

## Option Informatique en Spé MP et MP\*

### Autour du lemme de l'étoile : le corrigé

**Question 1** • La réponse est négative. Pour l'établir, utilisons le lemme de l'étoile. Supposons  $L_1$  rationnel. Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_1$ . Considérons alors la décomposition du mot  $u = a^N b^N$  : nécessairement,  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{N-i-j} b^N$ . Nous mettons en évidence une contradiction en notant que le mot  $uz = a^{N-j} b^N$  n'appartient pas à  $L_1$ .

• Autre méthode : nous allons montrer que  $L_1$  possède une infinité de résiduels. Il suffit de remarquer que le mot le plus court de  $(a^k)^{-1}L$  est  $b^k$  ; donc les résiduels de la forme  $(a^k)^{-1}L$  sont deux à deux distincts.

**Question 2** • Le langage  $L_2$  considéré est rationnel. Pour le prouver, nous pouvons remarquer que son complémentaire est  $(\Sigma^{2005})^*$  ; ou bien que  $L_2$  est la réunion des 2004 langages  $\Sigma^k(\Sigma^{2005})^*$ ,  $k \in [1, 2004]$ , lesquels sont tous rationnels.

**Question 3** • *Méthode 1* : il est bien connu que le langage  $M = \{a^n b^n \mid n \in \mathbb{N}\}$  n'est pas rationnel ; il en est donc de même de son complémentaire, qui est justement  $L_3$ .

• *Méthode 2* : exhibons une famille infinie de résiduels de  $L_3$  (il n'est pas nécessaire de les exhiber tous). Remarquons que  $(a^k)^{-1}L = \{a^q b^p : k + q \neq p\}$  donc  $a^{p-k} b^p \notin (a^k)^{-1}L_3$  quel que soit  $p \geq k$  ; en particulier,  $b^k \notin (a^k)^{-1}L_3$ . Par contre,  $b^j \in (a^k)^{-1}L$  quel que soit  $j \neq k$ . Donc  $((a^k)^{-1}L_3) \cap b^* = b^* \setminus \{b^k\}$ , ce qui montre que les résiduels  $(a^k)^{-1}L_3$  sont deux à deux distincts et par suite forment une famille infinie.

• *Méthode 3* : utilisons le lemme de l'étoile. Supposons  $L_3$  rationnel. Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_3$ . Considérons alors la décomposition du mot  $u = a^N b^{N!+N}$ , lequel appartient clairement à  $L_3$  : nécessairement,  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{N-i-j} b^{N!}$ . Il reste à exhiber  $k$  tel que  $xy^k z \notin L_3$ . Ceci se produit ssi  $N + kj = N! + N$ , ce qui est réalisé en prenant  $k = N!/j$  (licite car  $1 \leq j \leq N$ ).

**Question 4** • *Méthode 1* : utilisons le lemme de l'étoile. Supposons  $L_4$  rationnel. Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_4$ . Considérons alors la décomposition du mot  $u = a^{N+1} b^N$  : nécessairement,  $x = a^i$ ,  $y = a^j$  avec  $j > 0$  et  $z = a^{N+1-i-j} b^N$ . Nous mettons en évidence une contradiction en notant que le mot  $xz = a^{N+1-j} b^N$  n'appartient pas à  $L_4$ , puisque  $j > 0$  implique  $N + 1 - j < N + 1$ , soit  $N + 1 - j \leq N$ .

• *Méthode 2* : supposons  $L_4$  rationnel. Alors  $\widetilde{L}_4 = \{\widetilde{u} \mid u \in L_4\} = \{b^p a^n \mid p < n\}$  serait rationnel. Nous pouvons échanger les rôles des exposants  $a$  et  $b$ , et celui des lettres  $n$  et  $p$  : ainsi, le langage  $\widehat{L}_4 = \{a^n b^p \mid n < p\}$  serait lui aussi rationnel. Mais alors, le langage  $L_4 \cup \widehat{L}_4 = \{a^n b^p \mid n \neq p\}$  considéré à la question 3 serait rationnel : d'où la contradiction.

• *Méthode 3* : exhibons une famille infinie de résiduels de  $L_4$ . Pour  $k \geq 1$ , notons  $R_k = (a^k)^{-1}L_4$ . Le plus long mot de  $R_k$ , de la forme  $b^j$  est  $b^{k-1}$ . Ceci prouve que les  $R_k$  sont deux à deux distincts.

