Premiers pas en OCaml

1 Un nouveau langage

OCaml est un langage de programmation développé en France dans les années 1980 principalement pour la recherche et l'industrie, sous le nom "Objective Caml". Il est à la fois adapté à la programmation fonctionnelle, impérative et orientée objet (POO).

Remarque: Cependant, dans le cadre du programme de MP2I, nous utiliserons surtout ce langage de manière fonctionnelle, préférant le C pour la programmation impérative. La POO (objets, structures, méthodes...) est quant à elle hors-programme.

Ce langage peut être compilé et interprété (ce qui permet d'exécuter dynamiquement les programmes grâce à une transformation du code source en code bas niveau), comme suit (en bash):

```
• syntaxe de compilation : ocamlc source.ml -o exec
```

• syntaxe d'interprétation : utop

Lorsque l'on lance utop, on peut alors taper l'une des commandes suivantes :

définition;;	prend en compte la définition et affiche le type, la valeur et le nom de l'objet ainsi défini
expression;;	évalue l'expression et affiche son type et sa valeur
<pre>#use "source.ml";;</pre>	importe les définitions écrites dans le fichier source source.ml et évalue les expressions contenues dedans (interprétant donc, ligne à ligne, le fichier source)
exit 0;; (ou $Ctrl + D$)	sortie de utop

OCaml est un langage statiquement typé, c'est-à-dire que le type des valeurs et des fonctions est fixé par les définitions dans le code source (en particulier, avant l'appel de fonction et l'exécution).

Remarque : Le langage est muni d'un moteur d'inférence de type capable de déterminer le type d'une fonction, d'une valeur ou d'une expression, même quand celui-ci n'est pas précisé, et ceci grâce aux opérateurs, qui sont spécifiques à chaque type (*cf.* plus loin). Néanmoins, on veillera toujours dans ce cours à expliciter ces types.

Enfin, on met des parenthèses autour des sous-expressions.

```
Exemple: Ainsi, si f : int -> int est une fonction préalablement définie, alors "f 5 6-1" donne l'évaluation de f(5,6)-1 tandis que "f 5 (6-1)" donne celle de f(5,6-1).
```

2 Structure d'un fichier .ml

Un code OCaml écrit dans un fichier .ml est essentiellement constitué d'une suite de définitions, qui suivent la syntaxe décrite ci-dessous.

Syntaxe – définitions :

· Pour une définition simple, on a :

où nom est un identificateur, t un type et e une expression de type t.

· Pour une fonction, on écrira :

let nom_f (arg₁:t₁) ... (arg_n:t_n) :
$$t_s = expr_f$$

où $arg_1,...,arg_n$ et $t_1,...,t_n$ sont les arguments et leurs types, t_s est le type de sortie de la fonction, et $expr_f$ est une expression de type t_s .

Remarque : Dans la définition précédente, le type de nom_f est $t_1 \rightarrow t_2 \rightarrow \ldots \rightarrow t_n \rightarrow t_s$.

Ainsi, une valeur peut être vue comme une fonction d'arité nulle, c'est-à-dire qu'elle possède un nom, un type et une valeur. Les fonctions, quant à elles, on toujours une valeur mais qu'on peut qualifier de "fonctionnelle".

3 Expressions en OCaml

3.1 Qu'est-ce qu'une expression?

Définition (environnement):

À un instant donné, on appelle environnement courant l'ensemble des valeurs, fonctions et types définis au préalable ou prédéfinis dans le langage.

Définition (expression):

Pour un environnement donné, une expression est une suite de caractères à laquelle on peut attribuer un type et une valeur (au sens $large^{(*)}$)

Remarque: (*)Par exmple, la fonction $x \mapsto x + 1$ est une valeur, fonctionnelle, en tant qu'élément de $\mathbb{N}^{\mathbb{N}}$.

Remarque : Le type et la valeur d'une expression ne la caractérisent pas (prendre par exemple 3 et 1+2).

3.2 Les constantes

On donne ici quelques exemples des types prédéfinis du langage OCaml, qui constituent donc aussi des types d'expression possibles.

