

Computer Science 131

Programming Languages

September 12, 2000

Static and Dynamic Semantics

Syntax vs. Semantics

- Syntax
 - What phrases may occur where.
- Semantics:
 - What the phrases mean when put together.
 - What should the answer be?
 - How should execution proceed?

Purposes of a Language Definition

- For the programmer
 - Understanding the language
 - Reasoning about programs
- For the language implementor
 - Understanding what correct implementations must/may do
 - Deciding whether program transformations are correct
 - Facilitate multiple (compatible) implementations
- For the language designer
 - Recording design decisions
 - Understanding interaction between language features
 - Reasoning about the language

Formal Definitions?

- Why a formal semantics?
 - Informal definitions invariably contain ambiguities or errors.
 - Facilitates reasoning about the language
 - Facilitates reasoning about programs in the language
 - Facilitates reasoning about program transformations
 - May permit automatic generation of implementations
- Truth in advertising: still very hard to give a formal description of a full, real language
 - But can handle quite large subsets
 - Active research topic

Simple Arithmetic Expressions

- Abstract Syntax

$v ::= \underline{n}$	(values)
$e ::= v \mid e + e$	(expressions)
$p ::= e$	(programs)

- What does a program in this language mean?

Two Approaches to Formal Semantics

- Denotational semantics
 - The meaning of every expression is a mathematical object (a number, a function, etc.)
 - Compositionality: meaning of an expression is a function of the meanings of its sub-expressions.
 - E.g., the meaning of a while loop
while *b* do *e*
is calculated from the meanings of the guard expression ***b*** and of the loop body ***e***
 - Two expressions with the same meaning are interchangeable.

Two Approaches to Formal Semantics

- Operational semantics
 - Defines evaluation of (complete) programs
 - High-level specification of an interpreter
 - We can choose the level of abstraction
 - Which (if any) low-level machine details we want to describe
 - Data representations
 - Memory management
 - Which concepts considered primitive

Small-step Operational Semantics

- Defines a relation \rightarrow between programs corresponding to one step of execution

$$(3 + 4) + (5 + 3) \rightarrow 7 + (5 + 3)$$

$$7 + (5 + 3) \rightarrow 7 + 8$$

$$7 + 8 \rightarrow 15$$

Small-step Operational Semantics

- We define the relation \rightarrow^* to be the reflexive, transitive closure of \rightarrow .

$e \rightarrow^* e'$ if $e \rightarrow e'$.

$e \rightarrow^* e$ always.

$e \rightarrow^* e''$ if $e \rightarrow^* e'$ and $e' \rightarrow^* e''$
for some e' .

Small-step Operational Semantics

- We define the relation \rightarrow^* to be the reflexive, transitive closure of \rightarrow .

$$e \rightarrow^* e' \quad \text{if } e \rightarrow e' .$$

$$e \rightarrow^* e \quad \text{always.}$$

$$e \rightarrow^* e'' \quad \text{if } e \rightarrow^* e' \quad \text{and} \quad e' \rightarrow^* e'' \\ \text{for some } e' .$$

- For example, $(3+4)-(5-3) \rightarrow^* 5$
- A program p terminates with value v if $p \rightarrow^* v$
- A program p fails to terminate if $p \rightarrow p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow \dots$

Small-step Operational Semantics

- Equivalent definition: inference rules (hopefully familiar from CS 80)

$$\frac{e \rightarrow e'}{e \rightarrow^* e'} \qquad \frac{}{e \rightarrow^* e}$$

$$\frac{e \rightarrow^* e' \quad e' \rightarrow^* e''}{e \rightarrow^* e''}$$

Structured Operational Semantics

$$\frac{}{\underline{n}_1 + \underline{n}_2 \rightarrow \underline{n_1+n_2}}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 + e_2 \rightarrow e_1' + e_2}$$

$$\frac{e_2 \rightarrow e_2'}{e_1 + e_2 \rightarrow e_1 + e_2'}$$

Concurrency and Non-determinism?

$$\frac{}{\underline{n}_1 + \underline{n}_2 \rightarrow \underline{n_1+n_2}}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 + e_2 \rightarrow e_1' + e_2}$$

$$\frac{e_2 \rightarrow e_2'}{e_1 + e_2 \rightarrow e_1 + e_2'}$$

$$\begin{array}{l} (\underline{3} + \underline{4}) + (\underline{5} + \underline{3}) \rightarrow \underline{7} + (\underline{5} + \underline{3}) \\ (\underline{3} + \underline{4}) + (\underline{5} + \underline{3}) \rightarrow (\underline{3} + \underline{4}) + \underline{8} \end{array}$$

