Option informatique 2^{e} année

Lycée Louis-le-Grand, Paris

Corrigé

1 Exercice de logique

Question 1
L'affirmation de A peut s'exprimer sous la forme $(AV \Rightarrow BV) \wedge (AF \Rightarrow \neg BV)$; quant à l'affirmation de B elle peut s'exprimer sous la forme $(BV \Rightarrow \neg AV) \wedge (BF \Rightarrow AV)$. Il faut ajouter que A ne peut pas être à la fois un Pur et un Pire, de même que B. Autrement dit l'énoncé peut s'exprimer par la formule logique

$$(AV \Rightarrow BV) \wedge (AF \Rightarrow \neg BV) \wedge (BV \Rightarrow \neg AV) \wedge (BF \Rightarrow AV) \wedge \neg(AV \wedge AF) \wedge \neg(BV \wedge BF)$$

Question 2
La formule est logiquement équivalente à la formule obtenue en remplaçant toutes les occurrences de $p \Rightarrow q$ par $q \vee \neg p$. On peut au passage, pour préparer la suite, utiliser les lois de Morgan pour transformer $\neg(p \wedge q)$ en la formule équivalente $(\neg p \vee \neg q)$. La formule ci dessus est donc équivalente à la formule

$$(\neg AV \vee BV) \wedge (\neg AF \vee \neg BV) \wedge (\neg AV \vee \neg BV) \wedge (AV \vee \neg BF) \wedge (\neg AV \vee \neg AF) \wedge (\neg BV \vee \neg BF)$$

Question 3
La première possibilité est d'utiliser la distributivité de $\wedge$ par rapport à $\vee$. En considérant le premier et le troisième terme de la formule ci dessus, on a

$$(\neg AV \vee BV) \wedge (\neg AV \vee \neg BV) \equiv \neg AV \vee (BV \wedge \neg BV) \equiv \neg AV \vee 0 \equiv \neg AV$$

On peut alors incorporer le cinquième terme :

$$\neg AV \wedge (\neg AV \vee \neg AF) \equiv \neg AV$$

et la formule est donc équivalente à

$$\neg AV \wedge (\neg AF \vee \neg BV) \wedge (AV \vee \neg BF) \wedge (\neg BV \vee \neg BF)$$

Mais $\neg AV \wedge (AV \vee \neg BF) \equiv \neg AV \wedge \neg BF$. D'où la formule équivalente

$$\neg AV \wedge \neg BF \wedge (\neg AF \vee \neg BV) \wedge (\neg BV \vee \neg BF) \equiv \neg AV \wedge \neg BF \wedge (\neg AF \vee \neg BV)$$

qui est équivalente à la disjonction de conjonctions

$$(\neg AV \wedge \neg BF \wedge \neg AF) \vee (\neg AV \wedge \neg BF \wedge \neg BV)$$

autrement dit, soit A est un Versatile et B n'est pas un Pire, soit A n'est pas un Pur et B est un Versatile ; on peut alors simplifier en disant que soit A est un Versatile et B un Pur, soit A et B sont tous deux des Versatiles, soit A est un Pire et B un Versatile.
On peut alors obtenir la forme normale disjonctive en introduisant les variables manquantes :

$$(\neg AV \wedge \neg BF \wedge \neg AF) \equiv (\neg AV \wedge \neg BF \wedge \neg AF \wedge BV) \vee (\neg AV \wedge \neg BF \wedge \neg AF \wedge \neg BV)$$

$$(\neg AV \wedge \neg BF \wedge \neg BV) \equiv (\neg AV \wedge \neg BF \wedge \neg BV \wedge AF) \vee (\neg AV \wedge \neg BF \wedge \neg BV \wedge \neg AF)$$

et, comme on a deux fois la même conjonction de littéraux, l'expression équivalente

$$(\neg AV \wedge \neg BF \wedge \neg AF \wedge BV) \vee (\neg AV \wedge \neg BF \wedge \neg AF \wedge \neg BV) \vee (\neg AV \wedge \neg BF \wedge \neg BV \wedge AF)$$

On peut également parvenir à la forme normale à l'aide de la table de vérité

$$\begin{bmatrix} AV & AF & BV & BF & \text{expression} \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

qui permet de retrouver la forme normale disjonctive

$$(\neg AV \wedge AF \wedge \neg BV \wedge \neg BF) \vee (\neg AV \wedge \neg AF \wedge BV \wedge \neg BF) \vee (\neg AV \wedge \neg AF \wedge \neg BV \wedge \neg BF).$$

Remarquons qu'on peut condenser cette formule en :

$$(\neg AV \wedge AF \wedge \neg BV \wedge \neg BF) \vee (\neg AV \wedge \neg AF \wedge \neg BF),$$

mais il ne s'agit plus, *stricto sensu*, d'une forme normale disjonctive.

