

目次

1. 活動の概要
2. 関数型プログラミング言語
3. Standard ML
 - 3.1 Standard ML の特徴
 - 3.2 開発環境
 - 3.3 式の評価, 束縛, 関数, コメント, ファイルの使用, コンパイル
 - 3.4 主なデータ型とリテラル
 - 3.5 関数定義, パターンマッチング, if, 相互再帰/高階関数, 関数式, 部分適用, 中置演算子
 - 3.6 リスト操作関数, リストの畳み込み
 - 3.7 データ型の定義
 - 3.8 命令型言語の機能
 - 3.9 モジュール, ストラクチャ, シグネチャ, シグネチャ制約の透明性, ファンクタ
4. コンパイラの構成
 - 3.1 字句解析器
 - 3.2 構文解析器
5. 後期の活動について
6. 参考文献

1. 活動の概要

この班では, 関数型プログラミング言語 **Standard ML** を用いて, コンパイラを開発することを目標として活動した. 活動期間は通年であるが, 前期の活動としては, まず Standard ML の構文や言語機能についての勉強会を開くという形式をとった. 前期活動の終盤では, コンパイラ全体の構成についての勉強会を行い, 字句解析器や構文解析器について学んだ.

以下, 勉強会を通じて得られた知見をまとめる.

2. 関数型プログラミング言語

関数型プログラミング言語とは, 関数によってプログラムを構成するプログラミングスタイルを基本とする言語である. Standard ML の他に, LISP や Haskell, OCaml などが代表的な関数型プログラミング言語として知られている.

3. Standard ML

3.1 Standard ML の特徴

Standard ML は, プログラミング言語 ML の標準仕様, あるいは一方言として位置付けられる関数型プログラミング言語である. 以下のような特徴を持っている.

- 強力な型システム
- パターンマッチング機構

- 高度なモジュールシステム
- 命令型言語的な機能

関数がプログラムの主役となる関数型言語では、関数は第一級オブジェクト (first-class object) である。ここで、第一級オブジェクトとは、変数に格納したり、関数の引数や戻り値に使用できるデータのことをいう。

3.2 開発環境

Standard ML の処理系としては、コンパイラ **mlton** と、対話的実行環境 **Standard ML of New Jersey** (以下, **smlnj**) が存在する。そのため、ソースファイルを用いてプログラムを作成することも、ソースファイルを作成せずに式の評価や型システムの動作を対話的にテストすることもできる。

mlton と smlnj は、どちらも apt, pacman, brew といったパッケージ管理システム経由で導入が可能である。このプロジェクト活動においては、macOS 環境では brew 経由で、Ubuntu 環境では apt 経由で、Arch linux 環境では pacman 経由で導入した。

3.3 式の評価と束縛, 関数, コメント, ファイル

式の評価と束縛

smlnj は sml コマンドで起動する。式を入力して改行すれば評価される。複数行に渡る構文で記述している場合、改行して入力続けることができる。

```
- 2 * 3 + 4 * 5;  
val it = 26 : int  
- it - 20;  
val it = 6 : int
```

セミコロンは式の終端を表す。smlnj は式を1度評価することに、その結果を名前 **it** に束縛する。**it** を用いると、前回の式の評価結果を用いた新たな式の評価が可能である。

関数型言語では、代入は行わずに、異なる概念である**束縛**を行う。束縛は、式と名前を対応付ける操作を指す。例えば先ほど示したサンプルでは、最初は名前 **it** を **26** で束縛し、次は **6** で束縛している。

関数

以下で、2 つの int 型データを渡して和を得る add 関数を、2 種類の方法で定義する。

```
- fun add (x: int) (y: int): int = x + y;  
val add = fn : int -> int -> int  
- add 1 2;  
val it = 3 : int
```

```
- fun add x y = x + y;  
val add = fn : int -> int -> int
```

```
- add 1 2;  
val it = 3 : int
```

どちらも挙動は全く同じだが，前者では全ての引数と戻り値の型を明記し，後者では可能な限り型を省略している．ここでは，Standard ML の型推論を最大限利用している．型推論とは，コンパイル時に不足している型の情報を推論する機能である．

コメント

```
(* コメントは (* ネスト可能 *) *)
```

ファイル

```
(* file: test1.sml *)  
fun add x y = x + y  
val m = 2  
val n = 3
```