Type	Exemples	Remarques	
int	0, 1, -38		
float	0., 0.5, 12., 12.0	Il faut toujours mettre un point et préciser la partie entière (ne pas écrire ".5", par exemple).	
bool	true, false	Ce sont les deux seules valeurs de ce type.	
char	'a', '3', '\n', '\", '\\', '\"'		
string	"coucou", ""		
unit	()	C'est l'unique valeur de ce type.	

Remarque: Les booléens ne fonctionnement pas avec les entiers comme en C.

Remarque: Le type unit est l'équivalent du type void en C.

3.3 Les variables

Communément appelées (à tort) "variables", les valeurs définies dans l'environnement courant sont des expressions, ceci valant également pour les fonctions.

Exemple: Si on écrit "let a = 3", alors dans la suite a est une expression.

Remarque: Ces "variables" ressemblent donc plus aux variables mathématiques (muettes,...) que celles informatiques, qui désignent plutôt des espaces réservés en mémoire.

3.4 Les arguments

Dans une expression définissant une fonction, y compris une fonction anyonyme (cf. section suivante), il est possible d'utiliser les arguments de cette fonction. Un argument de fonction est donc une expression tant qu'on est dans le corps de la fonction.

Remarque : L'expression définissant la fonction possède donc un type, mais pas encore de valeur au sens d'une valeur dont le type est celui de sortie de la fonction (sa valeur est en fait fonctionnelle, par exemple : $fun x \rightarrow x+1$ qui est une fonction anonyme).

Remarque : Les expressions qu'on a vues jusqu'ici correspondent à des feuilles d'un arbre de syntaxe.

3.5 Les fonctions anonymes (ou abstractions)

En OCaml, il est possible de décrire une fonction par une simple expression, c'est-à-dire sans la définir, sans faire appel à **let** et sans lui donner de nom. Une telle fonction est dite anonyme.

Syntaxe – fonctions anonymes:

On utilise la syntaxe suivante :

fun
$$arg_1 arg_2 ... arg_n \rightarrow e$$

où $\arg_1,...,\arg_n$ sont des identificateurs et e est une expression qui peut utiliser ces arguments. Cette expression est de type $t_1 \rightarrow ... \rightarrow t_n \rightarrow t$ où $t_1,...,t_n$ sont les types des arguments

Cette expression est de type $t_1 \rightarrow \dots \rightarrow t_n \rightarrow t$ ou t_1,\dots,t_n sont les types des arguments et t est le type de e qu'on infère à partir de e (préciser les types n'est donc pas nécessaire).

3.6 Les appels de fonctions

3.6.1 Les opérateurs prédéfinis

Les opérateurs (arithmétiques, booléens, de comparaison) sont des cas particuliers de fonctions prédéfinies dans le langage, c'est pourquoi on les traite dans cette section.

Définition (notation préfixe, post-fixe, infixe):

On parle de notation préfixe (resp. post-fixe) lorsque l'opérateur se place avant (resp. après) ses opérandes. Pour les opérateurs binaires, il existe aussi la notation infixe où les opérandes se placent de part et d'autre des opérateurs.

Exemple : Le moins unaire (-) et la négation (\neg) sont notés de manière préfixe. La factorielle (!) est notée de façon post-fixe.

Les opérateurs prédéfinis en OCaml sont infixes s'ils sont binaires et préfixes s'ils sont unaires. On les présente dans le tableau ci-dessous :

Type t	Opérateurs binaires pour t	Opérateurs unaires pour t	Type de l'expression composée
int	+, -, *, /, mod	_	int
float	+.,, *., /., ** (puissance)		float
int ou float	>, >=, =, <, <=, <> ("différent de")		bool
bool	&&, , =	not	bool
char			
string	^ (concaténation)		string

Remarque: Pour les float, il faut utiliser l'opérateur -. pour les fonctions (par exemple: fun x -> -.x) mais ce n'est pas nécessaire pour les constantes (par exemple: -12.0).

Remarque : ** ne fonctionne pas sur les int, mais il existe des opérateurs de conversion.

