)

BNF was first used by John Backus and Peter Naur to describe Algol in 1963.

It consists of a set of **rules** or **productions** of the form:

<syntactic category> ::= a sequence of terminals,
syntactic categories, and the
symbol “|”

“::=” means “is defined as” (sometimes written as “→”)

juxtaposition means concatenation

“|” means alternation (or)

BNF Examples from Java

<integral type> ::=
 byte | short | int | long | char

<argument list> ::= <expression>
 | <argument list> , <expression>

<statement> ::= <statement without trailing substatement>
 | <labeled statement> | <if then statement>
 | <if then else statement> | <while statement >
 | <for statement>

<method declaration> ::=
 <method modifiers> <result type>
 <method declarator> <throws> <method body>
| <method modifiers> <result type>
 <method declarator> <method body>
| <result type> <method declarator>
 <throws> <method body>
| <result type> <method declarator> <method body>

Note that “|” corresponds to set union.

<type> ::= <primitive type> | <reference type>

The set of types is the union of the sets of primitive types and reference types.

Extended BNF (EBNF)

Since a BNF description of the syntax of a programming language relies heavily on recursion to provide lists of items, many definitions use these extensions:

- 1) `item?` or `[item]` means item is optional.
- 2) `item*` or `{ item }` means to take zero or more occurrences of item.
- 3) `item+` means to take one or more occurrences of item.
- 4) Parentheses are used for grouping.

Examples

```
<method declaration> ::=  
    <method modifiers>? <result type>  
    <method declarator> <throws>? <method body>
```

```
<method modifier> ::= public | protected | private | static  
    | abstract | final | synchronized | native
```

```
<method modifiers> ::= <method modifier>  
    | <method modifiers> <method modifier>
```

may be written

```
<method modifiers> ::= <method modifier>+
```

Or better, eliminate nonterminal `<method modifiers>` and write:

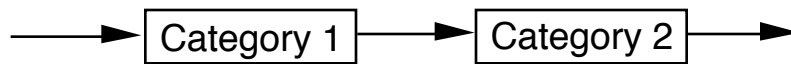
```
<method declaration> ::=  
    <method modifier>* <result type>  
    <method declarator> <throws>? <method body>
```

```
<argument list> ::= <expression> (, <expression>)*
```

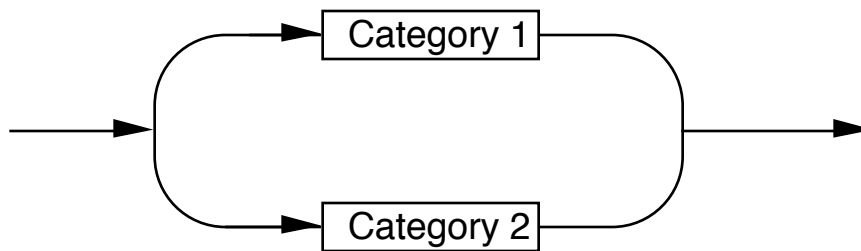
Syntax Diagrams (or Charts or Graphs)

Syntax diagrams are associated with Pascal, but they have been used to describe command languages and, recently, many other programming languages.

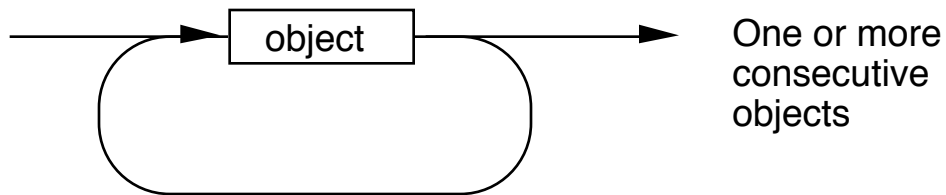
- 1) A terminal symbol is shown in a circle or oval.
- 2) A syntactic category is placed in a rectangle.
- 3) The concatenation of two objects is indicated by a flowline:



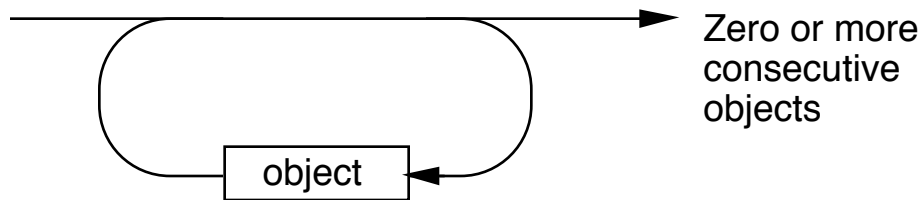
- 4) The alternation of two objects is shown by branching:



5) Repetition of objects is represented by a loop:



$\langle \text{thing} \rangle ::= \langle \text{object} \rangle \mid \langle \text{thing} \rangle \langle \text{object} \rangle$



$\langle \text{thing} \rangle ::= \langle \text{empty} \rangle \mid \langle \text{thing} \rangle \langle \text{object} \rangle$

$\langle \text{empty} \rangle ::=$

Example: Simple Expressions in Pascal

BNF

$\langle \text{simple expr} \rangle ::= \langle \text{term} \rangle \mid \langle \text{sign} \rangle \langle \text{term} \rangle \mid$

