École Normale Supérieure

Langages de programmation et compilation

examen 2016–2017

Jean-Christophe Filliâtre

20 janvier 2017

Les notes de cours manuscrites ou reprographiées sont les seuls documents autorisés.
Les deux problèmes sont indépendants. L'épreuve dure 3 heures.

---

1 Éloge de la simplicité

Dans ce problème, on considère la variante de mini-ML suivante, avec des entiers, une opération de soustraction et une construction `ifzero` qui permet de tester si une valeur est ou non l'entier 0.

e	::=	n	constante entière $n \in \mathbb{Z}$
	$\mid$	x	variable
	$\mid$	$e - e$	soustraction
	$\mid$	$\texttt{fun}\ x \to e$	fonction anonyme
	$\mid$	$e\ e$	application
	$\mid$	$\texttt{let}\ x = e\ \texttt{in}\ e$	variable locale
	$\mid$	$\texttt{ifzero}\ e\ \texttt{then}\ e\ \texttt{else}\ e$	conditionnelle

On munit ce langage d'une sémantique similaire à celle vue en cours, c'est-à-dire une stratégie d'appel par valeur avec évaluation de la gauche vers la droite. Les valeurs sont

v	::=	n	constante
	$\mid$	$\texttt{fun}\ x \to e$	fonction

Une sémantique opérationnelle à petits pas est donnée figure 1. On notera qu'elle est déterministe.

Comme on l'a vu en cours, il existe dans ce langage des expressions dont l'évaluation ne termine pas, comme $(\texttt{fun}\ x \to x\ x)\ (\texttt{fun}\ x \to x\ x)$. Mais on peut montrer en revanche que les expressions typées avec des *types simples* terminent toujours. C'est ce que nous allons faire dans les questions qui suivent. On se limite ici au seul type de base `int` et les types simples sont donc les suivants :

$$\tau \ ::= \ \texttt{int} \mid \tau \to \tau$$

Le jugement de typage est noté

$$\Gamma \vdash e : \tau$$

où l'environnement de typage Γ est une fonction des variables vers les types. Les règles d'inférence de ce jugement sont données figure 2. On admet les propriétés de progression (si $\vdash e : \tau$, alors soit e est une valeur, soit il existe e' tel que $e \to e'$) et de préservation (si $\vdash e : \tau$ et $e \to e'$ alors $\vdash e' : \tau$) pour ce langage.

contextes de réduction

$$
\begin{array}{rcl}
E & ::= & \square \\
& | & E\ e \\
& | & v\ E \\
& | & E - e \\
& | & v - E \\
& | & \texttt{let}\ x = E\ \texttt{in}\ e \\
& | & \texttt{ifzero}\ E\ \texttt{then}\ e\ \texttt{else}\ e
\end{array}
$$

réductions de tête

$$
\begin{array}{rcll}
(\texttt{fun}\ x \to e)\ v & \xrightarrow{\epsilon} & e[x \leftarrow v] & \\
\texttt{let}\ x = v\ \texttt{in}\ e & \xrightarrow{\epsilon} & e[x \leftarrow v] & \\
n_1 - n_2 & \xrightarrow{\epsilon} & n & \text{avec } n \text{ l'entier } n_1 - n_2 \\
\texttt{ifzero}\ 0\ \texttt{then}\ e_1\ \texttt{else}\ e_2 & \xrightarrow{\epsilon} & e_1 & \\
\texttt{ifzero}\ n\ \texttt{then}\ e_1\ \texttt{else}\ e_2 & \xrightarrow{\epsilon} & e_2 & \text{si } n \neq 0
\end{array}
$$

FIGURE 1 – Sémantique opérationnelle à petits pas.

