The test lasts 3 hours. Handwritten or printed course notes are the only documents allowed. The
questions are independent, in the sense that it is not necessary to have answered the previous questions in
order to deal with a question. On the other hand, questions can call on definitions or results introduced in
previous questions. Unless explicitly stated otherwise, all answers must be justified.
Feel free to answer in French or English.
Figures 2–3 are grouped together at the end of the subject on page 15. Suggestion: detach the last sheet.

Throughout this subject, we consider a small fragment of the Python language (different from
that of the project), whose abstract syntax is given in figure 2. This fragment includes the value
None, Booleans (False and True), integers and “lists”. Expressions are limited to constants (c),
variables (x), primitive operations (op), function calls (f) and a conditional expression. The latter,
e1 if e2 else e3, evaluates e2 then returns the value of e1 if e2 is true and the value of e3 otherwise.
A program is a sequence of function definitions (d), followed by the evaluation of a single expression
whose value is displayed with print. The body of a function is a sequence of assignments x = e,
followed by a return instruction. Each function can refer to previously defined functions or to
the function currently being defined (recursive function). Here’s an example of a program in this
fragment, written in concrete Python syntax, which calculates and displays a list of five integers:

```python
def aux(s):
    a = s[0]
    b = s[1]
    return [a+b] + s
def myst(n, s):
    return s if n==0 else myst(n-1, aux(s))
print(myst(3, [1]+[0]))
```

A big-step operational semantic is shown in figure 3. It expresses call-by-value. A value is denoted v.
The value of variables is given by an environment V, i.e. a function that maps variable names to
values. The semantics includes the definition of primitive operations (in the table at the bottom of
the figure). For each operation op, its semantics is given by a function [ [op] ] defined on values. This
function can be partial, i.e. not defined over all values. For example, the expression 1+None has no
value. The num operation is used to interpret a Boolean as an integer in certain operations.

**Question 1** What is the list calculated and displayed by the above program? (No justification is
required.)

**Correction**: This program computes the Fibonacci numbers and outputs the following
list:

```
[3, 2, 1, 1, 0]
```
Question 2  Implement a function `rev` that, when applied to a list \([v_0, v_1, \ldots, v_{n-1}]\), returns the list \([v_{n-1}, \ldots, v_1, v_0]\), with \(n \geq 0\). This function must be able to work with \(n = 0\), even though it is not possible to construct an empty list in our fragment. You can introduce an auxiliary function.

Correction: We use an auxiliary function, with an index \(i\) and an accumulator \(acc\). We handle the case \(n = 0\) separately, as we can’t build the empty list in this fragment.

```python
def revaux(s, i, acc):
    return acc if i==len(s) else revaux(s, i+1, [s[i]] + acc)
def rev(s):
    return s if len(s)==0 else revaux(s, 1, [s[0]])
```
Question 3 For each of the following expressions, give the derivation of its evaluation in an empty environment, when it exists, or justify that there is none. (Here we take the context of the two functions aux and myst given as examples at the beginning of the subject.)

- myst(0, [1]+[0])
- aux([0])
- myst(-1, [1]+[0])

Correction:

- myst(0, [1]+[0])
  Let bmyst be the body of function myst.

<table>
<thead>
<tr>
<th>1---&gt;1</th>
<th>0---&gt;0</th>
</tr>
</thead>
</table>

| [1]--->[1] | [0]--->[0] | V,n==0--->[True] | V,s-->[1,0] | 0--->[0] | [1]+[0]-->[1,0] | V={n:0,s:[1,0]},bmyst-->[1,0] |

myst(0,[1]+[0]) --> [1,0]

- aux([0])
  We have [0] -> [0]. Let V := \{s \mapsto [0]\}. We have V,s[0] -> 0. Then we seek to evaluate [1] in V' = \{s \mapsto [0]; a \mapsto 0\}. But [get] is only defined for an access within bounds, which is not the case here. So s[1] has no value in V' and thus the whole expression has no value.