上の内容のソースコードを，`test1.sml` という名前で保存する．

```
- use "test1.sml";  
[opening test1.sml]  
val add = fn : int -> int -> int  
val m = 2 : int  
val n = 3 : int  
val it = () : unit  
  
- add 4 5;  
val it = 9 : int
```

`smlnj` で `use "...";` を評価させると，文字列型データで指定したソースファイルをコンパイルして対話的実行環境にインポートする．

3.4 主なデータ型

基本データ型

型名	説明	リテラルの例
<code>bool</code>	論理型	<code>true</code> , <code>false</code>
<code>int</code>	整数型 (符号付き整数型)	<code>123</code> , <code>0x1a</code> , <code>~123</code> (負数)
<code>word</code>	ワード型 (符号なし整数型)	<code>0w123</code> , <code>0wx1a</code>

型名	説明	リテラルの例
real	実数型	3.14, ~3.14, 1.23e4, 1.23e~4
char	文字型	#"a", #"\n"
string	文字列型	"abc\n"
unit	ユニット型 (0-タプル)	()

複合型

型 (例)	説明	リテラルの例
int * int * string	タプルの例	(123, 456, "abc")
{ a: int, b: int, c: string }	レコード型	{ a = 123, b = 456, c = "abc" }
int list	リスト型 (値の列)	[3, 1, 4, 2, 5]

論理型

論理値の値は true または false である。論理型データの演算に用いる演算子として、not (否定), andalso (論理積), orelse (論理和) が挙げられる。

```
- not true;
val it = false : bool
- it orelse true;
val it = true : bool
```

数値型

数値型の演算には ~ (単項マイナス), + (加算), - (減算), * (乗算), abs (絶対値) などを用いることができる。除算演算子には / (real 型用) と div (int, word 型用) の 2 種類がある。剰余演算子は mod で、int 型と word 型で利用できる。

```
- 2 * 3;
val it = 6 : int
- ~ it;
val it = ~6 : int
- abs it;
val it = 6 : int
- 10.0 / 3.0;
val it = 3.33333333333 : real
- 10 div 3;
val it = 3 : int
- 10 mod 3;
val it = 1 : int
```

これらの算術演算子は、2 つの引数が同一の型でなければ利用できない。例えば `2 + 3.0` は型エラーとなる。`real(2) + 3.0` のように、適切な型に揃える必要がある。

実数型から整数型への変換には `ceil` (∞ 方向丸め)、`floor` ($-\infty$ 方向丸め)、`trunc` (0方向丸め)、`round` (四捨五入) といった関数を利用できる。

文字型/文字列型

`\n` のようなエスケープシーケンスが利用できる。また、文字列の連結には `^` 演算子を用いる。

```
- "abc" ^ "def\n"
val it = "abcdef\n" : string
```

タプル型/ユニット型

タプル (tuple) は、複数の値を組にしたものである。例えば、整数 `123` と文字列 `"abc"` のタプルは `(123, "abc")` のように書き、その型は `int * string` と書く。

タプルの `n` (≥ 1) 番目の要素を取り出すには `#n` という関数を利用する。

```
- (123, (456, "abc"));
val it = (123, (456, "abc")) : int * (int * string)
- #1 (#2 it);
val it = 456 : int
```

ユニット (unit) は、要素数が 0 のタプルで表現される。ユニット型の唯一の値は `()` であり、型名は `unit` である。ユニット型は、例えば、意味のある値を返さない関数の戻り値の型として用いられる。文字列を出力する `print` 関数の戻り値の型はユニット型である。

レコード型

レコード (record) はラベル付きの値の集合である。

```
- val taro = { name = "Taro", age = 25 };
val taro = {age=25,name="Taro"} : {age:int, name:string}
- #age taro
val it = 25 : int
```

上の例では、ラベルとして `name` と `age` を持つレコードを生成している。ここで作成したレコード `taro` の型は `{ age: int, name: string }` であり、各要素を取り出すには `#name` や `#age` のような名前関数 (セレクト関数) を用いる。

リスト型

リスト (list) は要素を一列に連結してデータ構造である。リテラルのシンタックスは `[1, 2, 3]` のようになっており、この場合の型名は `int list` となる。要素数 0 のリスト (空リスト) は `[]` ま

たは `[]` と書く．

```
- [1, 2, 3];
val it = [1,2,3] : int list
- nil;
val it = [] : 'a list
- [];
val it = [] : 'a list
```

多相型

以下のプログラムでは，恒等関数（引数の値をそのまま返す関数）を定義している．

```
- fun id x = x;
val id = fn : 'a -> 'a
```

