

# Agenda

## Type inference

- Untyped lambda-calculus
- Simply typed lambda-calculus
- System F
- Hindley-Milner typing
- Algorithm W



Universiteit Utrecht

[Faculty of Science  
Information and Computing Sciences]

## Compiler Construction

WWW: <http://www.cs.uu.nl/wiki/Cco>

Edition 2010/2011

2



Universiteit Utrecht

[Faculty of Science  
Information and Computing Sciences]



## 8. Type inference

3



Universiteit Utrecht

[Faculty of Science  
Information and Computing Sciences]



## 8.1 Untyped lambda-calculus

4



Universiteit Utrecht

[Faculty of Science  
Information and Computing Sciences]



- ▶ Church, Kleene (1930s).
- ▶ Formal system designed to investigate function definition, function application, and recursion.
- ▶ Idealised, minimalistic functional programming language.
- ▶ Only three language constructs: variables, lambda-abstraction, function application. (Other constructs can be encoded with these.)



## Alpha-equivalence and beta-substitution

$\lambda$  is a *binder*:  $\lambda x. t_1$  binds  $x$  in  $t_1$ . Unbound variables are called *free*.

**Alpha-equivalence:** terms that only differ in the names of their bound variables are considered equal. For example:  $\lambda x. x$  and  $\lambda y. y$  are alpha-equivalent.

**Alpha-conversion:** consistently renaming bound variables while avoiding free variables from being *captured*. For example:  $\lambda f. \lambda x. f x z$  can be alpha-converted into  $\lambda f. \lambda y. f y z$ , but not into  $\lambda f. \lambda z. f z z$ .

**Beta-substitution:** capture-avoiding substitution of free variables, performing alpha-conversion where necessary. For example:  $[z \mapsto y](\lambda f. \lambda x. f x z) = \lambda f. \lambda x. f x y$  and  $[z \mapsto x](\lambda f. \lambda y. f y x)$ .



$$\begin{array}{l} x \in \mathbf{Var} \quad \text{variables} \\ t \in \mathbf{Tm} \quad \text{terms} \end{array}$$

$$t ::= x \mid \lambda x. t_1 \mid t_1 t_2$$


## Semantics

$$v \in \mathbf{Val} \quad \text{values}$$

$$v ::= \lambda x. t$$

Big-step operational semantics:

$$t \Downarrow v \quad \text{evaluation}$$

$$\frac{}{\lambda x. t_1 \Downarrow \lambda x. t_1} \text{ [e-lam]}$$

$$\frac{t_1 \Downarrow \lambda x. t_{11} \quad t_2 \Downarrow v_2 \quad [x \mapsto v_2]t_{11} \Downarrow v}{t_1 t_2 \Downarrow v} \text{ [e-app]}$$

For example:  $(\lambda x. \lambda y. x) (\lambda x. x) (\lambda x. \lambda y. y) \Downarrow \lambda x. x$ .



Additional language constructs—such as local definitions, natural numbers, boolean constants, conditionals, arithmetic and relational operators, and even recursion—can be introduced as mere syntactic sugar.

For example:

$$\text{let } x = t_1 \text{ in } t_2 \text{ ni} \quad =_{\text{def}} \quad (\lambda x. t_2) t_1$$

In the sequel, we will just assume that some of these additional constructs are in fact added to the core calculus, so that we can use for example natural numbers in our example.



## Simple types

To study typing for the lambda-calculus, we extend the syntax of lambda-terms with mandatory type annotations for lambda-abstractions.