Left-to-Right Evaluation

$$\frac{}{\underline{n_1} + \underline{n_2} \rightarrow \underline{n_1+n_2}}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 + e_2 \rightarrow e_1' + e_2}$$

~~$$\frac{e_2 \rightarrow e_2'}{e_1 + e_2 \rightarrow e_1 + e_2'}$$~~

$$\frac{e_2 \rightarrow e_2'}{v_1 + e_2 \rightarrow v_1 + e_2'}$$

Adding Booleans

- Abstract Syntax

$v ::= \underline{n} \mid \underline{tt} \mid \underline{ff}$ (values)
 $e ::= v \mid e + e \mid e < e$ (expressions)
 $\mid \text{if } e \text{ then } e \text{ else } e$
 $p ::= e$ (programs)

- For example,

$\text{if } (\text{if } \underline{3} < \underline{5} \text{ then } \underline{ff} \text{ else } \underline{tt}) \text{ then } \underline{1+2} \text{ else } \underline{7+8}$

Dynamic Semantics

$$\frac{}{\underline{n}_1 + \underline{n}_2 \rightarrow \underline{n_1+n_2}}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 + e_2 \rightarrow e_1' + e_2}$$

$$\frac{e_2 \rightarrow e_2'}{v_1 + e_2 \rightarrow v_1 + e_2'}$$

$$\frac{}{\underline{n}_1 < \underline{n}_2 \rightarrow \underline{n_1 < n_2}}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 < e_2 \rightarrow e_1' < e_2}$$

$$\frac{e_2 \rightarrow e_2'}{v_1 < e_2 \rightarrow v_1 < e_2'}$$

Dynamic Semantics

$$\frac{e_1 \rightarrow e_1'}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 \rightarrow \text{if } e_1' \text{ then } e_2 \text{ else } e_3}$$

What other rules?

Dynamic Semantics

$$\frac{e_1 \rightarrow e_1'}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 \rightarrow \text{if } e_1' \text{ then } e_2 \text{ else } e_3}$$
$$\frac{}{\text{if } \underline{tt} \text{ then } e_2 \text{ else } e_3 \rightarrow e_2}$$
$$\frac{}{\text{if } \underline{ff} \text{ then } e_2 \text{ else } e_3 \rightarrow \underline{e_3}}$$

Stuck programs

- We now have the possibility of syntactically well-formed programs that have not yielded a value, but which cannot make progress!

```
3 + tt  
if tt < ff then 3 else 5
```

Stuck programs

- We now have the possibility of syntactically well-formed programs that have not yielded a value, but which cannot make progress!

3 + tt

if tt < ff then 3 else 5

- Answer 1: who cares?
 - Implementation-dependent behavior
 - Or, program should stop with error.

Stuck programs

- We now have the possibility of syntactically well-formed programs that have not yielded a value, but which cannot make progress!

3 + tt

if tt < ff then 3 else 5

- Answer 1: who cares?
 - Implementation-dependent behavior
 - Or, program should stop with error.
- Answer 2: a type system to prevent this

Typing

- We define a collection of types. For now, just
 $t ::= \text{Int} \mid \text{Bool}$
- We define the relation $e : t$
 - A collection of inference rules is used to define when this relation holds.
- The type system is frequently called the "static semantics" of the language.

Static Semantic Rules

$\frac{}{\underline{n} : \text{Int}}$ $\frac{}{\underline{tt} : \text{Bool}}$ $\frac{}{\underline{ff} : \text{Bool}}$

Static Semantic Rules

n : Int **tt** : Bool **ff** : Bool

???

e₁ + e₂ : Int

Static Semantic Rules

$$\frac{}{\underline{n} : \text{Int}} \quad \frac{}{\underline{tt} : \text{Bool}} \quad \frac{}{\underline{ff} : \text{Bool}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 + e_2 : \text{Int}}$$

Static Semantic Rules

$$\frac{}{\underline{n} : \text{Int}} \quad \frac{}{\underline{tt} : \text{Bool}} \quad \frac{}{\underline{ff} : \text{Bool}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 + e_2 : \text{Int}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 < e_2 : \text{Bool}}$$