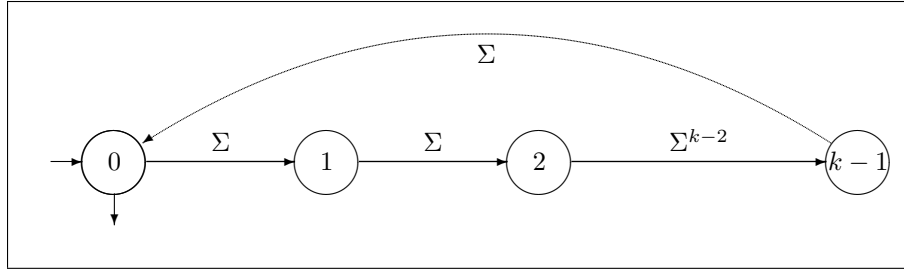
**Question 5** • Notons que  $L_\downarrow$  est la réunion d'une famille finie de langages rationnels :  $L_\downarrow = \bigcup_{k \leq 2005} L_\downarrow^k$ , avec

$L_\downarrow^k = \{a^n a^k b^k \mid n \in \mathbb{N}\}$ . Ainsi,  $L_\downarrow$  est rationnel.

**Question 6** • La réponse est négative. Pour l'établir, utilisons le lemme de l'étoile. Supposons  $L_\uparrow$  reconnu par un afdc  $\mathcal{A} = (Q, \delta, i, F)$ . Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_\uparrow$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_\uparrow$ . Notons  $M = \max(N, 2005)$  et considérons la décomposition du mot  $u = a^M b^M$  : nécessairement  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{M-i-j} b^M$ . Nous mettons en évidence une contradiction en notant que le mot  $xz = a^{M-j} b^M$  n'appartient pas à  $L_\uparrow$ .

• Autre méthode : le langage  $\{a^p b^q \mid p \geq q\}$  n'est pas rationnel (voir la question 4) ; or il est la réunion de  $L_\downarrow$  (qui est rationnel) et de  $L_\uparrow$ . Donc  $L_\uparrow$  n'est pas rationnel.

**Question 7** • L'afdc suivant reconnaît le langage  $L_7$  :



Nous avons représenté par une seule flèche la suite de transitions  $i \rightarrow i + 1$ ,  $i \in [2, k - 2]$ , toutes étiquetées par  $\Sigma$ . Avec la numérotation des états choisie, nous avons  $\delta^*(0, u) \equiv |u| \pmod{k}$ .

• Soit maintenant  $\mathcal{A} = (Q, \delta, i, F)$  un afdc ayant moins de  $k$  états; si  $\mathcal{A}$  reconnaît  $a^k$ , c'est qu'il existe un calcul d'étiquette  $a^k$  menant de l'état initial  $i$  à un état final  $q$ . Comme  $k$  est strictement supérieur au nombre d'états de  $\mathcal{A}$ , il existe (principe des tiroirs) des naturels  $j$  et  $j'$  tels que  $0 \leq j < j' < k$  et  $\delta^*(i, a^j) = \delta^*(i, a^{j'})$ . Alors  $\delta^*(i, a^{k+j'-j}) = \delta^*(i, a^k)$  appartient à  $F$ , donc  $a^{k+j'-j}$  est reconnu par  $\mathcal{A}$ . Or  $0 < j' - j < k$  et  $k + j' - j \equiv j' - j \pmod{k}$ ; donc  $a^{k+j'-j} \notin L_7$ .

**Question 8** • Soit  $u$  un mot de  $L_8$ , de longueur minimale. Soit  $\mathcal{A}$  un automate fini à  $n$  états reconnaissant  $L_8$ . Observons que, au cours de la lecture de  $u$ , chaque état de  $\mathcal{A}$  est visité au plus une fois: sinon, il existerait deux naturels  $i$  et  $j$  vérifiant  $0 \leq i < j \leq |u|$  et  $\delta^*(i, u_1 \dots u_i) = \delta^*(i, u_1 \dots u_j)$  et par suite le mot  $u_1 \dots u_i u_{j+1} \dots u_{|u|}$  serait encore dans  $L_8$ , ce qui contredirait l'hypothèse de minimalité. Nous en déduisons  $|u| < n$ .

**Question 9** • Les parties finies de  $L_9$  sont rationnelles. Montrons que ce sont les seules. Soit  $M$  une partie infinie de  $L_9$ ; supposons-la reconnue par un afdc  $\mathcal{A} = (Q, \delta, i, F)$ . Appliquons le lemme de l'étoile: il existe une constante  $N$  telle que tout mot  $u$  de  $M$ , de longueur au moins égale à  $N$ , se décompose en  $u = xyz$  avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset M$ . Choisissons alors  $u = a^n b^n$  avec  $n \geq N$  (un tel  $n$  existe puisque  $M$  est infini). Nécessairement,  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{n-i-j} b^n$ . Nous mettons en évidence une contradiction en notant que le mot  $xz = a^{n-j} b^n$  ne peut appartenir à  $M$ .