$$\frac{}{\Gamma \vdash n : \texttt{int}} \qquad \frac{x \in \mathrm{dom}(\Gamma)}{\Gamma \vdash x : \Gamma(x)}$$

$$\frac{\Gamma + x : \tau_1 \vdash e : \tau_2}{\Gamma \vdash \texttt{fun}\ x \to e : \tau_1 \to \tau_2} \qquad \frac{\Gamma \vdash e_1 : \tau_2 \to \tau_1 \quad \Gamma \vdash e_2 : \tau_2}{\Gamma \vdash e_1\ e_2 : \tau_1}$$

$$\frac{\Gamma \vdash e_1 : \texttt{int} \quad \Gamma \vdash e_2 : \texttt{int}}{\Gamma \vdash e_1 - e_2 : \texttt{int}} \qquad \frac{\Gamma \vdash e_1 : \tau_1 \quad \Gamma + x : \tau_1 \vdash e_2 : \tau_2}{\Gamma \vdash \texttt{let}\ x = e_1\ \texttt{in}\ e_2 : \tau_2}$$

$$\frac{\Gamma \vdash e_1 : \texttt{int} \quad \Gamma \vdash e_2 : \tau \quad \Gamma \vdash e_3 : \tau}{\Gamma \vdash \texttt{ifzero}\ e_1\ \texttt{then}\ e_2\ \texttt{else}\ e_3 : \tau}$$

FIGURE 2 – Règles de typage.

Pour un type τ donné, on définit un ensemble T_τ d'expressions closes de type τ par récurrence sur la structure de τ de la manière suivante :

1. $e \in T_{\texttt{int}}$ ssi l'évaluation de e termine ;
2. $e \in T_{\tau_1 \to \tau_2}$ ssi l'évaluation de e termine et quel que soit $e_1 \in T_{\tau_1}$, on a $e\ e_1 \in T_{\tau_2}$.

Notre objectif est de montrer que T_τ contient *toutes* les expressions closes de type τ.

Question 1 Montrer que si $\vdash e : \tau$ et $e \to e'$ alors $e \in T_\tau$ ssi $e' \in T_\tau$.

Question 2 Soit e telle que $x_1 : \tau_1 + \cdots + x_k : \tau_k \vdash e : \tau$ et $v_1, \ldots, v_k$ des valeurs closes de types respectifs $\tau_1, \ldots, \tau_k$ telles que $v_i \in T_{\tau_i}$ pour tout i. Montrer alors que

$$e[x_1 \leftarrow v_1] \cdots [x_k \leftarrow v_k] \in T_\tau.$$

Question 3 En déduire que l'évaluation de toute expression close typée termine.

Variables mutables. On étend maintenant notre langage avec la possibilité de modifier en place la valeur d'une variable. On ajoute au langage une nouvelle construction pour cela :

$$\begin{array}{rcll} e & ::= & \ldots & \\ & | & x \leftarrow e & \text{affectation} \end{array}$$

On suppose de plus que l'on distingue les variables introduites par `let` et les variables introduites par `fun` et que seules les variables introduites par `let` peuvent apparaître à gauche de l'affectation.

Pour donner une sémantique à ce nouveau langage, on introduit une notion d'adresse, notée a. On ne précise pas ce que sont les adresses mais on les suppose en nombre infini. Un état mémoire, noté M, est une fonction des adresses vers les valeurs. La notion de valeur ne change pas. En particulier, les adresses ne sont pas des valeurs. Pour définir une sémantique opérationnelle à petits pas sur ce nouveau langage, on ajoute les adresses à la syntaxe abstraite des expressions :

$$\begin{array}{rcll} e & ::= & \ldots & \\ & | & a & \text{accès à une adresse} \\ & | & a \leftarrow e & \text{affectation à une adresse} \end{array}$$

