

Bytecode Interpreters

Concepts of Programming Languages

Lecture 26

Outline

- » Course Evaluations!
- » Discuss **stack-based languages** and **stack machines**
- » Look briefly at how to compile **variables** and **functions**
- » Finish up the dang course

Course evaluations!

Recap

Recall: Abstract/Virtual Machines

Recall: Abstract/Virtual Machines

A **virtual machine** is a computational abstraction, like a Turing machine (but usually *easier to implement*)

Recall: Abstract/Virtual Machines

A **virtual machine** is a computational abstraction, like a Turing machine (but usually *easier to implement*)

Virtual machines are typically implemented as **bytecode interpreters**, where "programs" are streams of bytes and a command is represented as a byte (plus possibly some extra data)

Recall: Benefits of Stack Machines

Recall: Benefits of Stack Machines

Simplicity: Stacks aren't too complicated

Recall: Benefits of Stack Machines

Simplicity: Stacks aren't too complicated

Portability: Any OS should be able to handle a stream of bytes, so the machine dependent part of our programming language can be simplified

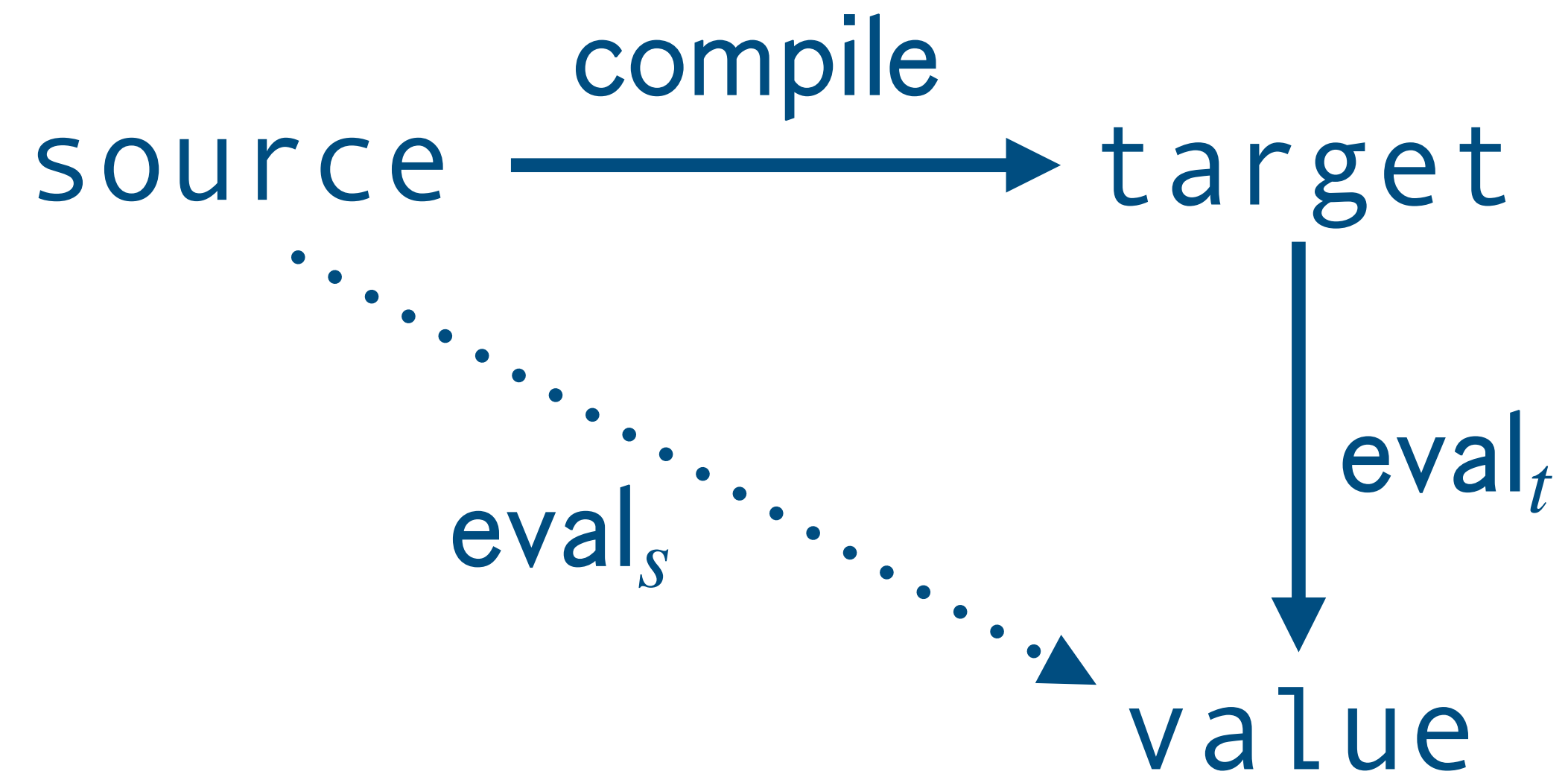
Recall: Benefits of Stack Machines

Simplicity: Stacks aren't too complicated

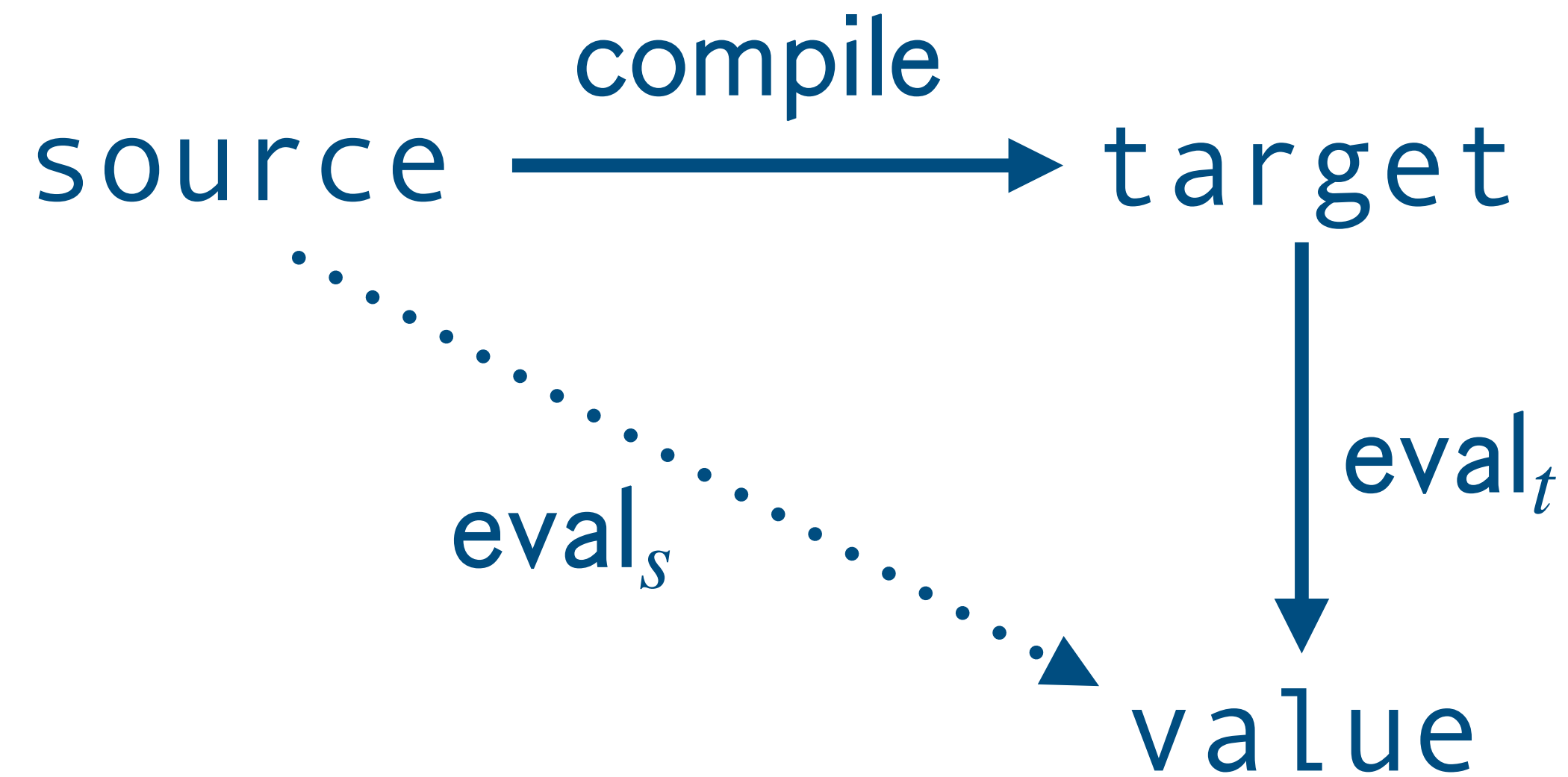
Portability: Any OS should be able to handle a stream of bytes, so the machine dependent part of our programming language can be simplified

Efficiency (sort of): They can be implemented in low-level languages, and so will generally be faster than the interpreters we build in this course (though not as fast as natively compiled code)

Recall: Compilation

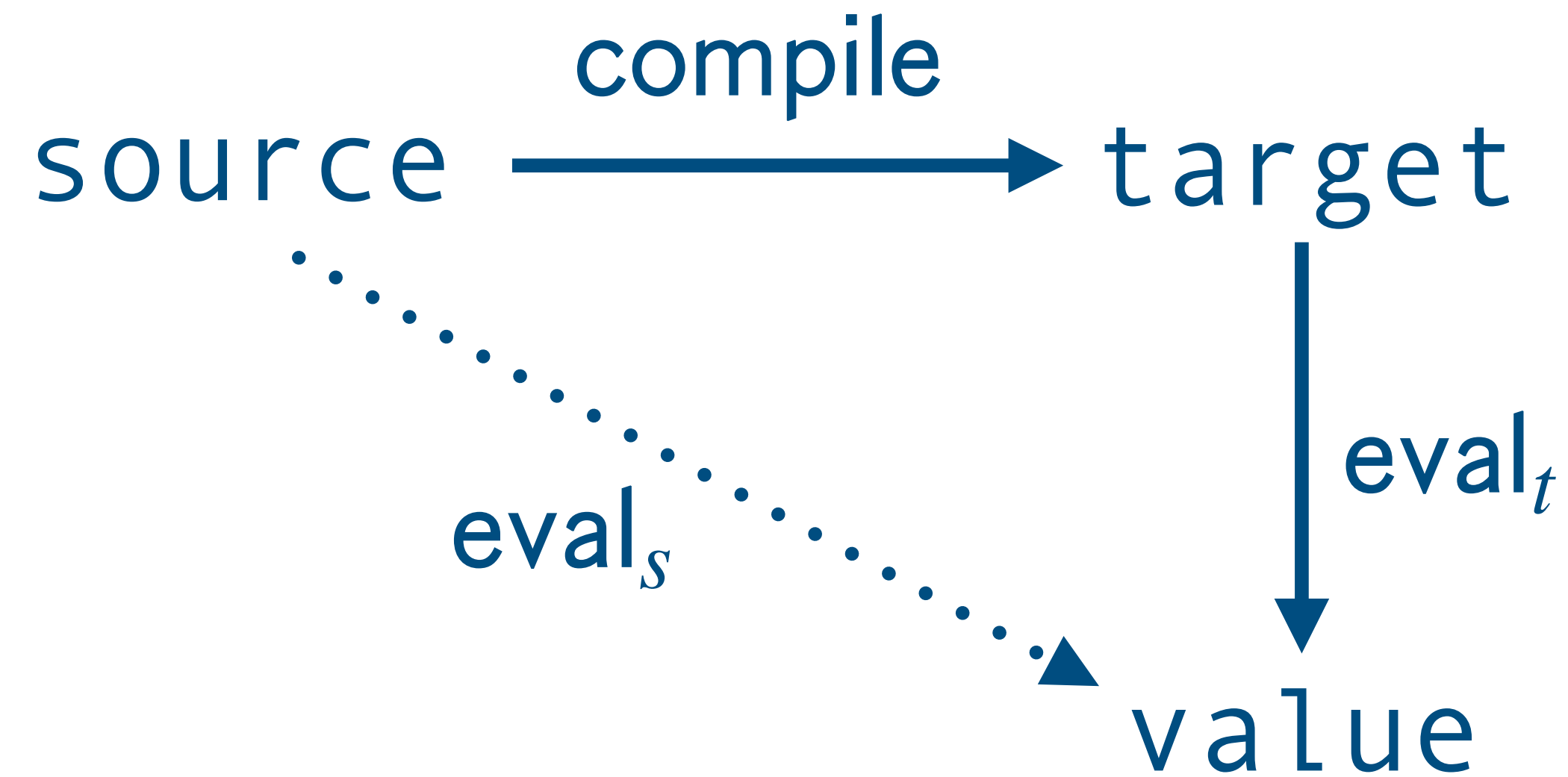


Recall: Compilation



Compilation is the process of translating a program in one language to another, maintaining semantic behavior

