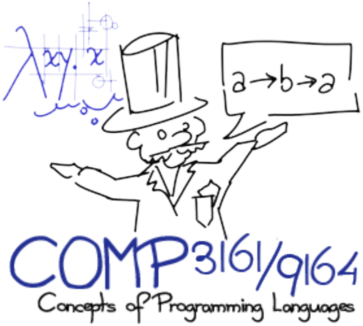

Polymorphism

Johannes Åman Pohjola
UNSW

Term 3 2022

Where we're at

- **Syntax Foundations** ✓
  Concrete/Abstract Syntax, Ambiguity, HOAS, Binding, Variables, Substitution
- **Semantics Foundations** ✓
  Static Semantics, Dynamic Semantics (Small-Step/Big-Step), (**Assignment 0**) Abstract Machines, Environments (**Assignment 1**)
- **Features**
  - Algebraic Data Types ✓
  - Polymorphism
  - Polymorphic Type Inference (**Assignment 2**)
  - Overloading
  - Subtyping
  - Modules
  - Concurrency

A Swap Function

Consider the humble `swap` function in Haskell:

$$\begin{array}{l} \mathit{swap} :: (t_1, t_2) \rightarrow (t_2, t_1) \\ \mathit{swap}\ (a, b) = (b, a) \end{array}$$

In our MinHS with algebraic data types from last lecture, we can't define this function.

Monomorphic

In MinHS, we're stuck copy-pasting our function over and over for every different type we want to use it with:

$$\begin{aligned}&\text{recfun } \mathit{swap_1} :: ((\texttt{Int} \times \texttt{Bool}) \to (\texttt{Bool} \times \texttt{Int}))\\ &\qquad\qquad p = (\text{snd } p, \text{fst } p)\end{aligned}$$

$$\begin{aligned}&\text{recfun } \mathit{swap_2} :: ((\texttt{Bool} \times \texttt{Int}) \to (\texttt{Int} \times \texttt{Bool}))\\ &\qquad\qquad p = (\text{snd } p, \text{fst } p)\end{aligned}$$

$$\begin{aligned}&\text{recfun } \mathit{swap_3} :: ((\texttt{Bool} \times \texttt{Bool}) \to (\texttt{Bool} \times \texttt{Bool}))\\ &\qquad\qquad p = (\text{snd } p, \text{fst } p)\\ &\ldots\end{aligned}$$

This is an acceptable state of affairs for some domain-specific languages, but not for general purpose programming.

Solutions

We want some way to specify that we don't care what the types of the tuple elements are.

$$swap :: (\forall a\ b.\ (a \times b) \rightarrow (b \times a))$$

This is called *parametric polymorphism* (or just *polymorphism* in functional programming circles). In Java and some other languages, this is called *generics* and polymorphism refers to something else. Don't be confused.

How it works

There are two main components to parametric polymorphism:

1. *Type abstraction* is the ability to define functions regardless of specific types (like the `swap` example before).In MinHS, we will write using **type** expressions like so: (the literature uses Λ)

$$\begin{array}{l} \mathit{swap} = \textbf{type}\ a.\\ \quad\quad \textbf{type}\ b.\\ \quad\quad\quad \textbf{recfun}\ \mathit{swap} :: (a \times b) \rightarrow (b \times a)\\ \quad\quad\quad\quad\quad\quad\quad p = (\mathsf{snd}\ p, \mathsf{fst}\ p) \end{array}$$

2. *Type application* is the ability to *instantiate* polymorphic functions to specific types. In MinHS, we use @ signs.

$$\mathit{swap}@\mathtt{Int}@\mathtt{Bool}\ (3, \mathtt{True})$$

Analogies

The reason they're called type abstraction and application is that they behave analogously to λ-calculus.
We have a β-reduction principle, but for types:

$$(\textbf{type}\ a.\ e)@\tau \quad \mapsto_\beta \quad (e[a := \tau])$$

Example (Identity Function)

$$\begin{array}{ll} & (\textbf{type}\ a.\ \textbf{recfun}\ f :: (a \to a)\ x = x)@\texttt{Int}\ 3 \\ \mapsto & (\textbf{recfun}\ f :: (\texttt{Int} \to \texttt{Int})\ x = x)\ 3 \\ \mapsto & 3 \end{array}$$

This means that **type** expressions can be thought of as functions from types to values.

Type Variables

What is the type of this?

$$(\textbf{type } a.\ \textbf{recfun } f :: (a \to a)\ x = x)$$

$$\forall a.\ a \to a$$

Types can mention *type variables* now[1].

If $id : \forall a.a \to a$, what is the type of id@`Int`?

$$(a \to a)[a := \texttt{Int}] = (\texttt{Int} \to \texttt{Int})$$

[1] Technically, they already could with recursive types.