`id` 関数の型は `'a -> 'a` と推論されていることがわかる．`'a` のようにアポストロフィで始まるパラメータは**型変数**（type variable）と呼ばれ，任意の型を表す．`'a -> 'a` は，`id` 関数が，例えば `int -> int` 型としても振る舞い，`char -> char` 型としても振舞うことを示している．

等値型

1 個のアポストロフィで始まる型変数は任意の型を表すが，2 個以上のアポストロフィで始まる型変数（`''a` など）は**等値型**（equality type）を表す．等値型とは，関数 `=` で等値比較可能な型である．`bool`，`int`，`char`，`string` などの離散的なデータ型は等値型であり，等値型から構成されるタプルやリストなどの複合型も等値型である．

```
fun id (x : ''a) = x;
val id = fn : ''a -> ''a
```

すべての等値型には，比較演算子 `=` および `<>` を用いることができる．

3.5 関数定義，パターンマッチング，if，相互再帰/高階関数，関数式，部分適用，中置演算子

関数定義

以下のプログラムは，自然数の冪乗（`x` の `y` 乗）を返す関数 `power` を定義している．

```
fun power x 0 = 1
  | power x y = x * power x (y - 1)
```

関数の宣言は `fun` キーワードから始まり、`=` の左に関数名と引数、右に式を記述する。複数のケースを書きたい場合、上記の例のように縦棒 `|` を使って関数の記述を継続することができる。

上記のソースコードを `power.sml` というファイル名で保存し、`smlnj` から読み込むと、次のように動作する。

```
- use "power.sml";
[opening power.sml]
val power = fn : int -> int -> int
val it = () : unit
- power 2 8;
val it = 256 : int
- power 2 0;
val it = 1 : int
```

この `power` 関数の宣言の 1 行目と 2 行目を入れ替えてしまうと、`power x y` がすべてのケースにマッチしてしまうため、関数の呼び出しは終了しない。

```
fun power x y = x * power x (y - 1)
  | power x 0 = 1      (* こちらの定義には到達しない *)
```

パターンマッチング

これまでの説明で既に使用しているが、パターンには、変数や整数リテラルの他に、タプルやリスト、レコードも書くことができる。

```
- fun length nil = 0
  | length (x::xs) = 1 + length xs;
val length = fn : 'a list -> int
- length [1, 2, 3, 4, 5];
val it = 5 : int
```

次の例のように、興味のない部分にはワイルドカードと呼ばれるパターン `_` を使うこともできる。

```
- fun second (_, y) = y;
val second = fn : 'a * 'b -> 'a
- second (123, "abc");
val it = "abc" : string
```

パターンは、`val` 宣言の左辺に使うこともできる。

```
- val (a, b) = (123, "abc");
val a = 123 : int
```

```
val b = "abc" : string
```

`case` 式を用いて、式の中でパターンマッチを行うこともできる。

```
- fun length xs =  
  case xs of nil => 0  
           | x::xs => 1 + length xs;  
val length = fn : 'a list -> int
```

if 式

パターンではなく論理式によって場合分けする場合には、`if` 式を用いる。あくまで式なので値を持つ。そのため、「`then` 式」「`else` 式」のどちらも省略できない。

```
- fun abs x = if x < 0 then ~x else x;  
val abs = fn : int -> int  
- abs ~5;  
val it = 5 : int
```

相互再帰関数

次のプログラムでは、引数が偶数のときに真を返す関数 `even` と、引数が奇数のときに真を返す `odd` を定義している。

```
- fun even 0 = true  
  | even n = odd (n - 1)  
  and odd 0 = false  
  | odd n = even (n - 1);  
val even = fn : int -> bool  
val odd = fn : int -> bool  
- even 2;  
val it = true : bool
```

相互再帰的な関数は、互いが互いを参照し合えるように、1 つの `fun` 宣言の中に記述する必要がある。そのために、この例のように `and` キーワードを用いる。

高階関数

関数を引数や戻り値に持つ関数は**高階関数**と呼ばれる。以下に示す `map` 関数が、標準で定義されている高階関数の代表的な例である。

```
- fun square x = x * x;  
val square = fn : int -> int
```



```
- map square [1, 2, 3, 4, 5];  
val it = [1, 4, 9, 16, 25] : int list
```

map 関数の定義は、例えば以下のように書くことができる。

```
fun map f nil = nil  
  | map f (x::xs) = f x :: map f xs
```

