Rapport sur l'épreuve "Mathématiques de l'Informatique

1. Description du problème

Le problème proposait l'étude des ensembles de mots définis par des propriétés de "comptage" de leurs sous-mots. Le comptage portait sur le nombre de façons distinctes d'écrire un mot u comme sous-mot d'un mot v. On envisageait dans le problème deux types de comptage. Le comptage "booléen" qui correspond à la distinction "u est sous-mot de v" ou "u n'est pas sous-mot de v", faisait l'objet des parties III et IV du problème. On démontrait que les langages définis par ce type de comptage sont précisément les langages reconnus par un monoïde J-trivial fini, ou encore par un monoïde ordonné fini. Le comptage modulo p, où p est un nombre premier, était étudié dans la cinquième partie : les langages définis par ce type de comptage sont exactement les langages reconnus par un p-groupe.

L'analogie entre monoïdes finis J-triviaux et p-groupes n'est d'ailleurs pas fortuite. Appelons matrice unitriangulaire une matrice triangulaire supérieure dont les coefficients diagonaux sont égaux à 1. On démontre alors que tout monoïde fini J-trivial -resp. tout p-groupe- est quotient d'un monoïde de matrices unitriangulaires à coefficients dans le semianneau de Boole -resp. dans le corps $\mathbb{Z}/p\mathbb{Z}$-

La première partie permettait de démontrer l'équivalence entre le formalisme des automates et celui des monoïdes. La deuxième partie décrivait un algorithme linéaire pour calculer l'automate minimal de l'ensemble des sous-mots d'un mot donné. Les parties III et IV aboutissaient à la démonstration du théorème de Imre Simon [3] en suivant la méthode proposée récemment en [4]. La principale difficulté consistait à prouver que tout monoïde fini J-trivial M est quotient d'un monoïde ordonné fini par récurrence sur le cardinal de M (Questions 5 à 8 du IV). Les résultats de la cinquième partie

sont dûs à S. Eilenberg et M.P. Schützenberger et l'énoncé suivait d'assez près l'exposé donné en [1]. On introduisait d'abord le "coefficient binômial" de deux mots dont on démontrait quelques propriétés élémentaires (voir [2] pour plus de détails). On utilisait ensuite les propriétés de l'algèbre d'un p-groupe pour conclure. A noter que des résultats plus précis peuvent être obtenus en utilisant le calcul des commutateurs dans le groupe libre [5].

[1] S. Eilenberg, Automata, Languages and Machines, Academic Press, Vol. B (1976) pp. 238-245.

[2] M. Lothaire, Combinatorics on Words, Addison Wesley, Encyclopedia of Mathematics 17 (1983) chapitre 6

[3] I. Simon, Piecewise testable events, Proc. 2nd GI Conf. Lecture Notes in Computer Science 33, springer (1975) 214-222.

[4] H. Straubing et D. Thérien, Partially ordered finite monoids and a theorem of I. Simon, à paraître

[5] D. Thérien, Subword counting and nilpotent groups, combinatorics on words, Progress and Perspectives, L. Cummings éd. Academic Press (1983) 297-305.

2. Commentaires sur les copies

La plupart des candidats ont abordé les deux premières parties du problème et ont cherché ensuite à grapiller des points dans les autres parties. Le problème comportait un certain nombre de questions faciles, voire très faciles, sur lesquelles de nombreux candidats semblent avoir perdu beaucoup de temps, sans pour autant les résoudre correctement. L'exemple le plus typique est le III 1°)b) où certains candidats ont besoin d'une page de calcul pour prouver que la composée de deux relations réflexives est réflexive.

La première partie ne comportait pas de difficulté et la notion d'automate semble être relativement bien connue des candidats. En revanche, la plupart des candidats ne connaissent ni la définition des langages rationnels ni l'énoncé du théorème de Kleene. La question 3°)a) n'a été

traitée correctement que par un tiers des copies environ.

Dans la seconde partie, les candidats se sont laissés abuser par l'aspect élémentaire du programme Pascal qui leur était proposé, et ont souvent confondu intuition et démonstration. De nombreux candidats semblent éprouver de réelles difficultés à mettre en place un raisonnement par récurrence, et la question 3°)b) n'a été traitée correctement que dans une seule copie !

