Programming Languages: Expressions

Ryan Stansifer

Florida Institute of Technology
Melbourne, Florida USA 32901

http://www.cs.fit.edu/~ryan/

16 March 2020
Expressions
Overview of Topics on Expressions

- Literals
- Constants
- Operators and their operations
- Order of evaluation
  - precedence, parentheses, associativity, short-circuit
- Rewriting
- Referential transparency
  - Scott, 4th ed., Section 6.2.1, page 230
  - Mitchell, Section 4.4, page 78.

A cursory discussion of data values occurs in the discussion of type following this part.
Definition of a Literal

Definition

A *literal* is a lexical element of the program language syntax (not an identifier) standing for a specific, unchanging value. As such the value is known at compile time.
### Various Literals

<table>
<thead>
<tr>
<th>Literal</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>365</td>
<td>typical integer</td>
</tr>
<tr>
<td>5.11E-8</td>
<td>real number in “scientific” notation</td>
</tr>
<tr>
<td>170_234</td>
<td>integer in Ada</td>
</tr>
<tr>
<td>TRUE</td>
<td>boolean value in Modula-3</td>
</tr>
<tr>
<td>’A’</td>
<td>typical character</td>
</tr>
<tr>
<td>$a$</td>
<td>character in Smalltalk</td>
</tr>
<tr>
<td>#symbol</td>
<td>symbol or atom in Smalltalk</td>
</tr>
<tr>
<td>’atom’</td>
<td>atom in LISP</td>
</tr>
<tr>
<td>&quot;string&quot;</td>
<td>typical string</td>
</tr>
<tr>
<td>5HHELLO</td>
<td>old “Hollerith” string in FORTRAN</td>
</tr>
<tr>
<td>[]</td>
<td>empty list in PROLOG and ML</td>
</tr>
<tr>
<td>()</td>
<td>unit in ML</td>
</tr>
</tbody>
</table>
Strings are, most commonly, sequences of characters delimited by the US-ASCII double quote character. However, Python allows short strings (not containing newlines) and long strings. Long strings begin and end with *three* quote characters. Furthermore, either the USA-ASCII double-quote or single-quote character and be used.
Raw or Verbatim Strings

A raw string, sometimes call a verbatim string especially in .Net and C# community, has fewer meta character sequences that stand for other characters and more characters stand just for themselves.

Leaning Toothpick Syndrom

We give three examples.

- C#
- Python
- Java a raw string syntax did not make Java 12, but is expected in Java 13
- Scala
Verbaim or Raw Strings

C#: Verbatim string literals @"..." (no escape sequences except the quote escape sequence—two quote marks). See Special Verbatim Character †.

Python: short string literal, and long string literals (triple quoted strings). A string literal with 'f' or 'F' in its prefix is a formatted string literal; see Formatted string literals. The 'f' may be combined with 'r'.

Python raw string literal. Unless an 'r' or 'R' prefix is present, escape sequences in strings are interpreted according to rules similar to those used by Standard C.

```
r"""
    \d +  # the integral part
    \.    # the decimal point
    \d *  # some fractional digits"""
```
Verbatim or Raw Strings

In the context of strings intended to be regular expressions, Java has a unique “quote-unquote” escape sections. This should not be confused with raw strings. This Java string

"\n\Q\n \Q\\\E\t"

(maybe?!) matches (but is not the same as) this:

<table>
<thead>
<tr>
<th>binary</th>
<th>oct</th>
<th>dec</th>
<th>Latin1</th>
<th>Unicode</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>1010</td>
<td>0012</td>
<td>10</td>
<td>U+000A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>\E</td>
<td>line feed</td>
</tr>
<tr>
<td>0101</td>
<td>1100</td>
<td>0134</td>
<td>92</td>
<td>U+005C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>\</td>
<td>backslash</td>
</tr>
<tr>
<td>0110</td>
<td>1110</td>
<td>0156</td>
<td>110</td>
<td>U+006E</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>latin small letter n</td>
</tr>
<tr>
<td>0010</td>
<td>0000</td>
<td>0040</td>
<td>32</td>
<td>U+0020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>\</td>
<td>space</td>
</tr>
<tr>
<td>0101</td>
<td>1100</td>
<td>0134</td>
<td>92</td>
<td>U+005C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>\</td>
<td>backslash</td>
</tr>
<tr>
<td>0101</td>
<td>0001</td>
<td>0121</td>
<td>81</td>
<td>U+0051</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Q</td>
<td>latin capital letter Q</td>
</tr>
<tr>
<td>0101</td>
<td>1100</td>
<td>0134</td>
<td>92</td>
<td>U+005C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>\</td>
<td>backslash</td>
</tr>
<tr>
<td>0000</td>
<td>1001</td>
<td>0011</td>
<td>9</td>
<td>U+0009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HT</td>
<td>horizontal tabulation</td>
</tr>
</tbody>
</table>
Verbaim or Raw Strings

Scala

raw"a\nb"
s"Hello, $name"
f"$name%s is $height%2.2f meters tall"

C++

R"(a\nb)"
String Interpolation

Python (> 3.6)
```python
sys.stdout.write(f"Area is {abs((x-z)*(y-w)):.2f} square cm.\n")
```
Scala
```scala
println(s"Area is ${abs((x-z)*(y-w))}.2f square cm.\n")
```
C#
```csharp
Console.WriteLine($"Area is ${abs((x-z)*(y-w)):.2f} square cm.\n")
```
String Interpolation

```cpp
defib(n)
    return (n > 2) ? (fib(n-1) + fib(n-2)) : 1;

def main()
    print(fib(20) = \texttt{{fib(20)}});
A *constant* is an identifier whose r-value does not change at run time. If the value of a constant can be determined at compile time, it is said to be a *static constant*, sometimes called a *compile-time constant* or *manifest constant*. Static constants are useful to the compiler for constant folding and other optimizations. An example in Ada:

```ada
Limit : constant Integer := 10_000;
Low_Limit : constant Integer := Limit/10;
Tolerance : constant Float := Dispersion(1.15);
```
Different Kinds of Constants

In C# you have both compile-time and run-time constants:

```csharp
public const int _Year = 2008;

public static readonly DateTime _Now = new DateTime();
```
Single Assignment Style

Single assignment style: each identifier is assigned a value only once.
An example in Java:

```java
final int golden = (year % 19) + 1;
final int century = (year / 100) + 1;
int epact = (11*golden+20+clavian-gregorian) % 30;
if (epact==24 || (epact==25&&golden==11)) epact++;
final int sunday = moon + 7 - ((extra+moon) % 7);
final String month = (sunday>31) ? "April": "March";
```
Broadly speaking *algebra* is the study of mathematical symbols and the rules for using these symbols.

An *expression* is a syntactically meaningful notation written in linear form and defined recursively by applying symbols for operators to zero or more other expressions as operands. Nullary operators themselves are expressions and are often called constants. It is often the case that a collection of distinct symbols used as variables are allowed as operands. In in abstract algebra, an *algebraic structure* is a collection of finitary operations; a set with this structure is also called an algebra.

Sane people will organize an algebraic structure around one or more clearly defined domains of values with operators of fixed arities (often 2) and types.
Expressions

Expression. An *expression* is a construct representing one (integer, boolean, real, etc.) value. Examples in Ada:

- 1_234: Integer
- 23 + 56: Integer
- 34 mod 2: Integer
- 4*X + 92: Integer
- Final='A' or else Mid='A': Boolean
- X mod 2 = 0: Boolean
- 2*X**3 + Y**2: Integer
- X / 34.0: Float
- 3*(X-5): Integer
- (((A-1)*(N-1))/4) mod 2 = 0: Boolean
The Typical Syntax Of Expressions

<expression> ::= <literal>
<expression> ::= <identifier>
<expression> ::= "(" <expression> ")"
<expression> ::= <expression> <operator> <expression>

This grammar has some defects, but captures the essence of expressions in most programming languages. Notice that knowing the operators in a programming language tells one pretty much all there is to know.
### Ada Operators

<table>
<thead>
<tr>
<th>Operator</th>
<th>Type (Associativity)</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>or</code></td>
<td>1 disjunction</td>
</tr>
<tr>
<td><code>xor</code></td>
<td>1 exclusive or</td>
</tr>
<tr>
<td><code>or else</code></td>
<td>1 conditional or</td>
</tr>
<tr>
<td><code>and</code></td>
<td>1 conjunction</td>
</tr>
<tr>
<td><code>and then</code></td>
<td>1 conditional and</td>
</tr>
<tr>
<td><code>=</code></td>
<td>2 equality</td>
</tr>
<tr>
<td><code>/=</code></td>
<td>2 inequality</td>
</tr>
<tr>
<td><code>&lt;</code></td>
<td>2 less than</td>
</tr>
<tr>
<td><code>&lt;=</code></td>
<td>2 less or equal</td>
</tr>
<tr>
<td><code>&gt;</code></td>
<td>2 greater than</td>
</tr>
<tr>
<td><code>&gt;=</code></td>
<td>2 greater or equal</td>
</tr>
<tr>
<td><code>in</code></td>
<td>2 membership</td>
</tr>
<tr>
<td><code>not in</code></td>
<td>2 not member</td>
</tr>
<tr>
<td><code>+</code></td>
<td>3 addition</td>
</tr>
<tr>
<td><code>-</code></td>
<td>3 subtraction</td>
</tr>
<tr>
<td><code>&amp;</code></td>
<td>3 concatenation</td>
</tr>
<tr>
<td><code>*</code></td>
<td>4 multiplication</td>
</tr>
<tr>
<td><code>/</code></td>
<td>4 division</td>
</tr>
<tr>
<td><code>mod</code></td>
<td>4 modulus</td>
</tr>
<tr>
<td><code>rem</code></td>
<td>4 remainder</td>
</tr>
<tr>
<td><code>**</code></td>
<td>5 exponentiation</td>
</tr>
<tr>
<td><code>abs</code></td>
<td>5 absolute value</td>
</tr>
<tr>
<td><code>not</code></td>
<td>5 negation</td>
</tr>
</tbody>
</table>

** is *non* associative; all others are left associative
<table>
<thead>
<tr>
<th>Operator</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>1</td>
</tr>
<tr>
<td>AND</td>
<td>2</td>
</tr>
<tr>
<td>NOT</td>
<td>3</td>
</tr>
<tr>
<td>=</td>
<td>4</td>
</tr>
<tr>
<td>#</td>
<td>4</td>
</tr>
<tr>
<td>&lt;</td>
<td>4</td>
</tr>
<tr>
<td>&lt;=</td>
<td>4</td>
</tr>
<tr>
<td>&gt;</td>