$\langle \text{simple expr} \rangle \langle \text{adding operator} \rangle \langle \text{term} \rangle$

$\langle \text{sign} \rangle ::= + \mid -$

$\langle \text{adding operator} \rangle ::= + \mid - \mid \mathbf{or}$

CBL

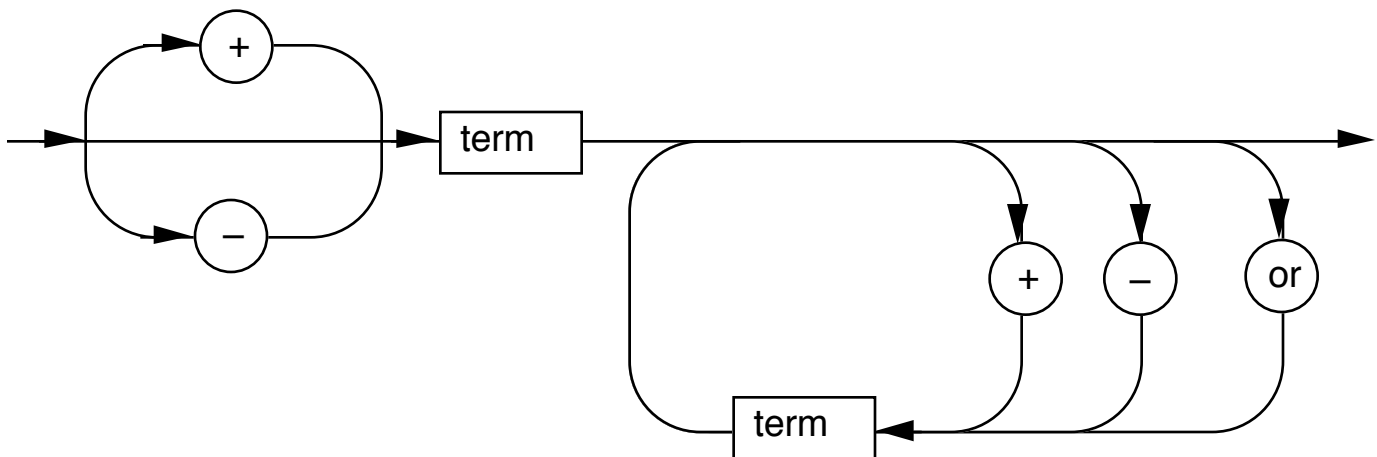
$$[\begin{matrix} + \\ - \end{matrix}] \langle \text{term} \rangle [\left\{ \begin{matrix} + \\ - \\ \text{or} \end{matrix} \right\} \langle \text{term} \rangle] \dots$$

EBNF

$\langle \text{simple expr} \rangle ::=$

$$[\langle \text{sign} \rangle] \langle \text{term} \rangle \{ \langle \text{adding operator} \rangle \langle \text{term} \rangle \}$$

Syntax Diagram



Exercise: According to the syntactic specification, which of these terminal strings are simple expressions, assuming that a, b, and c are legal terms:

- 1) a + b - c
- 2) - a or b + c
- 3) b - - c

Derivation or Parse Trees

The syntactic structure of a string of terminals as generated by a given grammar can be depicted as a **derivation tree**.

- Leaf nodes are labeled with terminal symbols.
- Each interior node is labeled by a nonterminal whose children represent the right side of a production for that nonterminal.
- The root is labeled with the syntactic category being derived, called the **start** symbol.

Example: A Restricted English Grammar

<sentence> ::= <subject> <predicate> .

<subject> ::= <simpsub> | <simpsub> **and** <subject>

<simpsub> ::= <qualifiers> <noun> | <noun>

<qualifiers> ::= <adjective> | <adjective> , <qualifiers>

<adjective> ::= **big** | **little**

<noun> ::= **dinners** | **books** | **Snoopy** | **Linus**

<predicate> ::=

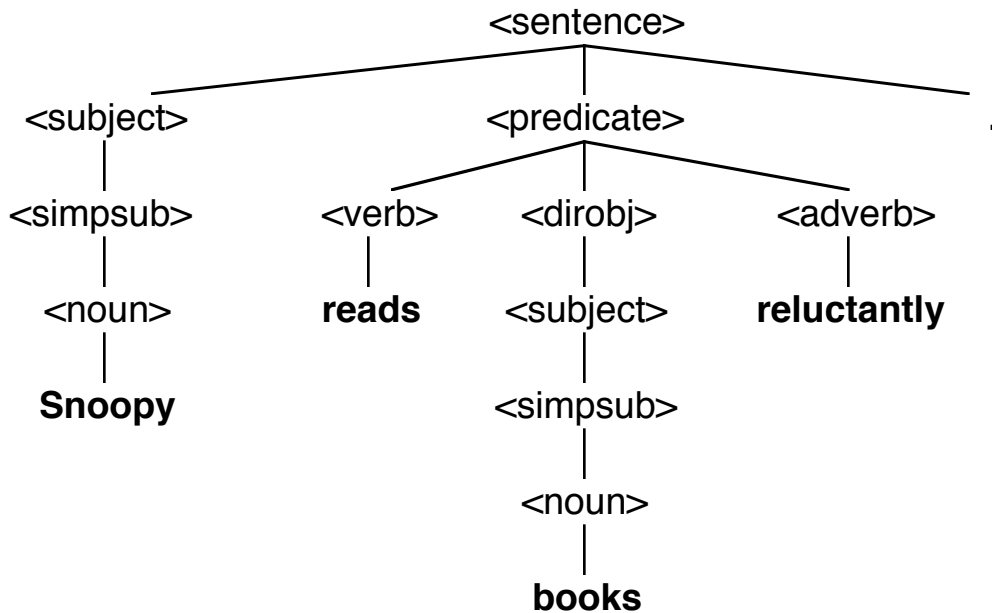
 <verb> | <verb> <diobj> | <verb> <diobj> <adverb>