For example:

$$\lambda x : \text{Nat}. x$$

$$(\lambda f : \text{Bool} \rightarrow \text{Nat}. \lambda x : \text{Bool}. f x) (\lambda y : \text{Bool}. 42)$$



## 8.2 Simply typed lambda-calculus



## Syntax

$$\begin{aligned} x &\in \mathbf{Var} && \text{variables} \\ \tau &\in \mathbf{Ty} && \text{types} \\ t &\in \mathbf{Tm} && \text{terms} \end{aligned}$$

$$\begin{aligned} \tau &::= \tau_1 \rightarrow \tau_2 \\ t &::= x \mid \lambda x : \tau. t_1 \mid t_1 t_2 \end{aligned}$$

☞ To render the sets of types inhabited, we add type constants such as *Nat* and *Bool*.



$$v \in \mathbf{Val} \quad \text{values}$$

$$v ::= \lambda x : \tau. t$$

Big-step operational semantics:

$$t \Downarrow v \quad \text{evaluation}$$

$$\frac{}{\lambda x : \tau. t_1 \Downarrow \lambda x : \tau. t_1} \text{ [e-lam]}$$

$$\frac{t_1 \Downarrow \lambda x : \tau. t_{11} \quad t_2 \Downarrow v_2 \quad [x \mapsto v_2] t_{11} \Downarrow v}{t_1 t_2 \Downarrow v} \text{ [e-app]}$$


## Type rules

$$\frac{\Gamma(x) = \tau}{\Gamma \vdash x : \tau} \text{ [t-var]}$$

$$\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x : \tau_1. t_1 : \tau_1 \rightarrow \tau_2} \text{ [t-lam]}$$

$$\frac{\Gamma \vdash t_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash t_2 : \tau_2}{\Gamma \vdash t_1 t_2 : \tau} \text{ [t-app]}$$


## Typing

Type environments map from variables to types:

$$\Gamma \in \mathbf{TyEnv} \quad \text{type environments}$$

$$\Gamma ::= [] \mid \Gamma_1[x \mapsto \tau]$$

As always, we write  $\Gamma(x) = \tau$  if the rightmost binding for  $x$  in  $\Gamma$  maps  $x$  to  $\tau$ .

The judgements of the typing relation read

$$\Gamma \vdash t : \tau \quad \text{typing}$$


## Eraseure

Writing  $[t]$  for the untyped lambda-term obtained from erasing all type annotations from the simply typed lambda-term  $t$ , we have:

if  $t \Downarrow v$ , then  $[t] \Downarrow [v]$ .

That is, types play no rôle at run-time.



## 8.3 System F



### Syntax

§8.3

$\alpha \in$	<b>TyVar</b>	type variables
$x \in$	<b>Var</b>	term variables
$\tau \in$	<b>Ty</b>	types
$t \in$	<b>Tm</b>	terms

$\tau ::=$	$\alpha \mid \tau_1 \rightarrow \tau_2 \mid \forall \alpha. \tau_1$
$t ::=$	$x \mid \lambda x : \tau. t_1 \mid t_1 t_2 \mid \Lambda \alpha. t_1 \mid t_1 [\tau]$



## Polymorphism

§8.3

Next, we add polymorphism to our language.

The system obtained is known as System F (Girard, 1972) or the second-order polymorphic lambda-calculus (Reynolds, 1974).

The main innovation with respect to the simply typed lambda-calculus is that, in addition to values, functions can also take types as arguments:

$$\Lambda \alpha. \lambda x : \alpha. x$$
$$(\Lambda \alpha. \Lambda \beta. \lambda x : \alpha. \lambda y : \beta. x) [Nat] [Bool] 2 \text{ false}$$


### Semantics

§8.3

$$v \in \mathbf{Val} \quad \text{values}$$
$$v ::= \lambda x : \tau. t \mid \Lambda \alpha. t$$

Big-step operational semantics:


$$t \Downarrow v \quad \text{evaluation}$$


$$\frac{}{\lambda x : \tau. t_1 \Downarrow \lambda x : \tau. t_1} \text{ [e-lam]}$$

$$\frac{t_1 \Downarrow \lambda x : \tau. t_{11} \quad t_2 \Downarrow v_2 \quad [x \mapsto v_2] t_{11} \Downarrow v}{t_1 t_2 \Downarrow v} \text{ [e-app]}$$

$$\frac{}{\Lambda \alpha. t_1 \Downarrow \Lambda \alpha. t_1} \text{ [e-tylam]}$$

$$\frac{t_1 \Downarrow \Lambda \alpha. t_{11} \quad [\alpha \mapsto \tau] t_{11} \Downarrow v}{t_1 [\tau] \Downarrow v} \text{ [e-tyapp]}$$

  $\Lambda$  is a binder for type variables.



