# Programming Languages and Compilers (CS 421)



2112 SC, UIUC

http://courses.engr.illinois.edu/cs421

Based in part on slides by Mattox Beckman, as updated by Vikram Adve and Gul Agha



- Data types play a key role in:
  - Data abstraction in the design of programs
  - Type checking in the analysis of programs
  - Compile-time code generation in the translation and execution of programs
    - Data layout (how many words; which are data and which are pointers) dictated by type

# Terminology

- Type: A type t defines a set of possible data values
  - E.g. short in C is  $\{x \mid 2^{15} 1 \ge x \ge -2^{15}\}$
  - A value in this set is said to have type t

 Type system: rules of a language assigning types to expressions

# 4

### Types as Specifications

- Types describe properties
- Different type systems describe different properties, eg
  - Data is read-write versus read-only
  - Operation has authority to access data
  - Data came from "right" source
  - Operation might or could not raise an exception
- Common type systems focus on types describing same data layout and access methods

# Sound Type System

If an expression is assigned type t, and it evaluates to a value v, then v is in the set of values defined by t

- SML, OCAML, Scheme and Ada have sound type systems
- Most implementations of C and C++ do not



# Strongly Typed Language

- When no application of an operator to arguments can lead to a run-time type error, language is strongly typed
  - Eg: 1 + 2.3;;
- Depends on definition of "type error"



### Strongly Typed Language

- C++ claimed to be "strongly typed", but
  - Union types allow creating a value at one type and using it at another
  - Type coercions may cause unexpected (undesirable) effects
  - No array bounds check (in fact, no runtime checks at all)
- SML, OCAML "strongly typed" but still must do dynamic array bounds checks, runtime type case analysis, and other checks



### Static vs Dynamic Types

- Static type: type assigned to an expression at compile time
- Dynamic type: type assigned to a storage location at run time
- Statically typed language: static type assigned to every expression at compile time
- Dynamically typed language: type of an expression determined at run time

# Type Checking

- When is op(arg1,...,argn) allowed?
- Type checking assures that operations are applied to the right number of arguments of the right types
  - Right type may mean same type as was specified, or may mean that there is a predefined implicit coercion that will be applied
- Used to resolve overloaded operations

# Type Checking

- Type checking may be done statically at compile time or dynamically at run time
- Dynamically typed (aka untyped) languages (eg LISP, Prolog) do only dynamic type checking
- Statically typed languages can do most type checking statically



### Dynamic Type Checking

- Performed at run-time before each operation is applied
- Types of variables and operations left unspecified until run-time
  - Same variable may be used at different types



### Dynamic Type Checking

- Data object must contain type information
- Errors aren't detected until violating application is executed (maybe years after the code was written)



### Static Type Checking

- Performed after parsing, before code generation
- Type of every variable and signature of every operator must be known at compile time



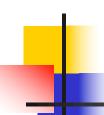
### Static Type Checking

- Can eliminate need to store type information in data object if no dynamic type checking is needed
- Catches many programming errors at earliest point
- Can't check types that depend on dynamically computed values
  - Eg: array bounds



# Static Type Checking

- Typically places restrictions on languages
  - Garbage collection
  - References instead of pointers
  - All variables initialized when created
  - Variable only used at one type
    - Union types allow for work-arounds, but effectively introduce dynamic type checks



### **Type Declarations**

- Type declarations: explicit assignment of types to variables (signatures to functions) in the code of a program
  - Must be checked in a strongly typed language
  - Often not necessary for strong typing or even static typing (depends on the type system)

# Type Inference

- Type inference: A program analysis to assign a type to an expression from the program context of the expression
  - Fully static type inference first introduced by Robin Miller in ML
  - Haskle, OCAML, SML all use type inference
    - Records are a problem for type inference

### Format of Type Judgments

A type judgement has the form

$$\Gamma$$
 |- exp :  $\tau$ 

- I is a typing environment
  - Supplies the types of variables (and function names when function names are not variables)
  - $\Gamma$  is a set of the form  $\{x:\sigma,\ldots\}$
  - For any x at most one  $\sigma$  such that  $(x : \sigma \in \Gamma)$
- exp is a program expression
- $\mathbf{r}$  is a type to be assigned to exp
- |- pronounced "turnstyle", or "entails" (or "satisfies" or, informally, "shows")



#### **Axioms - Constants**

 $\Gamma \mid -n : int$  (assuming *n* is an integer constant)