- myst(-1, [1]+[0]) has no value since evaluation does not terminate, which big-step semantics does not capture. More precisely, let us show that if e_1 -> n and e_2 -> v, and if myst(e_1, e_2) -> v', then n ≥ 0, by induction on the derivation of myst.
  - if n = 0, this is immediate;
  - otherwise, the derivation is as follows

| e1--->n | e2--->v | V={n->n,s->v},bmyst --> v' |

myst(e_1, e_2) --> v'

and by induction hypothesis applied to the sub-derivation if myst(n-1, aux(s)), we have n - 1 ≥ 0 and thus n > 0.
**Question 4**  We propose to program an interpreter of our language, in Java or in OCaml (your choice), following the big-step semantics of figure 3. Specify the Java/OCaml types and data structures used. Give the type of each Java/OCaml function involved in the interpreter. *You are not asked to write the code for these functions.* Recall that in Java (resp. OCaml) we obtain dictionaries whose keys are strings with `HashMap<String, ...>` (resp. module `StrMap = Map.Make(String)`).

**Correction :** We can setup the following types for the abstract syntax

```ocaml
type expr =
  | Enone
  | Ebool of bool
  | Eint of int
  | Evar of string
  | Eapp of string * expr list
  | Eite of expr * expr * expr

type stmt = string * expr

type def = string * string list * stmt list * expr

type program = def list * expr
```

and the following type for the values:

```ocaml
type value =
  | Vnone
  | Vbool of bool
  | Vint of int
  | Vlist of value array
```

Functions can be stored in a global hash table:

```ocaml
let funs : (string, string list * stmt list * expr) H.t = ...
```

The environment can be implemented with an immutable dictionary:

```ocaml
module M = Map.Make(String)

type env = value M.t
```

The interpreter can be implemented with the following three functions:

```ocaml
val num: value -> int
val expr: env -> expr -> value
val program: program -> unit (* évalue et imprime la valeur *)
```
\[
\begin{align*}
\Delta, \Gamma \vdash \text{None} : \text{none} & \quad \Delta, \Gamma \vdash b : \text{bool} & \quad \Delta, \Gamma \vdash n : \text{int} & \quad \Delta, \Gamma \vdash x : \Gamma(x) \\
\Delta, \Gamma \vdash e_1 : \tau_1 & \quad \Delta, \Gamma \vdash e_2 : \tau_2 & \quad \Delta, \Gamma \vdash e_3 : \tau_3 & \quad \Delta, \Gamma \vdash e_1 \text{ if } e_2 \text{ else } e_3 : \tau_1 \cup \tau_3 \\
\forall 0 \leq i < n, \Delta, \Gamma \vdash e_i : \tau_i & \quad \text{with } \Delta(f)(\tau_0, \ldots, \tau_{n-1}) \overset{\text{def}}{=} \bigcup_{f : \tau'_0 \times \cdots \times \tau'_{n-1} \rightarrow \tau \in \Delta} \tau
\end{align*}
\]

Environment $\Delta_{op}$:

<table>
<thead>
<tr>
<th>operator</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>add</td>
<td>${\text{list}} \times {\text{list}} \rightarrow {\text{list}}$</td>
</tr>
<tr>
<td>add</td>
<td>${\text{bool}, \text{int}} \times {\text{bool}, \text{int}} \rightarrow {\text{int}}$</td>
</tr>
<tr>
<td>sub</td>
<td>${\text{bool}, \text{int}} \times {\text{bool}, \text{int}} \rightarrow {\text{int}}$</td>
</tr>
<tr>
<td>len</td>
<td>${\text{list}} \rightarrow {\text{int}}$</td>
</tr>
<tr>
<td>mk</td>
<td>${\text{none}, \text{bool}, \text{int}, \text{list}} \rightarrow {\text{list}}$</td>
</tr>
<tr>
<td>get</td>
<td>${\text{list}} \times {\text{bool}, \text{int}} \rightarrow {\text{none}, \text{bool}, \text{int}, \text{list}}$</td>
</tr>
<tr>
<td>eq</td>
<td>${\text{none}, \text{bool}, \text{int}, \text{list}} \times {\text{none}, \text{bool}, \text{int}, \text{list}} \rightarrow {\text{bool}}$</td>
</tr>
</tbody>
</table>

Figure 1: Static typing of expressions.