<td>4</td>
</tr>
<tr>
<td>&gt;=</td>
<td>4</td>
</tr>
<tr>
<td>IN</td>
<td>4</td>
</tr>
<tr>
<td>+</td>
<td>5</td>
</tr>
<tr>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>&amp;</td>
<td>5</td>
</tr>
<tr>
<td>*</td>
<td>6</td>
</tr>
<tr>
<td>/</td>
<td>6</td>
</tr>
<tr>
<td>DIV</td>
<td>6</td>
</tr>
<tr>
<td>MOD</td>
<td>6</td>
</tr>
</tbody>
</table>

All infix operators are left associative.
C, C++, Java, C# Operators

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&amp;&amp;</td>
<td>2</td>
<td>conditional and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>^</td>
<td>4</td>
<td>exclusive or</td>
</tr>
<tr>
<td>&amp;</td>
<td>5</td>
<td>and</td>
</tr>
<tr>
<td>==, !=</td>
<td>6</td>
<td>equality and inequality</td>
</tr>
<tr>
<td>&lt;, &lt;=, &gt;, &gt;=</td>
<td>7</td>
<td>relational</td>
</tr>
<tr>
<td>instanceof</td>
<td>7</td>
<td>instance of class (Java)</td>
</tr>
<tr>
<td>&lt;&lt;, &gt;&gt;, &gt;&gt;&gt;</td>
<td>8</td>
<td>bit shift</td>
</tr>
<tr>
<td>+</td>
<td>9</td>
<td>addition (and concatenation in Java)</td>
</tr>
<tr>
<td>-</td>
<td>9</td>
<td>subtraction</td>
</tr>
<tr>
<td>*</td>
<td>10</td>
<td>multiplication</td>
</tr>
<tr>
<td>/</td>
<td>10</td>
<td>integer/real division</td>
</tr>
<tr>
<td>%</td>
<td>10</td>
<td>remainder</td>
</tr>
<tr>
<td>!</td>
<td>11</td>
<td>(unary) negation</td>
</tr>
</tbody>
</table>

All these infix operators are left associative.
Well-ingrained notions about the syntax of expressions:

- parenthesization,
- operator infix,
- operator precedence, and
- associativity of operators

cover the unfortunate difficulties in apprehending precisely the reading of compound expressions.
Among the important issues relating to the runtime evaluation of expressions is *short-circuit evaluation*. Short-circuit evaluation is possible with operators that may omit the evaluation of an operand (usually the second operand), when the final value is determined without knowing that operand’s value. The name comes from the analogy with an electrical systems that finds a circuit which is shorter than the one planned by the designer by jumping a narrow gap between wires or devices.
Short-circuit Evaluation

Some expressions denote “bad values” (runtime errors or nonterminating computations) only under certain circumstances. Knowing the order of evaluation allows the programmers to write simpler and clearer code.

if x<>0 and y/x < 1 then ... else ...

In Pascal, for example, this could lead to trouble when x has the value zero, for the language requires that the programmer assume the expression y/x may be evaluated. In Pascal the order of evaluation is left unspecified. Here is a selection from the Pascal report, page 21:

*The rules of Pascal neither require nor forbid the evaluation of the second part in such cases. This means that the programmer must assure that the second factor is well-defined, independent of the value of the first factor.*
Short-circuit Evaluation

if (very_unlikely() and very_costly()) then ... else ...

p = my_list;
while (p && p->key != val) p=p->next

p := my_list;
while (p <> nil) and (p^.key <> val) do p:=p^.next;

while (i<ub and A[i]<>' ') do i:=i+1;
fun square x = x * x;
fun sos (x,y) = (square x) + (square y);

sos (3,4)
  ==> (square 3) + (square 4) [Def’n of sos]
  ==> 3*3 + (square 4) [Def’n of square]
  ==> 9 + (square 4) [Def’n of *]
  ==> 9 + 4*4 [Def’n of square]
  ==> 9 + 16 [Def’n of *]
  ==> 25 [Def’n of +]

Expressions in imperative languages x or f(x) hardly ever denote the same value when they appear elsewhere. Because of :=, side effects, aliases, global variables, static variables, etc.
Significance

The significance of expressions is that they can often be evaluated in a simple manner.