関数式

関数を表す式は**関数式**と呼ばれ、しばしばラムダ抽象 (lambda abstraction), 匿名関数 (anonymous function) とも呼ばれる。

以下の例は、関数式を用いて square 関数を定義したものである。

```
- val square = fn x => x * x;  
val square = fn : int -> int  
- square 5;  
val it = 25 : int
```

関数式を用いると、map square [1, 2, 3, 4, 5] と等価な式を次のように書くこともできる。

```
- map (fn x => x * x) [1, 2, 3, 4, 5];  
val it = [1, 4, 9, 16, 25] : int list
```

階乗を求める関数 fact は、関数式を用いると容易に実装できる。再帰関数は関数式の中で左辺の名前を参照するため、次のように rec キーワードが必要になる。

```
- val rec fact = fn 0 => 1 | n => n * fact (n - 1);  
val fact = fn : int -> int  
- fact 5  
val it = 120 : int
```

部分適用

```
- fun add x y = x + y;  
val add = fn : int -> int -> int  
- val addTwo = add 2;  
val addTwo = fn : int -> int  
- addTwo 3;  
val it = 5 : int
```

関数適用は左結合であるため、例えば式 `add 2 3` は `(add 2) 3` のように評価される．ここで `add 2` は `add` 関数の**部分適用** (partial application) と呼ばれ、`int -> int` 型を持つ．`addTwo` 関数は、`add` 関数の部分適用を利用して定義されている．

関数の部分適用を考慮すると、`add` 関数の型は `int -> (int -> int)`、すなわち、引数が `int` 型で、戻り値が `int -> int` 型である関数の型と見ることができる．

今回の `add` 関数のように部分適用の可能な関数は、カーリー化された関数 (curried function) という．逆に、次に示す `add'` 関数のような関数は、非カーリー化された関数という．

```
- fun add' (x, y) = x + y;
val add' = fn : int * int -> int
- add' (2, 3);
val it = 5 : int
```

中置演算子

ここでは `pl` という中置演算子を宣言する．演算子名は、関数名として使えるものなら何でも構わない．次のように書くと、`pl` が優先順位 6 の (左結合の) 中置演算子として宣言される．

```
- infix 6 pl;
infix 6 pl
```

右結合の中置演算子を宣言するには、`infix` の代わりに `infixr` キーワードを用いる．中置演算子の定義は、普通の関数と同様に `fun` 宣言により記述する．例えば、中置演算子 `pl` の定義を次のように与えることが可能である．

```
- fun x pl y = x + y;
val pl = fn : int * int -> int
- 2 pl 3;
val it = 5 : int
```

中置演算子の直前に `op` キーワードを置くと、中置演算子を一時的に通常関数として扱うことができる．

```
- op pl (2, 3)
val it = 5 : int
```

中置演算子の結合の優先順位は 0 から 9 までの数字をとる (9が最も強い) ．トップレベルでは、以下のように演算子が宣言されている．

```
infix 0 before
infix 3 o :=
infix 4 = <> > < >= <=
```

```
infix 5 :: @
infix 6 :: + - ^
infix 7 * / div mod
```

3.6 リスト操作関数，リストの畳み込み

ML のリスト

ML では，複数のデータを扱う際には，配列よりもリストがよく使われる．リストを記述する方法は 2 ある．以下に示す 2 つの式は同一のリストを表している．

```
1 :: 2 :: 3 :: nil
[1, 2, 3]
```

`::` はしばしば **cons** と呼ばれ，第 1 引数に先頭要素，第 2 引数に後続のリストをとる．`::` は右結合であるから，`1 :: 2 :: 3 :: nil` の結合は `1 :: (2 :: (3 :: nil))` のようになる．

ML で扱うリストは，同一の型の要素のみで構成されるリストである．つまり，`[1, 1.0, #"a"]` のようなリストは許されない．型が `'a` である要素から構成されるリストの型を `'a list` と書く．

リストを扱う関数

Standard ML には，初めからリストを扱う関数が用意されている．

```
- [1, 2] @ [3, 4, 5]; (* リストの連結 *)
val it = [1, 2, 3, 4, 5] : int list
- List.filter (fn x => x mod 2 = 1) [1, 2, 3, 4, 5]; (* filter p xs : 述語
p が真を返す要素だけを残す *)
val it = [1, 3, 5] : int list
```