Troisième partie. Malgré le contre-exemple explicitement donné dans l'énoncé, certains candidats ont confondu les notions de facteurs et de sous-mots. Les questions relatives à l'index ont opéré une sélection très nette, de même que la deuxième question. Personne n'a su résoudre le 3°)b).

Les candidats ont surtout abordé les trois premières questions de la quatrième partie. Pour le 1°)a), la diversité des contre-exemples proposés est remarquable : on y retrouve pratiquement tous les exemples de groupes au programme de l'agrégation ! La question 2°) était difficile, mais a été partiellement résolue dans les meilleures copies. En revanche les questions 4 à 9, rarement abordées, n'ont jamais été résolues de façon satisfaisante.

L'aspect algébrique de la cinquième partie a attiré les meilleurs candidats qui ont souvent abordé les quatre premières questions. La fin du problème n'a été traitée que dans une seule copie.

Rappelons que les correcteurs tiennent le plus grand compte de la rédaction des copies, qui manque souvent de clarté et de rigueur. Enfin signalons l'orthographe du mot "occurrence"...

La répartition des notes, sur 82 copies corrigées, est la suivante

0 à 4	19	20 à 24	6
5 à 9	9	25 à 29	7
10 à 14	22	30 à 34	5
15 à 19	12	35 à 39	1
		40	1

3. Commentaires sur les questions

I) La première partie n'offrait guère de difficultés. La question 5°b), qui était la seule "question de cours" du problème, a cependant dérouté nombre de candidats. Rappelons donc que les langages rationnels de A^* forment le plus petit ensemble de langages de A^* contenant les langages finis et fermé pour les opérations union, produit et étoile. Le théorème de Kleene (qui figure, rappelons-le, au programme de l'option) affirme qu'un langage est rationnel si et seulement si il est reconnu par un automate fini. Le 5°b) découle directement du théorème de Kleene et des questions 4°)b et 5°)a.

II) 1°) Sur l'exemple l'éxécution de la procédure "automate" donnait le résultat suivant

L'ensemble des états est (0,...,5)
delta(5,a1) = 5
delta(4,a1) = 5
delta(5,a2) = 5
delta(4,a2) = 5
delta(3,a1) = 4
delta(3,a2) = 5
delta(2,a1) = 3
delta(2,a2) = 5
delta(1,a1) = 3
delta(1,a2) = 2
delta(0,a1) = 1
delta(0,a2) = 2

L'état initial est 0 ;
Tous les états sauf 5 sont des états finaux.

2°) La double inégalité $i+1 \leqslant \delta(i,a) \leqslant \delta(i+1,a)$ s'établissait facilement par récurrence sur m-i. En revanche la formule $\delta(i,b_{i+1}) = i+1$ pouvait être établie directement.

3°) Le a) se déduisait facilement du 2°). En revanche le b) était plus délicat et n'a été traité que dans une seule copie. Une solution consistait à établir d'abord par récurrence sur m-i la formule

$$\delta(i,a_j) = \begin{cases} k \text{ si } k \text{ est le plus petit entier tel que } k>i \text{ et } b_k \\ m+1 \text{ s'il n'existe aucun } k>i \text{ tel que } b_k = a_j \end{cases}$$

III) 1°) Beaucoup d'erreurs sur cette question.
Une classe modulo $\sim_n$ est caractérisée par un ensemble de mots de longueur inférieure ou égale à n. Comme il y a $N = 1 + \text{card} + \ldots + (\text{card}A)^n$ mots de longueur au plus n, il y a au plus 2^N classes distinctes. Si $A = \{a,b\}$, l'index de $\sim_1$ est 4 (considérer les mots 1,a,b et ab). Par ailleurs deux mots, équivalents pour $\sim_n$, et de longueur au plus n, sont nécessairement égaux, ce qui fournit $2^{n+1}-1$ classes distinctes. On pouvait ensuite observer que pour $n > o$, le mot $a^n b$ n'est équivalent à aucun mot de longueur inférieur ou égale à n.

