

## Analyse Syntaxique et Compilation, TP n° 4 : Menhir

Le manuel de Menhir se trouve ici : <http://gallium.inria.fr/~fpottier/menhir/manual.pdf>

Pour effectuer votre TP, il faut récupérer le répertoire `/ens/habermeh/Compil/tp4`. Pour chaque exercice, vous pourrez prendre les fichiers de ce répertoire comme point de départ pour travailler.

### Exercice 1 : Première spécification de grammaire Menhir.

Soit l'alphabet  $T \equiv \{\mathbf{true}, \mathbf{false}, \mathbf{id}, \vee, \wedge, (, )\}$ . Voici une grammaire décrivant la syntaxe des expressions booléennes et utilisant  $T$  pour ensemble de terminaux :

$$\begin{array}{l}
 e ::= \mathbf{id} \\
 \quad | \quad \mathbf{true} \\
 \quad | \quad \mathbf{false} \\
 \quad | \quad e \vee e \\
 \quad | \quad e \wedge e \\
 \quad | \quad (e)
 \end{array}$$

Indication : Essayer toujours d'abord de faire une partie simple de la grammaire (par exemple au début uniquement **true** et **false**) et étendre un par un ensuite.

1. Spécifier l'ensemble des terminaux à l'aide de la directive `%token` dans le fichier `parser.mly`.
2. Modifier le fichier `lexer.mll` pour reconnaître cet ensemble de non terminaux lors de l'analyse lexicale. Pour  $\wedge$ , on prendra par exemple `"/\`". Pour représenter les identificateurs de variables (terminal **id**), on pourra prendre par exemple le langage `['a'-'z']+`. N'oubliez pas d'ignorer les espaces et les retours à la ligne.
3. Modifier le fichier `parser.mly` en mimant la définition ci-dessus.
4. Lire attentivement le fichier `parser.conflicts` généré par `menhir`. Affecter des priorités raisonnables aux règles de production pour résoudre ces conflits. Pour cela, utiliser les directives `%left`, `%right` ou `%nonassoc` (voir la documentation pour comprendre leur fonctionnement).

### Exercice 2 : Un conflit classique.

On ajoute à  $T$  les terminaux `{if, then, else, =}`. On définit une syntaxe pour les instructions d'un langage très simple :

$$\begin{array}{l}
 i ::= \mathbf{if} \ e \ \mathbf{then} \ i \\
 \quad | \quad \mathbf{if} \ e \ \mathbf{then} \ i \ \mathbf{else} \ i \\
 \quad | \quad (i) \\
 \quad | \quad \mathbf{id} = e
 \end{array}$$

1. Modifier le fichier `lexer.mll` pour qu'il tienne compte des nouveaux terminaux.
2. Modifier aussi les fichiers `ast.ml` et `main.ml` convenablement.
3. Rajouter la définition de `i` dans le fichier `parser.mly`.
4. Comprendre le conflit détecté par `menhir` en donnant un exemple d'entrée qui peut être analysée de deux manières différentes.
5. Implémenter deux spécifications de grammaire correspondant à ces deux variantes.

**Exercice 3 : Une grammaire LR(2).**

On modifie ainsi la spécification de l'exercice précédent :

$$\begin{array}{l} i \quad ::= \dots \\ \quad \quad | \quad ma \\ ma \quad ::= a \\ \quad \quad | \quad a \wedge ma \\ a \quad ::= \mathbf{id} = e \end{array}$$

1. Faire les modifications nécessaires dans les fichiers (en particulier le fichier `parser.mly`) pour implémenter telle quelle cette définition.
2. Observer le conflit généré par `menhir`.
3. Modifier la structure de la grammaire pour qu'elle définisse le même langage mais soit LR(1) et donc adaptée à `menhir`.
4. Revenir à la grammaire initiale et utiliser le mot-clé `%inline` pour obtenir le même résultat que celui de la question 2.