Question 4
Si A est un Versatile et B un Pur, A a dit la vérité et n'est pourtant pas un Pur. Sinon, B est un Versatile et A n'est pas un Pur, si bien que B a dit la vérité tout en n'étant pas un Pur.

2 Problème sur les langages rationnels

2.1 Morphismes

Question 5
Comme $\varphi(\varepsilon) = \varphi(\varepsilon)\varphi(\varepsilon)$, on a nécessairement $\varphi(\varepsilon) = \varepsilon$.

Question 6
Considérant un mot $u = u_0 \dots u_{n-1}$ de $(n-1)$ lettres, on a : $\varphi(u) = \varphi(u_0) \dots \varphi(u_{n-1})$ donc φ est déterminé de façon unique par les images des caractères.
Inversement, une telle application φ est clairement un morphisme.

Question 7
On a $\varphi(abaab) = bbccdbbbcbbccdb$.

Question 8
L'automate considéré étant complet, on peut utiliser, quand $(q, u) \in Q \times \mathcal{B}^\star$, les notations $\delta^\star(q, u)$ pour désigner l'état où se retrouve l'automate, quand depuis l'état q on lit un mot u.
Soit alors l'automate fini déterministe défini (Q, q_0, F, Δ) défini par $\Delta(q, c) = \delta^\star(q, \varphi(c))$ pour tout état $q \in Q$ et tout caractère $c \in \mathcal{A}$. Cet automate est complet par construction.
Une récurrence immédiate (sur la longueur de u) prouve que, φ étant un morphisme, $\Delta^\star(q, u) = \delta^\star(q, \varphi(u))$.
Alors $\Delta^\star(q_0, u) \in F \Leftrightarrow \delta^\star(q_0, \varphi(u)) \in F$, c'est-à-dire que le nouvel automate reconnaît u si et seulement si l'ancien reconnaît $\varphi(u)$, ou encore $u \in \varphi^{-1}(M)$.

Question 9
Soit donc un nouvel automate fini $(Q', (q_0, 0, *), F', \Delta)$, où $*$ est une nouvelle lettre qu'on ajoute à l'alphabet $\mathcal{A}$, avec $Q' = Q \times \{0, 1, \dots, n-1\} \times \mathcal{A}$, $F' = F \times \{0\} \times \{*\}$ et où la fonction de transition Δ est définie de la façon suivante : pour tout état $q \in Q$ et tout caractère $c \in \mathcal{A}$, écrivons $\varphi(c) = u_1 \dots u_k \in \mathcal{B}^\star$ avec $k \leqslant n$, et ajoutons les transitions

$$(q, 0, *) \xrightarrow{\varepsilon} (q, 0, c) \xrightarrow{u_1} (q, 1, c) \xrightarrow{u_2} (q, 2, c) \cdots (q, k-2, c) \xrightarrow{u_{k-1}} (q, k-1, c) \xrightarrow{u_k} (\delta(q, c), 0, *).$$