2°) Il fallait observer d'une part que

$$\{v \in A^* \mid v \sim_n u\} = \bigcap_{w \text{ sous-mot de } u} S(w) \setminus \bigcup_{w \text{ non sous-mot de } u} S(w)$$

et d'autre part que $S(u)$ était saturé modulo $\sim_n$

3°) Pour le b), on se ramène facilement au cas où u et v sont de longueur strictement supérieure à n. u admet alors une factorisation $u = u_o a_1 u_1 \ldots a_n u_n$ (avec $a_i \in A$, $u_i \in A^*$) telle que $\eta(1) = \eta(u_o) \geqslant \eta(u_o a_1) = \eta(u_o a_1 u_1) \geqslant \ldots \geqslant \eta(u_o a_1 \ldots u_{n-1} a_n) = (u_o a_1 \ldots u_{n-1} a_n)$ et donc $\eta(u) = \eta(a_1 a_2 \ldots a_n)$. Comme $a_1 \ldots a_n$ est sous-mot de u, donc de v, on en déduit facilement $\eta(u) \leqslant \eta(v)$, et un raisonnement dual montre que $\eta(v) \leqslant \eta(u)$.

Pour le c) il suffisait d'appliquer le I1°) puis le I3°)a)

IV) 1°) La relation $\leqslant_J$ n'est transitive dans aucun groupe non trivial. Signalons toutefois qu'il existe des contre-exemples autres que les groupes, par exemple le monoïde multiplicatif des matrices $\begin{pmatrix}1&0\\0&1\end{pmatrix}$, $\begin{pmatrix}0&1\\0&1\end{pmatrix}$ et $\begin{pmatrix}1&0\\1&0\end{pmatrix}$.

Pour démontrer que M n'est pas un monoïde ordonné, il ne suffisait pas de prouver que la relation $\leqslant_J$ n'est pas compatible avec la multiplication !

2° i) $\Rightarrow$ ii) En prenant $y = 1$ on trouve $x^n = x^{n+1}$. Par ailleurs on observe que $(xy)^n = (xy)^n x = x(yx)^n = (yx)^n$

ii) $\Rightarrow$ i) On a $(xy)^n = (yx)^n = (yx)^{n+1} = y(xy)^n x$, d'où en itérant $(xy)^n = y^n(xy)^n x^n$ et donc $(xy)^n x = y^n(xy)^n x^{n+1} = y^n(xy)^n x^n = (xy)^n$.

(i) $\Rightarrow$ (iii) Si $x \leqslant_J y$ et $y \leqslant_J x$, il existe $s,t,u,v \in M$ tels que $x = syt$ et $y = uxv$, d'où $x = (su)x(vt)$, et en itérant $x = (su)^n x(vt)^n$. On conclut facilement en utilisant les égalités

$$(su)^n = u(su)^n \text{ et } (vt)^n = (vt)^n v.$$

(iii) $\Rightarrow$ (i) Soit x M. Puisque M est fini, la suite décroissante

$$1 \underset{J}{\geqslant} x \underset{J}{\geqslant} x^2 \ldots \underset{J}{\geqslant} x^n$$

est stationnaire et il existe $n_x > 0$ tel que $x^{n_x} = x^{n_x+1}$. En prenant $n = \max \{n_x \mid x \in M\}$, on a, pour tout $x \in M$, $x^n = x^{n+1}$. En particulier, si $x,y \in M$, il vient $(xy)^n \underset{J}{\geqslant} (xy)^n x \underset{J}{\geqslant} (xy)^{n+1} = (xy)^n$ d'où $(xy)^n = (xy)^n x$.

3°) Il suffisait d'utiliser les équations établies en 2°). Une preuve directe était possible mais assez délicate dans le cas du quotient !

