Module Langages Formels
TD 5 : Caractérisation logique et Automates Séquentiels

Exercice 1 `Caractérisation logique de Rat(Σ)`

Le but de cet exercice est d'illustrer le théorème de Büchi, qui donne une caractérisation logique de l'ensemble des langages reconnaissables :

> **Théorème (Büchi 1960, Elgot 1961)** :
> Un langage de mots finis est reconnaissable par automate fini si et seulement si il est définissable en logique MSO[S] (Monadique du second ordre avec successeur).

Une logique du premier ordre se construit avec
- Un ensemble de variables : $x, y \ldots$
- Des symboles de relations (unaires ou binaires dans notre cas) : $S, <, = \ldots$
- Des connecteurs logiques : $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$
- Des quantificateurs : $\forall, \exists$

Pour une logique du second ordre, on ajoute
- Des variables du second ordre : $X, Y \ldots$
- L'application d'une variable du premier ordre à une variable du second ordre (en tant que formule atomique) : $X(x)$

On parlera de logique monadique si on ne quantifie que des symboles de relation unaire.

On utilisera par la suite les variables x et y pour dénoter la position d'une lettre dans un mot. S correspondra à la relation successeur sur les entiers ($S : \begin{array}{l}[0,n] \to [0,n] \\ 0 \leq k < n \mapsto k+1\end{array}$, et $Q_a(x)$ signifiera que la lettre a est à la position x.

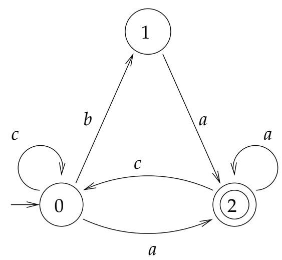
$\underline{w}$ est un modèle d'une formule si on connaît : $\mathsf{dom}(w),\ S_w,\ (Q_a)_{a\in\Sigma}$. (le mot w est considéré comme une fonction de $[1, n]$ dans Σ). Le langage décrit par une formule est **l'ensemble des modèles $\underline{w}$ vérifiant cette formule**.

Par exemple, la formule logique suivante signifie que "toute occurence de b est suivie d'une occurence de a" :

$$\forall x(Q_b(x) \Rightarrow (\exists y,\ Sxy \wedge Q_a(y)))$$

On considère l'automate suivant, dont le langage reconnu est tel que
- toute occurence de b est suivie d'une occurence de a ;

– aucune occurence de b ne suit immédiatement une occurence de a ;
– tout mot reconnu se termine par une occurence de a.



1.1. Construire la formule logique correspondant au langage reconnu par cet automate.

1.2. En supposant qu'il existe un prédicat X exprimant "*x est en position impaire*", donner la formule qui dénote l'ensemble des mots de longueur paire.

$$\exists X \left(\begin{array}{cc} & (\dots) \\ \wedge & (\dots) \\ \wedge & (\dots) \end{array} \right)$$

Soit L un langage reconnu par un automate fini $A = (Q, \Sigma, \delta, q_0, F)$, avec $n = |Q|$. On veut construire la formule logique dénotant L. Pour chaque état q_i, on supposera l'existence d'un prédicat X_i, $X_i(x)$ signifiant "*avant de lire la x^e lettre, l'automate est dans l'état q_i*".

1.3. Proposer une formule (1) signifiant que, lors de la lecture d'un mot, l'automate est toujours dans un et un seul état.

1.4. Proposer une formule (2) signifiant que, lors de la lecture d'un mot, l'automate commence toujours en q_0 (on notera $\mathsf{min}(x)$ la formule $\neg\exists z,\ Szx$).

1.5. Proposer une formule (3) signifiant que, lors de la lecture d'un mot chaque transition est une transition valide de l'automate.

1.6. Proposer une formule (4) signifiant que la lecture d'un mot de L termine toujours par un état final (on notera $\mathsf{max}(x)$ la formule $\neg\exists z,\ Sxz$).

La formule suivante caractérise donc exactement les mots du langage L :

$$\exists X_1 \exists X_2 \dots \exists X_n \left(\begin{array}{cc} & (1) \\ \wedge & (2) \\ \wedge & (3) \\ \wedge & (4) \end{array} \right)$$

Exercice 2 `Automates séquentiels`

Définition:
Soit A un alphabet d'entrée et B un alphabet de sortie. Un automate *séquentiel* (gauche) est un automate fini à 2 bandes $\mathcal{A} = (Q, A \times B^*, i, Q, \delta)$ tel que :

– la projection sur la bande d'entrée ((Q, A, i, Q, δ)) est un automate fini déterministe ;
– tout état est terminal.

Les étiquettes des flèches d'automate séquentiel sont des couples de $A \times B^*$. Si partant d'un état q on lit la lettre $a \in A$ sur l'entrée pour arriver dans l'état p, on écrit en sortie $v \in B^*$, tel que $q \overset{(a,v)}{\rightarrow} p$.

La relation de mots (finis) $R \subseteq A^* \times B^*$ est **réalisée par** $\mathcal{A}$ si R est l'ensemble des étiquettes des chemins finis commançant dans l'état i. En d'autres termes, un couple (f, g) est reconnu par $\mathcal{A}$ si il existe un état q tq $i \overset{(f,g)}{\rightarrow} q$.

La fonction φ est réalisable par un automate fini si la relation $(x, \varphi(x))$ est réalisable par automate fini.

Un automate séquentiel droit lit et écrit les mots de droite à gauche.

Un automate est *sous-séquentiel* si il existe une fonction *terminale* $w : Q \rightarrow B^*$. Un couple (f, g) est reconnu par $\mathcal{A}$ s'il existe un chemin

$$i \overset{f/g'}{\rightarrow} q$$

tel que $g = g'g''$ et $w(q) = g''$.

2.1. Donner un automate séquentiel qui efface les 0 en tête des mots (sur $\{0, 1\}$).

2.2.

1. Donner un automate séquentiel droit qui ajoute 1 à un entier en base 2 (pour l'alphabet d'entrée, prendre $\{0, 1, 2\}$ avec 2 qui marque la fin de l'entier).
2. Donner un automate sous-séquentiel droit qui fait de même avec cette fois-ci $\{0, 1\}$ pour alphabet d'entrée.

2.3.

1. Donner un automate séquentiel droit qui additionne deux entiers en base 2 (pour l'alphabet d'entrée, prendre $\{0, 1, 2\}^2$ -> le couple $(10, 3)$ s'écrira $(21010, 22211)$ et correspondra au mot $(2, 2)(1, 2)(0, 2)(1, 1)(0, 1)$).
2. Donner un automate sous-séquentiel droit qui fait de même avec cette fois-ci $\{0, 1\}^2$ pour alphabet d'entrée.

2.4.

1. Donner un automate séquentiel droit pour la soustraction en base 2 (pour l'alphabet d'entrée, prendre $\{2, 0, 1\}^2$).
2. Donner un automate sous-séquentiel droit qui fait de même avec cette fois-ci $\{0, 1\}^2$ pour alphabet d'entrée.

2.5. Donner un automate déterministe reconnaissant les mots finissant par le motif *abbab*, puis un automate séquentiel gauche supprimant toutes les occurrences de *abbab* dans un mot.