<diobj> ::= <subject>

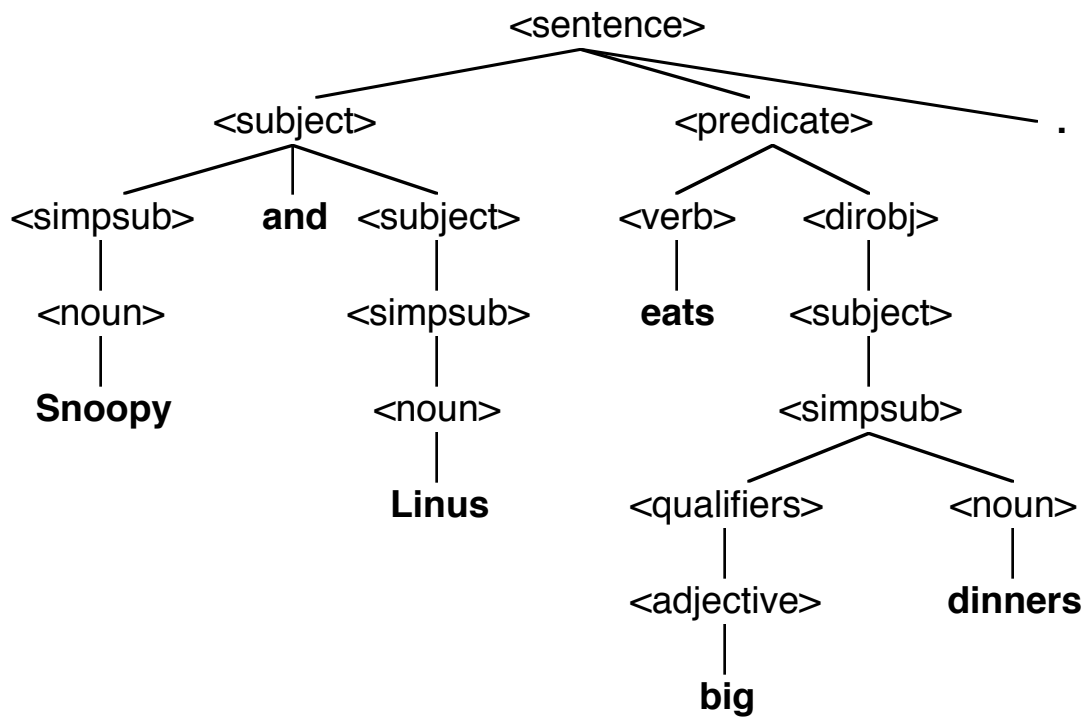
<verb> ::= **eats** | **reads**

<adverb> ::= **with feeling** | **reluctantly**

Snoopy reads books reluctantly.



Snoopy and Linus eats big dinners.



Exercises:

1. Draw a parse tree for the sentence:

little Snoopy eats Linus and books with feeling.

2. How many legal sentences are in the language?

3. Write descriptions of the restricted English grammar using:

a) CBL

b) EBNF

c) Syntax Diagrams

Expressions

Expressions are a vital component in programming languages.

Here is a simple language of expressions:

$\langle \text{expr} \rangle ::= \langle \text{name} \rangle \mid \langle \text{expr} \rangle \langle \text{operator} \rangle \langle \text{expr} \rangle$

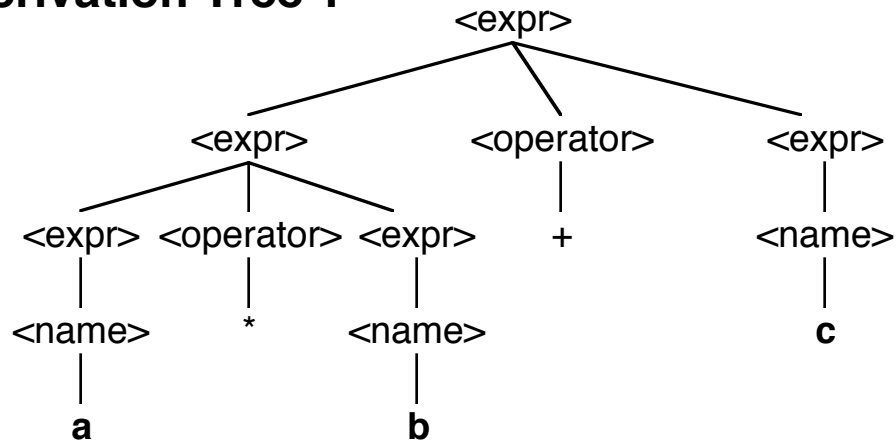
$\langle \text{name} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c}$