Static Semantic Rules

$$\frac{}{\underline{n} : \text{Int}} \quad \frac{}{\underline{tt} : \text{Bool}} \quad \frac{}{\underline{ff} : \text{Bool}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 + e_2 : \text{Int}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 < e_2 : \text{Bool}}$$
$$\frac{e_1 : ? \quad e_2 : ? \quad e_3 : ?}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 : ?}$$

Static Semantic Rules

$$\frac{}{\underline{n} : \text{Int}} \quad \frac{}{\underline{tt} : \text{Bool}} \quad \frac{}{\underline{ff} : \text{Bool}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 + e_2 : \text{Int}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 < e_2 : \text{Bool}}$$
$$\frac{e_1 : \text{Bool} \quad e_2 : ? \quad e_3 : ?}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 : ?}$$

Static Semantic Rules

$$\frac{}{\underline{n} : \text{Int}} \quad \frac{}{\underline{tt} : \text{Bool}} \quad \frac{}{\underline{ff} : \text{Bool}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 + e_2 : \text{Int}}$$
$$\frac{e_1 : \text{Int} \quad e_2 : \text{Int}}{e_1 < e_2 : \text{Bool}}$$
$$\frac{e_1 : \text{Bool} \quad e_2 : t \quad e_3 : t}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 : t}$$

Claim: Type Soundness

- Well-typed programs (i.e., programs that have some type) can't get stuck
 - That is, must either eventually reach a value or fail to terminate
- Why is this important? How to prove this?
 - See next lecture

But...

- Some programs wouldn't get stuck but still don't typecheck

`(if ff then tt else 4) + 1`

- For any interesting language, a type system preventing all bad programs also rejects programs that would run without problems.
- Research topic: type systems that catch as many errors as possible, but don't reject useful programs

Adding Local Definitions

- Abstract Syntax

$v ::= \underline{n} \mid \underline{tt} \mid \underline{ff}$ (values)
 $e ::= v \mid e + e \mid e < e$ (expressions)
 $\mid \text{if } e \text{ then } e \text{ else } e$
 $\mid x$
 $\mid \text{let } x \text{ be } e \text{ in } e$
 $p ::= e$ (programs)

Scoping For This Language

- In the expression **let x be e_1 in e_2**
 - The variable x is bound
 - The scope of x (where it can be referenced) is e_2

Scoping For This Language

- All α -equivalent expressions are identified
 - E.g., the following are the same expressions

let x be 3 in x + y

let z be 3 in z + y

- But not

let y be 3 in y + y

Changes to Dynamic Semantics

- Add the two rules:

$$\frac{e_1 \rightarrow e_1'}{\text{let } x \text{ be } e_1 \text{ in } e_2 \rightarrow \text{let } x \text{ be } e_1' \text{ in } e_2}$$

$$\frac{}{\text{let } x \text{ be } v_1 \text{ in } e_2 \rightarrow e_2[x \rightarrow v_1]}$$

Changes to Dynamic Semantics

- What if we had just this single rule instead?

$$\frac{}{\text{let } x \text{ be } e_1 \text{ in } e_2 \rightarrow e_2[x \rightarrow e_1]}$$

- Consider **let x be 1+2 in x+x**

Changes to the Static Semantics

- Typing is now *context-sensitive*
 - What is the type of **x** ?
 - Is **x+3** well-typed?

Changes to the Static Semantics

- Typing is now *context-sensitive*
 - What is the type of **x** ?
 - Is **x+3** well-typed?
 - Yes, in the program **let x be 4 in x+3**
 - No, in the program **let x be tt in x+3**
- Need to know the types of *free* variables

Changes to the Static Semantics

- 3-place typing relation $\Gamma \vdash e : t$
 - Here Γ is a *type environment*
 - Mapping from variables to their types
 - Type environments will be written as a list
 - e.g., **$x:\text{Int}, y:\text{Bool}, z:\text{Int}$**
 - The notation $\Gamma(\mathbf{x})$ gives the type of \mathbf{x} (lookup)
 - The notation
$$\Gamma, \mathbf{x}:\text{Int}$$
is the extension of Γ that maps \mathbf{x} to **Int** (insert)
 - We write $\vdash e : t$ when type env. is empty

Static Semantics

$$\frac{}{\Gamma \vdash \underline{n} : \text{Int}} \quad \frac{}{\Gamma \vdash \underline{tt} : \text{Bool}} \quad \frac{}{\Gamma \vdash \underline{ff} : \text{Bool}}$$