**Static Typing.** Although Python is a dynamically-typed language, we propose here to perform a little static typing on our language, with the dual aim of rejecting inconsistent programs and executing certain programs more efficiently. We give ourselves four kinds none, bool, int, and list, to represent respectively a None value, a Boolean value, an integer value or a list. A type $\tau$ is then a set of kinds, i.e.

$$\tau \subseteq \{\text{none}, \text{bool}, \text{int}, \text{list}\},$$

with the following interpretation: if an expression of type $\tau$ evaluates to a value, then this value will necessarily be of one of the kinds of $\tau$. In particular, a type can be the empty set $\emptyset$ (the expression cannot have a value) or the set $\{\text{none}, \text{bool}, \text{int}, \text{list}\}$ (the value can be any).

To type an expression, we give ourselves a context made up of two environments: an environment $\Gamma$ giving the type of the variables (a function from variables to types) and an environment $\Delta$ giving function types. (Primitive operations are seen as functions for typing.) The type of a function, noted $\sigma$, is of the form

$$\tau_0 \times \cdots \times \tau_{n-1} \rightarrow \tau$$

where $n$ is the number of function parameters. The environment $\Delta$ is a set of pairs $(f, \sigma)$ where $f$ is the name of a function and $\sigma$ is a function type. For a single function, there may be several different types in the $\Delta$ environment. The typing judgment for an expression is noted $\Delta, \Gamma \vdash e : \tau$. Figure 1 gives typing rules for expressions, as well as an environment $\Delta_{op}$ giving the types of primitive operations. Note that the add operation has two different types in $\Delta_{op}$, which is consistent with its two interpretations in figure 3.

To type an application $f(e_0, \ldots, e_{n-1})$, we make the union of all the types that can be obtained with the types of $f$ given by $\Delta$ and compatible with the types $\tau_i$ of the actual parameters $e_i$. In particular, the result may be the empty type $\emptyset$ if $\Delta$ contains no compatible function type.
Question 5  In an empty environment $\Gamma$ and an environment $\Delta$ containing only primitive operations, give

1. an expression of type $\emptyset$;
2. an expression of type $\{\text{bool}, \text{int}, \text{list}\}$.

Correction :
1. $1+\text{None}$
2. $\text{True}$ if $\text{True}$ else $1$ if $\text{True}$ else $[1]$

Question 6  For this question, the following environments are used:

\[
\Delta \overset{\text{def}}{=} \Delta_{op} \cup \{(f, \{\text{int}\} \times \{\text{list}\} \to \{\text{list}\}), (f, \{\text{int}\} \times \{\text{none}\} \to \{\text{none}\})\}
\]
\[
\Gamma \overset{\text{def}}{=} \{x \mapsto \{\text{list}\}\}
\]

For each of the following expressions, give its typing derivation in $\Delta, \Gamma$.

1. $f(1, x)$
2. $f(1, x[0])$  
3. $f(\text{None}, \text{None})$

Correction :
1. \[
\frac{1:\{\text{int}\} \quad x:\{\text{list}\}}{f(1, x):\{\text{list}\}}
\]
2. \[
\frac{1:\{\text{int}\}}{x[0]:\{\text{none, bool, int, list}\}}
\]
\[
\frac{f(1, x[0]):\{\text{none, list}\}}{}
\]
3. \[
\frac{\text{None}:\{\text{none}\} \quad \text{None}:\{\text{none}\}}{f(\text{None}, \text{None}):\emptyset}
\]
Question 7  What are the conditions over $\Delta, \Gamma$ and $e$ for the existence of a type $\tau$ such that $\Delta, \Gamma \vdash e : \tau$?

Correction : Let us show by induction on $e$ that $e$ has a type in $\Delta, \Gamma$ if and only if all the variables in $e$ are defined in $\Gamma$.

- if $e$ is a constant, this is immediate;
- if $e$ is a variable, the condition is precisely $x \in \text{dom}(\Gamma)$;
- if $e = e_1 \text{ if } e_2 \text{ else } e_3$, then by IH the three sub-expressions have a type iff their variables are defined in $\Gamma$, and thus $e$ has a type iff all its variables are defined in $\Gamma$;
- if $e = f(e_1, \ldots, e_n)$, then by IH all the sub-expressions $e_i$ have a type iff their variables are defined in $\Gamma$, and thus $e$ has a type iff all its variables are defined in $\Gamma$ (since the union is always defined, even if $f$ is not in $\Delta$ or does not have any compatible type).

