CIS 624: Structure of Programming Languages

Lecture 5 — Pseudo-Denotational Semantics

Boyana Norris

2017
Styles of formal semantics

Operational: Meanings for program phrases defined in terms of the steps of computation they can take during program execution.

Axiomatic: Meanings for program phrases defined indirectly via the axioms and rules of some logic of program properties (Hoare logic).

Denotational: Concerned with giving mathematical models of programming languages. Meaning for program phrases defined abstractly as elements of suitable mathematical structure.
Operational vs denotational semantics

Operational semantics defines an interpreter, from abstract syntax to abstract syntax. Metalanguage is inference rules (slides) or OCaml (interp.ml)

Denotational semantics defines a compiler (translator), from abstract syntax to a different language with known semantics

Target language is math, but we’ll make it a tiny core of OCaml (hence “pseudo”)

Metalanguage is math or OCaml (we’ll show both)
Operational vs denotational semantics (cont.)

Operational semantics is:

► simple
► of many flavors (natural/large-step, small-step, more or less abstract)
► good for: language definition, verifying language properties (e.g., IMP is deterministic), verifying correctness of tools that manipulate programs (e.g., interpreters, compilers, type checkers, etc.)
► disadvantages: awkward for verifying even simple programs (remember lecture 3?), not compositional

Denotational semantics is:

► mathematical (the meaning of a syntactic expression is a mathematical object)
► compositional

Denotational semantics is also called: fixed-point semantics, mathematical semantics, Scott-Strachey semantics.
Denotational semantics

Each phrase (i.e., part of a program), \( P \), is given a denotation, \( \llbracket P \rrbracket \) — a mathematical object representing the contribution of \( P \) to the meaning of any complete program in which it occurs.

The denotation of a phrase is determined just by the denotations of its sub phrases (i.e., the semantics is compositional).

What does it mean for a formal semantic system to be compositional? It means that the collection of mathematical objects we use to give denotations to program phases has to be sufficiently rich that it supports operations for modeling all the phrase-forming constructs of the program language in question (e.g., while loops can be tricky; more on that later).
The basic idea

A heap is a math/ML function from strings to integers:

\[ \text{string} \rightarrow \text{int} \]

An expression denotes a math/ML function from heaps to integers

\[ \text{den}(e) : (\text{string} \rightarrow \text{int}) \rightarrow \text{int} \]

A statement denotes a math/ML function from heaps to heaps

\[ \text{den}(s) : (\text{string} \rightarrow \text{int}) \rightarrow (\text{string} \rightarrow \text{int}) \]

Now just define \textit{den} in our metalanguage (math or ML), inductively over the source language abstract syntax
Expressions

\[ den(e) : (string \rightarrow int) \rightarrow int \]

\[
\begin{align*}
den(c) &= \text{fun } h \rightarrow c \\
den(x) &= \text{fun } h \rightarrow h \ x \\
den(e_1 + e_2) &= \text{fun } h \rightarrow (den(e_1) \ h) + (den(e_2) \ h) \\
den(e_1 \times e_2) &= \text{fun } h \rightarrow (den(e_1) \ h) \times (den(e_2) \ h)
\end{align*}
\]

In plus (and times) case, two “ambiguities”:

- “+” from meta language or target language?
  - Translate abstract + to OCaml +, (ignoring overflow)

- *When* do we denote \( e_1 \) and \( e_2 \)?
  - Not a focus of the metalanguage. At “compile time”. 
Switching metalanguage

With OCaml as our metalanguage, ambiguities go away

But it is harder to distinguish mentally between “target” and “meta”

If denote in function body, then source is “around at run time”

- After translation, should be able to “remove” the definition of the abstract syntax

- ML does not have such a feature, but the point is we no longer need the abstract syntax