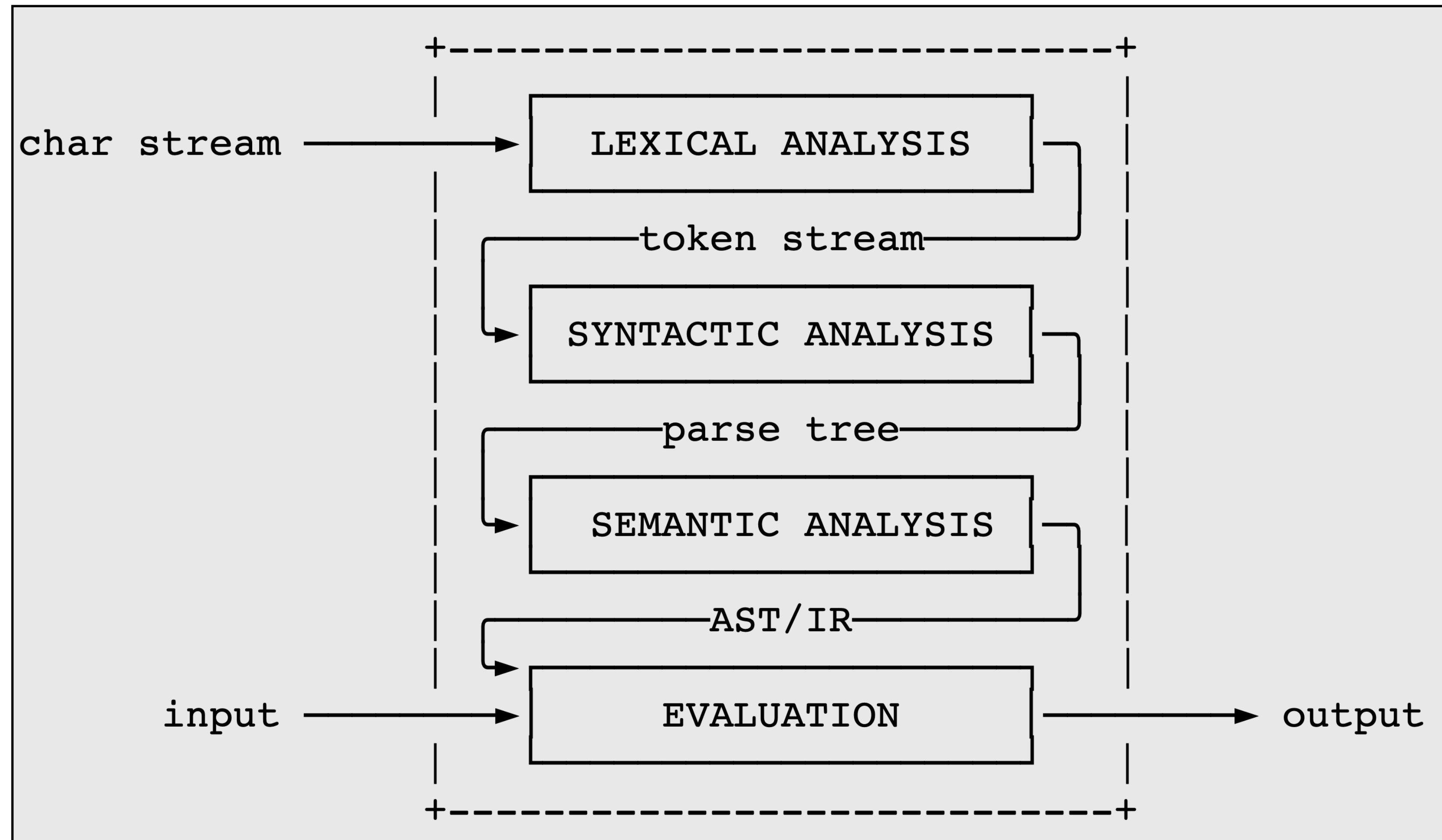
Recall: Compilation



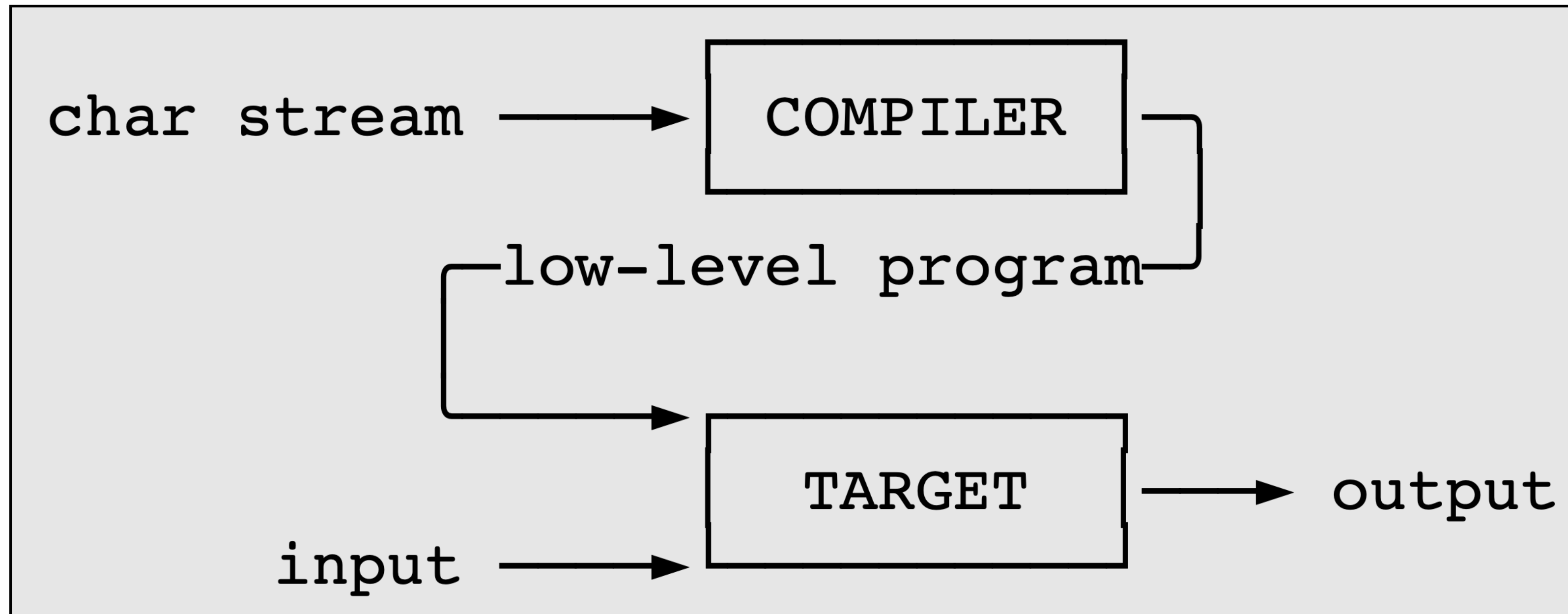
Compilation is the process of translating a program in one language to another, maintaining semantic behavior

Compilation can be a part of interpretation as well, like with **bytecode interpretation**