On associe ainsi à chaque transition étiquetée par un caractère c de l'automate qui reconnaît L un chemin étiqueté par les caractères consécutifs de $\varphi(c)$.
Notons que le nouvel automate n'est pas complet ni même déterministe.
Par construction même, on dispose de $\delta(q, c) = q' \Rightarrow (q', 0, *) \in \Delta^*((q, 0, *), \varphi(c))$. On en déduit d'abord que, pour tout mot $u \in \mathcal{B}^\star$, $\delta^\star(q, u) = q' \Rightarrow (q', 0, *) \in \Delta^*((q, 0, *), \varphi(u))$, puis que tout mot de $\varphi(L)$ est bien accepté par le nouvel automate : son état initial est $(q_0, 0, *)$ et ses états finals $(q_f, 0, *)$ avec $q_F \in F$.
Réciproquement, tout chemin du nouvel automate qui part de $(q_0, 0, *)$ et qui aboutit à un état (q, j, c) avec $j \geqslant 1$ passe nécessairement par l'état $(q, 0, *)$ et il n'y a qu'un chemin qui passe de $(q, 0, *)$ à un tel (q, j, c). D'autre part les mots u qui étiquettent $(q, 0, *) \xrightarrow{u} (q', 0, *)$ sont exactement les mots $u = \varphi(c)$ tels que $\delta^\star(q, c) = q'$.
Relisant un chemin réussi de $(q_0, 0, *)$ à $(q_f, 0, *)$ (où q_f est final), on repère les passages aux états de la forme $(q, 0, *)$, avec $q \in Q$: si w est un mot reconnu par notre nouvel automate, il existe une décomposition $w = w_1 \dots w_p$ où les w_k sont des **mots** sur $\mathcal{B}$ telle que $(q_0, 0, *) \xrightarrow{w_1} (q_1, 0, *) \cdots (q_{p-1}, 0, *) \xrightarrow{w_p} (q_f, 0, *)$.
Mais d'après notre remarque précédente chaque mot w_k est l'image par φ d'un caractère $c_k \in \mathcal{A}$ tel que $q_k = \delta(q_{k-1}, c_k)$.
Alors on a bien $w = \varphi(c_1 \dots c_p) \in \varphi(L)$ car $\delta^\star(q_0, c_1 \dots c_p) = \delta^\star(q_1, c_2 \dots c_p) = \delta(q_{p-1}, c_p) = q_f \in F$.

2.2 Exemples d'applications

Question 10
On sait que le langage $M = \{u, |u|_a = |u|_b\}$ n'est pas rationnel.
L est l'ensemble des mots contenant autant de a que de b mais sans les facteurs aaa ni bbb.
Remarquons que pour tout mot u, $|\varphi(u)|_a - |\varphi(u)|_b = |u|_a - |u|_b$ puisque $|aab|_a - |aab|_b = 1$ et pareillement $|abb|_a - |abb|_b = -1$.
Si donc $u \in L$, et si $\varphi(v) = u$, alors $|\varphi(v)|_a - |\varphi(v)|_b = |v|_a - |v|_b = 0$ donc $v \in M$: $\varphi^{-1}(L) \subset M$.
En outre, comme $\varphi(a)$ et $\varphi(b)$ commencent tous deux par a et finissent tous deux par b, aucune image $\varphi(u)$ d'un mot u quelconque ne contient les facteurs aaa ni bbb, donc $\varphi(M) \subset L$ et par conséquent $M \subset \varphi^{-1}(L)$.
Finalement $M = \varphi^{-1}(L)$ et si L était rationnel, M aussi : L n'est pas rationnel.

Question 11
On utilise l'automate produit, avec comme états finals les couples constitués de deux états finals.

Question 12
Posons $\mathcal{C} = \{\bar{a}, \bar{b}\}$.
On vérifiera qu'on retrouve les sous-mots en posant $K = \mathcal{A}^\star$ qui est rationnel ; les facteurs en posant $K = \mathcal{C}^\star.\mathcal{B}^\star.\mathcal{C}^\star$; les facteurs gauches en posant $K = \mathcal{B}^\star.\mathcal{C}^\star$ et les facteurs droits en posant $K = \mathcal{C}^\star.\mathcal{B}^\star$.
Intuitivement, on autorise l'effacement des lettres surmontées d'une barre.

Question 13
On garde les morphismes précédents, et on choisit cette fois $K = \mathcal{B}^\star.\bar{a}.\bar{b}.\mathcal{B}^\star$.

2.3 Mots sans carrés

Question 14
Cherchons à construire un mot sans facteur carré sur l'alphabet à deux lettres.
Supposons que le premier caractère est un a. Pour éviter le carré aa, le deuxième caractère doit être un b, puis pour éviter le carré bb le troisième caractère doit être un a. Mais alors, soit le quatrième caractère est un a et on a le facteur aa, soit c'est un b et le mot est $abab$ qui est un carré. Tout mot de longueur supérieur ou égal à 4 a donc un facteur carré.