Static Semantics

$$\frac{}{\Gamma \vdash \underline{n} : \text{Int}} \quad \frac{}{\Gamma \vdash \underline{tt} : \text{Bool}} \quad \frac{}{\Gamma \vdash \underline{ff} : \text{Bool}}$$
$$\frac{\Gamma \vdash e_1 : \text{Int} \quad \Gamma \vdash e_2 : \text{Int}}{\Gamma \vdash e_1 + e_2 : \text{Int}}$$
$$\frac{\Gamma \vdash e_1 : \text{Int} \quad \Gamma \vdash e_2 : \text{Int}}{\Gamma \vdash e_1 < e_2 : \text{Bool}}$$
$$\frac{\Gamma \vdash e_1 : \text{Bool} \quad \Gamma \vdash e_2 : t \quad \Gamma \vdash e_3 : t}{\Gamma \vdash \text{if } e_1 \text{ then } e_2 \text{ else } e_3 : t}$$

Static Semantics: The Interesting Rules

$$\Gamma \vdash \mathbf{x} : ?$$

Static Semantics: The Interesting Rules

$$\frac{}{\Gamma \vdash \mathbf{x} : \Gamma(\mathbf{x})}$$

Static Semantics: The Interesting Rules

$$\frac{}{\Gamma \vdash \mathbf{x} : \Gamma(\mathbf{x})}$$

$$\frac{\text{???}}{\Gamma \vdash \text{let } \mathbf{x} \text{ be } e_1 \text{ in } e_2 : t_2}$$

Static Semantics: The Interesting Rules

$$\frac{}{\Gamma \vdash \mathbf{x} : \Gamma(\mathbf{x})}$$

$$\frac{\Gamma \vdash \mathbf{e}_1 : \mathbf{t}_1 \quad \Gamma, \mathbf{x}:\mathbf{t}_1 \vdash \mathbf{e}_2 : \mathbf{t}_2}{\Gamma \vdash \text{let } \mathbf{x} \text{ be } \mathbf{e}_1 \text{ in } \mathbf{e}_2 : \mathbf{t}_2}$$

Claim: Type Soundness

- If a program is well-typed *in the empty type environment* (i.e., $\vdash p : t$ for some t) then it cannot get stuck
 - Note that such programs must have no free variables.

Adding Recursive Functions

- Abstract Syntax

$v ::= \underline{n} \mid \underline{tt} \mid \underline{ff}$ (values)
 $\mid \text{fun } x(x:t):t \text{ is } e$

$e ::= v \mid e + e \mid e < e$ (expressions)
 $\mid \text{if } e \text{ then } e \text{ else } e$
 $\mid x \mid \text{let } x \text{ be } e \text{ in } e$
 $\mid e e$

$p ::= e$ (programs)

WARNING

- This is not SML!
- The expression **fun f(x:t₁):t₂ is e**
 - is a function value, like SML's **fn x => e** except that it can be recursive.
 - The equivalent SML code would be

```
let  
    fun f(x:t1):t2 = e  
in  
    f  
end
```

Scoping

- In the expression **fun $f(x:t_1):t_2$ is e**
 - Both **f** and **x** as bound variables
 - The name of the argument is **x**
 - The *local* name of the function is **f**
 - Both **f** and **x** have **e** as their scope

α -conversion

- The following are the exact same programs:

```
fun f(x:Int):Int is f(h(x+y))
```

```
fun g(z:Int):Int is f(h(x+y))
```

- But not:

```
fun h(y:Int):Int is h(h(y+y))
```

Example Code

```
(fun f(x:Int):Int is x+1) 3
```

Example Code

let

```
fact be (fun g(y:Int):Int is  
        if (y=0) then 1  
        else y*(g(y-1)))
```

in

```
(fact 3) + (fact 4)
```

Dynamic Semantics

- Add the rules

$$\frac{\mathbf{e}_1 \rightarrow \mathbf{e}_1'}{\mathbf{e}_1 \ \mathbf{e}_2 \rightarrow \mathbf{e}_1' \ \mathbf{e}_2}$$

$$\frac{\mathbf{e}_2 \rightarrow \mathbf{e}_2'}{\mathbf{v}_1 \ \mathbf{e}_2 \rightarrow \mathbf{v}_1 \ \mathbf{e}_2'}$$