Question 8  Show that, for an environment $\Delta, \Gamma$ and an expression $e$, there exists at most one type $\tau$ such that $\Delta, \Gamma \vdash e : \tau$.

Correction : Let us assume $\Delta, \Gamma \vdash e : \tau$ and $\Delta, \Gamma \vdash e : \tau'$, and let us show $\tau = \tau'$ by structural induction on the typing derivation.

- if $e$ is a constant, this is immediate;
- if $e$ is a variable, we have $\tau = \tau' = \Gamma(x)$;
- if $e = f(e_1, \ldots, e_n)$, then by IH the three types for the three sub-expressions are the same, hence the result;
- if $e = f(e_1, \ldots, e_n)$, then by IH all the sub-expressions $e_i$ have the same types in the two derivations, and the union gives the same type for $e$.

Question 9  With regard to typing, is there a difference between the following two environments?

$$
\Delta = \{(f, \{\text{bool}\} \times \{\text{bool}\} \to \{\text{int}\}) ; (f, \{\text{bool}\} \times \{\text{int}\} \to \{\text{int}\})\}
$$

and

$$
\Delta' = \{(f, \{\text{bool}\} \times \{\text{bool}\} \to \{\text{int}\})\}
$$

Same question with the following two environments:

$$
\Delta = \{(f, \{\text{bool}\} \times \{\text{bool}\} \to \{\text{int}\}) ; (f, \{\text{bool}\} \times \{\text{int}\} \to \{\text{list}\})\}
$$

and

$$
\Delta' = \{(f, \{\text{bool}\} \times \{\text{bool}\} \to \{\text{int, list}\})\}
$$

Correction :
1. In the first case, there is no difference. Indeed, if the types of the two arguments of $f$ intersect $\{\text{bool}\}$ and, either $\{\text{bool}\}$, or $\{\text{int}\}$, then we have $\{\text{int}\}$ for the application, and otherwise $\emptyset$, in both cases.

2. In the second case, however, there is a difference, since we could get less precise types. Example: $f(\text{False}, \text{False})$ is now given the type $\text{list}$.

**Typing a function.** To type a function definition, we introduce the judgment $\Delta \vdash f : \sigma$ which means “in environment $\Delta$, the definition of function $f$ admits the function type $\sigma$”. We propose the following rule for this judgment:

$$
\begin{align*}
\text{let } f(x_0, \ldots, x_{n-1}) & \overset{\text{def}}{=} x_n = e_n; \ldots; x_{m-1} = e_{m-1}; \text{return } e \\
\Gamma_n & \overset{\text{def}}{=} \{x_0 \mapsto \tau_0; \ldots; x_{n-1} \mapsto \tau_{n-1}\} \\
& \quad \forall n \leq i < m, \Delta, \Gamma_i \vdash e_i : \tau_i \quad \Gamma_{i+1} \overset{\text{def}}{=} \Gamma_i[x_i \mapsto \tau_i] \\
\Delta & \vdash f : \tau_0 \times \cdots \times \tau_{n-1} \rightarrow \tau
\end{align*}
$$

(Note its similarity to the semantics rule.) In particular, this rule allows a recursive definition to be typed, by showing $\Delta \vdash f : \sigma$ for an environment $\Delta$ containing one or more types for $f$.

**Question 10** For the $\text{aux}$ function at the beginning of the subject, show that we have

$$
\Delta_{op} \vdash \text{aux} : \{\text{list}\} \rightarrow \{\text{list}\}.
$$

Give the complete typing derivation.