4°) Le monoïde M_n est en fait le quotient de M par la congruence qui identifie 0 et n.

a) Il ne suffit pas de dire qu'il existe un élément minimal, il faut encore vérifier son unicité.

b) La question était plus difficile qu'il n'y parait. Il fallait distinguer les cas (a) $xy \neq n$ et $yz \neq n$ (b) $xyz = n$ et (c) ($xy = n$ ou $yz = n$) et $xyz \neq n$ et observer dans le cas (c)

que $xyz \lesssim_J xy$ et donc que $xyz = 0$

c) Là encore, plusieurs cas à distinguer

5°) On vérifiait sans trop de peine que la propriété d'être quotient d'un monoïde ordonné fini est stable par passage au sous-monoïde, au monoïde quotient et au produit de deux monoïdes.

a) Si Card $N(M) \geqslant 2$, M est isomorphe à un sous-monoïde du produit $\prod_{n \in N(M)} M_n$, ce qui permet de conclure par récurrence puisque Card $M_n <$ Card M

b) On suppose $N(M) = \{n\}$ et $n^2 = n$. Si $x \neq 0$, on a $x \geqslant_J n$ par minimalité de n et il existe $r,s \in M$ tels que $n = rx$ Il vient

$$n \geqslant_J nr \geqslant_J nrx \geqslant_J nrxs = n^2 = n$$

d'où $n = nr = nrx$ et $n = nx$. Donc si $x \neq 0$ et $y \neq 0$, $nxy = ny = n$ et $xy \neq 0$. Donc $M \setminus \{0\}$ est un sous-monoïde de M. Par récurrence $M \setminus \{0\}$ est quotient d'un monoïde ordonné T. On en déduit que M est quotient d'un monoïde ordonné en adjoignant un zéro à T.

8°)

a) La classe d'un mot u est caractérisée par $\alpha(u)$ et par l'ensemble des triplets $(\alpha(u_0), a, \alpha(u_1))$ tels que $u_0 a u_1 = u$ et $a \in A \cup \{1\}$.

On en déduit facilement que $\sim_A$ est d'index fini

b) C'était la question la plus difficile du problème. Posons $I = \{n,o\}$. On se ramenait facilement au cas où $\varphi(u)$ et $\varphi(v)$ sont éléments de I et on traitait successivement les cas suivants

(1) Il existe une factorisation $u = u_0 u_1$ avec $\varphi(u_0)$ et $\varphi(u_1)$ dans $M \setminus I$

(1') Même condition que (1) en substituant v à u.

(2) Les conditions (1) et (1') sont exclues mais il existe une factorisation $u = u_o a u_1$ avec $a \in A$, $\varphi(u_o) \in M \setminus I$ et $\varphi(u_1) \in M \setminus I$

(2') Même condition que (2) en substituant v à u.

(3) La cas restant : toutes les factorisations $u = u_o u_1$ ou $u = u_o a u_1$ (resp. $v = v_o v_1$ ou $v = v_o a v_1$) satisfont $\varphi(u_o) \in I$ ou $\varphi(u_1) \in I$ (resp. $\varphi(v_o) \in I$ ou $\varphi(v_1) \in I$).

Dans le cas (1), on considère une factorisation $u = u_o u_1$ (avec $\varphi(u_o)$, $\varphi(u_1) \in M \setminus I$) telle que le couple $(\alpha(u_o), \alpha(u_1))$ soit maximal pour l'ordre produit dans T x T. Puisque $u \sim_A v$, il existe des factorisations $v = v_o v_1$ et $u = u'_o u'_1$ telles que $\alpha(u_o) \leqslant \alpha(v_o) \leqslant \alpha(u'_o)$ et $\alpha(u_1) \leqslant \alpha(v_1) \leqslant \alpha(u'_1)$. On peut supposer que u_o est facteur gauche de u'_o et que u'_1 est facteur droit de u_1 (l'autre cas se traite de manière analogue). On en déduit d'une part $\alpha(u'_o) \leqslant \alpha(u_o)$ d'où $\varphi(u_o) = \varphi(v_o) = \varphi(u'_o)$ car $\varphi(u_o) \in M \setminus I$ et d'autre part $\varphi(u_1) \leqslant_J \varphi(u'_1)$ d'où $\varphi(u'_1) \in M \setminus I$ puisque $\varphi(u_1) \in M \setminus I$. D'après la maximalité de la factorisation choisie, on a nécessairement $\alpha(u_1) = \alpha(v_1) = \alpha(u'_1)$ puis $\varphi(u_1) = \varphi(v_1)$ et finalement $\varphi(u) = \varphi(v)$
Le cas (1') est analogue et les cas (2) et (2') se traitent par une méthode analogue. Enfin dans le cas (3), on note u_o le facteur gauche de u (éventuellement vide) de longueur maximale tel que $\varphi(u_o) \notin I$, et on pose $u = u_o a u_1$ avec $a \in A$. On a alors $\varphi(u_o a) \in I$ et $\varphi(u_1) \in I$ d'où $\varphi(u) \in I.I = \{n,o\}\{n,o\} = O$. De même $\varphi(v) = O$ donc $\varphi(u) = \varphi(v)$.