 $\Gamma$  |- true : bool

 $\Gamma$  |- false : bool

- These rules are true with any typing environment
- $\blacksquare$   $\Gamma$ , n are meta-variables



### Axioms – Variables (Monomorphic Rule)

Notation: Let  $\Gamma(x) = \sigma$  if  $x : \sigma \in \Gamma$ 

Note: if such of exits, its unique

Variable axiom:

$$\Gamma \mid -x : \sigma$$
 if  $\Gamma(x) = \sigma$ 



# Simple Rules - Arithmetic

Primitive Binary operators ( $\oplus \in \{+, -, *, ...\}$ ):

$$\Gamma \mid -e_1:\tau_1 \qquad \Gamma \mid -e_2:\tau_2 \quad (\oplus):\tau_1 \to \tau_2 \to \tau_3 \\
\Gamma \mid -e_1 \oplus e_2:\tau_3$$

Special case: Relations (~∈ { < , > , =, <=, >= }):

$$\Gamma \mid -e_1 : \tau \quad \Gamma \mid -e_2 : \tau \quad (\sim) : \tau \rightarrow \tau \rightarrow \text{bool}$$

$$\Gamma \mid -e_1 \quad \sim \quad e_2 : \text{bool}$$

For the moment, think  $\tau$  is int

# -

### Example: $\{x:int\} | -x + 2 = 3 : bool$

What do we need to show first?

$$\{x:int\} \mid -x + 2 = 3 : bool$$

# 4

### Example: $\{x:int\} | -x + 2 = 3 : bool$

What do we need for the left side?

$$\{x : int\} \mid -x + 2 : int$$
  $\{x : int\} \mid -3 : int$   $\{x : int\} \mid -x + 2 = 3 : bool$ 

### Example: $\{x:int\} | -x + 2 = 3 : bool$

How to finish?

```
\{x:int\} \mid -x:int \mid \{x:int\} \mid -2:int \mid \{x:int\} \mid -x+2:int \mid \{x:int\} \mid -3:int \mid \{x:int\} \mid -x+2=3:bool
```

# 4

### Example: $\{x:int\} | -x + 2 = 3 : bool$

#### Complete Proof (type derivation)



## Simple Rules - Booleans

#### Connectives

$$\Gamma \mid -e_1 : bool$$
  $\Gamma \mid -e_2 : bool$   $\Gamma \mid -e_1 & e_2 : bool$ 

$$\Gamma \mid -e_1 : bool$$
  $\Gamma \mid -e_2 : bool$   $\Gamma \mid -e_1 \mid e_2 : bool$ 

# 4

### Type Variables in Rules

If\_then\_else rule:

```
\Gamma \mid -e_1 : bool \quad \Gamma \mid -e_2 : \tau \quad \Gamma \mid -e_3 : \tau
\Gamma \mid -(if e_1 then e_2 else e_3) : \tau
```

- $\mathbf{r}$  is a type variable (meta-variable)
- Can take any type at all
- All instances in a rule application must get same type
- Then branch, else branch and if\_then\_else must all have same type

# 4

### **Function Application**

Application rule:

$$\frac{\Gamma \mid -e_1 : \tau_1 \rightarrow \tau_2 \quad \Gamma \mid -e_2 : \tau_1}{\Gamma \mid -(e_1 e_2) : \tau_2}$$

If you have a function expression  $e_1$  of type  $\tau_1 \rightarrow \tau_2$  applied to an argument  $e_2$  of type  $\tau_1$ , the resulting expression  $e_1 e_2$  has type  $\tau_2$ 

# Fun Rule

- Rules describe types, but also how the environment \(\Gamma\) may change
- Can only do what rule allows!
- fun rule:

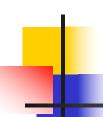
$$\{x \colon \tau_1\} + \Gamma \mid -e \colon \tau_2$$

$$\Gamma \mid -\text{fun } x -> e \colon \tau_1 \to \tau_2$$

# -

### Fun Examples

```
\{y : int \} + \Gamma \mid -y + 3 : int \}
 \Gamma \mid -fun y -> y + 3 : int \rightarrow int \}
```



### (Monomorphic) Let and Let Rec

#### let rule:

$$\Gamma \mid -e_1 : \tau_1 \quad \{x : \tau_1\} + \Gamma \mid -e_2 : \tau_2$$

$$\Gamma \mid -(\text{let } x = e_1 \text{ in } e_2) : \tau_2$$

#### let rec rule:

$$\{x: \tau_1\} + \Gamma \mid -e_1:\tau_1 \{x: \tau_1\} + \Gamma \mid -e_2:\tau_2$$
  