**Question 10** • La réponse est négative. Pour l'établir, utilisons le lemme de l'étoile. Supposons  $L_{10}$  reconnu par un afdc  $\mathcal{A} = (Q, \delta, i, F)$ . Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_{10}$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_{10}$ . Considérons alors la décomposition du mot  $u = a^{2^N} b^N$ : comme  $2^N > N$ , nous avons nécessairement  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{2^N-i-j} b^N$ . Nous mettons en évidence une contradiction en notant que le mot  $uz = a^{2^N-j} b^N$  n'appartient pas à  $L_{10}$ .

**Question 11** • La réponse est négative. Pour l'établir, utilisons le lemme de l'étoile. Supposons  $L_{11}$  reconnu par un afdc  $\mathcal{A} = (Q, \delta, i, F)$ . Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_{11}$ . Considérons alors la décomposition associée au mot  $u = a^{N^2}$ ; nous aurons  $0 < |y| \leq N$ , donc  $N^2 - N < |xz| < N^2$ ; à plus forte raison,  $N^2 - 2N + 1 < |xz| < N^2$ . Donc  $|xz|$  n'est pas un carré parfait, si bien que  $xz$  n'appartient pas à  $L_{11}$ .

*Remarque:* si l'on a résolu la question 17, il suffit de remarquer que la suite des longueurs des mots du langage est à lacunes non bornées, puisque la différence entre deux carrés consécutifs  $n^2$  et  $(n + 1)^2$  peut être rendue aussi grande que l'on veut.

**Question 12** • Soit  $L$  un langage infini. Choisissons dans  $L$  un mot  $u_0 \neq \varepsilon$ . Puis, une fois obtenu le mot  $u_n$ , choisissons dans  $L$  un mot  $u_{n+1}$  tel que  $|u_{n+1}| \geq |u_n| + 2^n$ . Une récurrence immédiate montre que  $|u_n| \geq 2^n$ . Le langage  $M = \{u_n | n \in \mathbb{N}\}$  n'est pas rationnel. En effet, s'il l'était, en vertu du lemme de l'étoile, il devrait contenir un langage de la forme  $xy^*z$ , avec  $y \neq \varepsilon$ . La suite des longueurs de ces mots est arithmétique, or on ne peut extraire une suite arithmétique de la suite des longueurs des mots de  $M$ .

*Remarque:* si l'on a résolu l'exercice 17, il suffit de noter que la suite des longueurs des mots de  $M$  est à lacunes non bornées.

**Question 13** • Appliquons le lemme de l'étoile: si  $L_P$  était rationnel, Il existerait une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_P$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L$ . Choisissons un nombre premier  $p \geq n$ ; sa décomposition est  $a^p = a^i a^j a^k$ , avec  $j \neq 0$  et  $u_n = a^i a^{nj} a^k \in L_P$  quel que soit  $n \in \mathbb{N}$ . Si  $i = k = 0$ , alors  $|u_2| = 2p$ ; si  $i + k = 1$ , alors  $|u_{p+1}| = p^2$ ; si  $i + k \geq 2$ , alors  $|u_{i+k}| = (i + k)(j + 1)$ . Dans tous les cas, nous avons exhibé un mot de la forme  $a^q$ , avec  $q$  non premier.