De telles adresses ne font pas partie du langage dans lequel l'utilisateur écrit des programmes, mais sont introduites uniquement pour les besoins de la sémantique opérationnelle. La relation de réduction en un pas prend la forme

$$M_1/e_1 \rightsquigarrow M_2/e_2$$

et s'interprète comme « l'état mémoire M_1 et l'expression e_1 se réduisent en un pas de calcul vers l'état mémoire M_2 et l'expression e_2 ». Les contextes de réduction traduisent toujours une stratégie d'appel par valeur avec évaluation de la gauche vers la droite. Ce sont donc les mêmes qu'à la figure 1, avec l'addition d'un nouveau contexte pour réduire à droite d'une affectation :

$$\begin{array}{rcll} E & ::= & \ldots & \\ & | & a \leftarrow E & \end{array} \qquad \frac{M/e \overset{\epsilon}{\rightsquigarrow} M'/e'}{M/E(e) \rightsquigarrow M'/E(e')}$$

Les réductions de tête sont données figure 3.

$$
\begin{array}{rcll}
M/a & \overset{\epsilon}{\rightsquigarrow} & M/M(a) & \\
M/(\texttt{fun}\ x \to e)\ v & \overset{\epsilon}{\rightsquigarrow} & M/e[x \leftarrow v] & \\
M/\texttt{let}\ x = v\ \texttt{in}\ e & \overset{\epsilon}{\rightsquigarrow} & M[a \mapsto v]/e[x \leftarrow a] & \text{avec } a \text{ une adresse fraîche} \\
M/a \leftarrow v & \overset{\epsilon}{\rightsquigarrow} & M[a \mapsto v]/v & \\
M/n_1 - n_2 & \overset{\epsilon}{\rightsquigarrow} & M/n & \text{avec } n \text{ l'entier } n_1 - n_2 \\
M/\texttt{ifzero}\ 0\ \texttt{then}\ e_1\ \texttt{else}\ e_2 & \overset{\epsilon}{\rightsquigarrow} & M/e_1 & \\
M/\texttt{ifzero}\ n\ \texttt{then}\ e_1\ \texttt{else}\ e_2 & \overset{\epsilon}{\rightsquigarrow} & M/e_2 & \text{si } n \neq 0
\end{array}
$$

FIGURE 3 – Sémantique opérationnelle avec état mémoire.

Question 4 Donner la séquence de réductions de l'expression

$$
\begin{array}{l}
\texttt{let}\ x = 1\ \texttt{in} \\
\texttt{let}\ f = \texttt{fun}\ y \to x \leftarrow x - y\ \texttt{in} \\
(x \leftarrow 0) - (f\ 42)
\end{array}
$$

dans un état mémoire initialement vide.

Typage. On type ce nouveau langage avec les mêmes types simples que précédemment. L'environnement de typage Γ, qui donne le type des variables, reste le même. Pour typer les adresses, on se donne un second environnement, à savoir une fonction Σ des adresses vers les types. Le jugement de typage prend la forme $\Gamma, \Sigma \vdash e : \tau$. Les règles de typage de la figure 2 sont inchangées, si ce n'est que Σ est ajouté à côté de tout environnement.

Question 5 Donner les règles de typage pour les nouvelles constructions, à savoir une adresse a, une affectation $x \leftarrow e$ et une affectation $a \leftarrow e$.

Correction du typage. Pour adapter la preuve de correction du typage vue en cours, il faut mettre en relation les états mémoire et les environnements. On dit qu'un état mémoire M est bien typé dans un environnement Γ, Σ, ce que l'on note $\Gamma, \Sigma \vdash M$, si $\mathrm{dom}(M) = \mathrm{dom}(\Sigma)$ et $\Gamma, \Sigma \vdash M(a) : \Sigma(a)$ pour tout $a \in \mathrm{dom}(M)$.

Question 6 Montrer la préservation du typage par la réduction, c'est-à-dire que, si $\Gamma, \Sigma \vdash e : \tau$, $\Gamma, \Sigma \vdash M$ et $M/e \rightsquigarrow M'/e'$, alors il existe un environnement Σ' tel que $\Gamma, \Sigma' \vdash e' : \tau$ et $\Gamma, \Sigma' \vdash M'$. (On admet les résultats de permutation, d'affaiblissement et de préservation par substitution analogues à ceux vus en cours.)