Question 15
On a $T_1 = ab$, $T_2 = abba$, $T_3 = abbabaab$, $T_4 = abbabaabbaababba$. Le nombre de caractères de T_n double à chaque étape donc $|T_n| = 2^n$.
Montrons par récurrence que T_n est un préfixe de T_{n+1}.
C'est vrai si $n = 0$, et si c'est vrai pour $n-1$, on a $T_n = T_{n-1}R_{n-1}$, d'où en appliquant θ, $T_{n+1} = \theta(T_n) = \theta(T_{n-1})\theta(R_{n-1}) = T_nR_n$, ce qui montre que T_n est un préfixe de T_{n+1}.
NDLR : ceci permet de parler de la réunion (croissante) des T_n comme la suite infinie de Thue-Morse, qui est l'unique point fixe de la substitution θ ; en dehors des propriétés étudiées ici, elle a de nombreuses propriétés remarquables.

Question 16
On utilise un programme récursif : si $n = 0$ la réponse est évidente, sinon on applique la transformation θ au mot T_{n-1}. La transformation θ est-elle même écrite sous forme récursive, et a un coût linéaire en la longueur du mot qu'on lui soumet.

```
let rec thue n =
    let rec theta = function
       | `a` :: q -> `a` :: `b` :: (theta q)
       | `b` :: q -> `b` :: `a` :: (theta q)
       | [] -> []
    in match n with
       | 0 -> [ `a` ]
       | n -> theta (thue (n-1)) ;;
```

Question 17
Supposons que T_n contienne un facteur aaa.
Il apparaît soit à une place paire sous la forme $aaax$ (on numérote les caractères à partir de 0), mais ceci est impossible car on devrait avoir $aaax = \theta(uv)$ pour un facteur uv de T_{n-1}, soit $aa = \theta(u)$ qui est impossible ; s'il apparaît à une place impaire, on aurait $xaaa = \theta(uv)$, soit $aa = \theta(v)$ ce qui est encore impossible. Le raisonnement est le même avec le facteur bbb.
Donc $T_n \notin \mathcal{A}^*\{aaa, bbb\}\mathcal{A}^*$.

Question 18
Supposons donc que T_n contient un mot de la forme $uvuvu$ où u n'est pas le mot vide. Considérons un tel facteur de longueur minimale, apparaissant pour la première fois à la position p dans T_n.

a) Soit v n'est pas le mot vide et $|uvuvu| = 3|u| + 2|v| \geqslant 3 + 2 = 5$, soit v est le mot vide, mais comme T_n ne contient ni le mot aaa ni le mot bbb, on a $|u| \geqslant 2$, soit $|uvuvu| \geqslant 6$.

b) Soit w un facteur de T_n de longueur au moins 5. Il apparait comme facteur de longueur 5 dans un facteur de longueur 8 de T_n de la forme $\theta^3(x)$ pour un caractère x de T_{n-3}. Mais, si $x = a$, $\theta^3(x) = T_3 = abbabaab$ dont les 4 facteurs de longueur 5 contiennent soit aa, soit bb. Et si $x = b$, alors $\theta^3(x) = baababba$, mot miroir de $\theta^3(a) = T_3$ qui a la même propriété.
Donc tout facteur de longueur 5, et en particulier $uvuvu$ contient soit le facteur aa soit le facteur bb.
Comme $uvuvu = (uvu)vu = uv(uvu)$, soit ce facteur aa apparaît dans le premier uvu, soit dans le deuxième facteur uv (donc dans uvu), soit le facteur aa ou bb est coupé par la première parenthèse fermante, mais comme u n'est pas vide, il n'est pas coupé par la deuxième parenthèse ouvrante.

Dans tous les cas ce facteur est un facteur de uvu.
c) Les mots aa et bb n'apparaissent qu'à des positions impaires dans T_n (les facteurs de longueur 2 à des positions paires sont des images de caractères de T_{n-1}). Il s'ensuit que les positions des occurrences de uvu dans le mot T_n ont toutes la même parité, en particulier p et $p + |uv|$ ont la même parité et donc $|uv|$ est paire.
d) Soit u' le mot formé par les lettres de u qui ont une position paire dans T_n et v' construit de manière similaire à partir de v. Alors T_{n-1} contient le mot $u'v'u'v'u'$ de longueur strictement inférieure à celle de $uvuvu$ ce qui contredit la minimalité de $uvuvu$, à moins que u' ne soit le mot vide, autrement dit que u soit réduit à un caractère apparaissant à une position impaire. Mais dans ce cas $uvuvu = xvxvx$ (où v n'est pas le mot vide puisque T_n ne contient ni aaa, ni bbb) ; soit alors v'' le mot formé par les lettres de v apparaissant à des positions impaires ; alors T_{n_1} contient le mot $xv''xv''x$ ce qui contredit encore la minimalité de $uvuvu$.
On est ainsi parvenu à la contradiction demandée.