Remarque: Notons que contrairement au C, les tests d'égalité se font avec un simple =. En fait, il existe aussi en OCaml un opérateur ==, mais celui teste l'égalité physique, c'est-à-dire liée aux adresses dans la mémoire.

On a de façon générale : $op^1 e_1$, ou bien $e_1 op^2 e_2$, où :

- op¹ est un opérateur unaire
- \cdot op² est un opérateur binaire
- e_1 et e_2 sont des expressions de types respectifs t_1 et t_2 .

3.6.2 Les fonctions en général

Syntaxe – appel de fonction:

Pour appeler une fonction f, on utilise la syntaxe suivante :

$$f e_1 e_2 \ldots e_n$$

où $e_1,...,e_n$ sont des expressions du type attendu pour les arguments de la fonction. L'expression entière est alors du type de sortie de f.

Remarque : OCaml évalue l'expression en lisant les arguments de la droite vers la gauche.

3.6.3 L'alternative

Syntaxe – branche conditionnelle et alternative :

Les branches conditionnelles en OCaml suivent la syntaxe suivante :

```
if cond then e_1 else e_2
```

où cond est une expression de type bool et e_1 et e_2 sont des expressions de même type t. Cette expression est de type t et de valeur celle de e_1 si la condition vaut true, celle de e_2 sinon.

Remarque: Le "else" est obligatoire.

3.6.4 Les n-uplets

Syntaxe – tupples:

Un *n*-uplet ou tupple est de la forme (e_1, e_2, \ldots, e_n) où pour tout $i \in [1..n]$, e_i est de type t_i . Cette expression est alors de type $t_1 * t_2 * \ldots * t_n$.

Exemple: "(3, "a", (1,2))" est de type - : int * string * (int * int).

3.6.5 Les définitions locales

Syntaxe – définition locale :

Si l'on veut définir localement une variable à l'intérieur d'une expression, on utilise:

```
let def in e
```

où let def suit la syntaxe des définitions et e est une expression de type t qui peut utiliser les variables définies dans def. L'expression globale reste de type t.

Cette commande est s'avère utile, notamment lorsque l'on veut déconstruire un tupple, c'est-à-dire récupérer individuellement ses composantes.

```
Exemple : let couple = (3,4);;
let (c,d) = couple in a+b
```

4 Définition récursive et mutuellement récursive

4.1 Fonctions récursives simples

Dans une déclaration classique de fonction, l'expression \mathbf{e} peut faire intervenir les arguments de la fonction, \mathbf{a}_1 jusqu'à \mathbf{a}_n , des fonctions préalablement définies, mais pas la fonction \mathbf{f} elle-même. Afin de réaliser des appels récursifs sur une fonction, il faut définir celle-ci comme étant une fonction récursive.

Syntaxe – fonction récursive :

Une fonction récursive est définie comme suit :

```
let rec f (a_1:t_1) ... (a_n:t_n) : t_s = e
```

où e est une expression de type t_s pouvant faire appel à f.

Exemple: Voyons deux façons possibles de coder la factorielle de manière récursive.

4.2 Fonctions mutuellement récursives

Syntaxe – fonctions mutuellement récursives :

```
On peut définir des fonctions qui soient récursives entre elles comme suit :
```

```
let rec f<sup>1</sup> (a<sub>1</sub><sup>1</sup>:t<sub>1</sub><sup>1</sup>) ... (a<sub>n<sub>1</sub></sub><sup>1</sup>:t<sub>n<sub>1</sub></sub><sup>1</sup>) : t<sub>s</sub><sup>1</sup> = expr<sub>1</sub> and f<sup>2</sup> (a<sub>1</sub><sup>2</sup>:t<sub>1</sub><sup>2</sup>) ... (a<sub>n<sub>2</sub></sub><sup>2</sup>:t<sub>n<sub>2</sub></sub><sup>2</sup>) : t<sub>s</sub><sup>2</sup> = expr<sub>2</sub> : and f<sup>k</sup> (a<sub>2</sub><sup>k</sup>:t<sub>1</sub><sup>k</sup>) ... (a<sub>n<sub>k</sub></sub><sup>k</sup>:t<sub>n<sub>k</sub></sub><sup>k</sup>) : t<sub>s</sub><sup>k</sup> = expr<sub>k</sub>
```

où pour tout $i \in [1..k]$, $\exp \mathbf{r}_i$ est une expression de type \mathbf{t}_s^i pouvant faire intervenir \mathbf{f}^j pour tout $j \in [1..k]$ et a_r^i pour tout $r \in [1..n_i]$.