Static Semantics

$t ::= \text{Int} \mid \text{Bool} \mid t \rightarrow t$

Static Semantics

$t ::= \text{Int} \mid \text{Bool} \mid t \rightarrow t$

$$\frac{\Gamma \vdash e_1 : ? \quad \Gamma \vdash e_2 : ?}{\Gamma \vdash e_1 e_2 : t}$$

Static Semantics

$t ::= \text{Int} \mid \text{Bool} \mid t \rightarrow t$

$$\frac{\Gamma \vdash e_1 : t_2 \rightarrow t \quad \Gamma \vdash e_2 : t_2}{\Gamma \vdash e_1 e_2 : t}$$

Static Semantics

$t ::= \text{Int} \mid \text{Bool} \mid t \rightarrow t$

$$\frac{\Gamma \vdash e_1 : t_2 \rightarrow t \quad \Gamma \vdash e_2 : t_2}{\Gamma \vdash e_1 e_2 : t}$$

???

$\Gamma \vdash \text{fun } f(x:t_1):t_2 \text{ is } e : ???$

Static Semantics

$t ::= \text{Int} \mid \text{Bool} \mid t \rightarrow t$

$$\frac{\Gamma \vdash e_1 : t_2 \rightarrow t \quad \Gamma \vdash e_2 : t_2}{\Gamma \vdash e_1 e_2 : t}$$

???

$$\Gamma \vdash \text{fun } f(x:t_1):t_2 \text{ is } e : t_1 \rightarrow t_2$$

Static Semantics

$t ::= \text{Int} \mid \text{Bool} \mid t \rightarrow t$

$$\frac{\Gamma \vdash e_1 : t_2 \rightarrow t \quad \Gamma \vdash e_2 : t_2}{\Gamma \vdash e_1 e_2 : t}$$

$$\frac{\Gamma, x:t_1, f:t_1 \rightarrow t_2 \vdash e : t_2}{\Gamma \vdash \text{fun } f(x:t_1):t_2 \text{ is } e : t_1 \rightarrow t_2}$$

Dynamic Semantics

- Add the rules

$$\frac{\mathbf{e}_1 \rightarrow \mathbf{e}_1'}{\mathbf{e}_1 \ \mathbf{e}_2 \rightarrow \mathbf{e}_1' \ \mathbf{e}_2}$$

$$\frac{\mathbf{e}_2 \rightarrow \mathbf{e}_2'}{\mathbf{v}_1 \ \mathbf{e}_2 \rightarrow \mathbf{v}_1 \ \mathbf{e}_2'}$$

Dynamic Semantics

- Add the rules

$$\frac{e_1 \rightarrow e_1'}{e_1 e_2 \rightarrow e_1' e_2}$$

$$\frac{e_2 \rightarrow e_2'}{v_1 e_2 \rightarrow v_1 e_2'}$$

$$(\text{fun } f(x:t_1):t_2 \text{ is } e_2) v \rightarrow ???$$

Dynamic Semantics

- Add the rules

$$\frac{e_1 \rightarrow e_1'}{e_1 \ e_2 \rightarrow e_1' \ e_2}$$

$$\frac{e_2 \rightarrow e_2'}{v_1 \ e_2 \rightarrow v_1 \ e_2'}$$

$$\frac{}{(\text{fun } f(x:t_1):t_2 \text{ is } e_2) \ v \rightarrow e_2[x \rightarrow ?, f \rightarrow ?]}$$

Dynamic Semantics

- Add the rules

$$\frac{e_1 \rightarrow e_1'}{e_1 \ e_2 \rightarrow e_1' \ e_2}$$

$$\frac{e_2 \rightarrow e_2'}{v_1 \ e_2 \rightarrow v_1 \ e_2'}$$

$$\frac{}{(\text{fun } f(x:t_1):t_2 \text{ is } e_2) \ v \rightarrow (e_2[x \rightarrow v])[f \rightarrow ?]}$$

Dynamic Semantics

- Add the rules

$$\frac{e_1 \rightarrow e_1'}{e_1 \ e_2 \rightarrow e_1' \ e_2}$$

$$\frac{e_2 \rightarrow e_2'}{v_1 \ e_2 \rightarrow v_1 \ e_2'}$$

$$\begin{array}{l} (\text{fun } f(x:t_1):t_2 \text{ is } e) \ v \ \rightarrow \\ (e_2[x \rightarrow v])[f \rightarrow (\text{fun } f(x:t_1):t_2 \text{ is } e)] \end{array}$$