**Exercise:** investigate erasure for System F.



$$\frac{\Gamma(x) = \tau}{\Gamma \vdash x : \tau} \text{ [t-var]}$$

$$\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x : \tau_1. t_1 : \tau_1 \rightarrow \tau_2} \text{ [t-lam]}$$

$$\frac{\Gamma \vdash t_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash t_2 : \tau_2}{\Gamma \vdash t_1 t_2 : \tau} \text{ [t-app]}$$

$$\frac{\Gamma \vdash t_1 : \tau_1}{\Gamma \vdash \Lambda \alpha. t_1 : \forall \alpha. \tau_1} \text{ [t-tylam]}$$

$$\frac{\Gamma \vdash t_1 : \forall \alpha. \tau_1}{\Gamma \vdash t_1 [\tau_0] : [\alpha \mapsto \tau_0] \tau_1} \text{ [t-tyapp]}$$



Note: functions can take polymorphic functions as arguments.

For example:

$$\lambda f : \forall \alpha. \alpha \rightarrow \text{Nat}. f [\text{Nat}] 2 + f [\text{Bool}] \text{false}$$

This function takes a “normal” polymorphic function (i.e. of rank 1) as argument and so it has itself a rank-2 type. Its type reads  $(\forall \alpha. \alpha \rightarrow \text{Nat}) \rightarrow \text{Nat}$ . In general, if a function takes a function with a rank- $n$  type as argument, it has itself a rank- $(n + 1)$  type.



## 8.4 Hindley-Milner typing



### The Hindley-Milner system §8.4

The Hindley-Milner (Hindley, 1969; Milner, 1978) type system is a compromise between the full power of System F and the desire to leave out type annotations.

Hindley-Milner typing comes with two crucial restrictions:

1. All types are of at most rank 1, i.e., functions cannot take polymorphic functions as arguments.
2. Functions can only have a polymorphic type if they are directly bound in a local definition (let-polymorphism).

The resulting type system is at the heart of languages like Haskell and ML and allows that for each well-typed term a so-called principal (i.e., most polymorphic) type can be inferred.



## Type inference

For both the simply typed lambda-calculus and System F, implementing a type checker is straightforward. (Exercise: ...)

Although programming languages based on System F are very powerful, writing type annotation on every function parameter is very tedious—especially if types become more involved due to polymorphism.

So, the question is: can we derive an algorithm that takes an erased System-F term as argument and that *infers* all missing type annotations? That way, we can have the full power for System F, without the burden of having to write possibly complex or otherwise tiresome type annotations.

The answer is **no**, type inference for System F is undecidable (Wells, 1994).



### Syntax §8.4

$x \in \mathbf{Var}$	term variables
$t \in \mathbf{Tm}$	terms

$t ::= x \mid \lambda x. t_1 \mid t_1 t_2 \mid \mathbf{let } x = t_1 \mathbf{ in } t_2 \mathbf{ ni}$
--

Local definitions play a crucial rôle in typing now and so, rather than syntactic sugar, they form a true language construct now.



$$v \in \mathbf{Val} \quad \text{values}$$

$$v ::= \lambda x. t$$

Big-step operational semantics:

$$t \Downarrow v \quad \text{evaluation}$$


## Typing

Implementing the rank-1 restriction, the type language is stratified into two levels: types and type schemes.

$$\begin{array}{ll} \alpha \in \mathbf{TyVar} & \text{type variables} \\ \tau \in \mathbf{Ty} & \text{types} \\ \sigma \in \mathbf{TyScheme} & \text{type schemes} \end{array}$$

$$\begin{array}{l} \tau ::= \alpha \mid \tau_1 \rightarrow \tau_2 \\ \sigma ::= \tau \mid \forall \alpha. \sigma_1 \end{array}$$