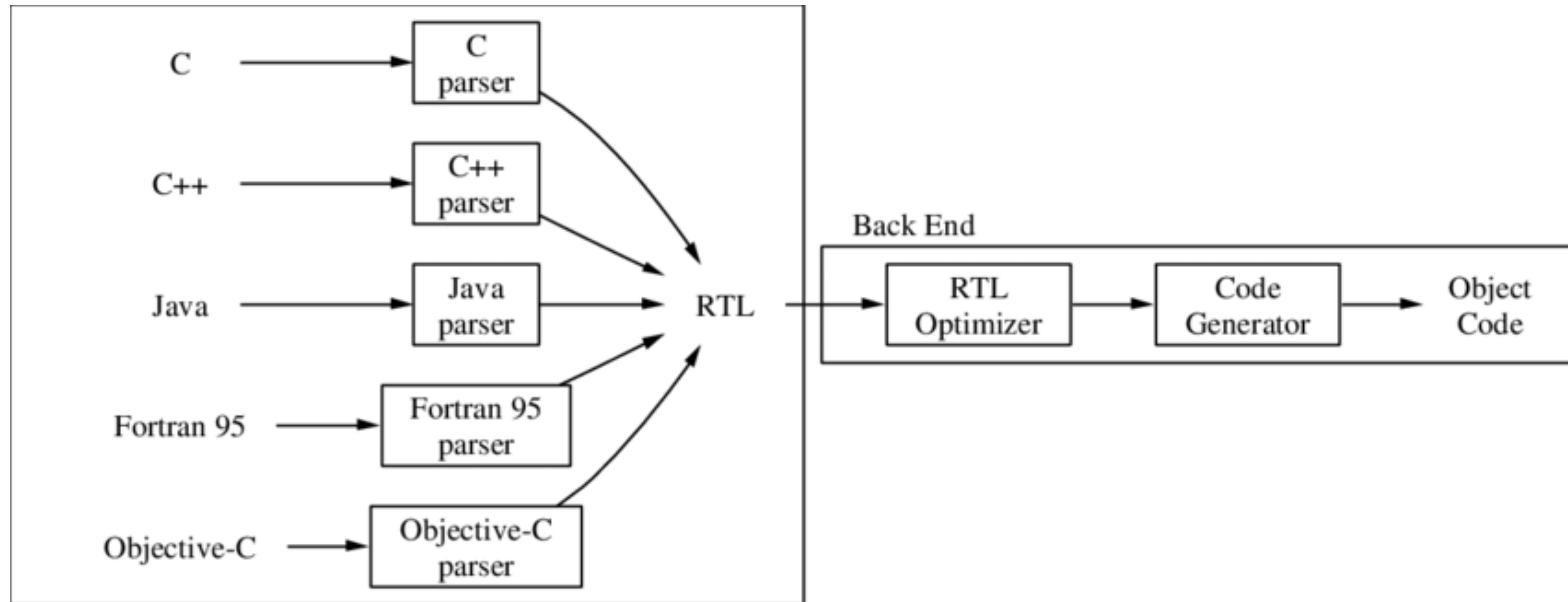
Recall: The Interpretation Pipeline



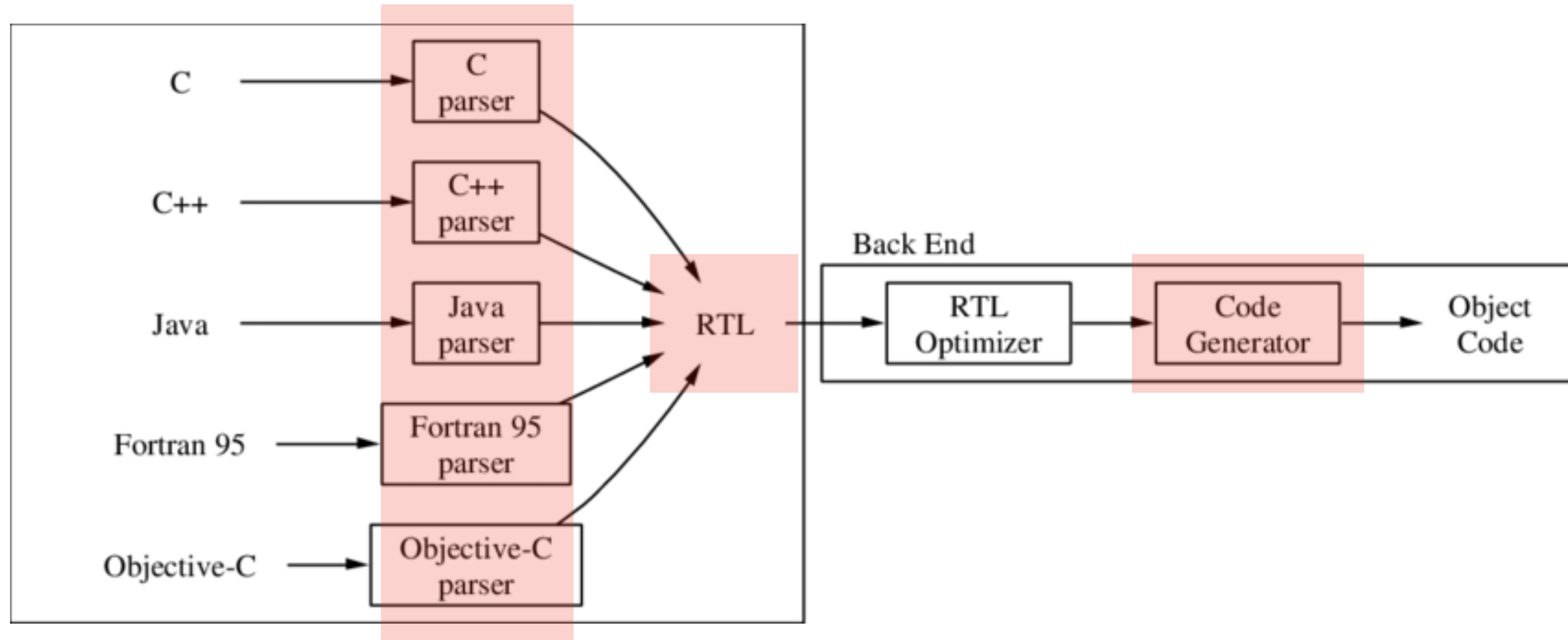
The Compiler Pipeline



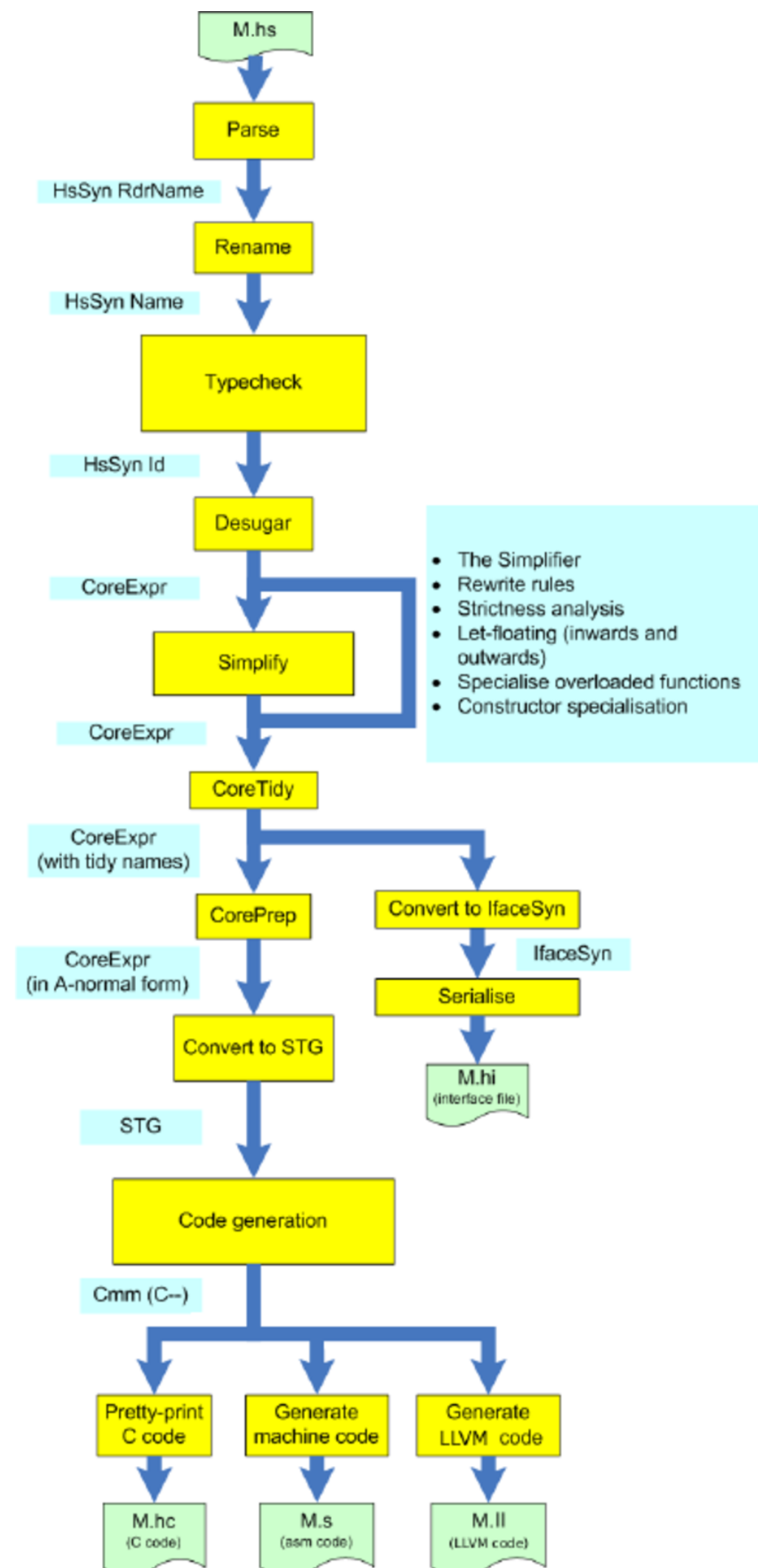
Example Pipelines: gcc



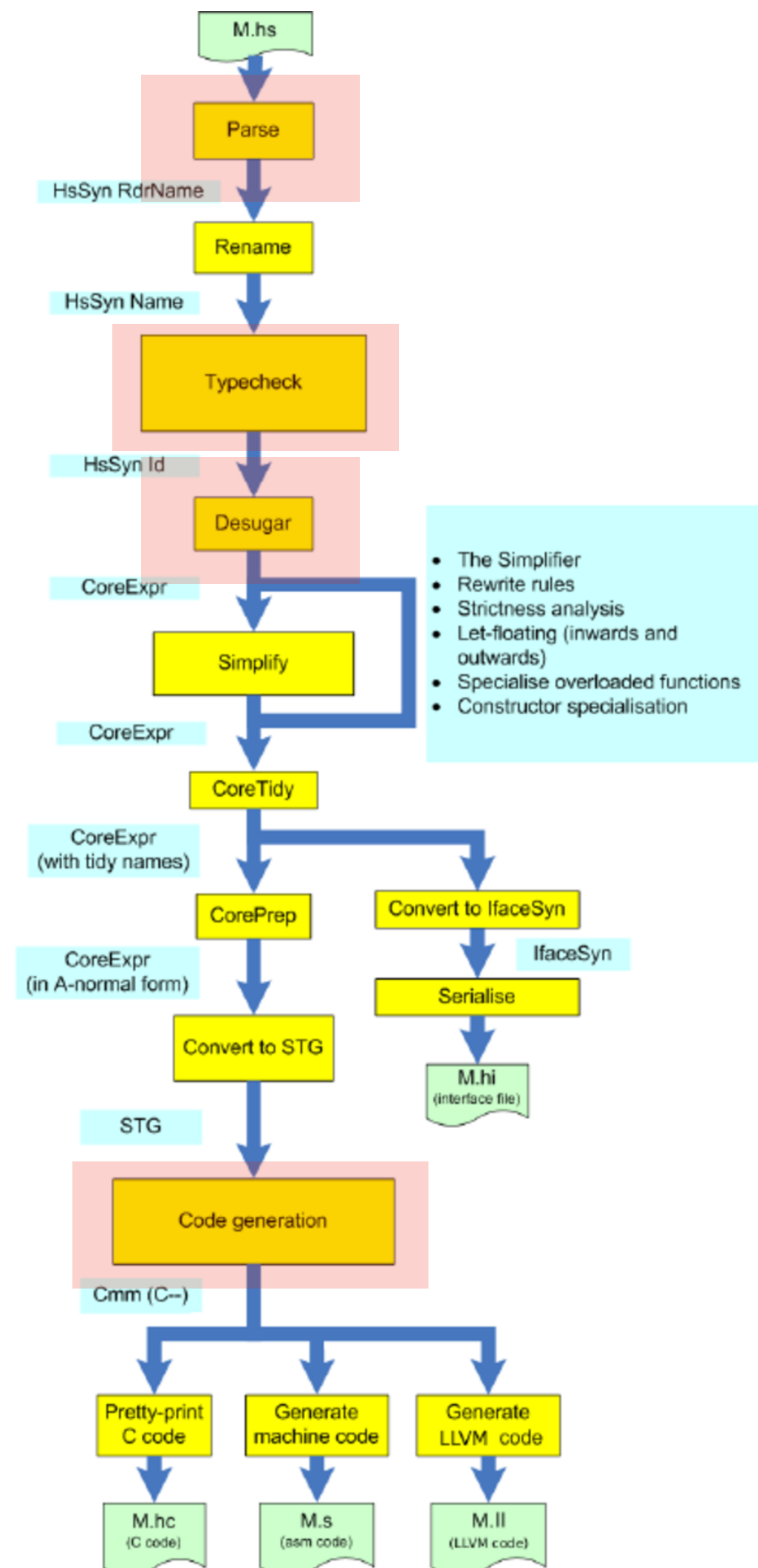
Example Pipelines: gcc



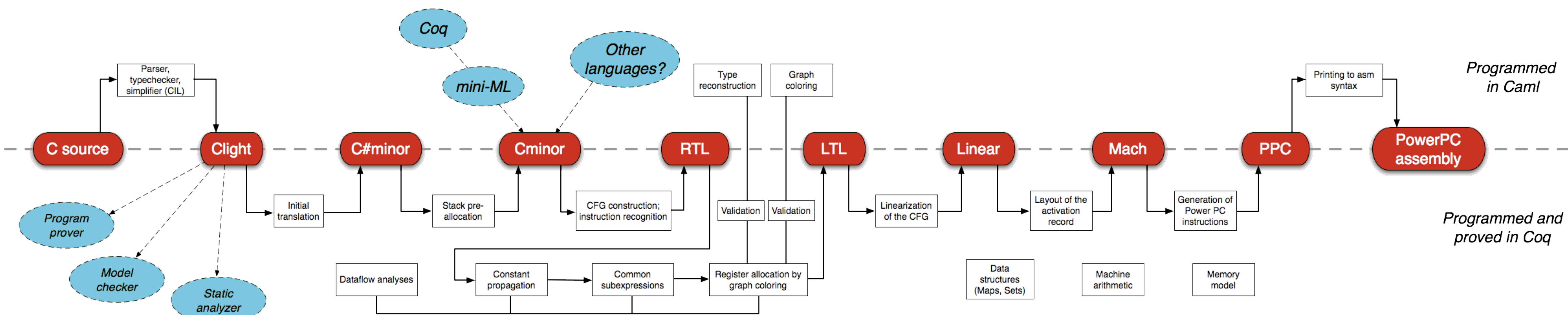
Example Pipelines: GHC (Haskell)