Typing Rules Sketch

We would like rules that look something like this:

$$\frac{\Gamma \vdash e : \tau}{\Gamma \vdash \textbf{type}\ a.\ e : \forall a.\ \tau}$$

$$\frac{\Gamma \vdash e : \forall a.\ \tau}{\Gamma \vdash e@\rho : \tau[a := \rho]}$$

But these rules don't account for what type variables are available or in scope.

Type Wellformedness

With variables in the picture, we need to check our types to make sure that they only refer to well-scoped variables.

$$\frac{t\ \textbf{bound} \in \Delta}{\Delta \vdash t\ \mathsf{ok}} \qquad \frac{}{\Delta \vdash \texttt{Int}\ \mathsf{ok}} \qquad \frac{}{\Delta \vdash \texttt{Bool}\ \mathsf{ok}}$$

$$\frac{\Delta \vdash \tau_1\ \mathsf{ok} \quad \Delta \vdash \tau_2\ \mathsf{ok}}{\Delta \vdash \tau_1 \rightarrow \tau_2\ \mathsf{ok}} \qquad \frac{\Delta \vdash \tau_1\ \mathsf{ok} \quad \Delta \vdash \tau_2\ \mathsf{ok}}{\Delta \vdash \tau_1 \times \tau_2\ \mathsf{ok}}$$

(etc.)

$$\frac{\Delta, a\ \textbf{bound} \vdash \tau\ \mathsf{ok}}{\Delta \vdash \forall a.\ \tau\ \mathsf{ok}}$$

Typing Rules, Properly

We add a second context of type variables that are bound.

$$\frac{a\ \textbf{bound}, \Delta; \Gamma \vdash e : \tau}{\Delta; \Gamma \vdash \textbf{type}\ a.\ e : \forall a.\ \tau}$$

$$\frac{\Delta; \Gamma \vdash e : \forall a.\ \tau \qquad \Delta \vdash \rho\ \mathsf{ok}}{\Delta; \Gamma \vdash e@\rho : \tau[a := \rho]}$$

(the other typing rules just pass Δ through)

Dynamic Semantics

First we evaluate the LHS of a type application as much as possible:

$$\frac{e \quad \mapsto_M \quad e'}{e@\tau \quad \mapsto_M \quad e'@\tau}$$

Then we apply our β-reduction principle:

$$\frac{}{(\textbf{type}\ a.\ e)@\tau \quad \mapsto_M \quad e[a := \tau]}$$

Curry-Howard

Previously we noted the correspondence between types and logic:

$\times$	$\wedge$
$+$	$\vee$
$\rightarrow$	$\Rightarrow$
1	$\top$
0	$\bot$
$\forall$	$\forall$

Curry-Howard

The type quantifier $\forall$ corresponds to a universal quantifier $\forall$, but it is not the same as the $\forall$ from first-order logic. What's the difference?

First-order logic quantifiers range over a set of *individuals* or values, for example the natural numbers:

$$\forall x.\ x + 1 > x$$

These quantifiers range over propositions (types) themselves. It is analogous to *second-order logic*, not first-order:

$$\forall A.\ \forall B.\ A \land B \Rightarrow B \land A$$
$$\forall A.\ \forall B.\ A \times B \rightarrow B \times A$$

The first-order quantifier has a type-theoretic analogue too (type indices), but this is not nearly as common as polymorphism.

Generality

If we need a function of type $\texttt{Int} \rightarrow \texttt{Int}$, a polymorphic function of type $\forall a.\ a \rightarrow a$ will do just fine, we can just instantiate the type variable to $\texttt{Int}$. But the reverse is not true. This gives rise to an ordering.

Generality

A type τ is *more general* than a type ρ, often written $\rho \sqsubseteq \tau$, if type variables in τ can be instantiated to give the type ρ.

Example (Functions)

$$\texttt{Int} \rightarrow \texttt{Int} \quad \sqsubseteq \quad \forall z.\ z \rightarrow z \quad \sqsubseteq \quad \forall x\ y.\ x \rightarrow y \quad \sqsubseteq \quad \forall a.\ a$$

Implementation Strategies

Our simple dynamic semantics belies an implementation headache.

We can easily define functions that operate uniformly on multiple types. But when they are compiled to machine code, the results may differ depending on the size of the type in question.

There are two main approaches to solve this problem.

Template Instantiation

Key Idea

Automatically generate monomorphic copies of each polymorphic function, based on the types applied to it.

For example, if we defined our polymorphic swap function:

$$\begin{array}{l} \mathit{swap} = \textbf{type}\ a.\ \textbf{type}\ b. \\ \qquad \textbf{recfun}\ \mathit{swap} :: (a \times b) \rightarrow (b \times a) \\ \qquad\qquad p = (\mathsf{snd}\ p, \mathsf{fst}\ p) \end{array}$$

Then a type application like swap@`Int`@`Bool` would be replaced statically by the compiler with the monomorphic version:

$$\begin{array}{l} \mathit{swap}_{\mathtt{IB}} = \textbf{recfun}\ \mathit{swap} :: (\mathtt{Int} \times \mathtt{Bool}) \rightarrow (\mathtt{Bool} \times \mathtt{Int}) \\ \qquad p = (\mathsf{snd}\ p, \mathsf{fst}\ p) \end{array}$$

A new copy is made for each unique type application.