**Correction :**

we set $G := \{s:\text{list}\}$ and any=$\{$none, bool, int, list$\}$

$$
\begin{align*}
a & : \text{any} \quad b & : \text{any} \\
\hline
s & : \text{list} \quad 0 & : \text{int} \quad s & : \text{list} \\
\hline
\hline
\hline
\text{G}|-s[0] & : \text{any} \quad \text{G}+\{a:\text{any}\}-s[1] & : \text{any} \quad \text{G}+\{a,b:\text{any}\}|-[-a+b]+s & : \{\text{list}\}
\end{align*}
$$

$$
\begin{align*}
\hline
\hline
\hline
\hline
\rightarrow \text{ aux } : \{\text{list}\} \rightarrow \{\text{list}\}
\end{align*}
$$

**Question 11** Propose at least two different types $\sigma$ such that, for each, we have

$$
\Delta_{op} \cup \{(\text{aux} : \{\text{list}\} \rightarrow \{\text{list}\}); (\text{myst}, \sigma)\} \vdash \text{myst} : \sigma
$$

for the $\text{myst}$ function at the beginning of the subject. (No justification is requested, i.e., we don’t ask for the typing derivations, but only for the two types $\sigma$.)
Correction:

\[
\text{myst} : \{\text{int}\} \times \{\text{list}\} \rightarrow \{\text{list}\} \\
\text{myst} : \{\text{bool}, \text{int}\} \times \{\text{list}\} \rightarrow \{\text{list}\}
\]

---

**Question 12** Propose a type for the function

```python
def loop(x):
    return loop(x)
```

that is as informative as possible.

Correction:

\[
\text{loop} : \{\text{none}, \text{bool}, \text{int}, \text{list}\} \rightarrow \emptyset
\]

---

**Question 13** We would like to show the type safety property in the following sense: if \(V, e \rightarrow v\) and \(\Delta, \Gamma \vdash e : \tau\), then \(T(v) \in \tau\) where function \(T\) gives the type of a semantic value and is unsurprisingly defined as

\[
\begin{align*}
T(\text{None}) &= \text{none} \\
T(\text{b}) &= \text{bool} \\
T(\text{n}) &= \text{int} \\
T([v_0, \ldots, v_{n-1}]) &= \text{list}
\end{align*}
\]

Give necessary conditions on \(V\) and \(\Delta, \Gamma\) for type safety to be possible. *But we don’t ask to show type safety.*

Correction: A first obvious condition is consistency between \(\Gamma\) and \(V\):

for all \(x \in \text{dom}(V)\), we have \(T(V(x)) \in \Gamma(x)\)

Another condition is consistency between \(\Delta\) and operations:

if \((\text{op}, \tau_0 \times \ldots \times \tau_{n-1} \rightarrow \tau) \in \Delta\),
then for all \(v_0, \ldots, v_{n-1}\) such that \(T(v_i) \in \tau_i\),
we have \(T([\text{op}](v_0, \ldots, v_{n-1})) \in \tau\)

We can check that it is indeed the case with the contents of figures 3 and 1. Last, we need every function to be well-typed wrt \(\Delta\), that is, for all \((f, \sigma) \in \Delta\), we have \(\Delta \vdash f : \sigma\).
**Question 14**  Our typing rule for function is not algorithmic: it only allows us to check the definition of a function \( f \) with respect to \( \Delta \), not to find a type for \( f \). Propose an algorithm to infer a set of types for a function whose definition we have.

**Correction:** Although not very efficient, one solution is to successively give the parameters all possible combinations of singleton types, then type the function in this environment. If you obtain a type \( \emptyset \) during typing, you fail, i.e. you eliminate this case. We then give the function all the types we’ve obtained.

When two types differ only in the type of the result, or in the type of a single parameter, we can combine them into a single type with a union (and then start again). This improves type readability and, above all, typing efficiency, since less work is required to type an application.

There is, however, one difficulty in typing a recursive function \( f \). We start by giving it a type \( \text{all} \times \cdots \times \text{all} \rightarrow \emptyset \) to perform a first iteration (during which we do not fail on a type \( \emptyset \) if this is due to a recursive call, directly or indirectly). Once you’ve a set of types has been found for \( f \), we start again with these types. The process converges.