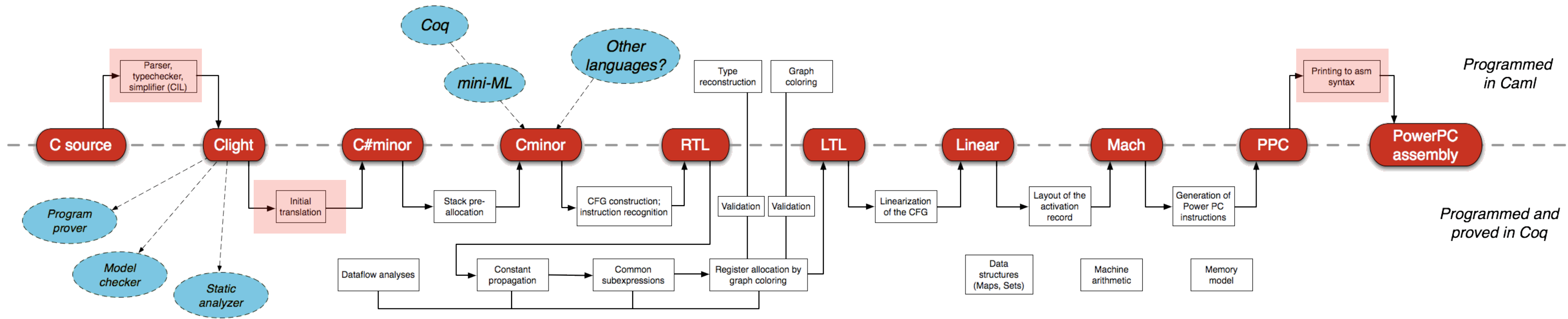
Example Pipelines: GHC (Haskell)



Example Pipelines: CompCert (C)



Example Pipelines: CompCert (C)



Stack-Based Arithmetic

Stack-Based Arithmetic (Syntax)

$\langle \text{prog} \rangle ::= \{ \langle \text{com} \rangle \}$

$\langle \text{com} \rangle ::= \text{ADD} \mid \text{SUB} \mid \text{MUL} \mid \text{DIV} \mid \text{PUSH } \langle \text{num} \rangle$

$\langle \text{num} \rangle ::= \mathbb{Z}$

Stack-Based Arithmetic (Semantics)

$$\langle \mathcal{S}, P \rangle$$

A **value** is an integer (\mathbb{Z})

A **configuration** is made up of a stack (\mathcal{S}) of values and a program (P) given by **<prog>**

Stack-Based Arithmetic (Semantics)

$$\frac{}{\langle m :: n :: \mathcal{S}, \text{ADD } P \rangle \longrightarrow \langle (m + n) :: \mathcal{S}, P \rangle} \text{ (add)}$$

$$\frac{}{\langle m :: n :: \mathcal{S}, \text{SUB } P \rangle \longrightarrow \langle (m - n) :: \mathcal{S}, P \rangle} \text{ (sub)}$$

$$\frac{}{\langle m :: n :: \mathcal{S}, \text{MUL } P \rangle \longrightarrow \langle (m \times n) :: \mathcal{S}, P \rangle} \text{ (mul)}$$

$$\frac{n \neq 0}{\langle m :: n :: \mathcal{S}, \text{DIV } P \rangle \longrightarrow \langle (m/n) :: \mathcal{S}, P \rangle} \text{ (div)}$$

$$\frac{}{\langle \mathcal{S}, \text{PUSH } n \ P \rangle \longrightarrow \langle n :: \mathcal{S}, P \rangle} \text{ (push)}$$

Example (Evaluation)

PUSH 2 PUSH 3 SUB PUSH 4 MUL

demo
(stack machine)

Compiling Arithmetic Expressions

	n	\Rightarrow	PUSH n
e_1	+ e_2	\Rightarrow	$\mathcal{C}(e_2)$ $\mathcal{C}(e_1)$ ADD
e_1	- e_2	\Rightarrow	$\mathcal{C}(e_2)$ $\mathcal{C}(e_1)$ SUB
e_1	* e_2	\Rightarrow	$\mathcal{C}(e_2)$ $\mathcal{C}(e_1)$ MUL
e_1	/ e_2	\Rightarrow	$\mathcal{C}(e_2)$ $\mathcal{C}(e_1)$ DIV

We need a procedure \mathcal{C} for converting an arithmetic expression into a stack program. *Note the order!*

Example (Compilation)

4 * (2 - 3)

demo

(compiling arithmetic expressions)

Variables

Variables (Syntax)

$\langle \text{prog} \rangle ::= \{ \langle \text{com} \rangle \}$

$\langle \text{com} \rangle ::= \text{ADD} \mid \text{SUB} \mid \text{MUL} \mid \text{DIV} \mid \text{PUSH } \langle \text{num} \rangle$
 $\mid \text{ASSIGN } \langle \text{var} \rangle \mid \text{LOOKUP } \langle \text{var} \rangle$

$\langle \text{num} \rangle ::= \mathbb{Z}$

$\langle \text{var} \rangle ::= \mathbb{I}$

Variables (Semantics)

$$\langle \mathcal{S}, \mathcal{E}, P \rangle$$

A **value** is an integer (\mathbb{Z})

A **configuration** is made up of a stack \mathcal{S} of values, an environment \mathcal{E} (mapping of identifiers to values), and a program P given by **<prog>**

Variables (Semantics)

$$\frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{ADD } P \rangle \longrightarrow \langle (m + n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{(add)} \quad \frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{SUB } P \rangle \longrightarrow \langle (m - n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{(sub)}$$

$$\frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{MUL } P \rangle \longrightarrow \langle (m \times n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{(mul)} \quad \frac{n \neq 0}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{DIV } P \rangle \longrightarrow \langle (m/n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{(div)}$$

$$\frac{}{\langle \mathcal{S}, \mathcal{E}, \text{PUSH } n P \rangle \longrightarrow \langle n :: \mathcal{S}, \mathcal{E}, P \rangle} \text{(push)}$$

$$\frac{}{\langle n :: \mathcal{S}, \mathcal{E}, \text{ASSIGN } x P \rangle \longrightarrow \langle \mathcal{S}, \mathcal{E}[x \mapsto n], P \rangle} \text{(asn)} \quad \frac{}{\langle n :: \mathcal{S}, \mathcal{E}, \text{LOOKUP } x P \rangle \longrightarrow \langle \mathcal{E}(x) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{(lkp)}$$

Variables (Semantics)

basically the same

$$\frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{ADD } P \rangle \longrightarrow \langle (m + n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (add)} \quad \frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{SUB } P \rangle \longrightarrow \langle (m - n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (sub)}$$

$$\frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{MUL } P \rangle \longrightarrow \langle (m \times n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (mul)} \quad \frac{n \neq 0}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{DIV } P \rangle \longrightarrow \langle (m/n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (div)}$$

$$\frac{}{\langle \mathcal{S}, \mathcal{E}, \text{PUSH } n P \rangle \longrightarrow \langle n :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (push)}$$

$$\frac{}{\langle n :: \mathcal{S}, \mathcal{E}, \text{ASSIGN } x P \rangle \longrightarrow \langle \mathcal{S}, \mathcal{E}[x \mapsto n], P \rangle} \text{ (asn)} \quad \frac{}{\langle n :: \mathcal{S}, \mathcal{E}, \text{LOOKUP } x P \rangle \longrightarrow \langle \mathcal{E}(x) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (lkp)}$$

Variables (Semantics)

basically the same

$$\frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{ADD } P \rangle \longrightarrow \langle (m + n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (add)} \quad \frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{SUB } P \rangle \longrightarrow \langle (m - n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (sub)}$$

$$\frac{}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{MUL } P \rangle \longrightarrow \langle (m \times n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (mul)} \quad \frac{n \neq 0}{\langle m :: n :: \mathcal{S}, \mathcal{E}, \text{DIV } P \rangle \longrightarrow \langle (m/n) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (div)}$$

$$\frac{}{\langle \mathcal{S}, \mathcal{E}, \text{PUSH } n P \rangle \longrightarrow \langle n :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (push)}$$

new rules

$$\frac{}{\langle n :: \mathcal{S}, \mathcal{E}, \text{ASSIGN } x P \rangle \longrightarrow \langle \mathcal{S}, \mathcal{E}[x \mapsto n], P \rangle} \text{ (asn)} \quad \frac{}{\langle n :: \mathcal{S}, \mathcal{E}, \text{LOOKUP } x P \rangle \longrightarrow \langle \mathcal{E}(x) :: \mathcal{S}, \mathcal{E}, P \rangle} \text{ (lkp)}$$

Example (Evaluation)

PUSH 2 ASSIGN x PUSH 3 ASSIGN y
LOOKUP x LOOKUP y ADD

Compiling Let-Expressions (Attempt)

x \implies **LOOKUP** x

let $x = e_1$ **in** e_2 \implies $\mathcal{C}(e_1)$ **ASSIGN** x $\mathcal{C}(e_2)$

Compiling Let-Expressions (Attempt)

x \implies **LOOKUP** x

let $x = e_1$ **in** e_2 \implies $\mathcal{C}(e_1)$ **ASSIGN** x $\mathcal{C}(e_2)$

Except this isn't quite right

Example

let $y = 1$ in

let $x = \text{let } y = 2 \text{ in } y$ in

y

Scoping

```
let y = 1 in  
let x = let y = 2 in y in  
y
```

Scoping

```
let y = 1 in  
let x = let y = 2 in y in  
y
```

The language we've just described is only good for compiling from languages with **dynamic scoping**

Scoping

```
let y = 1 in  
let x = let y = 2 in y in  
y
```

The language we've just described is only good for compiling from languages with **dynamic scoping**

We can use closures to deal with lexical scoping (and functions)!

Functions

The Rough Picture

```
let k = fun x -> fun y -> x in  
let a = k 2 in  
a 3
```

*Compilation is just a big sequence of
transformations*

The Rough Picture

```
(fun k ->  
  let a = k 2 in  
  a 3)  
(fun x -> fun y -> x)
```

We can simulate let expressions with functions!
(we did this in lab)

The Rough Picture

```
(fun k ->  
  (fun a -> a 3)  
  (k 2))  
(fun x -> fun y -> x)
```

and again...

The Rough Picture

```
[ (fun k ->  
  (fun a -> a 3)  
  (k 2))  
  (fun x -> fun y -> x) ]
```

that was just to make the next part more convenient...
think of `[expr]` as as `compile(expr)`

The Rough Picture

```
[ (fun x -> fun y -> x) ]  
[ (fun k -> (fun a -> a 3) (k 2)) ]  
CALL
```

*We introduce as **CALL** command to call functions
Note the order, function/argument will go on a stack!*

The Rough Picture

```
FUN ? X
  [fun y -> x]
  RETURN
[ (fun k -> (fun a -> a 3) (k 2)) ]
CALL
```

*We introduce a **FUN** command to define functions
and a **RETURN** command to return from functions*

The Rough Picture

```
FUN ? X
  FUN ? Y
    [x]
    RETURN
  RETURN
[ (fun k -> (fun a -> a 3) (k 2)) ]
CALL
```

and again...

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
[ (fun k -> (fun a -> a 3) (k 2)) ]
CALL
```

*The familiar **LOOKUP** command...*
And functions let us know how many commands they have

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  [ (fun a -> a 3) (k 2) ]
  RETURN
CALL
```

and we can keep going...