Evaluating Template Instatiation

This approach has a number of advantages:

1. Little to no run-time cost
2. Simple mental model
3. Allows for custom specialisations (e.g. list of booleans into bit-vectors)
4. Easy to implement

However the downsides are just as numerous:

1. Large binary size if many instantiations are used
2. This can lead to long compilation times
3. Restricts the type system to statically instantiated type variables.

Languages that use Template Instantiation: Rust, C++, some ML dialects

Polymorphic Recursion

Consider the following Haskell data type:

$$\textbf{data}\ \mathit{Dims}\ a = \mathrm{Step}\ a\ (\mathit{Dims}\ [a])\ |\ \mathrm{Epsilon}$$

This describes a list of matrices of increasing dimensionality, e.g:

$$\mathrm{Step}\ 1\ (\mathrm{Step}\ [1,2]\ (\mathrm{Step}\ [[1,2],[3,4]]\ \mathrm{Epsilon})) :: \mathit{Dims}\ \mathtt{Int}$$

We can write a sum function like this:

$$\begin{array}{l} \mathit{sumDims} :: \forall a.\ (a \to \mathtt{Int}) \to \mathit{Dims}\ a \to \mathtt{Int} \\ \mathit{sumDims}\ f\ \mathrm{Epsilon} = 0 \\ \mathit{sumDims}\ f\ (\mathrm{Step}\ a\ t) = (f\ a) + \mathit{sumDims}\ (\mathit{sum}\ f)\ t \end{array}$$

How many different instantiations of the type variable a are there?
We'd have to run the program to find out.

HM Types

Template instantiation can't handle all polymorphic programs.

In practice a statically determined subset can be carved out by restricting what sort of programs can be written:

1. Only allow $\forall$ quantifiers on the outermost part of a type declaration (not inside functions or type constructors).
2. Recursive functions cannot call themselves with different type parameters.

This restriction is sometimes called *Hindley-Milner* polymorphism. This is also the subset for which *type inference* is both complete and tractable.

Boxing

An alternative to our copy-paste-heavy template instantiation approach is to make all types represented the same way. Thus, a polymorphic function only requires one function in the generated code.

Typically this is done by *boxing* each type. That is, all data types are represented as a pointer to a data structure on the heap. If everything is a pointer, then all values use exactly 32 (or 64) bits of stack space.

The extra indirection has a run-time penalty, but it results in smaller binaries and unrestricted polymorphism.

Languages that use boxing: Haskell, Java, C♯, OCaml

Constraining Implementations

How many possible implementations are there of a function of the following type?

$$\texttt{Int} \rightarrow \texttt{Int}$$

How about this type?

$$\forall a.\ a \rightarrow a$$

Polymorphic type signatures constrain implementations.

Parametricity

Definition

The principle of parametricity states that the result of polymorphic functions cannot depend on values of an abstracted type.
More formally, suppose I have a polymorphic function g that takes a type parameter. If run any arbitrary function $f : \tau \rightarrow \tau$ on some values of type τ, then run the function $g@\tau$ on the result, that will give the same results as running $g@\tau$ first, then f.

Example

$$foo :: \forall a.\ [a] \rightarrow [a]$$

We know that **every** element of the output occurs in the input.
The parametricity theorem we get is, for all f:

$$foo \circ (map\ f) = (map\ f) \circ foo$$

More Examples

$$head :: \forall a.\ [a] \to a$$

What's the parametricity theorems?

Example (Answer)

For any f:

$$f\ (head\ \ell) = head\ (map\ f\ \ell)$$

More Examples

$$(+\!\!+) :: \forall a.\ [a] \to [a] \to [a]$$

What's the parametricity theorem?

Example (Answer)

$$map\ f\ (a +\!\!+ b) = map\ f\ a +\!\!+ map\ f\ b$$

More Examples

$$\mathit{concat} :: \forall a.\ [[a]] \rightarrow [a]$$

What's the parametricity theorem?

Example (Answer)

$$\mathit{map}\ f\ (\mathit{concat}\ \mathit{ls}) = \mathit{concat}\ (\mathit{map}\ (\mathit{map}\ f)\ \mathit{ls})$$

Higher Order Functions

$$filter :: \forall a.\ (a \rightarrow Bool)\ \rightarrow [a] \rightarrow [a]$$

What's the parametricity theorem?

Example (Answer)

$$filter\ \ p\ (map\ \ f\ \ ls) = map\ \ f\ \ (filter\ (p \circ f)\ \ ls)$$

Parametricity Theorems

Follow a similar structure. In fact it can be mechanically derived, using the *relational parametricity* framework invented by John C. Reynolds, and popularised by Wadler in the famous paper, "Theorems for Free!"[2].

[2] `https://people.mpi-sws.org/~dreyer/tor/papers/wadler.pdf`