See denote.ml
Statements, w/o while

\[ den(s) : (\text{string} \rightarrow \text{int}) \rightarrow (\text{string} \rightarrow \text{int}) \]

\[ den(\text{skip}) = \text{fun } h \rightarrow h \]
\[ den(x := e) = \text{fun } h \rightarrow (\text{fun } v \rightarrow \text{if } x=v \text{ then } den(e) h \text{ else } h v) \]
\[ den(s_1; s_2) = \text{fun } h \rightarrow den(s_2) (den(s_1) h) \]
\[ den(\text{if } e \ s_1 \ s_2) = \text{fun } h \rightarrow \text{if } den(e) h > 0 \text{ then } den(s_1) h \text{ else } den(s_2) h \]

Same ambiguities; same answers

See denote.ml
While

\[ \text{den(while} \ e \ s) = \begin{cases} \text{let rec } f \ h = & \text{let d1=denote_exp e in} \\ & \text{let d2=denote_stmt s in} \\ & \text{if (den(e) h)>0} \\ & \text{let rec } f \ h = \\ & \text{then f (den(s) h)} \\ & \text{else h in} \\ & f \\ \end{cases} \]

The function denoting a while statement is inherently recursive!

Good thing our target language has recursive functions!

Why doesn’t \( \text{den(while} \ e \ s) = \text{den(if} \ e \ (s; \text{while} \ e \ s) \text{skip}) \) make any sense?
Two common mistakes

A denotational semantics should “eagerly” translate the entire program

▶ E.g., both branches of an if

But a denotational semantics should “terminate”

▶ I.e., avoid any circular definitions in the translating
▶ The result of the translation can use (well-founded) recursion
▶ E.g., compiling a while-loop should not produce an infinite amount of code
Finishing the story

```ocaml
define denote_prog s =
  let d = denote_stmt s in
  fun () -> (d (fun x -> 0)) "ans"
```

Compile-time: let x = denote_prog (parse file)

Run-time: print_int (x ())

In-between: We have a OCaml program using only functions, variables, ifs, constants, +, *, >, etc.

- Does not use any constructors of exp or stmt (e.g., Seq)
The real story

For “real” denotational semantics, target language is math

(And we write $[s]$ instead of $\text{den}(s)$)

Example: $[x := e][H] = [H][x \mapsto [e][H]]$

There are two major problems, both due to while:

1. Math functions do not diverge, so no function denotes while 1 skip
2. The denotation of loops cannot be circular
The elevator version, which we will not pursue

For (1), we “lift” the semantic domains to include a special $\perp$

$$den(s) : \text{string} \rightarrow \text{int} \rightarrow ((\text{string} \rightarrow \text{int}) \cup \perp)$$

- Have to change meaning of $den(s_2) \circ den(s_1)$ appropriately

For (2), we use `while e s` to define a (meta)function $f$ that given a lifted heap-transformer $X$ produces a lifted heap-transformer $X'$:

- If $den(e)(den(H)) = 0$, then $den(H)$
- Else $X \circ den(s)$

Now let $den(\text{while } e \ s)$ be the least fixed-point of $f$

- An hour of math to prove the least fixed-point exists
- Another hour to prove it is the limit of starting with $\perp$ and applying $f$ over and over (i.e., any number of loop iterations)
- Keywords: monotonic functions, complete partial orders, Knaster-Tarski theorem
Practical applications

Denotational semantics is used in practice as the basis for **static analysis** of programs (all compilers and many other tools do this!)

- Used for discovering properties of all possible dynamic executions (e.g., for code optimization, bug finding, security, ...)

Active area of research, key conferences:

- Principles of Programming Languages (POPL), Computer-Aided Verification (CAV), Programming Language Design and Implementation (PLDI).

CIS 410/510 Program Analysis and Transformation covers some of this theory and applies it to real software.)
Where we are

- Have seen operational and denotational semantics
- Connection to interpreters and compilers
- Useful for rigorous definitions and proving properties
- Next: Equivalence of semantics
  - Crucial for compiler writers
  - Crucial for code maintainers
- Then: Leave IMP behind and consider functions

But first: Will any of this help write an O/S service?