$\langle \text{operator} \rangle ::= + \mid *$

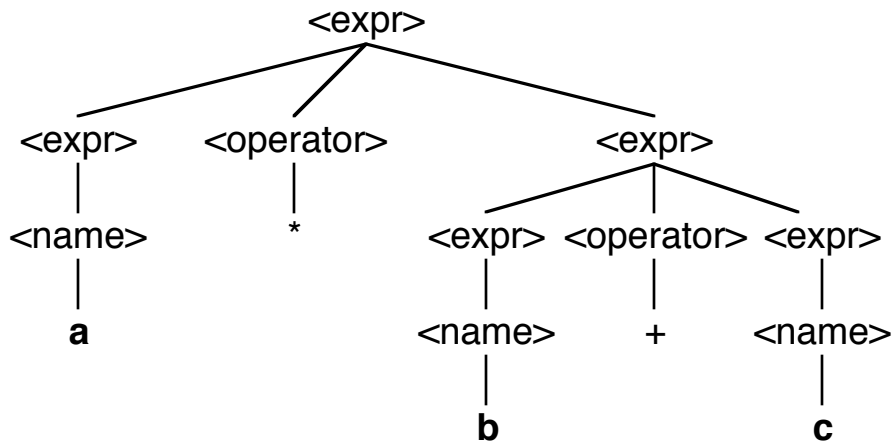
Consider the terminal string "**a * b + c**".

It allows two derivations in this language of expressions:

Derivation Tree 1



Derivation Tree 2



A grammar is **ambiguous** if some terminal string has two or more distinct structural descriptions (parse trees).

Revision 1: $\langle \text{expr} \rangle ::= \langle \text{name} \rangle \mid \langle \text{expr} \rangle \langle \text{operator} \rangle \langle \text{name} \rangle$

- called left recursion
- gives left-to-right evaluation of operators
- yields the first derivation tree

Revision 2: $\langle \text{expr} \rangle ::= \langle \text{name} \rangle \mid \langle \text{name} \rangle \langle \text{operator} \rangle \langle \text{expr} \rangle$

- called right recursion
- gives right-to-left evaluation of operators
- yields the second derivation tree
- this is the order of execution of all binary operators in APL

Note: Some (very strange) languages are inherently ambiguous—every grammar for the language is ambiguous.

Exercise: Look at $\langle \text{expression} \rangle$ in Java

- a) How are precedence levels provided for operators?
- b) How is the order of execution within a level determined?
- c) How are parentheses handled?

Another Syntactic Ambiguity: The dangling “else”

What is the meaning of this Java or C statement:

if (E₁) **if** (E₂) S₁; **else** S₂;

This ambiguity can be solved in several ways:

Java: A semantic solution specified using English—the statement is considered equivalent to

if (E₁) { **if** (E₂) S₁; **else** S₂; }

Algol: A syntactic solution

<conditional statement> ::=

if <Bool expr> **then** <unconditional statement> |

if <Bool expr> **then** <unconditional statement>

else <statement>

Note: A compound statement is unconditional.

Ada: A syntactic solution

if <condition> **then**

<sequence of statements>

[**else** <sequence of statements>]

end if;