リストの畳み込み

`foldr` は，二項演算子を使って右からリストを畳み込む関数である．二項演算子を `+` とすると，式 `foldr + 0 [1, 2, 3]` は `1+(2+(3+0))` と展開される．

```
- foldr (op +) 0 [1, 2, 3]; (* 3 + (2 + (1 + 0)) = 6 *)
val it = 6 : int
- foldr (op -) 0 [1, 2, 3]; (* 3 - (2 - (1 - 0)) = 2 *)
val it = 2 : int
```

畳み込み関数は，リスト操作関数の実装によく用いられる．例えば，`map` 関数は `foldr` を使って以下のように書くこともできる．

```
- fun map' f = foldr (fn (x, s) => f x :: s) nil;
val map' = fn : ('a -> 'b) -> 'a list -> 'b list
- map' (fn x => x * x) [1, 2, 3, 4, 5];
val it = [1, 4, 9, 16, 25] : int list
```

3.7 データ型の定義

datatype 宣言

次のプログラムは、長方形の大きさを表す `rectangle` 型と、長方形の面積を求める `area` 関数を定義したものである。

```
- datatype rectangle = RECT of real * real;
datatype rectangle = RECT of real * real
- fun area (RECT (w, h)) = w * h;
val area = fn : rectangle -> real
- val r = RECT (3.0, 4.0);
val r = RECT (3.0, 4.0) : rectangle
- area r;
val it = 12.0 : real
```

datatype キーワードで始まる宣言により、新しいユーザー定義型を定義できる。**datatype** 宣言で定義されるデータ型は、**代数的データ型** (algebraic data type) と呼ばれる。

この例で定義されている `rectangle` は**型構成子** (type constructor), `RECT` は**値構成子** (value constructor) と呼ばれる。型構成子と値構成子には、同じ名前を用いることもできる。

一つの型に複数の値構成子を定義することもできる。次のプログラムは、円または長方形を表す `shape` 型を定義したものである。

```
- datatype shape = CIRC of real
                | RECT of real * real
datatype shape = CIRC of real | RECT of real * real
- fun area (CIRC r) = 3.14 * r * r
    | area (RECT (w,h)) = w * h;
val area = fn : shape -> real
- val c = CIRC 2.0;
val c = CIRC 2.0 : shape
- area c;
val it = 12.56 : real
```

次のプログラムでは、曜日を表すデータ型 `dayOfWeek` を宣言したものである。このように、引数のない値構成子のみからなるデータ型は、特に**列挙型** (enumerated type) と呼ばれる。

```

- datatype dayOfWeek = SUM | MON | TUE | THU | FRI | SAT;
datatype dayOfWeek = FRI | MON | SAT | SUN | THU | TUE | WED
- fun isHoliday SAT = true
  | isHoliday SUN = true
  | isHoliday _ = false;
val isHoliday = fn : dayOfWeek -> bool
- isHoliday MON;
val it = false : bool

```

次のプログラムは、リスト型を模した型を宣言したものである。このように、型引数を導入して多相型を構成することもできる。

```

- datatype 'a list' = NIL | CONS of 'a * 'a list';
datatype 'a list' = CONS of 'a * 'a list' | NIL

- fun length' NIL = 0
  | length' (CONS (x, xs)) = 1 + length' xs;
val length' = fn : 'a list' -> int

- length' (CONS(1, CONS(2, CONS(3, NIL)))); (* length [1,2,3] 相当 *)
val it = 3 : int

```

型シノニム

存在するデータ型に別名を与えるには、以下のように `type` 宣言を用いる。

```

- type 'a collection = 'a list;
type 'a collection = 'a list

```

抽象データ型

値構成子が隠蔽された型を**抽象データ型** (abstract data type) という。これを用いて、ユーザーに対してデータへのアクセス手段のみを提供することができる。

次のプログラムは、スタックを表すデータ型 `'a stack` と、スタックを操作する関数群を宣言したものである。このプログラムでは、`'a stack` を抽象データ型として宣言している。

3.8 命令型言語の機能

3.9 モジュールシステム

4. コンパイラの構成

4.1 字句解析器

4.2 構文解析器

5. 後期の活動について

6. 参考文献

- Andrew W. Appel 著, 神林 靖 監修/編集, 滝本 宗宏 編集 『最新コンパイラ構成技法』
- 『ウォークスルー Standard ML』 <http://walk.northcol.org/sml/>
- 『お気楽 Standard ML of New Jersey 入門』
http://www.geocities.jp/m_hiroi/func/index.html#sml