  $\Gamma \mid -(\text{let rec } x = e_1 \text{ in } e_2):\tau_2$ 

# Example

Which rule do we apply?

```
|- (let rec one = 1 :: one in let x = 2 in fun y \rightarrow (x :: y :: one)) : int \rightarrow int list
```

# Example

```
(2) {one : int list} |-
Let rec rule:
                             (let x = 2 in
                         fun y -> (x :: y :: one)
{one : int list} |-
(1 :: one) : int list
                              : int \rightarrow int list
 |- (let rec one = 1 :: one in
     let x = 2 in
      fun y -> (x :: y :: one)) : int \rightarrow int list
```

# Proof of 1

Which rule?

{one : int list} |- (1 :: one) : int list

# 4

### Proof of 1

Binary Operator

```
(3) (4) {one : int list} |- {one : int list} |- 1: int one : int list {one : int list}
```

where (::): int  $\rightarrow$  int list  $\rightarrow$  int list

10/7/21

# Proof of 1

3

4

```
Constant Rule
```

{one : int list} |-

1: int

Variable Rule

{one : int list} |-

one: int list

{one : int list} |- (1 :: one) : int list

10/7/21

Let Rule  $\{x:int; one : int list\} \mid -fun y -> (x :: y :: one))$   $\{one : int list\} \mid -2:int : int \rightarrow int list\}$   $\{one : int list\} \mid -(let x = 2 in fun y -> (x :: y :: one)) : int \rightarrow int list\}$ 

Constant

```
{x:int; one : int list} |-
                                                                                                                                                                                                                                                                                                                                                                   fun y ->
                                                                                                                                                                                                                                                                                                                                                                                                (x :: y :: one))
\{\text{one : int list}\}\ | -2: \text{int} : \text{int} \rightarrow \text{int list}\}
                                       \{one : int list\} \mid - (let x = 2 in let x =
                                                                             fun y -> (x :: y :: one)) : int \rightarrow int list
```

?

```
{x:int; one : int list} |- fun y -> (x :: y :: one))
: int \rightarrow int list
```

?

```
{y:int; x:int; one : int list} |- (x :: y :: one) : int list

{x:int; one : int list} |- fun y -> (x :: y :: one))

: int \rightarrow int list

By the Fun Rule
```

40

```
6
```

```
7
```

```
{y:int; x:int; one:int list} {y:int; x:int; one:int list} 

|- x:int |- (y :: one) : int list 

{y:int; x:int; one : int list} |- (x :: y :: one) : int list 

{x:int; one : int list} |- fun y -> (x :: y :: one)) 

: int \rightarrow int list 

By BinOp where ( :: ) : int \rightarrow int list \rightarrow int list
```

```
Constant Rule
{y:int; x:int; one:int list}
                                  {y:int; x:int; one:int list}
                                   |- (y :: one) : int list
  - x:int
{y:int; x:int; one : int list} |- (x :: y :: one) : int list
    \{x:int; one : int list\} | -fun y -> (x :: y :: one) \}
                                 : int \rightarrow int list
```

Binary Operation Rule

By BinOp where  $(::): int \rightarrow int list \rightarrow int list$ 

	Variable Rule
Variable Rule	{; one:int list;}
{y:int;}  - y:int	- one : int list
{y:int; x:int; one:	int list} - (y :: one) : int list

10/7/21 44



### Curry - Howard Isomorphism

- Type Systems are logics; logics are type systems
- Types are propositions; propositions are types
- Terms are proofs; proofs are terms

 Function space arrow corresponds to implication; application corresponds to modus ponens



### Curry - Howard Isomorphism

Modus Ponens

$$\frac{\mathsf{A} \Rightarrow \mathsf{B} \quad \mathsf{A}}{\mathsf{B}}$$

Application

$$\Gamma \mid -e_1 : \alpha \to \beta \quad \Gamma \mid -e_2 : \alpha$$

$$\Gamma \mid -(e_1 e_2) : \beta$$

# Mea Culpa

- The above system can't handle polymorphism as in OCAML
- No type variables in type language (only metavariable in the logic)
- Would need:
  - Object level type variables and some kind of type quantification
  - let and let rec rules to introduce polymorphism
  - Explicit rule to eliminate (instantiate) polymorphism