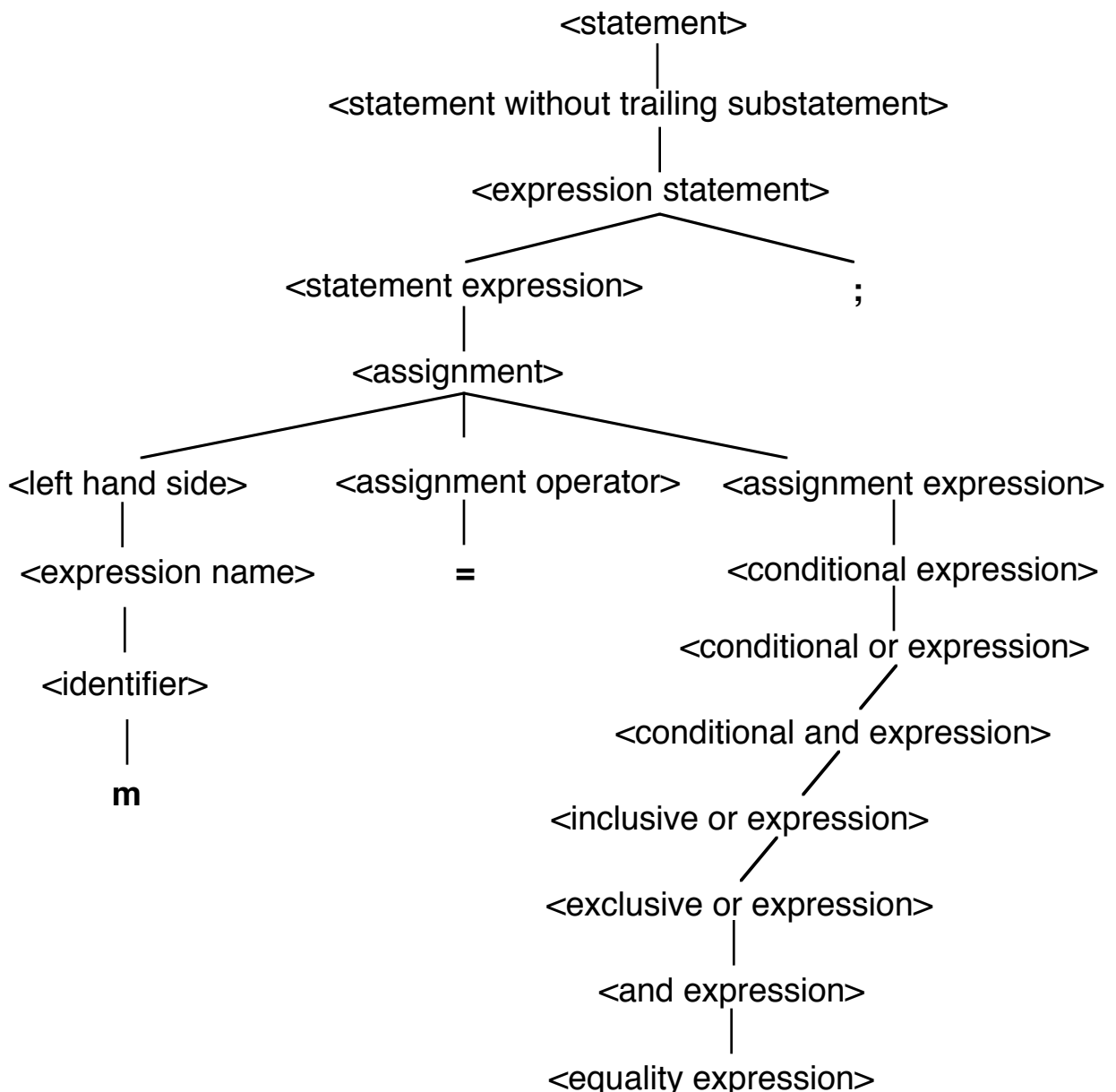
Each **if** needs an **end if**.