**Question 14** • Utilisons le lemme de l'étoile. Supposons  $L_{14}$  reconnu par un afdc. Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_{14}$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L$ . Considérons alors la décomposition du mot  $u = \mathbf{10}^N \mathbf{20}^N \mathbf{1}$ , qui est l'écriture décimale de  $(10^{N+1} + 1)^2$  et appartient donc à  $L_{14}$ . D'après le lemme de l'étoile  $u$  admet au moins une décomposition de la forme  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L$ .  $xy$  doit être un préfixe de  $\mathbf{10}^{N-1}$ ; le cas  $x = \varepsilon$ ,  $y = \mathbf{10}^j$ ,  $z = \mathbf{0}^{N-j} \mathbf{20}^N \mathbf{1}$  avec  $0 < j < N$  est exclu: en effet,  $xz = \mathbf{0}^{N-j} \mathbf{20}^N \mathbf{1}$  commence par un  $\mathbf{0}$  et ne peut donc appartenir à  $L$ . Ainsi,  $x = \mathbf{10}^i$ ,  $y = \mathbf{0}^j$ ,  $z = \mathbf{0}^{N-i-j} \mathbf{20}^N \mathbf{1}$  avec  $j > 0$ . Le mot  $xy^3z = \mathbf{10}^{N+2j} \mathbf{20}^N \mathbf{1}$  devrait appartenir à  $L$ . Mais ceci est impossible: en effet, notant  $p$  le nombre dont  $xy^3z$  est l'écriture décimale, nous avons  $p = 10^{2N+2j+2} + 2 \cdot 10^{N+1} + 1$ ; nous en déduisons  $p > 10^{2N+2j+2} = (10^{N+j+1})^2$  et  $p < 10^{2N+2j+2} + 2 \cdot 10^{N+j+1} + 1 = (10^{N+j+1} + 1)^2$ . Alors  $p$ , strictement compris entre deux carrés parfaits consécutifs, ne peut être un carré parfait.

• Remarque: le résultat subsiste, même si l'on considère que les écritures décimales peuvent commencer par un ou plusieurs zéros. En effet, le nombre qui s'écrit  $\mathbf{0}^{N-j} \mathbf{20}^N \mathbf{1}$  est multiple de 3, mais pas de 9: ce ne peut donc pas être un carré parfait.

**Question 15** • Supposons  $L_{15}$  rationnel, et soit  $\mathcal{A} = (Q, \delta, i, F)$  un afdc reconnaissant  $L_{15}$ . Notons  $n$  le nombre des états de  $\mathcal{A}$ . D'après le principe des tiroirs, il existe des naturels  $j$  et  $k$  vérifiant  $0 \leq j < k \leq n$  tels que  $\delta^*(i, a^j) = \delta^*(i, a^k)$ ; nous en déduisons  $\delta^*(i, a^{n+k-j} b(bc)^n) = \delta^*(i, a^n b(bc)^n)$ , donc  $a^{n+k-j} b(bc)^n \in L_{15}$  ce qui est contradictoire.

• Soient  $p$  et  $q$  des naturels tels que  $uvw = a^p b(bc)^p$  et  $uw = a^q (bc)^q$ . Nécessairement,  $q \leq p$ ; si  $q < p$ , alors  $u$ , facteur gauche commun à  $uvw$  et  $uw$ , est de la forme  $a^i$  avec  $i \leq q$ ; de même,  $w$  doit être de la forme  $(bc)^j$  avec  $j \leq q$ , ou  $c(bc)^j$  avec  $j < q$ ; dans le premier cas,  $uw = a^i (bc)^j \notin L_{15}$ ; dans le deuxième cas,  $uw = a^i c(bc)^j \notin L_{15}$ . Ayant ainsi mis en évidence une contradiction, nous pouvons conclure:  $p = q$ , donc  $v = \varepsilon$ .

• Supposons que  $L_{15}$  contienne un rationnel infini  $M$ . Appliquons le lemme de l'étoile: il existe un entier  $N$  tel que tout mot de  $M$  de longueur au moins  $N$  s'écrit  $xyz$  avec  $y \neq \varepsilon$ , et  $xy^*z \subset M$ . Soit  $xyz$  un tel mot; notant  $u = xy$ ,  $v = y$  et  $w = z$ , nous mettons en évidence deux mots  $uw$  et  $uvw$  qui appartiennent à  $M$ , donc à  $L$ , avec  $v \neq \varepsilon$ , d'où une contradiction.

**Question 16** • Utilisons le lemme de l'étoile. Supposons  $L_{16}$  reconnu par un afdc  $\mathcal{A} = (Q, \delta, i, F)$ . Il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_{16}$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_{16}$ . Considérons alors la décomposition du mot  $u = a^{5N} b^{3N}$ : nécessairement  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{5N-i-j} b^{3N}$ . Nous mettons en évidence une contradiction en notant que le mot  $uz = a^{5N-j} b^{3N}$  n'appartient pas à  $L_{16}$ .

**Question 17** • Appliquons le lemme de l'étoile à  $L_{17}$ : il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_{17}$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_{17}$ .  $L_{17}$  étant infini, un tel mot  $u$  existe. Notons  $p = |xz|$  et  $q = |y|$ . Observons la suite arithmétique de premier terme  $p$  et de raison  $q$ : ses termes sont tous des longueurs de mots de  $L_{17}$ . Comme  $q$  est strictement positif, la longueur d'une lacune de  $L_{17}$  est strictement inférieure à  $\max(p+1, q)$ ; par suite,  $L_{17}$  est à lacunes bornées.