5 Définition de types

5.1 Renommer un type

Syntaxe – définition d'un nouveau type :

En OCaml, il est possible de définir des nouveaux types grâce à la syntaxe suivante :

```
type nomtype = expr_de_type
```

où nomtype est un identificateur et expr de type est une expression de type connue.

```
Exemple: type entier = int;;
     type mon_triplet = entier * float * entier;;
```

Les expressions de type peuvent être simples ou composées, on liste dans ce qui suit les principales sortes de type.

5.1.1 Les types simples

On distingue trois sortes de types simples :

- · Les types prédéfinis : int, float, bool, char, string, unit.
- Les types préalablement définis : après une définition de type de la forme type nomtype = expr, nomtype est une expression de type simple.
- · Les variables de type : une variable de type est désignée par 'identificateur.

Exemple: fun x -> x;; dans utop donne a' -> 'a = <fun>: 'a est une variable de type.

5.1.2 Les types produit cartésien

Les types produit cartésien sont les $t_1 * t_2 * \dots * t_n$ où t_1, \dots, t_n sont des expressions de type.

Un type produit cartésien désigne donc un type de *n*-uplet, ressemblant aux structures en C où chaque champ pouvait avoir un type différent, à ceci près qu'ici, les composantes du *n*-uplet n'ont pas de nom. Il existe néanmoins un autre type appelé type produit (à ne pas confondre avec le type produit cartésien présenté ici), qui permet de nommer les champs (*cf.* plus loin).

```
Exemple: type faux triplet = int * (int * int)
```

Remarque: $t_1*t_2*t_3$, $(t_1*t_2)*t_3$ et $t_1*(t_2*t_3)$ ne représentent pas le même type.

Pour accéder aux différentes composantes d'un n-uplet, on utilise let in.

Exemple:

5.1.3 Les types fonctionnels

Le type $t_e \rightarrow t_s$ désigne le type d'une fonction à un seul argment de type t_e , et dont le type de sortie est t_s . Plus généralement, si f est une fonction de type de sortie t_s dont les arguments sont de types respectifs $t_1,...,t_n$, alors le type de f est : $t_1 \rightarrow \ldots \rightarrow t_n \rightarrow t_s$.

Exemple:

En OCaml, une fonction de deux arguments est vue comme une fonction qui associé au premier argument une fonction du deuxième argument. Ainsi, $t_1->t_2->t_3$ et $t_1->(t_2->t_3)$ représentent le même type.

Exemple:

Remarque:

Remarque: Attention en revanche, les types $t_1 \rightarrow t_2 \rightarrow t_3$ et $(t_1 \rightarrow t_2) \rightarrow t_3$ sont différents.

5.2 Définir un type paramétré

5.2.1 Type paramétré à un seul paramètre

Syntaxe – type paramétré simple :

On définit un type paramétré comme suit :

```
type 'a nomtype = expr de type
```

où a et nomdutype sont des identificateurs et expr_de_type est une expression qui peut faire intervenir la variable 'a.

Après une telle définition, t nomdutype est une expression de type pour n'importe quelle expression de type t. À l'oral, on désignera parfois 'a par "alpha".

```
Exemple: type 'a quadruplet = 'a*'a*'a*'a;;
    let a:bool quadruplet = (true, true, true, false);;
    type quad_bool = bool quadruplet;;
```

Remarque: Attention à ne pas oublier de définir le type 'a après.