Type environments map from variables to type schemes:

$$\Gamma \in \mathbf{TyEnv} \quad \text{type environments}$$

$$\Gamma ::= [] \mid \Gamma_1[x \mapsto \sigma]$$


$$\frac{}{\lambda x. t_1 \Downarrow \lambda x. t_1} \text{ [e-lam]}$$

$$\frac{t_1 \Downarrow \lambda x. t_{11} \quad t_2 \Downarrow v_2 \quad [x \mapsto v_2] t_{11} \Downarrow v}{t_1 t_2 \Downarrow v} \text{ [e-app]}$$

$$\frac{t_1 \Downarrow v_1 \quad [x \mapsto v_1] t_2 \Downarrow v}{\text{let } x = t_1 \text{ in } t_2 \text{ ni } \Downarrow v} \text{ [e-let]}$$


## Free type variables

$\forall$  is a binder for type variables:  $\alpha$  is bound in  $\forall \alpha. \sigma_1$ .

We write  $ftv(\sigma)$  for the set of type variables that appear free in  $\sigma$ .

Similarly, we write  $ftv(\Gamma)$  for the set of type variables that appear free in the codomain of  $\Gamma$ .





The judgements of the typing relation take the form

$$\Gamma \vdash t : \sigma \quad \text{typing}$$

The typing relation is defined by a natural deduction system comprised from six rules: one for each of the four term constructors and two for dealing with polymorphism.



## Typing rules (cont'd)

$$\frac{\Gamma \vdash t : \sigma_1 \quad \alpha \notin \text{ftv}(\Gamma)}{\Gamma \vdash t : \forall \alpha. \sigma_1} \quad [t\text{-gen}]$$

$$\frac{\Gamma \vdash t : \forall \alpha. \sigma_1}{\Gamma \vdash t : [\alpha \mapsto \tau_0] \sigma_1} \quad [t\text{-inst}]$$

The premise  $\alpha \notin \text{ftv}\{\Gamma\}$  in  $[t\text{-gen}]$  is needed because we do not have any binders for type variables in our term language.

(Exercise: show that without this premise we can derive  $\lambda x. x : \forall \alpha. \forall \beta. \alpha \rightarrow \beta$ .)



$$\frac{\Gamma(x) = \sigma}{\Gamma \vdash x : \sigma} \quad [t\text{-var}]$$

$$\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x. t_1 : \tau_1 \rightarrow \tau_2} \quad [t\text{-lam}]$$

$$\frac{\Gamma \vdash t_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash t_2 : \tau_2}{\Gamma \vdash t_1 t_2 : \tau} \quad [t\text{-app}]$$

$$\frac{\Gamma \vdash t_1 : \sigma_1 \quad \Gamma[x \mapsto \sigma_1] \vdash t_2 : \tau}{\Gamma \vdash \text{let } x = t_1 \text{ in } t_2 \text{ ni} : \tau} \quad [t\text{-let}]$$



## 8.5 Algorithm W



Algorithm W (Damas and Milner, 1982) establishes a procedure for obtaining a principal type, for each well-typed term in the Hindley-Milner system.

Intuitively, a principal type is the most polymorphic type that can be assigned to a given term.



## Strategy

- ▶ Algorithm W proceeds by initially “guessing” a fresh type variable for every parameter type and by incrementally refining these guesses as more information on the use of parameters becomes available.
- ▶ Algorithm W uses a syntax-directed variation of the Hindley-Milner type rules in which generalisation only occurs at let-bindings and instantiation only occurs at the use-sites of variables.

**Syntax-directed:** for every term, at most one rule applies.



- ▶ We have to somehow “guess”, for every lambda-abstraction, what the type of its formal parameter is.
- ▶ The rules  $[t\text{-}gen]$  and  $[t\text{-}inst]$  can be applied to terms of any form. We have to decide when to apply them.