---

**Parsing.** We wish to perform a syntactic analysis of our small language. A grammar for a subset of the expressions is as follows:

\[
E ::= \text{id} \\
| E \text{ if } E \text{ else } E \\
| E [ E ]
\]

**Question 15** Show that this grammar is ambiguous.

**Correction:** We have two derivation trees for the expression \( \text{id if id else id[id]} \), namely

\[
\begin{array}{c}
E \\
| \text{----------------} \\
E \text{ [ E ]} \\
| \text{----------------} \\
E \text{ if E else E} \\
| \text{id} \\
| \text{id id} \\
\end{array}
\quad
\begin{array}{c}
E \text{ if E else E} \\
| \text{id} \\
| \text{id id} \\
\end{array}
\]

**Question 16** This grammar is fed to a tool such as CUP or Menhir, as follows:

\[
\text{expr:} \\
| \text{IDENT } \{ \ldots \} \\
| \text{expr } \text{IF } \text{expr } \text{ELSE } \text{expr } \{ \ldots \} \\
| \text{expr } \text{LBR } \text{expr } \text{RBR } \{ \ldots \}
\]
Correction : These are shift/reduce conflicts:

- the conflict \( E \text{ if } E \text{ else } E \ . \ [E] \) (reduce if-else or shift of \([\) )
- the conflict \( E \text{ if } E \text{ else } E . \text{ if } E \text{ else } E \) (reduce if-else or shift of if)

We can declare priorities as follows

\%
\nonassoc ELSE
\%
\nonassoc IF
\%
\nonassoc LBR

that is, strongest priority for the shifting of \([\), then for shifting of \(IF\) last for the reduction of \(IF-ELSE\) (the priority is that of the rightmost token in the rule to be reduced, that is ELSE here).

**Question 17** Another grammar is proposed for this language:

\[
E ::= A \\
    | A \text{ if } E \text{ else } E \\
A ::= \text{ident} \\
    | A \ [ E ]
\]

Show that this grammar is SLR(1). (Reminder: SLR(1) means that there is no conflict anymore in the deterministic LR(0) automaton when reduction of non-terminal symbol \(X\) is only performed for lookahead symbols in \(\text{FOLLOW}(X)\).)

Correction : We have \(\text{NULL}(A) = \text{NULL}(E) = \text{false}\). We have \(\text{FIRST}(A) = \text{FIRST}(E) = \{\text{ident}\}\). We have \(\text{FOLLOW}(A) = \{\#, [,], \text{else}, \text{if}\}\) and \(\text{FOLLOW}(E) = \{\#, [\}\). The SLR(1) automaton has ten states:

state 0:  
\[ \text{S \to} . \ E \ # \]  
\[ \text{E \to} . \ A \]  
\[ \text{E \to} . \ A \text{ if } E \text{ else } E \]  
\[ \text{A \to} . \ \text{ident} \]  
\[ \text{A \to} . \ A \ [ E ] \]

state 1:  
\[ \text{A \to} \text{ident} . \]

state 2:  
\[ \text{S \to} E . \ # \]

state 3: // no conflict here, since if and [ are not in \(\text{Follow}(E)\)  
\[ \text{E \to} A . \]
E → A . if E else E
A → A . [ E ]

state 4:
E → . A
E → . A if E else E
A → . ident
A → . A [ E ]
A → A [ . E ]

state 4:
A → A [ E . ]

state 6:

state 7:
A → . ident
A → . A [ E ]
E → . A
E → . A if E else E
E → A if . E else E

state 8:
E → A if E . else E

state 9:
A → . ident
A → . A [ E ]
E → . A
E → . A if E else E
E → A if E else . E

state 10:
E → A if E else E .

and no state contains a conflict.

---

**Compiling to x86-64 assembly.** We propose to compile our little language to x86-64 assembler. (A cheat sheet is given in the appendix.) We assume that the integers in our language are limited to signed 64-bit integers. We adopt the same representation as in the project, where any value is the address of a memory block allocated on the heap, whose size is a multiple of 64 bits, with the following form:

| None | 0 0 |
| Boolean False | 1 0 |
| Boolean True | 1 1 |
| integer $n$ | 2 $n$ |
| list | 3 $n$ $v_0$ $v_1$ ... $v_{n-1}$ |

The first word is an integer indicating the kind of value.
**Question 18**  Discuss this choice of representation. In particular, could `None` be represented by the null pointer? If so, would there be any interest in doing so? And could we represent an integer directly by its value?

**Correction:** The choice is guided by the fact that, even if we did a little bit of static typing, we do not always know statically the type of a value. In particular, some Python functions are “polymorphic” in the sense that they have to accommodate values of different types. Thus, a first benefit of this representation is uniformity (one value = one address = one word).