### Support for Polymorphic Types

- Monomorpic Types  $(\tau)$ :
  - Basic Types: int, bool, float, string, unit, ...
  - Type Variables:  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , ε
  - Compound Types:  $\alpha \rightarrow \beta$ , int \* string, bool list, ...
- Polymorphic Types:
  - Monomorphic types τ
  - Universally quantified monomorphic types
  - $\forall \alpha_1, \ldots, \alpha_n$  .  $\tau$
  - Can think of  $\tau$  as same as  $\forall \cdot \tau$

#### **Example FreeVars Calculations**

- Vars('a -> (int -> 'b) -> 'a) ={'a , 'b}
- FreeVars (All 'b. 'a -> (int -> 'b) -> 'a) =
- {'a, 'b} {'b}= {'a}
- FreeVars {x : All `b. <u>a</u> -> (int -> `b) -> <u>a</u>,
- id: All 'c. 'c -> 'c,
- y: All 'c. 'a -> 'b -> 'c} =
- {'a} U {} U {'a, 'b} = {'a, 'b}

#### Support for Polymorphic Types

- Typing Environment \(\Gamma\) supplies polymorphic types (which will often just be monomorphic) for variables
- Free variables of monomorphic type just type variables that occur in it
  - Write FreeVars(τ)
- Free variables of polymorphic type removes variables that are universally quantified
  - FreeVars( $\forall \alpha_1, \dots, \alpha_n \cdot \tau$ ) = FreeVars( $\tau$ ) { $\alpha_1, \dots, \alpha_n$  }
- FreeVars( $\Gamma$ ) = all FreeVars of types in range of  $\Gamma$

### Monomorphic to Polymorphic

- Given:
  - type environment
  - monomorphic type τ
  - τ shares type variables with Γ
- Want most polymorphic type for τ that doesn't break sharing type variables with Γ
- Gen $(\tau, \Gamma) = \forall \alpha_1, ..., \alpha_n \cdot \tau$  where  $\{\alpha_1, ..., \alpha_n\} = \text{freeVars}(\tau) \text{freeVars}(\Gamma)$

#### Polymorphic Typing Rules

A type judgement has the form

$$\Gamma$$
 |- exp :  $\tau$ 

- I uses polymorphic types
- τ still monomorphic
- Most rules stay same (except use more general typing environments)
- Rules that change:
  - Variables
  - Let and Let Rec
  - Allow polymorphic constants
- Worth noting functions again



### Polymorphic Let and Let Rec

#### let rule:

$$\Gamma \mid -e_1 : \tau_1 \{x : Gen(\tau_1, \Gamma)\} + \Gamma \mid -e_2 : \tau_2 \}$$

$$\Gamma \mid -(let x = e_1 in e_2) : \tau_2$$

#### let rec rule:

$$\{x: \tau_1\} + \Gamma \mid -e_1:\tau_1 \{x: Gen(\tau_1, \Gamma)\} + \Gamma \mid -e_2:\tau_2 \Gamma_1$$
  
 $\Gamma \mid -(let rec x = e_1 in e_2):\tau_2$ 

### Polymorphic Variables (Identifiers)

#### Variable axiom:

$$\Gamma \mid -x : \varphi(\tau)$$
 if  $\Gamma(x) = \forall \alpha_1, ..., \alpha_n . \tau$ 

- Where  $\varphi$  replaces all occurrences of  $\alpha_1, \ldots, \alpha_n$  by monotypes  $\tau_1, \ldots, \tau_n$
- Note: Monomorphic rule special case:

$$\Gamma \mid -x : \tau$$
 if  $\Gamma(x) = \tau$ 

Constants treated same way



### Fun Rule Stays the Same

fun rule:

$$\{x \colon \tau_1\} + \Gamma \mid -e \colon \tau_2$$

$$\Gamma \mid -\text{ fun } x -> e \colon \tau_1 \to \tau_2$$

- Types  $\tau_1$ ,  $\tau_2$  monomorphic
- Function argument must always be used at same type in function body

#### Polymorphic Example

- Assume additional constants and primitive operators:
- hd : $\forall \alpha$ .  $\alpha$  list ->  $\alpha$
- tl:  $\forall \alpha$ .  $\alpha$  list ->  $\alpha$  list
- is\_empty :  $\forall \alpha$ .  $\alpha$  list -> bool
- (::) :  $\forall \alpha. \alpha \rightarrow \alpha \text{ list } \rightarrow \alpha \text{ list}$
- $\blacksquare$  [] :  $\forall \alpha$ .  $\alpha$  list