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  [k 2]
  [fun a -> a 3]
  CALL
  RETURN
CALL
```

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  [2]
  [k]
  CALL
  [fun a -> a 3]
  CALL
  RETURN
CALL
```


The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  PUSH 2
  [k]
  CALL
  [fun a -> a 3]
  CALL
  RETURN
CALL
```

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  PUSH 2
  LOOKUP K
  CALL
  [fun a -> a 3]
  CALL
  RETURN
CALL
```

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  PUSH 2
  LOOKUP K
  CALL
  FUN ? A
    [a 3]
    RETURN
  CALL
  RETURN
CALL
```

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  PUSH 2
  LOOKUP K
  CALL
  FUN ? A
    [3]
    [a]
    CALL
    RETURN
  CALL
  RETURN
CALL
```

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN ? K
  PUSH 2
  LOOKUP K
  CALL
  FUN ? A
    PUSH 3
    [a]
    CALL
    RETURN
  CALL
  RETURN
CALL
```

The Rough Picture

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN 10 K
  PUSH 2
  LOOKUP K
  CALL
  FUN 4 A
    PUSH 3
    LOOKUP A
    CALL
    RETURN
  CALL
  RETURN
CALL
```

The Rough Picture

```
let k = fun x -> fun y -> x in
let a = k 2 in
a 3
```



```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN 10 K
  PUSH 2
  LOOKUP K
  CALL
  FUN 4 A
    PUSH 3
    LOOKUP A
    CALL
    RETURN
  CALL
  RETURN
CALL
```

Compilation is just a big sequence of transformations

The Rough Picture

```
let k = fun x -> fun y -> x in
let a = k 2 in
a 3
```



*Byte-code interpretation additionally
maps each command to a byte value*

```
7 4
7 2
10 1
9
9
7 10
0 2
10 0
8
7 4
0 3
10 0
8
9
8
9
8
```


Syntax

```
<prog> ::= { <com> }  
<com>  ::= PUSH <int>  
        | ADD   | SUB   | MUL   | DIV  
        | FUN <ident> <int> | CALL | RETURN  
        | LOOKUP
```

Semantics (Configurations)

$$\langle \mathcal{S}, \mathcal{E}, P \rangle$$

A **value** is an integer (\mathbb{Z}) or a closure (\mathbb{C}) of the form (\mathcal{E}, x, P)

A **configuration** is made up of a stack \mathcal{S} of values, an environment \mathcal{E} (mapping of identifiers to values) and a program P given by **<prog>**

Semantics (Functions)

$$\frac{}{\langle \mathcal{S}, \mathcal{E}, \text{FUN } x \ n \ P \rangle \longrightarrow \langle (\mathcal{E}, x, P[1..n]) :: \mathcal{S}, \mathcal{E}'[x \mapsto v], P[n+1..] \rangle} \text{ (fun)}$$

Function definitions carry a **parameter name** and an **offset**, which we use to construct the closure

Semantics (Continuation Passing)

$$\frac{}{\langle (\mathcal{E}', x, P') :: v :: \mathcal{S}, \mathcal{E}, \text{CALL } P \rangle \longrightarrow \langle (\mathcal{E}, _, P) :: \mathcal{S}, \mathcal{E}'[x \mapsto v], P' \rangle} \text{ (call)}$$

$$\frac{}{\langle v :: (\mathcal{E}', _, P') :: \mathcal{S}, \mathcal{E}, \text{RETURN } P \rangle \longrightarrow \langle v :: \mathcal{S}, \mathcal{E}', P' \rangle} \text{ (ret)}$$

One challenge: when we call a function, where to we "return" to?

Answer: We put the information on the stack itself in a closure!

Example

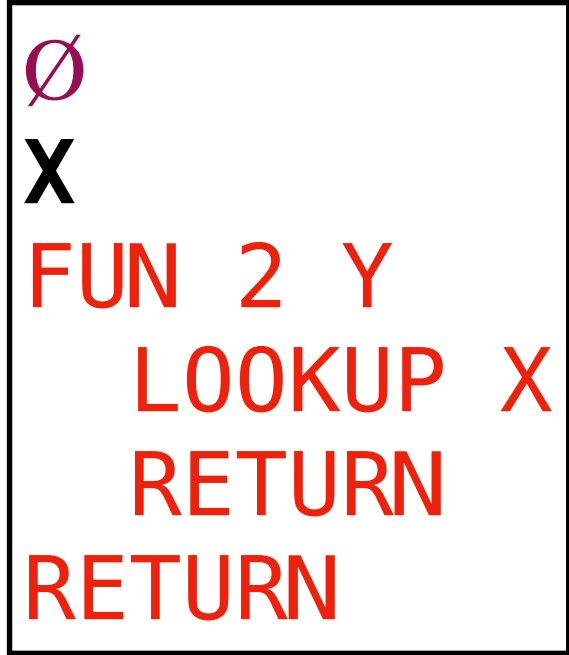
Stack:

Env:

```
FUN 4 X
  FUN 2 Y
    LOOKUP X
    RETURN
  RETURN
FUN 10 K
  PUSH 2
  LOOKUP K
  CALL
  FUN 4 A
    PUSH 3
    LOOKUP A
    CALL
    RETURN
  CALL
  RETURN
CALL
```

Example

Stack:



Env:

FUN 10 K
PUSH 2
LOOKUP K
CALL
FUN 4 A
PUSH 3
LOOKUP A
CALL
RETURN
CALL
RETURN
CALL

Example

Stack:

Ø

X

FUN 2 Y

LOOKUP X

RETURN

RETURN

Ø

K

PUSH 2

LOOKUP K

CALL

FUN 4 A

PUSH 3

LOOKUP A

CALL

RETURN

CALL

RETURN

Env:

CALL

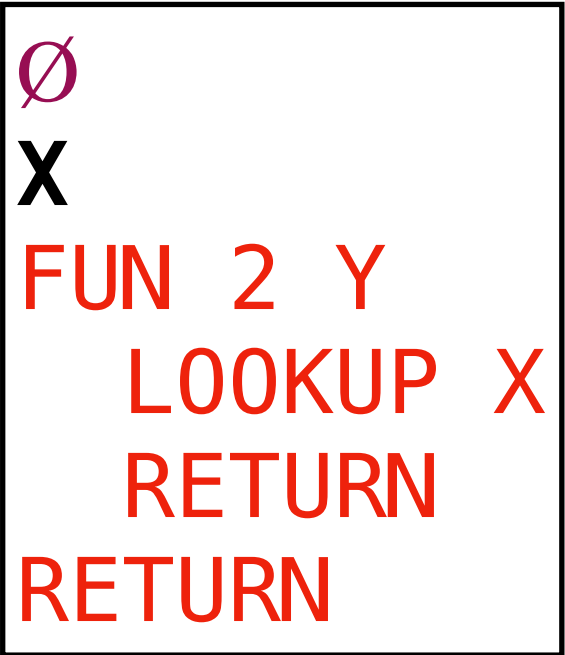
Example

Stack:



Env:

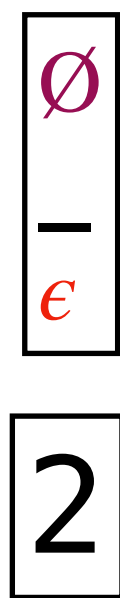
K \mapsto



PUSH 2
LOOKUP K
CALL
FUN 4 A
 PUSH 3
 LOOKUP A
 CALL
 RETURN
CALL
RETURN

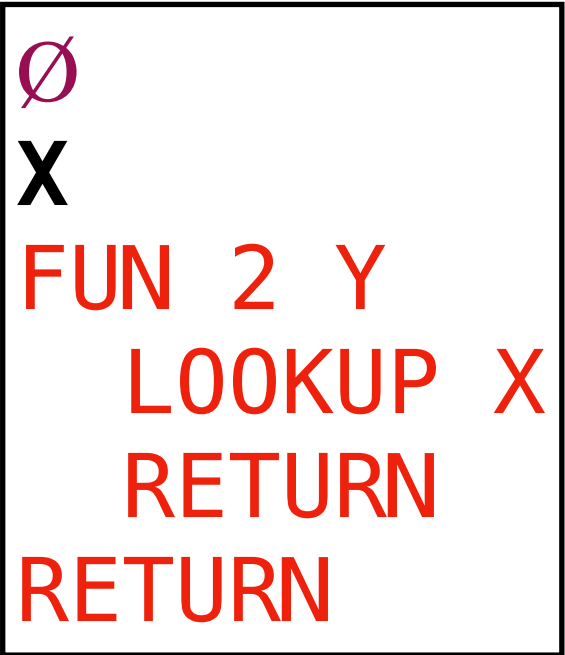
Example

Stack:



Env:

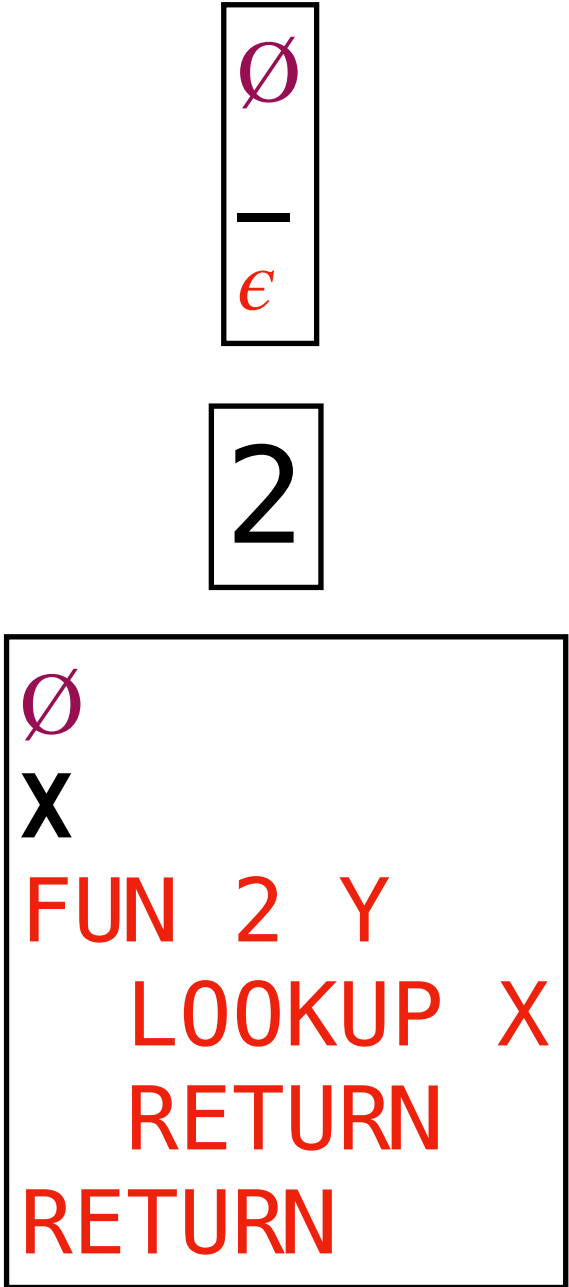
K \mapsto



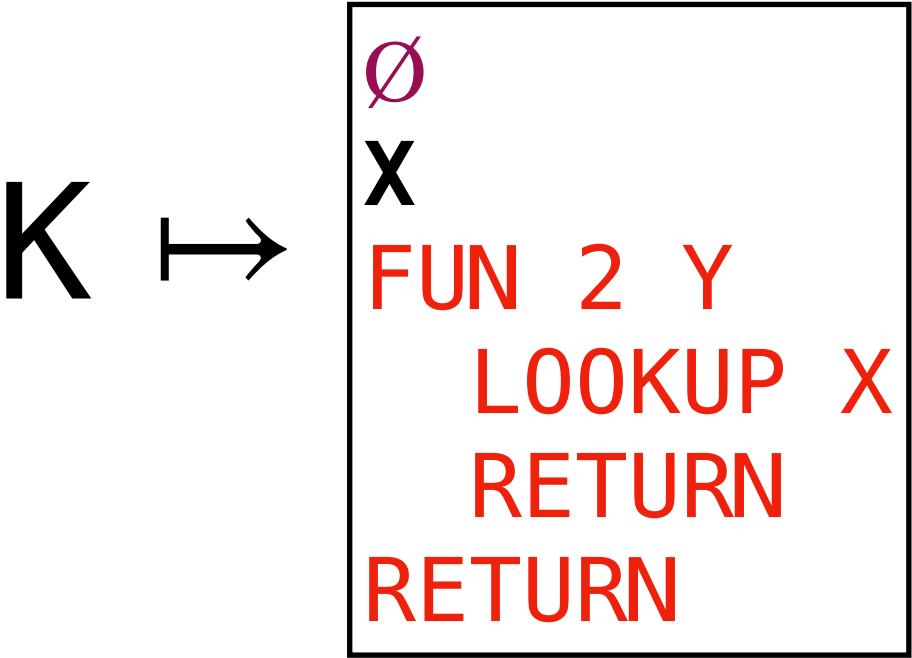
LOOKUP K
CALL
FUN 4 A
 PUSH 3
 LOOKUP A
 CALL
 RETURN
CALL
RETURN