## Generalisation and instantiation

The syntax-directed type rules are defined in terms of metaoperations  $gen$ . and  $inst$ :

$$\begin{aligned}
 gen &: \text{TyEnv} \rightarrow \text{Ty} \rightarrow \text{TyScheme} \\
 gen_{\Gamma}(\tau) &= \\
 &\text{let } \{\alpha_1, \dots, \alpha_n\} = \text{ftv}(\tau) \setminus \text{ftv}(\Gamma) \\
 &\text{in } \forall \alpha_1. \dots \forall \alpha_n. \tau
 \end{aligned}$$

$$\begin{aligned}
 inst &: \text{TyScheme} \rightarrow \text{Ty} \\
 inst(\forall \alpha_1. \dots \forall \alpha_n. \tau_1) &= \\
 &\text{let } \alpha'_1, \dots, \alpha'_n \text{ be fresh} \\
 &\text{in } [\alpha_1 \mapsto \alpha'_1] \dots [\alpha_n \mapsto \alpha'_n] \tau_1
 \end{aligned}$$



$$\frac{\Gamma(x) = \sigma_0}{\Gamma \vdash x : inst(\sigma_0)} [t-var]$$

$$\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x. t_1 : \tau_1 \rightarrow \tau_2} [t-lam]$$

$$\frac{\Gamma \vdash t_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash t_2 : \tau_2}{\Gamma \vdash t_1 t_2 : \tau} [t-app]$$

$$\frac{\Gamma \vdash t_1 : \tau_1 \quad \Gamma[x \mapsto gen_{\Gamma}(\tau_1)] \vdash t_2 : \tau}{\Gamma \vdash let\ x = t_1\ in\ t_2\ ni : \tau} [t-let]$$

☞ All judgements have the form  $\Gamma \vdash t : \tau$  (rather than  $\Gamma \vdash t : \sigma$ ).



# Applying type substitutions

Applying a type substitution to a type scheme:

$$\begin{aligned} id\sigma &= \sigma \\ [\alpha \mapsto \tau_0]\alpha &= \tau_0 \\ [\alpha \mapsto \tau_0]\alpha_0 &= \alpha_0 && \text{if } \alpha \neq \alpha_0 \\ [\alpha \mapsto \tau_0](\tau_1 \rightarrow \tau_2) &= [\alpha \mapsto \tau_0]\tau_1 \rightarrow [\alpha \mapsto \tau_0]\tau_2 \\ [\alpha \mapsto \tau_0](\forall \alpha. \sigma_1) &= \forall \alpha. \sigma \\ [\alpha \mapsto \tau_0](\forall \alpha_0. \sigma_1) &= \forall \alpha_0. [\alpha \mapsto \tau_0]\sigma_1 && \text{if } \alpha \neq \alpha_0 \\ (\theta_1 \circ \theta_2)\sigma &= \theta_1\theta_2\sigma \end{aligned}$$

Applying a type substitution to a type environment:

$$\begin{aligned} \theta[] &= [] \\ \theta(\Gamma_1[x \mapsto \sigma]) &= \theta\Gamma_1[x \mapsto \theta\sigma] \end{aligned}$$



Algorithm W makes use of **type substitutions**:

$$\theta \in \mathbf{TySubst} \quad \text{type substitutions}$$

$$\theta ::= id \mid [\alpha \mapsto \tau] \mid \theta_1 \circ \theta_2$$



# Unification

Algorithm W makes use of Robinson's unification algorithm (1965).

$$\mathcal{U} : \mathbf{Ty} \times \mathbf{Ty} \rightarrow \mathbf{TySubst}$$

$\mathcal{U}$  provides a partial function that, for any two types  $\tau_1$  and  $\tau_2$ , constructs a *most general unifier*, i.e., a type substitution  $\theta$ , such that  $\theta\tau_1 = \theta\tau_2$  and, for all  $\theta'$  with  $\theta'\tau_1 = \theta'\tau_2$  there is a  $\theta''$  with  $\theta' \approx \theta'' \circ \theta$ . (Where  $\theta_1 \approx \theta_2$  iff,  $\theta_1\sigma = \theta_2\sigma$  for all  $\sigma$ .)

If  $\tau_1$  and  $\tau_2$  are not unifiable,  $\mathcal{U}$  fails.