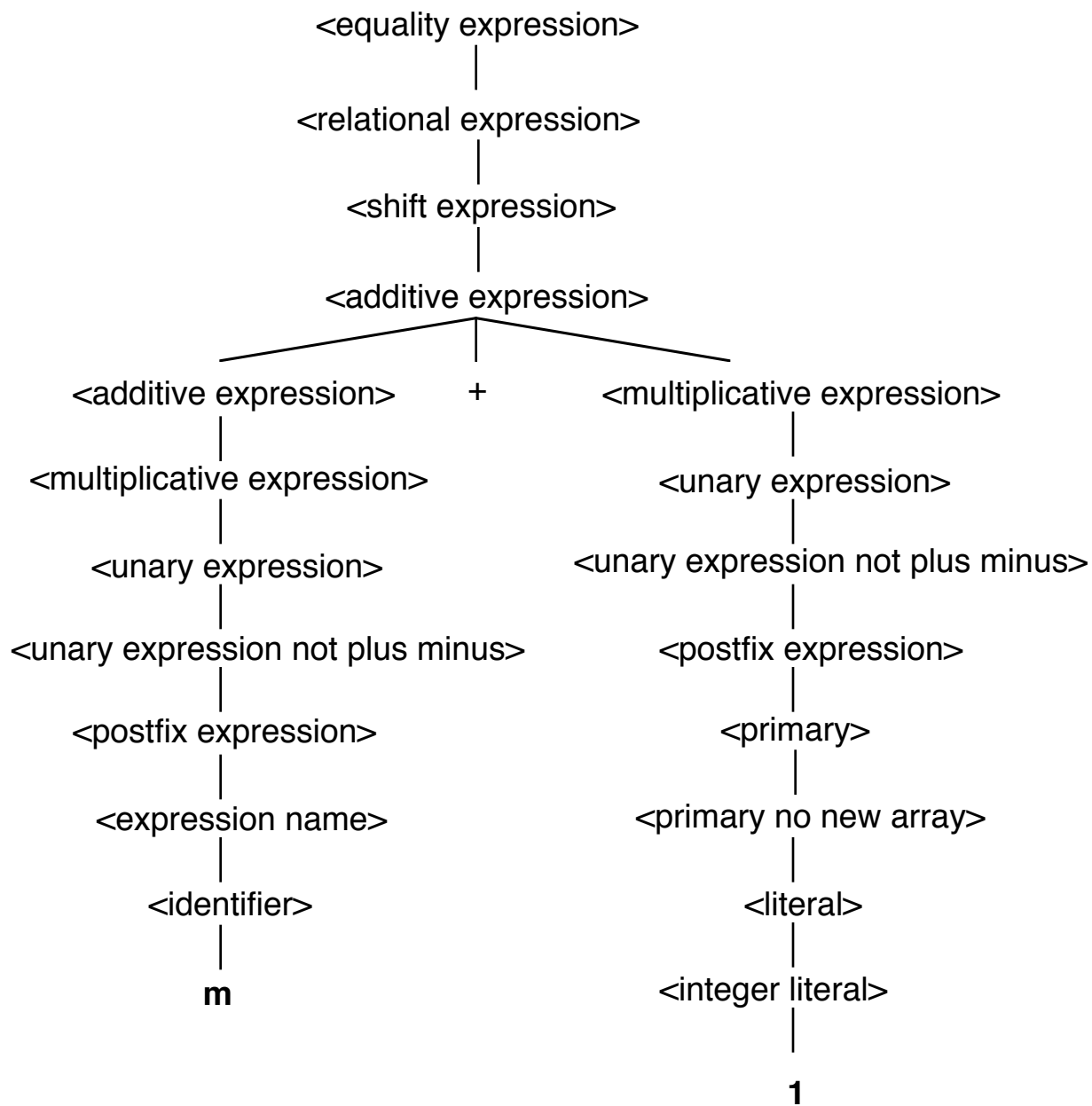
Parse Tree in Java

Each correctly formed construct in Java has a derivation tree defined by its BNF syntax.

This example uses the Concrete Syntax of Java at the end of this part.

Example: `m = m + 1;` is a correct `<statement>`





Limitations of Syntactic Definitions

- The Concrete Syntax for Java is an incomplete description of Java—not all the terminal strings generated are regarded as valid Java programs, although all legal programs can be derived from the BNF grammar.
- In fact, there is no BNF (or EBNF or Syntax Diagram) grammar that generates only the legal Java programs.
- A BNF grammar defines a **context-free** language: the left-hand side of each BNF rule contains only one syntactic category—it is replaced by one of its alternative definitions regardless of the context in which it occurs.
- Example of a **context-sensitive** grammar:

$\langle \text{sentence} \rangle ::= \mathbf{abc} \mid \mathbf{a}\langle \text{thing} \rangle \mathbf{bc}$

$\langle \text{thing} \rangle \mathbf{b} ::= \mathbf{b}\langle \text{thing} \rangle$

$\langle \text{thing} \rangle \mathbf{c} ::= \langle \text{other} \rangle \mathbf{bcc}$

$\mathbf{a}\langle \text{other} \rangle ::= \mathbf{aa} \mid \mathbf{aa}\langle \text{thing} \rangle$

$\mathbf{b}\langle \text{other} \rangle ::= \langle \text{other} \rangle \mathbf{b}$

The language generated = { **abc**, **aabbcc**, **aaabbbccc**, ... }

Sample Derivation:

$\langle \text{sentence} \rangle \Rightarrow \mathbf{a}\langle \text{thing} \rangle \mathbf{bc} \Rightarrow \mathbf{ab}\langle \text{thing} \rangle \mathbf{c}$

$\Rightarrow \mathbf{ab}\langle \text{other} \rangle \mathbf{bcc} \Rightarrow \mathbf{a}\langle \text{other} \rangle \mathbf{bbcc} \Rightarrow \mathbf{aabbcc}$

Exercise: Derive the $\langle \text{sentence} \rangle$ **aaabbbccc**

- The set of Java programs is not a context-free language; it has context-sensitive features.
- The context-sensitive features may be informally described as a set of restrictions or context conditions.
- They deal primarily with the declarations of identifiers, the types of variables and expressions, and the compatibility of objects and operations.
- Some examples:
 - 1) Operands for the **&&** operation must be of boolean type.
 - 2) In a method definition of a function, the return expression must be compatible with the return type of the method.
 - 3) When a method is called, the actual parameters must match the formal parameters in number and type.

Exercise: List four more context conditions in Java or C++.

Basic Parts of a Programming Language

1. Expression

Construct that produces (defines) a value.

2. Command (statement)

Construct that changes the state of the computer.

- change in memory
- produce output of some form
- change control state (position of execution in the program)

3. Declaration

Construct that binds an identifier (a name) to a programming language artifact.

Java Bindings

Identifier

Binding

variable

memory location

method

method definition (code segment)

class, interface

Class object

package

set of classes and interfaces
and directory path

constant

value

Assignment

The BNF for Java shows that an assignment in Java is really an expression.

Its value is the value assigned to the variable to the left of the equal sign.

Having assignment be an expression allows code such as the following:

```
m = n = 0;  
if ((count = count - 1) > 0) ...  
add(b = new JButton("Press"));
```

Other Statement Expressions

The other statement expressions can be viewed as normal expressions.

When used as a command, the value produced by such an expression is simply discarded.

Examples

```
num++;  
--count;  
Math.sqrt(100.0);    // a useless command  
new JFrame();
```

C (and C++)

In C *all* expressions can act as commands.

