

J. P. PÉCUCHE

**Automates boustrophedon, semi-groupe de
Birget et monoïde inversif libre**

RAIRO. Informatique théorique, tome 19, n° 1 (1985), p. 71-100

http://www.numdam.org/item?id=ITA_1985__19_1_71_0

© AFCET, 1985, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Informatique théorique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

AUTOMATES BOUSTROPHEDON, SEMI-GROUPE DE BIRGET ET MONOÏDE INVERSIF LIBRE (*)

par J. P. PÉCUCHE (1)

Communiqué par J.-E. PIN

Résumé. — On montre que la puissance des automates boustrophédon reste inchangée pour divers modes de reconnaissance. On en déduit certaines propriétés des parties rationnelles du semi-groupe de Birget et du monoïde inversif libre.

Abstract. — The power of two-way automata is showed to be unaltered by various recognition rules. Some properties of the rational sets in the Birget's semigroup and the free inverse monoïd are deduced.

1. INTRODUCTION

Un automate boustrophédon est un automate à une bande de lecture pouvant lire dans les deux sens. Il accepte un mot si une lecture commencée sur le bord gauche dans un état initial se termine sur le bord droit dans un état final.

Shepherdson [7] a montré que les langages reconnus par de tels automates sont les langages rationnels. Ainsi, contrairement à ce qui se passe avec les automates à pile (*cf.* [1]), on ne change pas la puissance d'un automate fini en lui autorisant une lecture dans les deux sens.

Nous montrons ici que cette stabilité des automates finis est même bien plus grande qu'on ne l'imaginait. On peut ainsi leur autoriser plus de liberté en leur laissant accepter les mots pour lesquels il existe une lecture qui commence à l'intérieur du mot dans un état initial et se termine à l'intérieur du mot dans un état final après avoir examiné tout le mot. On peut d'autre part leur imposer plus de restrictions en leur réclamant une lecture couvrante pour tous les points de départ et tous les points d'arrivée dans le mot. Nous

(*) Reçu novembre 1983, révisé juillet 1984.

(1) L.I.T.P., Laboratoire d'Informatique, Faculté des Sciences et des Techniques, B.P. n° 67, 76130 Mont-Saint-Aignan, France.

montrons que dans les deux cas les boustrophédon reconnaissent encore les langages rationnels.

On peut de façon naturelle interpréter un automate boustrophédon travaillant sur l'alphabet A comme un automate fini (à un seul sens de lecture) travaillant sur un alphabet $A \cup \bar{A}$ où \bar{A} est une copie disjointe de A . On est ainsi amené à rechercher les rapports existant entre le langage de A^* reconnu par le boustrophédon et le langage de $(A \cup \bar{A})^*$ reconnu par l'automate fini sous-jacent. Nous montrons que l'on peut calculer le premier à partir du second à l'aide des réductions du semi-Dyck ou de la réduction de Birget. Les formules obtenues mettent en évidence des propriétés inattendues de ces réductions. Elles montrent également, comme le montrait déjà la comparaison avec les automates à pile, que la rationalité des langages boustrophédon est en soi surprenante.

Un semi-groupe S (ou un monoïde) est dit régulier si l'équation $sxs = s$ admet une solution pour tout s et inversif si le système $(sxs = s, xsx = x)$ admet une unique solution pour tout s . L'étude algébrique des semi-groupes inversifs repose sur celle des semi-groupes inversifs libres et constitue un point charnière entre l'étude des groupes et celle des semi-groupes. D'autre part Birget a montré que l'étude générale des semi-groupes peut dans la plupart des cas se ramener à celle plus simple des semi-groupes réguliers. Il associe pour cela à (presque) tout semi-groupe S un semi-groupe régulier $(S)_{\text{reg}}$ que nous appellerons semi-groupe de Birget de S .

Dans la dernière partie de l'article nous utiliserons les rapports mis en évidence entre les automates boustrophédon et les automates finis sous-jacents pour obtenir certaines propriétés de stabilité des parties rationnelles du semi-groupe de Birget de A^* et du monoïde inversif libre sur A .

2. AUTOMATES BOUSTROPHÉDON

Nous introduisons dans ce paragraphe les notations et définitions adoptées concernant les automates boustrophédon.