**Question 18** • Ce langage n'est pas rationnel. Pour le prouver, nous appliquerons le lemme de l'étoile à  $L_{18}$ : il existe une constante  $N \geq 0$  telle que tout mot  $u$  de  $L_{18}$  de longueur supérieure ou égale à  $N$  se décompose en  $u = xyz$ , avec  $|xy| \leq N$ ,  $y \neq \varepsilon$  et  $xy^*z \subset L_{18}$ . Choisissons  $u = a^N c b^N$ : alors  $x = a^i$ ,  $y = a^j$  avec  $j \neq 0$  et  $z = a^{N-i-j} c b^N$ . Le mot  $xz = a^{N-j} c b^N$  n'appartient pas à  $L_{18}$ .

**Question 19** • On pourrait croire que ce langage n'est pas rationnel, mais il n'en est rien! En fait  $L_{19} = \Sigma^*$ : on devine ce résultat, en choisissant plusieurs mots «au hasard», et en constatant que chacun admet la décomposition indiquée.

Soit donc  $u \in \Sigma^*$ ; notons  $n = |u|$ . Pour alléger la rédaction, nous noterons  $u[i..j]$  le mot  $u_i u_{i+1} \dots u_{j-1} u_j$ , et ce pour  $0 \leq i \leq j \leq n$ . Définissons  $\varphi: k \in \llbracket 0, n \rrbracket \mapsto |u[1..k]|_a - |u[k+1..n]|_b$ . Nous remarquons que  $\varphi(0) = -|u|_b \leq 0$ , tandis que  $\varphi(n) = |u|_a \geq 0$ . Soit  $k \in \llbracket 0, n-1 \rrbracket$ ; observons comment varie  $\varphi$  lorsque l'on passe de la position  $k$  à la position  $k+1$  immédiatement à droite:

- si  $u_{k+1} = a$ , alors  $|u[1..k+1]|_a = |u[1..k]|_a + 1$  et  $|u[k+2..n]|_b = |u[k+1..n]|_b$ ;
- si  $u_{k+1} = b$ , alors  $|u[1..k+1]|_a = |u[1..k]|_a$  et  $|u[k+2..n]|_b = |u[k+1..n]|_b - 1$ .

Nous constatons que, dans tous les cas,  $\varphi(k+1) - \varphi(k)$  est égal à  $+1$ . Donc  $\varphi(k) = \varphi(0) + k = k - |u|_b$  s'annule pour  $k = |u|_b$ . Ainsi, en prenant  $x = u[1..k]$  et  $y = u[k+1..n]$ , nous avons  $u = xy$  avec  $|x|_a = |y|_b$ , ce qui montre que  $u \in L_{19}$ .

**Question 20** • La preuve est très semblable à celle du lemme de l'étoile. Soit  $\mathcal{A} = (Q, \delta, i, F)$  un afdc reconnaissant  $L_{20}$ . Prenons  $N = 35|Q|$ , et soit  $x$  un mot de  $L_{20}$  de longueur au moins égale à  $N$ . Pour  $0 \leq k \leq |Q|$ , notons  $x[1..35k]$  le préfixe de  $x$  de longueur  $35k$ , et  $q_k = \delta^*(i, x[1..35k])$ . D'après le principe des tiroirs, il existe deux indices  $k$  et  $k'$  appartenant à  $[[0, |Q|]]$ , tels que  $k < k'$  et  $\delta^*(i, x[1..35k]) = \delta^*(i, x[1..35k'])$ . Notons alors  $u = x[1..35k]$ ,  $v = x[35k + 1..35k']$  et  $w = x[35k' + 1..|x|]$ . Constatons que  $|uv| = 35k' \leq N$ ,  $|v| = 35(k' - k) \neq 0$ ,  $|u|$  est multiple de 5,  $|v|$  est multiple de 7 et  $uv^i w \in L_{20}$  quel que soit  $i \in \mathbb{N}$ . Seul ce dernier point nécessite une explication :  $\delta^*(q_k, v) = q_k$ , donc  $\delta^*(q_k, v^i) = q_k$  puis  $\delta^*(uv^i) = \delta^*(\delta^*(i, u), v^i) = \delta^*(q_k, v^i) = q_k$  et donc :

$$\delta^*(uv^i w) = \delta^*(\delta^*(i, uv^i), w) = \delta^*(q_k, w) = \delta^*(\delta^*(i, uv), w) = \delta^*(i, uvw) \in F$$

**FIN**