### Polymorphic Example

Show:

?

```
{} |- let rec length =
    fun I -> if is_empty I then 0
        else 1 + length (tl I)
    in length (2 :: []) + length(true :: []) : int
```

### -

#### Polymorphic Example: Let Rec Rule

```
■ Show: (1)
                                    (2)
{length: \alpha list -> int} {length: \forall \alpha. \alpha list -> int}
|- fun | -> ...
                           |- length (2 :: []) +
                              length(true :: []) : int
 : \alpha list -> int
{} |- let rec length =
       fun I -> if is empty I then 0
                  else 1 + length (tl l)
 in length (2 :: []) + length(true :: []) : int
```



### Polymorphic Example (1)

Show:

?

```
{length:\alpha list -> int} |-
fun | -> if is_empty | then 0
else 1 + length (tl | l)
```

:  $\alpha$  list -> int

### Polymorphic Example (1): Fun Rule

```
Show:
                (3)
{length: \alpha list -> int, I: \alpha list } |-
if is_empty I then 0
    else length (hd l) + length (tl l) : int
{ length: \alpha list -> int } |
fun I -> if is empty I then 0
                  else 1 + length (tl l)
: \alpha list -> int
```



### Polymorphic Example (3)

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

?

 $\Gamma$ |- if is\_empty | then 0 else 1 + length (tl | l) : int

### Polymorphic Example (3):IfThenElse

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

```
(4) (5) (6) \Gamma|-\text{ is\_empty I} \quad \Gamma|-\text{ 0:int} \quad \Gamma|-\text{ 1 + length (tl I)} : bool : int
```

 $\Gamma$ |- if is\_empty | then 0 else 1 + length (tl | ) : int



### Polymorphic Example (4)

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

 $\Gamma$  - is\_empty I : bool



### Polymorphic Example (4):Application

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

?

 $\Gamma$ |- is\_empty :  $\alpha$  list -> bool  $\Gamma$ |- I :  $\alpha$  list

 $\Gamma$ |- is\_empty | : bool

### Polymorphic Example (4)

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

```
By Const since \alpha list -> bool is instance of \forall \alpha. \alpha list -> bool
```

```
\Gamma|- is_empty : \alpha list -> bool \Gamma|- I : \alpha list
```

 $\Gamma$ |- is\_empty | : bool

### Polymorphic Example (4)

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

By Const since  $\alpha$  list -> bool is instance of  $\forall \alpha$ .  $\alpha$  list -> bool  $\Gamma(I) = \alpha$  list

By Variable

 $\Gamma$ |- is\_empty :  $\alpha$  list -> bool

 $\Gamma$ |-|:  $\alpha$  list

 $\Gamma$ |- is empty | : bool

This finishes (4)



### Polymorphic Example (5):Const

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

By Const Rule

 $\Gamma$ |- 0:int



### Polymorphic Example (6):Arith Op

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

 $\Gamma$ |-1 + length (tl l) : int



### Polymorphic Example (7):App Rule

- Let  $\Gamma = \{ length : \alpha list -> int, l: \alpha list \}$
- Show

$$\Gamma$$
 | - tl :  $\alpha$  list ->  $\alpha$  list

By Variable

$$\Gamma$$
 - I :  $\alpha$  list

$$\Gamma$$
 |- (tl l) :  $\alpha$  list

By Const since  $\alpha$  list ->  $\alpha$  list is instance of  $\forall \alpha$ .  $\alpha$  list ->  $\alpha$  list

### Polymorphic Example: (2) by ArithOp

- Let  $\Gamma'$  = {length:  $\forall \alpha$ .  $\alpha$  list -> int}
- Show:

```
(8) (9) \Gamma' |- \Gamma' |- \Gamma' |- length (2 :: []) : int length(true :: []) : int {length: \forall \alpha. \ \alpha \ list \ -> int}  |- length (2 :: []) + length(true :: []) : int
```



### Polymorphic Example: (8)AppRule

- Let  $\Gamma' = \{length: \forall \alpha. \alpha | list -> int\}$
- Show:

```
\Gamma' |- length : int list -> int \Gamma' |- (2 :: []) :int list \Gamma' |- length (2 :: []) :int
```