105. <statement> ::= <expression statement>

106. <expression statement> ::= <expression> ;

This design decision leads to some very strange programs.

```
main ()
{
    int n;
    n = 5;
    n++;
    n+1;
    123;
    printf("n = %d\n",n);
}
```

/* Output:

n = 6

*****/

Concrete Syntax for Java

Programs

1. `<goal> ::= <compilation unit>`
2. `<compilation unit> ::= <package declaration>? <import declarations>?
<type declarations>?`

Declarations

3. `<package declaration> ::= package <package name>;`
4. `<import declarations> ::= <import declaration>
| <import declarations> <import declaration>`
5. `<import declaration> ::= <single type import declaration>
| <type import on demand declaration>`
6. `<single type import declaration> ::= import <type name>;`
7. `<type import on demand declaration> ::= import <package name> . * ;`
8. `<type declarations> ::= <type declaration>
| <type declarations> <type declaration>`
9. `<type declaration> ::= <class declaration> | <interface declaration> | ;`
10. `<class declaration> ::= <class modifiers>? class <identifier> <super>?
<interfaces>? <class body>`
11. `<class modifiers> ::= <class modifier> | <class modifiers> <class modifier>`
12. `<class modifier> ::= public | abstract | final`
13. `<super> ::= extends <class type>`
14. `<interfaces> ::= implements <interface type list>`
15. `<interface type list> ::= <interface type> | <interface type list> , <interface type>`
16. `<class body> ::= { <class body declarations>? }`
17. `<class body declarations> ::= <class body declaration>
| <class body declarations> <class body declaration>`
18. `<class body declaration> ::= <class member declaration>
| <static initializer> | <constructor declaration>`
19. `<class member declaration> ::= <field declaration> | <method declaration>`

20. <static initializer> ::= **static** <block>
21. <constructor declaration> ::= <constructor modifiers>? <constructor declarator>
<throws>? <constructor body>
22. <constructor modifiers> ::= <constructor modifier>
| <constructor modifiers> <constructor modifier>
23. <constructor modifier> ::= **public** | **protected** | **private**
24. <constructor declarator> ::= <simple type name> (<formal parameter list>?)
25. <formal parameter list> ::= <formal parameter>
| <formal parameter list> , <formal parameter>
26. <formal parameter> ::= <type> <variable declarator id>
27. <throws> ::= **throws** <class type list>
28. <class type list> ::= <class type> | <class type list> , <class type>
29. <constructor body> ::= { <explicit constructor invocation>? <block statements>? }
30. <explicit constructor invocation> ::= **this** (<argument list>?)
| **super** (<argument list>?)
31. <field declaration> ::= <field modifiers>? <type> <variable declarators>;
32. <field modifiers> ::= <field modifier>
| <field modifiers> <field modifier>
33. <field modifier> ::= **public** | **protected** | **private** | **static** | **final** | **transient** | **volatile**
34. <variable declarators> ::= <variable declarator>
| <variable declarators> , <variable declarator>
35. <variable declarator> ::= <variable declarator id>
| <variable declarator id> = <variable initializer>
36. <variable declarator id> ::= <identifier> | <variable declarator id> []
37. <variable initializer> ::= <expression> | <array initializer>
38. <method declaration> ::= <method header> <method body>
39. <method header> ::=
<method modifiers>? <result type> <method declarator> <throws>?
40. <result type> ::= <type> | **void**
41. <method modifiers> ::= <method modifier>
| <method modifiers> <method modifier>

- 42. <method modifier> ::= **public** | **protected** | **private** | **static** | **abstract** | **final**
| **synchronized** | **native**
- 43. <method declarator> ::= <identifier> (<formal parameter list>?)
- 44. <method body> ::= <block> | ;
- 45. <interface declaration> ::= <interface modifiers>? **interface** <identifier>
| <extends interfaces>? <interface body>
- 46. <interface modifiers> ::= <interface modifier>
| <interface modifiers> <interface modifier>
- 47. <interface modifier> ::= **public** | **abstract**
- 48. <extends interfaces> ::= **extends** <interface type>
| <extends interfaces> , <interface type>
- 49. <interface body> ::= { <interface member declarations>? }
- 50. <interface member declarations> ::= <interface member declaration>
| <interface member declarations> <interface member declaration>
- 51. <interface member declaration> ::= <constant declaration>
| <abstract method declaration>
- 52. <constant declaration> ::= <constant modifiers> <type> <variable declarator>
- 53. <constant modifiers> ::= **public** | **static** | **final**
- 54. <abstract method declaration> ::=
<abstract method modifiers>? <result type> <method declarator> <throws>? ;
- 55. <abstract method modifiers> ::= <abstract method modifier>
| <abstract method modifiers> <abstract method modifier>
- 56. <abstract method modifier> ::= **public** | **abstract**
- 57. <array initializer> ::= { <variable initializers>? , ? }
- 58. <variable initializers> ::= <variable initializer>
| <variable initializers> , <variable initializer>
- 59. <variable initializer> ::= <expression> | <array initializer>

Types

- 60. <type> ::= <primitive type> | <reference type>
- 61. <primitive type> ::= <numeric type> | **boolean**

- 62. <numeric type> ::= <integral type> | <floating-point type>
- 63. <integral type> ::= **byte** | **short** | **int** | **long** | **char**
- 64. <floating-point type> ::= **float** | **double**
- 65. <reference type> ::= <class or interface type> | <array type>
- 66. <class or interface type> ::= <class type> | <interface type>
- 67. <class type> ::= <type name>
- 68. <interface type> ::= <type name>
- 69. <array type> ::= <type> []