V)

1°) et 2°) Pas de difficulté particulière.

3°) On pouvait observer que μ est l'inverse de $-\sigma$, et que σ est une involution.

4°)a) Les formules s'établissaient par récurrence sur la longueur de v en utilisant le 2°).

b) résultait immédiatement de la formule $\mu(uv) = \mu(u)\mu(v)$

5°)b) Par récurrence sur $|w|$. Le cas $|w| = O$ ne présente pas de difficulté. Si $|w| = n+1$ et si v est facteur de w,

on a

$$\binom{u^{p^{n+1}}}{v} = \sum_{v_1 \cdots v_p = v} \binom{u^{p^n}}{v_1} \cdots \binom{u^{p^n}}{v_p}$$

Or par hypothèse de récurrence $\binom{u^{p^n}}{v_i} \equiv 0 \bmod p$ si $0 < |v_i| \leqslant n$ ce qui montre que $\binom{u^{p^{n+1}}}{v} \equiv 0 \bmod p$ si $0 < |v| \leqslant n$. Si $v = 1$, on a $\binom{u^{p^{n+1}}}{1} \equiv 1 \bmod_p$ et si $v = w$, il y a exactement p terme non nuls (modulo p) dans la sommation ci-dessus, correspondant aux p factorisations pour lesquelles l'un des v_i est égal à w. Donc

$$\binom{u^{p^{n+1}}}{w} = p \binom{u^{p^n}}{w} \equiv 0 \mod p$$

c) On a $x^{p^n} = 1$ pour tout élément x de $G = A^* |\equiv_w$. On en déduit facilement que G est un groupe fini, puis -mais c'est moins évident que ne l'ont affirmé certains candidats - que G est un p-groupe (en utilisant par exemple la formule des classes)

8°)a) On pouvait observer que pour tout mot u et pour toute lettre a

$$(1-ua) = (1-u)+u(1-a)$$

9°)a) Résultat classique, qui découle lui aussi de la formule des classes.

d) Puisque $\binom{p}{i} \equiv 0 \bmod p$ pour $1 \leqslant i \leqslant p-1$, on obtenait $(1-x)^p = 1+(-1)^p$. Pas de problème si p est impair. Et si $p = 2$ $1+1 = 0 \bmod 2$! Comme x est dans le centre de G, K^p est l'idéal engendré par $(1-x)^p$, d'où le résultat.

12°)b) Si $L \in \mathcal{G}_p$, L est reconnu par un p-groupe d'après 5°)c) Réciproquement, si L est reconnu par un p-groupe G, L est union de $\equiv_n$-classes (où n = Card G-1) et chaque $\equiv_n$-classe est intersection de $\equiv_w$-classes. Enfin chaque $\equiv_w$-classe est combinaison booléenne des $L_{w,i}$ et $L \in \mathcal{G}_p$.

A noter que cette dernière question conduit à un algorithme pour décider si un langage rationnel donné L est élément de $\mathcal{G}_p$. On calcule d'abord l'automate minimal de L, puis le monoïde M associé à cet automate (décrit dans la première partie). Alors L est élément de $\mathcal{G}_p$ si, et seulement si, M est un p-groupe. De la même façon, on peut décider si un langage rationnel donné est élément de J : il faut et il suffit que M soit J-trivial.