Question 7 Montrer la propriété de progression, c'est-à-dire que si $\emptyset, \Sigma \vdash e : \tau$ alors soit e est une valeur, soit pour tout M tel que $\emptyset, \Sigma \vdash M$ il existe M' et e' tels que $M/e \rightsquigarrow M', e'$.

Question 8 En déduire la sûreté du typage, à savoir que si $\emptyset, \Sigma \vdash e : \tau$, si $\emptyset, \Sigma \vdash M$ et si $M/e \rightsquigarrow^\star M', e'$ avec M'/e' irréductible, alors e' est une valeur.

Question 9 Donner une expression close e de type $\texttt{int} \to \texttt{int}$ correspondant à la fonction factorielle sur les entiers naturels, c'est-à-dire telle que $e\ n$ s'évalue en $n!$ pour tout $n \in \mathbb{N}$.

Question 10 Donner une expression close e de type `int` dont l'évaluation ne termine pas.

Question 11 Proposer un schéma de compilation pour ce langage. On discutera notamment de la possibilité d'allouer certaines variables sur la pile.

Capture par valeur. On souhaite ajouter à notre langage la possibilité de capturer une variable *par valeur* à l'intérieur d'une fermeture. Pour cela, on modifie la syntaxe des fonctions de la manière suivante :

$$\begin{array}{rcll} e & ::= & \ldots & \\ & | & \texttt{fun}\ x\ \{x_1, \ldots, x_n\} \to e & \text{fonction} \end{array}$$

L'ensemble $\{x_1, \ldots, x_n\}$ représente le sous-ensemble des variables libres de e que l'on souhaite capturer par valeur. Lorsque cet ensemble est vide, la construction coïncide avec la construction $\texttt{fun}\ x \to e$ précédente. Une variable capturée par valeur reste une variable mutable à l'intérieur du corps de la fonction ; mais elle représente une *nouvelle* variable. Ainsi, les deux expressions suivantes

$$\begin{array}{l} \texttt{let}\ x = 0\ \texttt{in} \\ \texttt{let}\ f = \texttt{fun}\ y \to x \leftarrow y - x\ \texttt{in} \\ (f\ 42) - x \end{array} \qquad \begin{array}{l} \texttt{let}\ x = 0\ \texttt{in} \\ \texttt{let}\ f = \texttt{fun}\ y\ \{x\} \to x \leftarrow y - x\ \texttt{in} \\ (f\ 42) - x \end{array}$$

s'évaluent respectivement en 0 et 42.

Question 12 Proposer une traduction de ce troisième langage vers le précédent, c'est-à-dire une élimination de la construction $\texttt{fun}\ x\ \{x_1, \ldots, x_n\} \to e$ au profit de la construction $\texttt{fun}\ x \to e$ précédente, qui respecte la sémantique.

Question 13 En supposant que *toutes* les variables sont capturées par valeur, montrer que de nouveau l'évaluation de toute expression close bien typée termine. (Attention : ce n'est pas le même langage qu'au départ, puisque les variables restent mutables.)

---

2 Un visiteur, venu d'ailleurs [1]

Dans ce problème, on s'intéresse à l'écriture d'un interprète dans un langage orienté objet, en l'occurrence Java. Prenons l'exemple minimaliste d'un langage d'expressions réduites à des constantes entières et des soustractions, c'est-à-dire la syntaxe abstraite suivante :

$$\begin{array}{rcll} e & ::= & n & \text{constante entière } n \in \mathbb{Z} \\ & | & e - e & \text{soustraction} \end{array}$$

Une telle syntaxe abstraite peut être représentée en Java par trois classes, dont on omet ici les constructeurs :

```
abstract class Expr { }
class Ecte extends Expr { int n; }
class Esub extends Expr { Expr e1, e2; }
```

La classe `Ecte` représente une constante entière n et la classe `Esub` une expression de la forme $e_1 - e_2$. On peut écrire un interprète pour ces expressions en introduisant une méthode `interp` dans la classe `Expr`, de la manière suivante

```
abstract class Expr { abstract int interp(); }
```

et en la définissant dans chacune des deux sous-classes.