Four rules / equations:

fun len x = foldr (const (+1)) 0 x

fun const x y = x

fun foldr f v [] = v
    foldr f v (x:xs) = f x (foldr f v xs)
Rewriting

\[
\text{len ("ab":"":"z":[])}
\]
Rewriting

len ("ab":"":"z":[])

foldr (const (+1)) 0 ("ab":":"":"z":[]) -- len rule
Rewriting

len ("ab":"":"z":[])  
foldr (const (+1)) 0 ("ab":"":"z":[]) -- len rule  
(const (+1)) "ab" (foldr (const (+1) 0) ("":"z":[]))
Rewriting

\[
\text{len } \left("\text{ab}"::"\text{z}"::[]\right)
\]

\[
\text{foldr } (\text{const } (+1)) \ 0 \ \left(\"\text{ab}"::\"\text{z}"::[]\right) \quad \text{-- len rule}
\]

\[
(\text{const } (+1)) \ \"\text{ab}\" \ \left(\text{foldr } (\text{const } (+1) \ 0) \ (\"\"::\"z"::[]))\right)
\]

\[
(+1) \ \left(\text{foldr } (\text{const } (+1) \ 0) \ (\"\"::\"z"::[]))\right) \quad \text{-- const rule}
\]
Rewriting

(+1) (((const (+1)) "") (foldr (const (+1)) 0 ("z" : [])))
Rewriting

(+1) (((const (+1)) "") (foldr (const (+1)) 0 ("z":[])))

(+1) ((+1) (foldr (const (+1)) 0 ("z":[])))
Rewriting

(+1) (((const (+1)) "") (foldr (const (+1)) 0 ("z":[]))))
(+1) ((+1) (foldr (const (+1)) 0 ("z":[]))))
(+1) ((+1) ((const (+1) "z") (foldr (const (+1)) 0 ([])))))
Rewriting

(+1) (((const (+1)) "") (foldr (const (+1)) 0 ("z":[])))
(+1) ((+1) (foldr (const (+1)) 0 ("z":[])))
(+1) ((+1) ((const (+1) "z") (foldr (const (+1)) 0 ([]))))
(+1) ((+1) ((+1) (foldr (const (+1)) 0 ([]))))
(+1) ((+1) ((+1) 0))
(+1) ((+1) 1)
(+1) 2
3
Referential Transparency

The Leibnizean principle:

*Eadem sunt, quorum unum potest substitui alteri salva veritate.*

*Those things are identical of which one can be substituted for the other without loss of truth*
Willard Quine uses the phrase to refer to the substitutivity of identities. For example, in the sentence

*Tully was a Roman.*

the word “Tully” may be replaced by “Cicero,” which was another name of the same man. But the sentence

*William Rufus was so-called because of the colour of his hair.*

becomes untrue if we replace “William Rufus” by another description of the same man, “King William II.”
We see the same idea reflected in Frege’s “Über Sinn und Bedeutung”, *Zeitschrift für Philosophie und philosophische Kritik*, new series, 100, 1892, pages 25–50.

The meaning of a sentence must remain unchanged when a part of the sentence is replaced by an expression having the same meaning.

die Bedeutung eines Satzes sein Wahrheitswert ist, so muß dieser unverändert bleiben, wenn ein Satzteil durch einen Ausdruck von derselben Bedeutung, aber anderem Sinne ersetzt wird. Und das ist in der Tat der Fall. Leibniz erklärt geradezu: "Eadem sunt, quae sibi mutuo substitui possunt, salva veritate".
The identity of indiscernibles:

*Things are the same which can be substituted for each other without loss of truth.*

*If our supposition that the reference of a sentence is its truth value is correct, the latter must remain unchanged when a part of the sentence is replaced by an expression having the same reference. And this is in fact the case. Leibniz gives the definition: 'Eadem sunt, quae sibi mutuo substitui possunt, salva veritate.' What else but the truth value could be found, that belongs quite generally to every sentence if the reference of its components is relevant, and remains unchanged by substitutions of the kind in question?*

Translation by Max Black.
A language is referentially transparent if any sub-expression can be replaced with any sub-expression of equal value anywhere in the language.
Can you find a context in a programming language in which 1+2 is not equal to three?
Can you find a context in a programming language in which \(1+2\) is not equal to three?

// Comment: 1+2 is three
"1+2"
1+2*3