Second, functions have to test types dynamically, e.g. to perform a + operation or to print a value (making a distinction

We could indeed represent `None` with the null pointer. Yet, the proposed representation makes no difference, as we can allocate it statically (and only once).

On the contrary, we could not represent an integer by its value, for we could mistake it for a pointer (e.g. to a list).

**Question 19**  Explain how the static typing described in the previous questions can be used locally to produce more efficient assembly code. Give examples.

**Correction:** The information provided by static typing allows us to avoid, in some cases, checking for the value type at runtime. For instance, if we compile the expression `e1+e2` and we know that both `e1` and `e2` have type `{bool, int}`, then we can directly call the function that performs the addition of two integers, instead of first checking whether we have to perform an addition or a concatenation.

Example: the compilation of function

```python
def opt(x, y):
    return 1+x+y
```

whose type is `{bool, int} × {bool, int} → {int}`.

We can go further and not allocate intermediate values on the heap, storing them directly in registers or on the stack (unboxing). In the example above, we do not have to allocate the intermediate value `1+x` on the heap, but only the final value. We can do so because the sub-expression `1+x` has type `{bool, int}`.

**Compiling a small function.** We consider the following Python function

```python
def mult(x, y, z):
    return z if x==0 else mult(dec(x), y, add(y, z))
```

where `dec` (not shown) implements the function `n → n − 1` for an integer `n` and `add` (not shown) implements the addition of two integers. We assume that function `mult` is given the type

```
{int} × {int} × {int} → {int}
```

that is, we only use it with integer parameters and we get a result that is an integer. We intend to compile function `mult` to some optimized x86-64 assembly code.
**Question 20**  Give some RTL code for function `mult`.

**Correction :**

```plaintext
#4 mult(#1, #2, #3)
1: load 8(#1) #5 -> 2
2: jz #5 -> 3, 4
3: mov #3 #4 -> exit
4: #6 <- call dec(#1) -> 5
5: #7 <- call add(#2, #3) -> 6
6 #4 <- call mult(#6, #2, #7) -> exit
```

**Question 21**  Give some optimized assembly code for function `mult`. In particular, if a tail call can be optimized, do it. You are free to proceed the way you want, i.e., you do not have to follow the optimized code generation scheme described in the course. We assume that functions `dec` and `add` are available, and that they follow the x86-64 calling conventions (but you don’t have to align the stack before calling them).

**Correction :**

```assembly
multadd:pushq %r12 # we use 3 callee-save registers for x,y,z
    movq %rdi, %r12
    pushq %r13
    movq %rsi, %r13
    pushq %r14
    movq %rdx, %r14
1: testq %r12, %r12 # TCO => loop
    jz 2f
    movq %r12, %rdi
    call dec
    movq %rax, %r12
    movq %r13, %rdi
    movq %r14, %rsi
    call add
    movq %rax, %r14
    jmp 1b
2: movq %r14, %rax # return z and restore the callee-save
    popq %r14
    popq %r13
    popq %r12
    ret
```
\[ e ::= c \quad \text{constant} \quad c ::= \text{None} \]
| \( x \) | variable | \( b \in \{\text{True}, \text{False}\} \)
| \( \text{op}(e, \ldots, e) \) | operation | \( n \in \mathbb{Z} \)
| \( f(e, \ldots, e) \) | application |
| \( e \text{ if } e \text{ else } e \) | conditional |

\[
\begin{align*}
\text{len} & : \text{def} \quad \text{len}(\text{expr}) = \text{len}(\text{expr}) + \text{len}(\text{expr}) \\
\text{add} & : \text{def} \quad \text{add}((\text{expr}, \ldots, \text{expr}), \text{expr}) = \text{add}(\text{expr}, \text{expr}) + \text{add}(\text{expr}, \text{expr}) \\
\text{sub} & : \text{def} \quad \text{sub}(\text{expr}, \text{expr}) = \text{sub}(\text{expr}, \text{expr}) - \text{sub}(\text{expr}, \text{expr}) \\
\text{print} & : \text{def} \quad \text{print}((\text{expr}, \ldots, \text{expr})) = \text{print}(\text{expr}, \text{expr})
\end{align*}
\]