For example:

$$\begin{aligned} \mathcal{U}(\alpha \rightarrow Bool \rightarrow \alpha, Nat \rightarrow \beta \rightarrow \gamma) \\ = [\gamma \mapsto Nat] \circ [\beta \mapsto Bool] \circ [\alpha \mapsto Nat] \end{aligned}$$



$$\begin{aligned}
 \mathcal{U}(\alpha, \alpha) &= id \\
 \mathcal{U}(\alpha_1, \tau_2) &= [\alpha_1 \mapsto \tau_2] && \text{if } \alpha_1 \notin \text{ftv}(\tau_2) \\
 \mathcal{U}(\tau_1, \alpha_2) &= [\alpha_2 \mapsto \tau_1] && \text{if } \alpha_2 \notin \text{ftv}(\tau_1) \\
 \mathcal{U}(\tau_{11} \rightarrow \tau_{12}, \tau_{21} \rightarrow \tau_{22}) &= \text{let } \theta_1 = \mathcal{U}(\tau_{11}, \tau_{21}) \\
 &\quad \theta_2 = \mathcal{U}(\theta_1 \tau_{12}, \theta_1 \tau_{22}) \\
 &\quad \text{in } \theta_2 \circ \theta_1 \\
 \mathcal{U}(\tau_1, \tau_2) &= fail && \text{in all other cases}
 \end{aligned}$$

☞ The side conditions  $\alpha_1 \notin \text{ftv}(\tau_2)$  and  $\alpha_2 \notin \text{ftv}(\tau_1)$  are known as the “occurs check” and prevent the construction of infinite types.



## Algorithm W: variables

$$\begin{aligned}
 \mathcal{W}(\Gamma, x) &= \\
 &\text{if } x \in \text{dom}(\Gamma) \\
 &\text{then } (\text{inst } (\Gamma(x)), id) \\
 &\text{if } fail
 \end{aligned}$$



$$\mathcal{W} : \text{TyEnv} \times \text{Tm} \rightarrow \text{Ty} \times \text{TySubst}$$

Algorithm W takes a type environment  $\Gamma$  and a term  $t$ , and produces, if  $t$  is well-typed in  $\Gamma$ , a type  $\tau$  and a type substitution  $\theta$ , such that  $\theta\Gamma \vdash t : \text{gen}_{\theta\Gamma}(\tau)$ . Moreover,  $\text{gen}_{\theta\Gamma}(\tau)$  is then a principal type for  $t$  in  $\theta\Gamma$ .

If  $t$  is ill-typed in  $\Gamma$ , Algorithm W fails.



## Algorithm W: lambda-abstractions

$$\begin{aligned}
 \mathcal{W}(\Gamma, \lambda x. t_1) &= \\
 &\text{let } \alpha_1 \text{ be fresh} \\
 &\quad (\tau_2, \theta_1) = \mathcal{W}(\Gamma[x \mapsto \alpha_1], t_1) \\
 &\text{in } (\theta_1 \alpha_1 \rightarrow \tau_2, \theta_1)
 \end{aligned}$$



$$\begin{aligned} \mathcal{W}(\Gamma, t_1 t_2) = & \\ \text{let } \alpha \text{ be fresh} & \\ (\tau_1, \theta_1) = \mathcal{W}(\Gamma, t_1) & \\ (\tau_2, \theta_2) = \mathcal{W}(\theta_1 \Gamma, t_2) & \\ \theta_3 = \mathcal{U}(\theta_2 \tau_1, \tau_2 \rightarrow \alpha) & \\ \text{in } (\theta_3 \alpha, \theta_3 \circ \theta_2 \circ \theta_1) & \end{aligned}$$


$$\begin{aligned} \mathcal{W}(\Gamma, \text{let } x = t_1 \text{ in } t_2 \text{ ni}) = & \\ \text{let } (\tau_1, \theta_1) = \mathcal{W}(\Gamma, t_1) & \\ (\tau, \theta_2) = \mathcal{W}(\theta_1 \Gamma[x \mapsto \text{gen}_{\theta_1 \Gamma}(\tau_1)], t_2) & \\ \text{in } (\tau, \theta_2 \circ \theta_1) & \end{aligned}$$