On appelle *automate boustrophédon* (en abrégé $2FA$ pour 2-way finite automaton) sur l'alphabet (fini) A tout quadruplet $\mathcal{A} = (Q, Q_-, Q_+, F)$ constitué d'un ensemble fini d'états Q disjoint de A , de deux parties de Q , Q_- (les états initiaux) et Q_+ (les états finaux), et d'une partie F de $Q \times (A \cup \bar{A}) \times Q$ contenant les flèches, où $\bar{A} = \{\bar{a}/a \in A\}$ est une copie de A disjointe de A et de Q .

On appelle *section* de $w \in A^*$ tout couple $(u, v) \in A^* \times A^*$ vérifiant $uv = w$ et *description instanée* (d'un calcul) de w tout mot $uqv \in A^*QA^*$ dans lequel

(u, v) est une section de w . On définit sur les descriptions instantanées une relation binaire notée \vdash par :

$$\left\{ \begin{array}{ll} uqav \vdash uaq'v & \text{si } (q, a, q') \in F, \\ uaqv \vdash uq'av & \text{si } (q, \bar{a}, q') \in F. \end{array} \right\} \quad (2.1)$$

On appelle *chemin* dans \mathcal{A} tout mot de F^* de la forme :

$$c = (q_0, x_0, q_1)(q_1, x_1, q_2) \dots (q_{n-1}, x_{n-1}, q_n).$$

Son *étiquette* est $|c| = x_0 \dots x_{n-1} \in (A \cup \bar{A})^*$, l'étiquette du chemin vide étant le mot vide.

On appelle *calcul* de $w \in A^*$ dans \mathcal{A} toute suite $(w_i, q_i, w'_i) \ i=0 \dots n$ de descriptions instantanées de w vérifiant $w_i q_i w'_i \vdash w_{i+1} q_{i+1} w'_{i+1}$ pour $i=0 \dots n-1$. A un tel calcul les relations (2.1) permettent d'associer un unique chemin dont l'étiquette est appelée une *lecture* de w . On dit que le calcul (ou la lecture associée) est *réussi* si $q_0 \in Q_-$ et $q_n \in Q_+$ et qu'il est *couvrant* s'il existe $i, j \in \{0, \dots, n\}$ tels que $w_i = w'_j = 1$.

On note \vdash^* la fermeture réflexive et transitive de \vdash . On a donc $uqv \vdash^* u'q'v'$ ssi il existe un calcul de $w = uv = u'v'$ de premier terme uqv et de dernier terme $u'q'v'$. Si c est un chemin associé à un tel calcul on notera $uqv \vdash_c^* u'q'v'$.

On définit trois modes de reconnaissance par l'automate \mathcal{A} .

Le premier (classique) consiste à accepter les mots $w \in A^*$ admettant un calcul réussi de section initiale $(1, w)$ et finale $(w, 1)$. Le langage reconnu selon le mode 1 est donc

$$\mathcal{L}^1(\mathcal{A}) = \{ w \in A^* / \exists (q_-, q_+) \in Q_- \times Q_+ \text{ tels que } q_- \vdash^* w q_+ \}.$$

Le deuxième mode de reconnaissance consiste à accepter les mots $w \in A^*$ admettant un calcul réussi et couvrant. Le langage reconnu selon le mode 2 est donc :

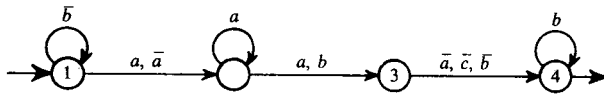
$$\mathcal{L}^2(\mathcal{A}) = \{ w \in A^* / \exists (w_i, w'_i) \in A^{*2}, \exists (q_-, q_+, p_1, p_2) \in Q_- \times Q_+ \times Q^2, \\ \text{tels que } w_1 q_- w'_1 \vdash^* p_1 w \vdash^* p_2 w_2 q_+ w'_2 \\ \text{ou } w_1 q_- w'_1 \vdash^* p_1 w \vdash^* p_2 w \vdash^* w_2 q_+ w'_2 \}.$$

Le troisième mode de reconnaissance consiste à accepter les mots $w \in A^*$ admettant des calculs réussis et couvrants de sections initiales et finales arbitraires. Le langage reconnu selon le mode 3 est donc :

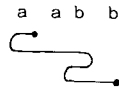
$$\mathcal{L}^3(\mathcal{A}) = \{ w \in A^* / \forall (w_i, w'_i) \in A^{*2} \text{ tels que } w = w_i w'_i (i=1, 2) \\ \exists (q_-, q_+, p_1, p_2) Q_- \times Q_+ \times Q^2 \\ \text{tels que } w_1 q_- w'_1 \vdash^* p_1 w \vdash^* p_2 w_2 q_+ w'_2 \\ \text{ou } w_1 q_- w'_1 \vdash^* p_1 w \vdash^* p_2 w \vdash^* w_2 q_+ w'_2 \}.$$

Enfin nous noterons $\mathcal{L}^0(\mathcal{A})$ le langage reconnu par \mathcal{A} lorsqu'il est considéré comme un automate fini (au sens habituel) sur l'alphabet $A \cup \bar{A}$.