\[
\begin{align*}
\forall 0 \leq i < n, \quad V, e_i & \rightarrow v_i \\
f(x_0, \ldots, x_{n-1}) & \text{def} x_n = e_n; \ldots; x_{m-1} = e_{m-1}; \text{return } e \\
V_n & \text{def} \{x_0 \mapsto v_0; \ldots; x_{m-1} \mapsto v_{n-1}\} \\
\forall n \leq i < m, \quad V, e_i & \rightarrow v_i \\
V_{i+1} & \text{def} V_i[x_i \mapsto v_i] \\
V_m, e & \rightarrow v
\end{align*}
\]

\[
\begin{align*}
\text{add}((\text{expr}, \ldots, \text{expr}), \text{expr}) & = \text{add}(\text{expr}, \text{expr}) + \text{add}(\text{expr}, \text{expr}) \\
\text{sub}(\text{expr}, \text{expr}) & = \text{sub}(\text{expr}, \text{expr}) - \text{sub}(\text{expr}, \text{expr}) \\
\text{print}((\text{expr}, \ldots, \text{expr})) & = \text{print}(\text{expr}, \text{expr})
\end{align*}
\]

\[
\begin{align*}
\forall 0 \leq i < n, \quad V, e_i & \rightarrow v_i \\
f(x_0, \ldots, x_{n-1}) & \text{def} x_n = e_n; \ldots; x_{m-1} = e_{m-1}; \text{return } e \\
V_n & \text{def} \{x_0 \mapsto v_0; \ldots; x_{m-1} \mapsto v_{n-1}\} \\
\forall n \leq i < m, \quad V, e_i & \rightarrow v_i \\
V_{i+1} & \text{def} V_i[x_i \mapsto v_i] \\
V_m, e & \rightarrow v
\end{align*}
\]

---

Figure 2: Abstract Syntax.

---

\[
\begin{align*}
\forall 0 \leq i < n, \quad V, e_i & \rightarrow v_i \\
f(x_0, \ldots, x_{n-1}) & \text{def} x_n = e_n; \ldots; x_{m-1} = e_{m-1}; \text{return } e \\
V_n & \text{def} \{x_0 \mapsto v_0; \ldots; x_{m-1} \mapsto v_{n-1}\} \\
\forall n \leq i < m, \quad V, e_i & \rightarrow v_i \\
V_{i+1} & \text{def} V_i[x_i \mapsto v_i] \\
V_m, e & \rightarrow v
\end{align*}
\]

---

Figure 3: Big-step Operational Semantics.
Appendix: x86-64 cheat sheet

A fragment of the x86-64 instruction set is given here. You are free to use any other part of the x86-64 assembler. In the following, \( r_i \) designates a register, \( n \) an integer constant and \( L \) a label.

- `mov r_2, r_1` copies register \( r_2 \) into register \( r_1 \)
- `mov $n, r_1` loads constant \( n \) into register \( r_1 \)
- `mov $L, r_1` loads the address of label \( L \) into register \( r_1 \)
- `sub r_2, r_1` computes \( r_1 - r_2 \) and stores it into \( r_1 \)
- `mov n(r_2), r_1` loads \( r_1 \) with the value contained in memory at address \( r_2 + n \)
- `mov r_1, n(r_2)` writes in memory at address \( r_2 + n \) the value of \( r_1 \)
- `push r_1` pushes the value of \( r_1 \) on the stack
- `pop r_1` pops a value from the stack and stores it into register \( r_1 \)
- `test r_2, r_1` sets the flags according to the value of \( r_1 \land r_2 \)
- `jz L` jumps to address \( L \) if flags signal a zero value
- `jmp L` jumps to address \( L \)
- `call L` pushes the return address to the stack and jumps to address \( L \)
- `ret` pops an address from the stack and jumps there

Calling conventions:

- up to six arguments are passed via registers \( %rdi, %rsi, %rdx, %rcx, %r8, %r9 \);
- other arguments are passed on the stack, if any;
- the returned value is put in \( %rax \);
- registers \( %rbx, %rbp, %r12, %r13, %r14, %r15 \) are callee-saved: they won’t be clobbered by a call;
- the other registers are caller-saved: they may be clobbered by a call;
- \( %rsp \) is the stack pointer, \( %rbp \) the frame pointer.