Question 19
On a
$$U_1 = abc, U_2 = abcacb, U_3 = abcacbabcbac, U_4 = abcacbabcbacabcacbacabcb$$
Soit a_n, b_n et c_n le nombre d'occurences de a, b ou c dans le mot U_n.
On a $a_{n+1} = a_n + b_n$, $b_{n+1} = a_n + c_n$ et $c_{n+1} = a_n + b_n$.
Montrons par récurrence que $\forall n \geqslant 1$, $a_n = b_n = c_n$.
C'est vrai pour $n = 1$ et les formules précédentes montrent que si c'est vrai pour n, c'est vrai pour $n + 1$.
On en déduit que pour $n \geqslant 1$, $a_{n+1} = 2a_n$, soit $a_n = 2^{n-1}a_1 = 2^{n-1}$ et donc $|U_n| = 3 \times 2^{n-1}$.
La même démonstration par récurrence que pour T_n montre que U_n est un préfixe de U_{n+1}.

Question 20
On a $\theta \circ \pi(a) = \theta(abb) = abbaba$ et $\pi \circ \mu(a) = \pi(abc) = abbaba$. De même $\theta \circ \pi(b) = \theta(ab) = abba$ et $\pi \circ \mu(b) = \pi(ac) = abba$. On a donc $\theta \circ \pi = \pi \circ \mu$.
On en déduit par récurrence que $\forall n \in \mathbb{N}$, $\theta^n \circ \pi = \pi \circ \mu^n$, et en particulier $\theta^n \circ \pi(a) = \pi \circ \mu^n(a)$, soit encore $\theta^n(abb) = \pi(U_n)$. Mais $\theta^n(ab) = T_{n+1}$ et $\theta^n(abba) = T_{n+2}$.
Ceci montre que $\pi(U_n)$ admet comme préfixe T_{n+1} et est un préfixe de T_{n+2}, soit $T_{n+1} \preccurlyeq \pi(U_n) \preccurlyeq T_{n+2}$ (pour la relation d'ordre : est un préfixe de ...).
Supposons que U_n admette un facteur carré de la forme uu donc (quitte à remplacer U_n par U_{n+1}) un facteur de la forme uux pour un certain caractère x, alors $\pi(U_n)$ admet le facteur $\pi(u)\pi(u)\pi(x)$, donc il en est de même de T_{n+2}. Mais $\pi(u) = av$ pour un certain v et $\pi(x) = aw$ si bien que $\pi(u)\pi(u)\pi(x) = avavaw$, ce qui montre que T_{n+2} contient un facteur $avava$, ce qui est exclu par la question 18.
Donc U_n est sans carré.

Question 21
Supposons que le langage L soit reconnu par un automate déterministe (Q, p_0, F, δ) à n états et soit $u \in \{a, b, c\}^*$ de longueur supérieure à n et sans carré, ce qui est possible d'après les questions précédentes.
Considérons le chemin pour le mot udu : $p_0 \xrightarrow{u} p \xrightarrow{du} q$.
Comme $udu \notin L$, on a $q \notin F$.
Comme la longueur de u est supérieure à n, le chemin contient une boucle $p_0 \xrightarrow{u_1} p_1 \xrightarrow{u_2} p_1 \xrightarrow{u_3} p \xrightarrow{du} q$ avec $u_2 \neq \varepsilon$. Mais en supprimant cette boucle, on obtient le mot u_1u_3du qui lui est dans L puisque $u_1u_3 \neq u$, ce qui exige que $q \in F$, d'où la contradiction.
Le langage L ne peut donc pas être reconnu par un automate.
Par contre, on ne peut pas appliquer le lemme de pompage (encore appelé lemme de l'étoile), car pour toute décomposition d'un mot $udu \in L$ sous la forme $w_1w_2w_3$, il existe une décomposition de w_2 sous la forme $u_1u_2u_3$ telle que $w_1u_1u_2^nu_3w_3$ appartienne à L pour tout $n \in \mathbb{N}$.