Example

Stack:



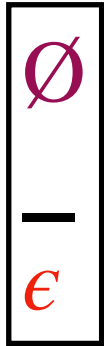
Env:



```
CALL
FUN 4 A
    PUSH 3
    LOOKUP A
    CALL
    RETURN
CALL
RETURN
```

Example

Stack:



K \mapsto {...}
—
FUN 4 A
 PUSH 3
 LOOKUP A
 CALL
 RETURN
CALL
RETURN

Env:

X \mapsto 2

FUN 2 Y
 LOOKUP X
 RETURN
RETURN

Example

Stack:

Ø

—

ϵ

K ↦ {...}

—

FUN 4 A

PUSH 3

LOOKUP A

CALL

RETURN

CALL

RETURN

X ↦ 2

Y

LOOKUP X

RETURN

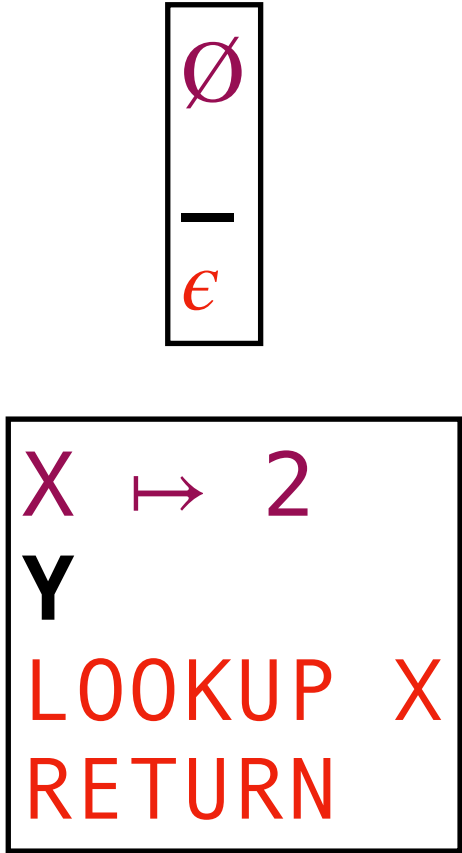
Env:

X ↦ 2

RETURN

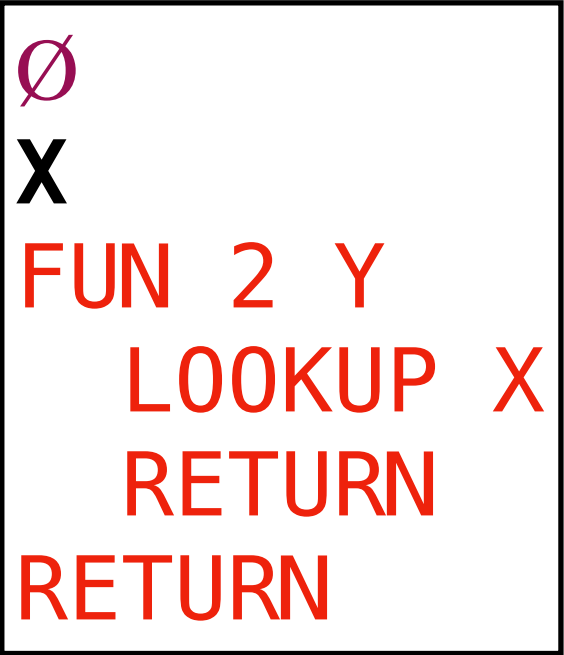
Example

Stack:



Env:

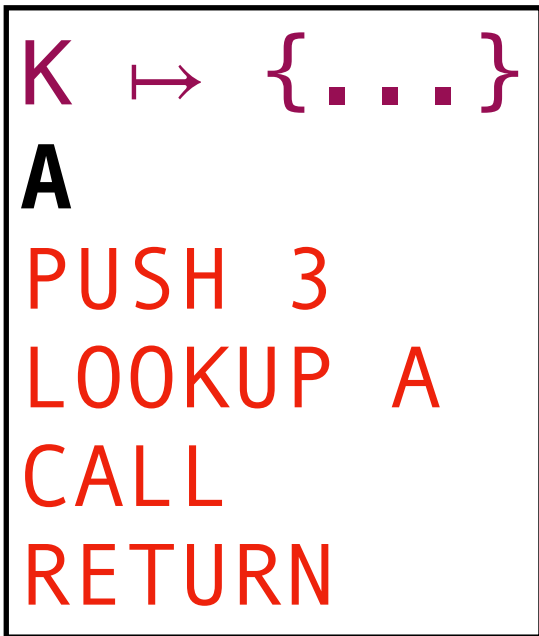
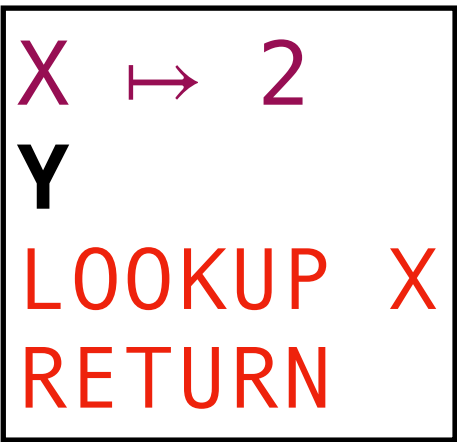
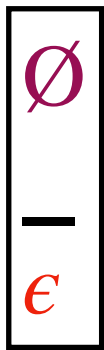
K \mapsto



FUN 4 A
PUSH 3
LOOKUP A
CALL
RETURN
CALL
RETURN

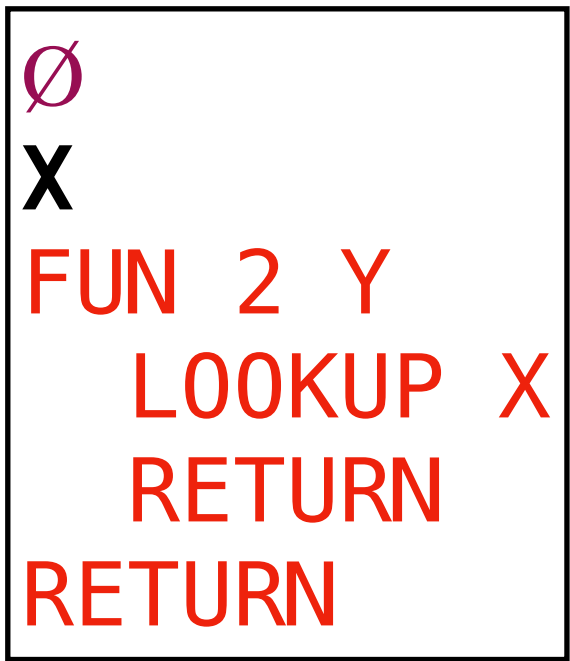
Example

Stack:



Env:

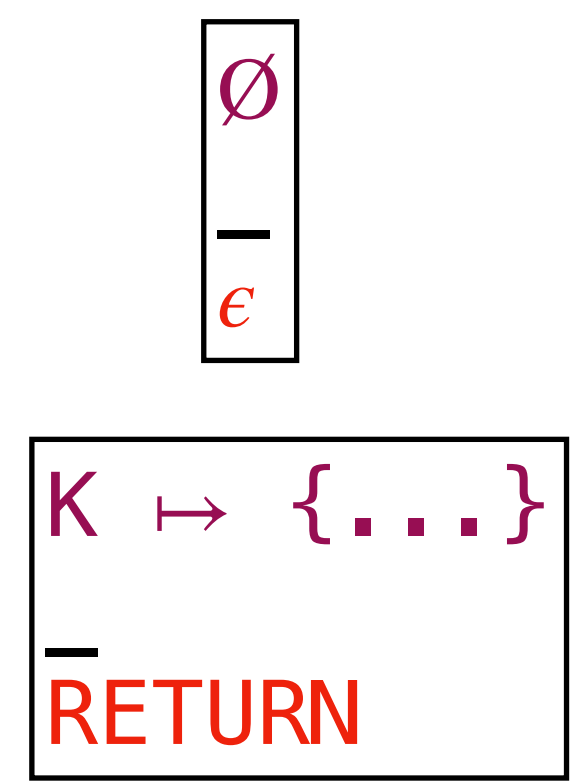
K ↦



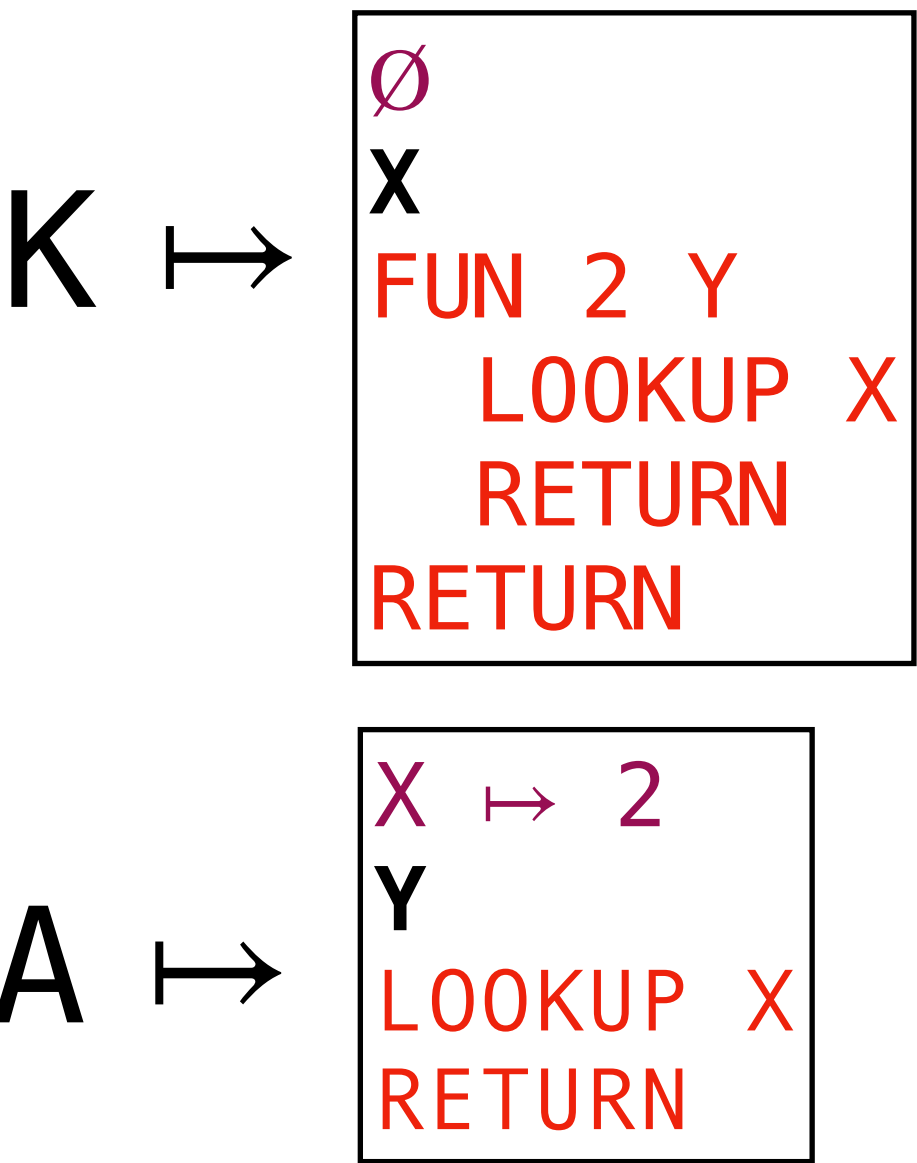
CALL
RETURN

Example

Stack:



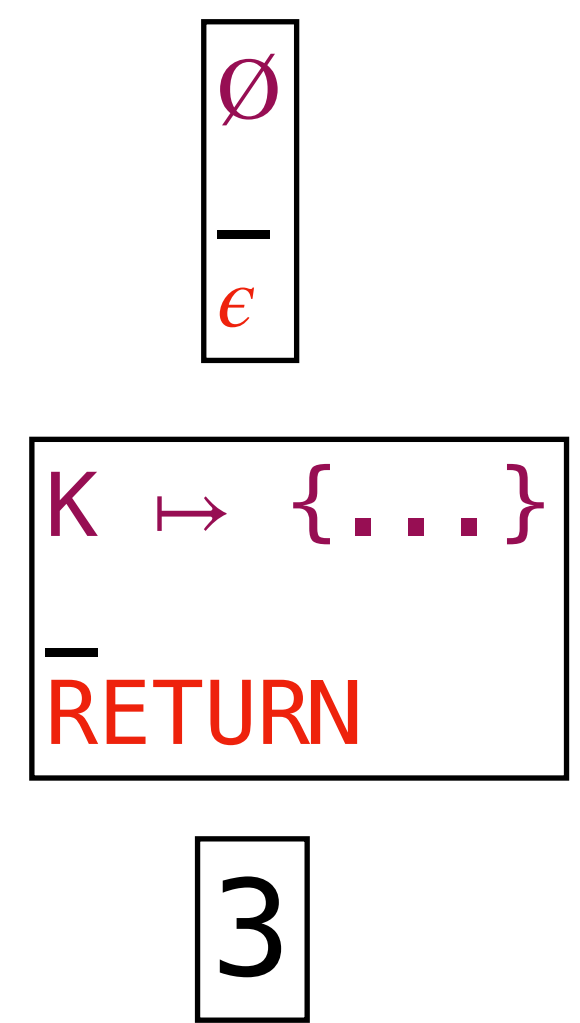
Env:



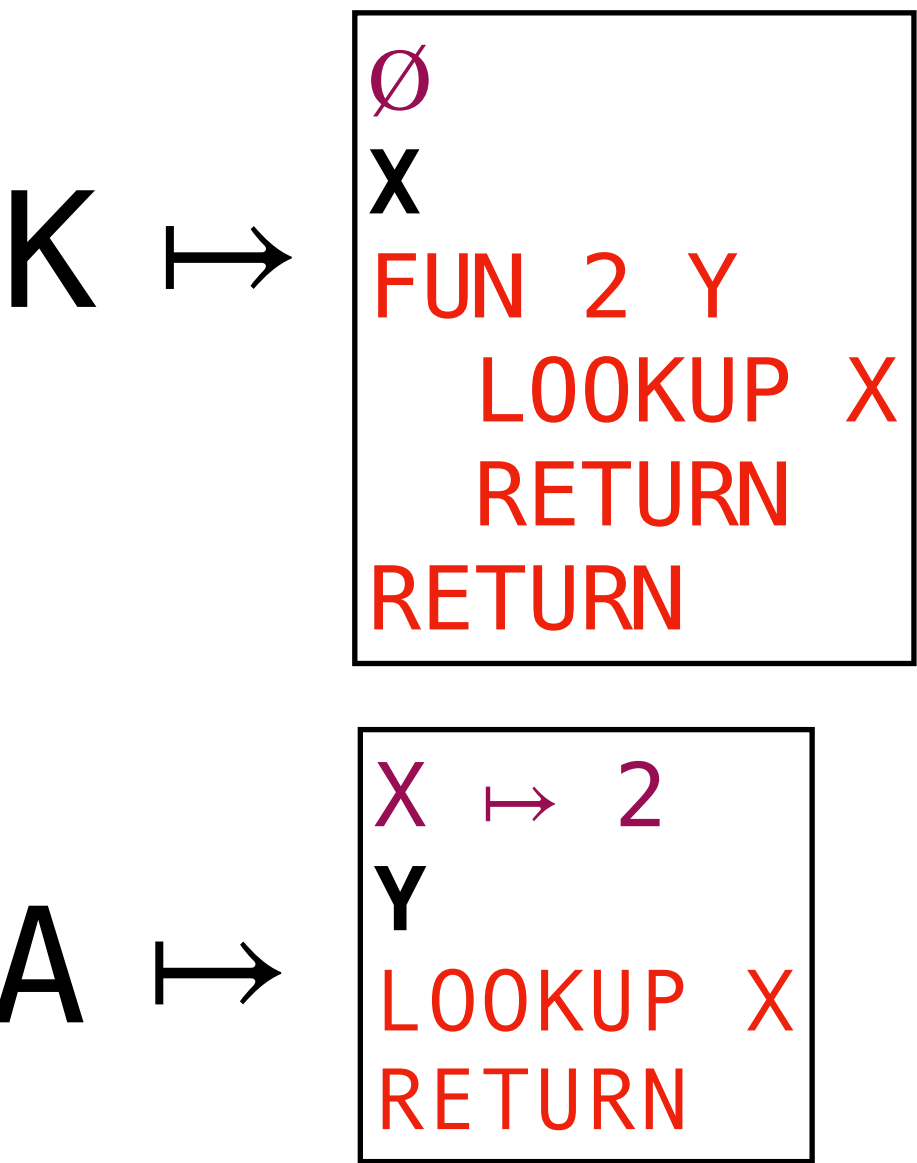
PUSH 3
LOOKUP A
CALL
RETURN

Example

Stack:



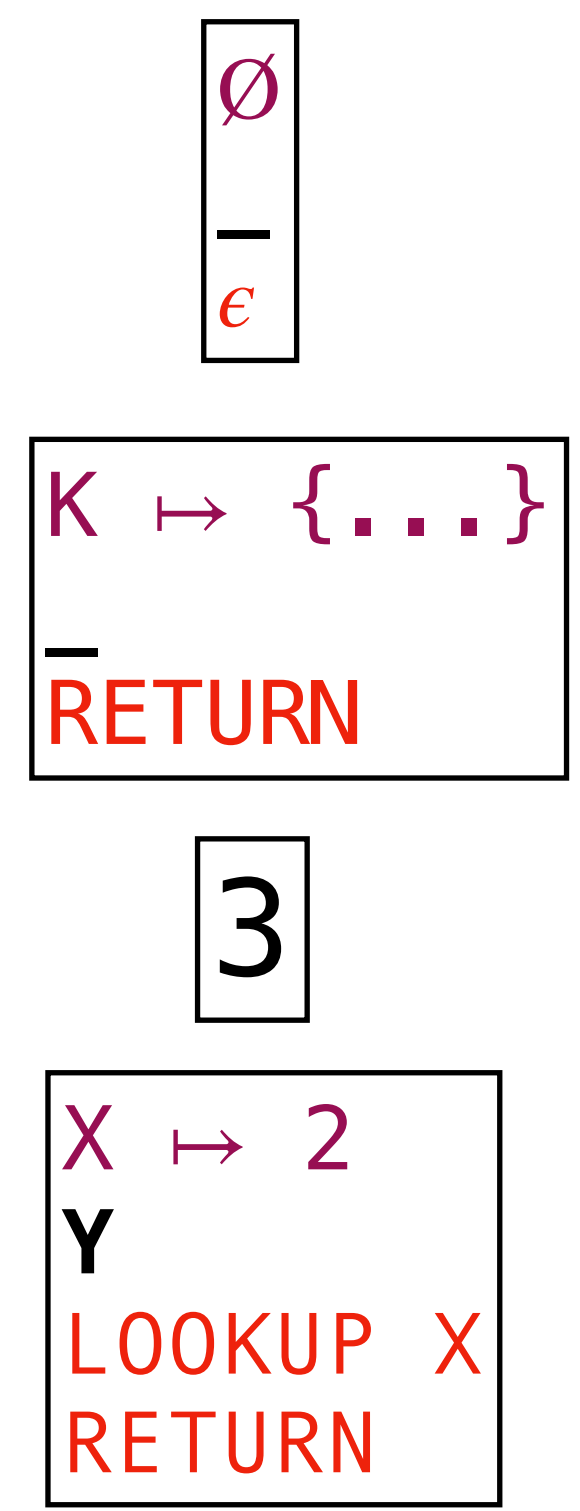
Env:



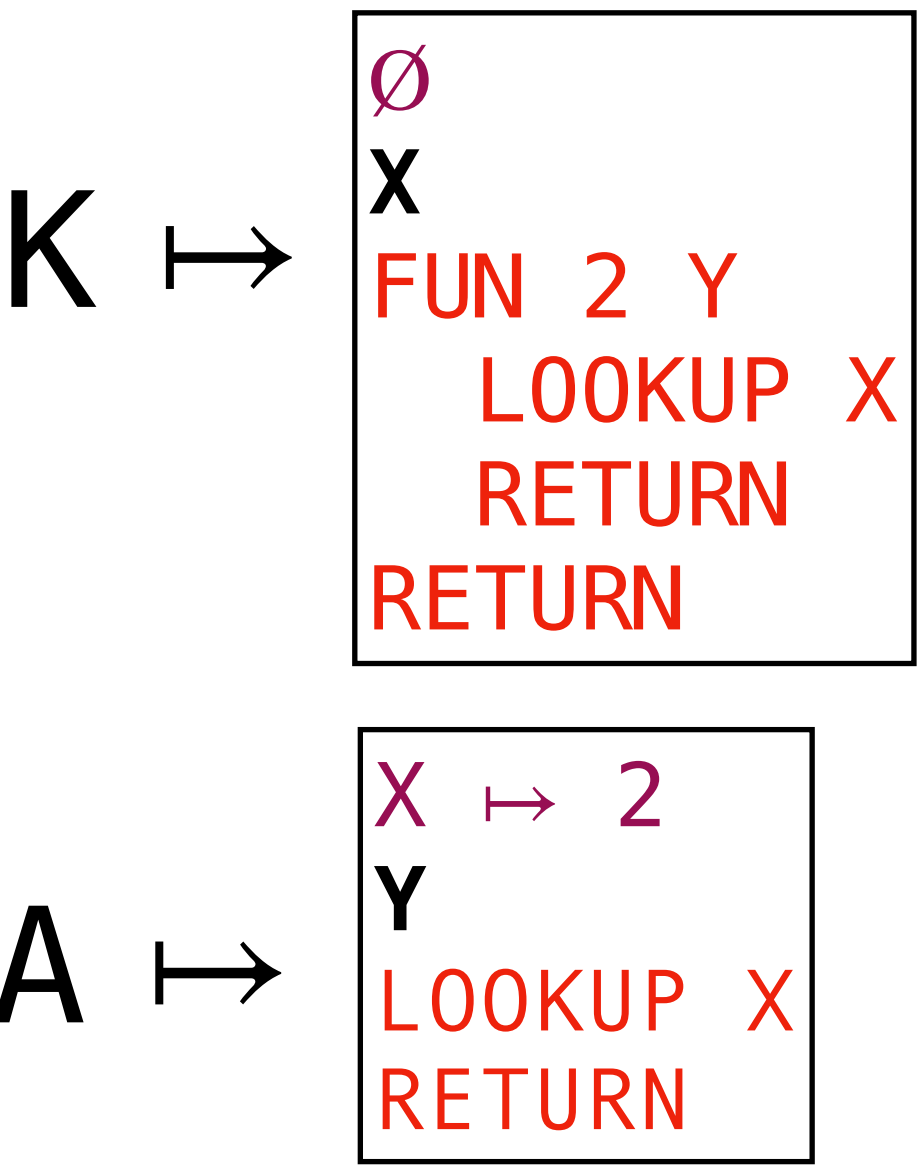
LOOKUP A
CALL
RETURN

Example

Stack:



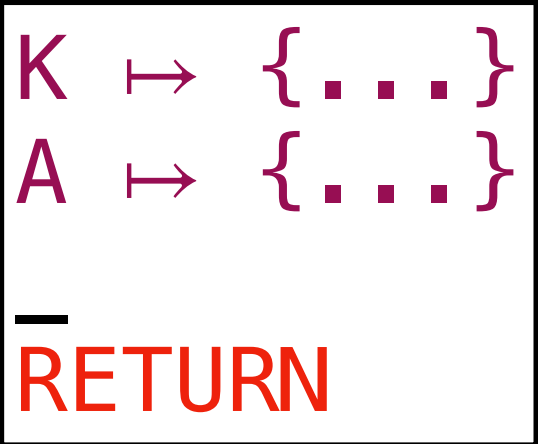
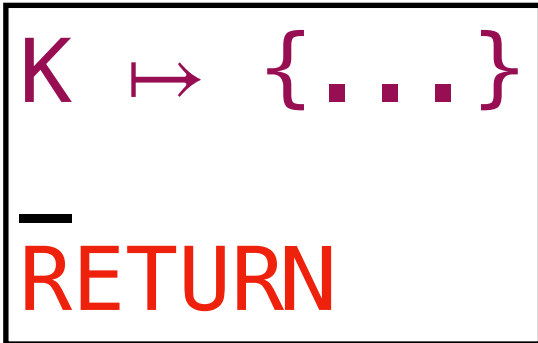
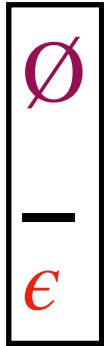
Env:



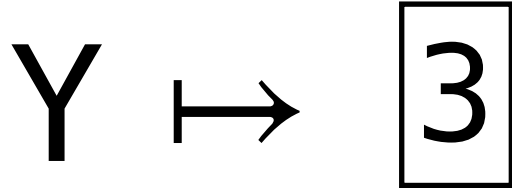
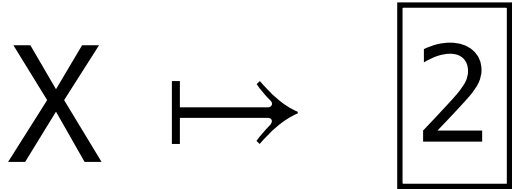
CALL
RETURN

Example

Stack:



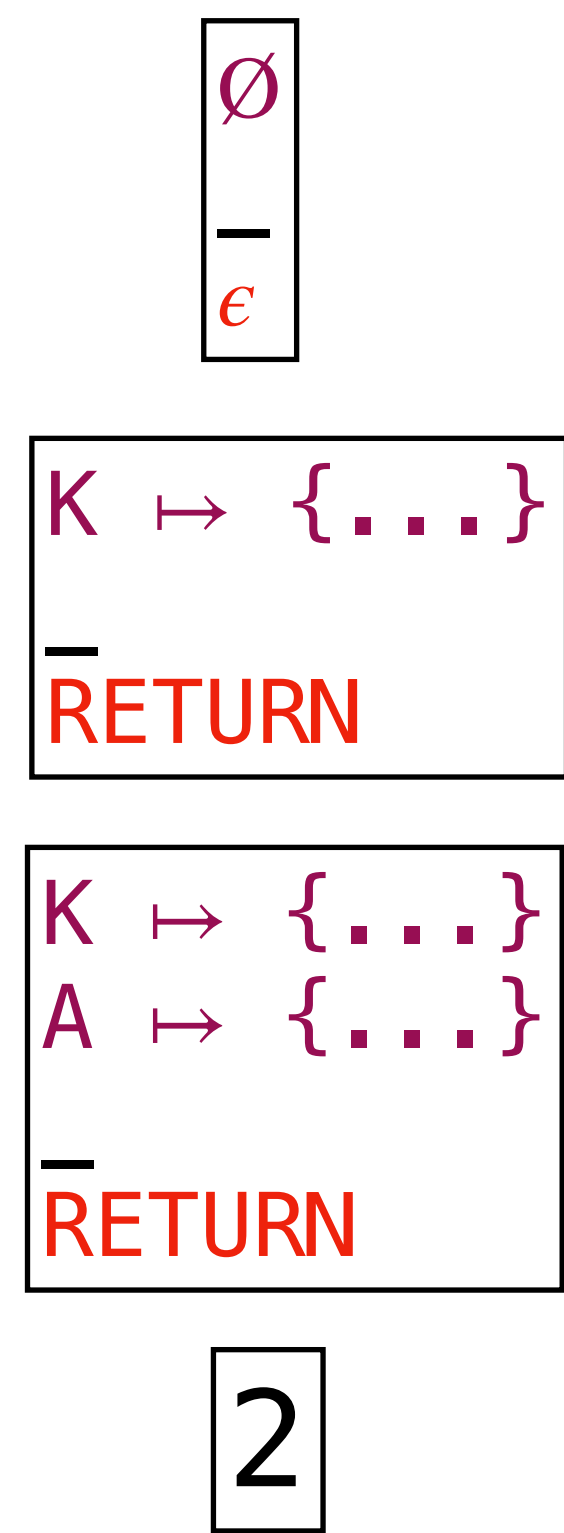
Env:



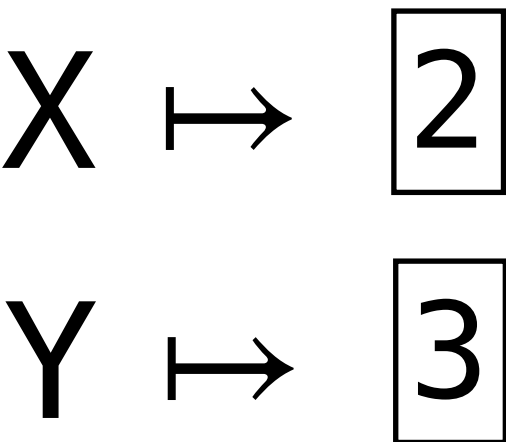
LOOKUP X
RETURN

Example

Stack:



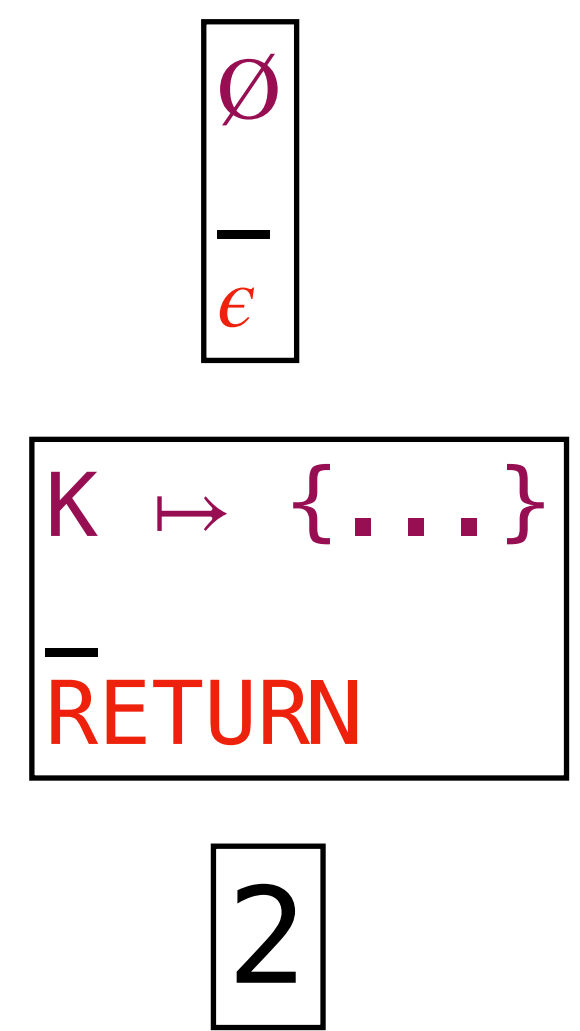
Env:



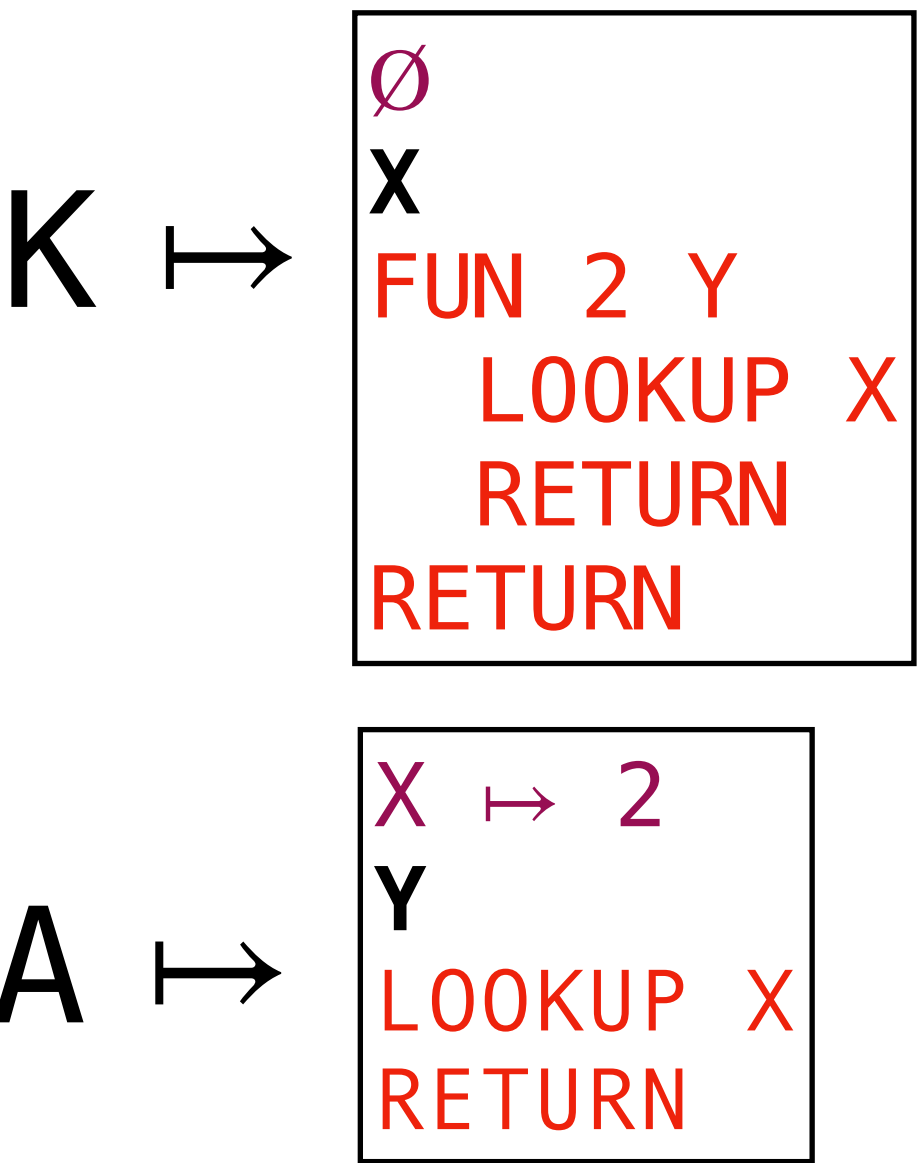
RETURN

Example

Stack:



Env:



RETURN

Example

Stack:

∅
—
ϵ

2

Env:

K \mapsto

∅
x
FUN 2 Y
 LOOKUP X
 RETURN
RETURN

RETURN

Example

Stack:

2

Env:

ϵ

demo
(show-and-tell)

What's next?

More OCaml:

- » Modules, functional data structures, mutability
- » GADTS, effects, parallelism
- » applications in ML, linear algebra, scientific computing

More PL:

- » Come learn Rust (and linear types) with me next semester!
- » Learn Haskell, Elm, Scala

More Math/Type Theory:

- » Go learn about session types with Professor Das next semester!
- » Category theory (functors, monads, comonads), Logic, Type theory

More Computers:

- » Compilers, Linkers, LLVM
- » Formal verification
- » embedded systems programming, tensor program compilation

fin

(thanks everyone)