Question 14 Donner le code Java de la méthode `interp` pour chacune des deux sous-classes `Ecte` et `Esub`.

Une autre solution. Bien que parfaitement correcte, cette façon de procéder a le défaut d'être intrusive, car elle mélange les données (c'est-à-dire ici les types, champs et constructeurs définissant la syntaxe abstraite) et l'algorithme (c'est-à-dire ici le code de l'interprète). Il existe une autre façon de procéder, connue sous le nom de *visiteur*, permettant de séparer données et algorithme. L'idée est d'écrire le code de l'interprète dans une classe à part `Interp`

```
class Interp {
  int visit(Ecte e) { ... }
  int visit(Esub e) { ... }
}
```

et de se contenter d'une intrusion minimale dans la syntaxe abstraite sous la forme d'une méthode « acceptant » le visiteur.

```
abstract class Expr { abstract int accept(Interp i); }
class Ecte extends Expr { ...; int accept(Interp i) { return i.visit(this); }
class Esub extends Expr { ...; int accept(Interp i) { return i.visit(this); }
```

Question 15 Donner le code Java de la classe `Interp`.

Question 16 Le code de la méthode `accept` étant rigoureusement le même dans les classes `Ecte` et `Esub`, on peut être tenté de l'écrire dans la classe `Expr`, afin que les classes `Ecte` et `Esub` en héritent. (Java permet en effet de définir une méthode dans une classe abstraite, exactement pour cette raison.) Expliquer pourquoi ce n'est pas possible ici.

1. Hommage à *Toy Story*, évidemment.

Question 17 Donner le code assembleur x86-64 de la méthode `accept` de la classe `Ecte` ainsi que des deux méthodes `visit` de la classe `Interp`. On suppose un schéma de compilation analogue à celui du cours : les objets sont des pointeurs vers le tas et leur premier champ contient un pointeur vers le descripteur de classe ; pour une méthode, l'objet `this` est passé dans `%rdi` et l'argument dans `%rsi` (il n'y a qu'un argument ici à chaque fois).

Annexe : aide-mémoire x86-64

On donne ici un fragment du jeu d'instructions x86-64. Vous êtes libre d'utiliser tout autre élément de l'assembleur x86-64. Dans ce qui suit, r_i désigne un registre, n une constante entière et L une étiquette.

`mov` r_2`,` r_1	copie le registre r_2 dans le registre r_1
`mov` n`,` r_1	charge la constante n dans le registre r_1
`add` r_2`,` r_1	calcule la somme de r_1 et r_2 dans r_1 (on a de même `sub` et `imul`)
`mov` n`(`r_2`),` r_1	charge dans r_1 la valeur contenue en mémoire à l'adresse $r_2 + n$
`mov` r_1`,` n`(`r_2`)`	écrit en mémoire à l'adresse $r_2 + n$ la valeur contenue dans r_1
`push` r_1	empile la valeur contenue dans r_1
`pop` r_1	dépile une valeur dans le registre r_1
`jmp` L	saute à l'adresse désignée par l'étiquette L
`call` L	saute à l'adresse désignée par l'étiquette L, après avoir empilé l'adresse de retour
`jmp` `*`o	saute à l'adresse désignée par l'opérande o
`call` `*`o	saute à l'adresse désignée par l'opérande o, après avoir empilé l'adresse de retour
`ret`	dépile une adresse et y effectue un saut

On alloue de la mémoire sur le tas avec un appel à `malloc`, qui attend un nombre d'octets dans `%rdi` et renvoie l'adresse du bloc alloué dans `%rax`.