5.2.2 Type paramétré à plusieurs paramètres

Syntaxe – type paramétré à plusieurs paramètres :

On les définit comme suit :

```
type ('a_1, 'a_2,..., 'a_n) nomdutype = expr_de_type
```

où $a_1,...,a_n$ et nomdutype sont des identificateurs et expr_de_type est une expression de type pouvant faire intervenir ' a_1 , ' $a_2,...$,' a_n .

```
Exemple : type ('a,'b) quadruplet = 'a*'b*'a*'b;;
    let a:(float,int) quadruplet = (2.0,-1,3.0,4);;
    type quad_bool = (bool,bool) quadruplet;;
    type 'd semi_bool_quad = (bool,'d) quadruplet;;
```

5.3 Définir un type somme

5.3.1 Définition du type somme

Syntaxe - type somme :

Un type somme est un type dont on définit préalablement les valeurs possibles. Comme les types simples, il peut être paramétré, ce que l'on obtient avec la syntaxe :

```
type ('a<sub>1</sub>, 'a<sub>2</sub>,..., 'a<sub>n</sub>) typeSomme =

| Cons<sub>1</sub> [of t<sub>1</sub>]

| Cons<sub>2</sub> [of t<sub>2</sub>]

:

| Cons<sub>m</sub> [of t<sub>m</sub>]
```

où $\mathtt{Cons}_1,...,\mathtt{Cons}_m$ sont des identificateurs appelés constructeurs et $\mathtt{t}_1,...,\mathtt{t}_n$ sont des expressions de type (que l'on met facultativement), qui peuvent faire intervenir $\mathtt{a}_1,...,\mathtt{a}_n$.

Voici une syntaxe alternative qui met en valeur le caractère facultatif de ces types :

Après une telle définition :

- n'importe quelle expression de la forme $\mathtt{et}_1,\mathtt{et}_2,\ldots,\mathtt{et}_n$ typeSomme est une expression de type si $\mathtt{et}_1,\ldots,\mathtt{et}_1$ sont des expressions de type
- chaque $Cons_0^i$ est une expression dont le type est ('a₁, 'a₂,..., 'a_n) typeSomme
- chaque $Cons_1^j$ e est une expression dont le type est ('a₁, 'a₂,..., 'a_n) typeSomme ou une spécialisation de ce type, si e est une expression de type t^j .

Remarque: Une fonction dont les arguments appartiennent à un type paramétré (donc elle-même de type paramétré) est appelée fonction polymorphe. Lorsque l'un ou plusieurs des paramètres de type sont définis, on parle alors de spécialisation pour cette fonction.

5.3.2 Filtrage associé

Syntaxe - filtrage:

Le filtrage sur une expression issue d'un type somme permet de réaliser des instructions ou retourner des valeurs différentes en fonction de la valeur de cette expression. Ceci se fait en listant de manière exhaustive les cas possibles ainsi que la valeur associée à chaque cas :

où e est une expression dont le type est ('a₁, 'a₂,..., 'a_n) typeSomme ou une spécialisation, valⁱ est une expression de type t^1 (une valeur) et varⁱ est un identificateur (pour une variable).

On peut aussi écrire une syntaxe plus générale :

Remarque: · L'ordre des motifs est important.

- · Il ne faut pas qu'il y ait de motif redondant.
- · Les motifs doivent couvrir toutes les valeurs possibles du type somme.

Les warnings possibles dans les filtrages sont donc la redondance et la non exhaustivité.

Remarque : On peut n'est pas limité à une seule valeur lorsqu'on traite le cas $Cons_1^i$: on peut aussi le filtrer successivement selon plusieurs valeurs $val^{i,j}$ différentes.

Exemple : On peut modéliser un jeu de cartes à l'aide de types sommes : *cf.* fichier "carte-bis.ml".

5.3.3 Un type somme récursif : implémentation de la pile en OCaml

Comme on a le choix, pour chaque constructeur d'un type somme, de préciser ou non un type associé à ce constructeur, il est possible de définir des types sommes récursifs. On illustre cela dans l'exemple suivant, dédié à une implémentation possible de la structure de pile en OCaml.