Blocks and Commands

- 70. <block> ::= { <block statements>? }
- 71. <block statements> ::= <block statement> | <block statements> <block statement>
- 72. <block statement> ::= <local variable declaration statement> | <statement>
- 73. <local variable declaration statement> ::= <local variable declaration>;
- 74. <local variable declaration> ::= <type> <variable declarators>
- 75. <statement> ::= <statement without trailing substatement>
 | <labeled statement> | <if then statement> | <if then else statement>
 | <while statement> | <for statement>
- 76. <statement no short if> ::= <statement without trailing substatement>
 | <labeled statement no short if> | <if then else statement no short if>
 | <while statement no short if> | <for statement no short if>
- 77. <statement without trailing substatement> ::= <block> | <empty statement>
 | <expression statement> | <switch statement> | <do statement>
 | <break statement> | <continue statement> | <return statement>
 | <synchronized statement> | <throws statements> | <try statement>
- 78. <empty statement> ::= ;
- 79. <labeled statement> ::= <identifier> : <statement>
- 80. <labeled statement no short if> ::= <identifier> : <statement no short if>
- 81. <expression statement> ::= <statement expression>;
- 82. <statement expression> ::= <assignment> | <preincrement expression>
 | <postincrement expression> | <predecrement expression>

- 106. <try statement> ::= **try** <block> <catches> | **try** <block> <catches>? <finally>
- 107. <catches> ::= <catch clause> | <catches> <catch clause>
- 108. <catch clause> ::= **catch** (<formal parameter>) <block>
- 109. <finally > ::= **finally** <block>

Expressions

- 110. <constant expression> ::= <expression>
- 111. <expression> ::= <assignment expression>
- 112. <assignment expression> ::= <conditional expression> | <assignment>
- 113. <assignment> ::= <left hand side> <assignment operator> <assignment expression>
- 114. <left hand side> ::= <expression name> | <field access> | <array access>
- 115. <assignment operator> ::= = | *= | /= | %= | += | -= | <<= | >>=
 - | >>>= | &= | ^= | |=
- 116. <conditional expression> ::= <conditional or expression>
 - | <conditional or expression> ? <expression> : <conditional expression>
- 117. <conditional or expression> ::= <conditional and expression>
 - | <conditional or expression> || <conditional and expression>
- 118. <conditional and expression> ::= <inclusive or expression>
 - | <conditional and expression> && <inclusive or expression>
- 119. <inclusive or expression> ::= <exclusive or expression>
 - | <inclusive or expression> | <exclusive or expression>
- 120. <exclusive or expression> ::= <and expression>
 - | <exclusive or expression> ^ <and expression>
- 121. <and expression> ::= <equality expression>
 - | <and expression> & <equality expression>
- 122. <equality expression> ::= <relational expression>
 - | <equality expression> == <relational expression>
 - | <equality expression> != <relational expression>
- 123. <relational expression> ::= <shift expression>
 - | <relational expression> < <shift expression>
 - | <relational expression> > <shift expression>
 - | <relational expression> <= <shift expression>
 - | <relational expression> >= <shift expression>
 - | <relational expression> **instanceof** <reference type>

124. <shift expression> ::= <additive expression>
 | <shift expression> << <additive expression>
 | <shift expression> >> <additive expression>
 | <shift expression> >>> <additive expression>
125. <additive expression> ::= <multiplicative expression>
 | <additive expression> + <multiplicative expression>
 | <additive expression> - <multiplicative expression>
126. <multiplicative expression> ::= <unary expression>
 | <multiplicative expression> * <unary expression>
 | <multiplicative expression> / <unary expression>
 | <multiplicative expression> % <unary expression>
127. <cast expression> ::= (<primitive type>) <unary expression>
 | (<reference type>) <unary expression not plus minus>
128. <unary expression> ::= <preincrement expression>
 | <predecrement expression>
 | + <unary expression> | - <unary expression>
 | <unary expression not plus minus>
129. <predecrement expression> ::= -- <unary expression>
130. <preincrement expression> ::= ++ <unary expression>
131. <unary expression not plus minus> ::=
 <postfix expression> | ~ <unary expression>
 | ! <unary expression> | <cast expression>
132. <postdecrement expression> ::= <postfix expression> --
133. <postincrement expression> ::= <postfix expression> ++
134. <postfix expression> ::= <primary> | <expression name>
 | <postincrement expression> | <postdecrement expression>
135. <method invocation> ::= <method name> (<argument list>?)
 | <primary> . <identifier> (<argument list>?)
 | **super** . <identifier> (<argument list>?)
136. <field access> ::= <primary> . <identifier> | **super** . <identifier>
137. <primary> ::= <primary no new array> | <array creation expression>
138. <primary no new array> ::= <literal> | **this** | (<expression>)
 | <class instance creation expression> | <field access>
 | <method invocation> | <array access>
139. <class instance creation expression> ::= **new** <class type> (<argument list>?)
140. <argument list> ::= <expression> | <argument list> , <expression>