Voici donc les définitions des fonctions élémentaires :

```
Accès au sommet (sans filtrage)
                             let sommet (p:'a pile) : 'a =
                                 (* hyp : not(est pile vide p) *)
                                 let PileNonVide (elem,pp) = p in elem
Accès au sommet (avec filtrage)
                             let sommet bis (p:'a pile) : 'a =
                                 (* hyp : not(est_pile_vide p) *)
                                 match p with
                                      | PileVide -> failwith "pile vide"
                                      | PileNonVide (elem, ) -> elem
                 Empilement
                             let empiler (p:'a pile) (elem:'a) : 'a pile =
                                 PileNonVide (elem,p)
                 Dépilement
                             let depiler (p:'a pile) : 'a pile =
                                 (* hyp : not(est_pile_vide p) *)
                                 match p with
                                      | PileVide -> failwith "pile vide"
                                     | PileNonVide ( ,pp) -> pp
```

À cela, on rajoute les définitions de deux fonctions non élémentaires : l'une permettant de sommer les valeurs d'une pile d'entiers (donc, **int** pile ici), l'autre permettant de concaténer les chaînes contenue dans une pile de chaîne de caractères (**string** pile), la chaîne au sommet de la pile se plaçant au début.

```
Somme (récursif non terminal)
                            let rec somme (p:int pile) : int =
                                match p with
                                     | PileVide -> 0
                                     | PileNonVide (e,pp) -> e + (somme pp)
   Somme (récursif terminal)
                            let somme_rt (p:int pile) : int =
                                let rec aux (pp:int pile) (s:int) =
                                     (* calcule s + la somme des éléments de pp *)
                                     match pp with
                                         | PileVide -> s
                                         | PileNonVide (e,spp) -> aux spp (s+e)
                                in aux p 0
Concaténation (non temrinal)
                            let rec conca (p:string pile) : string =
                                match p with
                                     | PileVide -> ""
                                     | PileNonVide (e,pp) -> e^(conca pp)
    Concaténation (temrinal)
                            let conca_rt (p:string pile) : string =
                                let rec aux (pp:string pile) (s:string) =
                                    match pp with
                                         | PileVide -> s
                                         | PileNonVide (e,spp) -> aux spp (s^e)
                                in aux p ""
```

Remarque: Remarquons bien que l'opération de concaténation ne se fait pas dans le même sens pour conca et conca rt, même si ces deux fonctions sont bien équivalentes (réfléchir à

la façon dont se construit la chaîne totale dans les deux cas).

5.3.4 Le type option : un type somme paramétré prédéfini

Le type option est un type particulier, il est défini comme suit :

Il est notamment utile si l'on veut s'autoriser à ne "rien" retourner dans une fonction : alors, on retournera $\texttt{Some}\ v$ lorsqu'il est possible de trouver une valeur v convenable, et $\texttt{None}\ dans$ le cas contraire.

6 Définir un type produit

Les types produit, aussi appelés types enregistrement ou "record", sont l'équivalent des **struct** en C : ils permettent de rassembler des données de types différents et d'y accéder grâce à des noms.

Syntaxe – type produit:

Un type produit peut, tout comme un type somme, être paramétré. On le définit comme suit :

où $\operatorname{champ}_1,...,\operatorname{champ}_n$ sont des identificateurs et $\operatorname{t}_1,...,\operatorname{t}_n$ sont des expressions de type pouvant utiliser $\operatorname{a}_1,...,\operatorname{a}_n$.

• Lorsqu'on définit alors une variable de type typeProduit, on adopte la syntaxe suivante :

```
let var : (s_1,...,s_n) typeProduit = \{champ_1 = val_1,...,champ_m = val_m\} où s_1,...,s_n sont des expressions de type et val_1,...,val_m sont de types respectifs t_1,...,t_m (avec la spécialisation induite par s_1,...,s_n).
```

• Pour $i \in [1..m]$, on peut alors accéder au champ champ_i de var avec : var.champ_i.

Remarque: Contrairement au C, il est nécessaires de réécrire les noms des champs lors de la définition d'une variable, mais on peut par contre les initialiser dans un ordre différent de celui indiqué dans la définition du type produit.