Exemple 2.2 : Considérons l'automate boustrophédon \mathcal{A} représenté avec les conventions habituelles par le schéma suivant :



Le mot $aabb$ admet par l'automate la lecture $\bar{a}aabb\bar{b}b$ représentée ci-dessous :



et associée au calcul (réussi et couvrant) suivant :

$$a1abb \vdash 2aabb \vdash a2abb \vdash aa2bb \vdash aab3b \vdash aa4bb \vdash aab4b \vdash aabb4.$$

Le lecteur vérifiera que l'on a $\mathcal{L}^1(\mathcal{A}) = a^+ b^+$, $\mathcal{L}^2(\mathcal{A}) = a^+ b^*$, $\mathcal{L}^3(\mathcal{A}) = ab^*$, $\mathcal{L}^0(\mathcal{A}) = \bar{b}^* (a + \bar{a}) a^* (a + b) (\bar{a} + \bar{b} + \bar{c}) b^*$. On a toujours, évidemment, $\mathcal{L}^1(\mathcal{A}) \cup \mathcal{L}^3(\mathcal{A}) \subseteq \mathcal{L}^2(\mathcal{A})$.

REMARQUE 2.3 : On peut toujours supposer, sans perte de généralité, qu'un automate boustrophédon contient un seul état initial. Si en effet $\mathcal{A} = (Q, Q_-, Q_+, F)$ en contient plus d'un, il reconnaît le même quadruplet de langages que l'automate $\mathcal{A}' = (Q \cup i, i, Q_+, F')$ obtenu en adjoignant à Q un nouvel état i , en étendant F en :

$$F' = F \cup \{ (i, x, q) / \exists q_- \in Q_- (q_-, x, q) \in F \} \quad (x \in A \cup \bar{A})$$

et en posant $Q'_+ = Q_+$ si $Q_- \cap Q_+ = \emptyset$ et $Q'_+ = Q_+ \cup i$ sinon.

Nous dirons qu'un automate boustrophédon $A = (Q, Q_-, Q_+, F)$ est *déterministe* (en abrégé *2DFA* pour 2-way deterministic finite automaton) s'il vérifie les trois conditions suivantes :

$$(1) Q = \{i\};$$

(2) pour tout $p \in Q$ et tout $x \in A \cup \bar{A}$ il existe au plus une flèche $(p, x, q) \in F$;

(3) pour tout $p \in Q$ les flèches $(p, x, q) \in F$ d'origine p sont toutes étiquetées dans A ($x \in A$) ou toutes étiquetées dans \bar{A} ($x \in \bar{A}$).

REMARQUE 2.4 : Un automate fini [respectivement fini déterministe] $\mathcal{A} = (Q, Q_-, Q_+, F)$ sur l'alphabet A peut être considéré comme un automate boustrophédon [respectivement boustrophédon déterministe] sur A ne contenant aucune flèche étiquetée dans \bar{A} . Dans ce cas tout calcul couvrant est du type $qw^* \dagger wq'$ ce qui montre que :

$$\mathcal{L}^0(\mathcal{A}) = \mathcal{L}^1(\mathcal{A}) = \mathcal{L}^2(\mathcal{A}) = \mathcal{L}(\mathcal{A}) \quad \text{et} \quad \mathcal{L}^3(\mathcal{A}) = \mathcal{L}(\mathcal{A}) \cap \{1\},$$

où $\mathcal{L}(\mathcal{A})$ désigne le langage reconnu au sens classique du terme par \mathcal{A} .

3. RECONNAISSANCE DES MOTS PAR LES AUTOMATES BOUSTROPHEDON

Nous étudions dans ce paragraphe les langages reconnus par les automates boustrophédon selon chacun des trois modes de reconnaissance décrits plus haut. Les preuves que nous donnons sont constructives et montrent que, sauf pour les boustrophédons déterministes utilisés selon le mode 3, les langages reconnus sont toujours exactement les langages rationnels.

Dans ce qui suit l'alphabet fini A sera fixé. Nous noterons $\text{Rat}(A^*)$ la famille des langages rationnels de A^* et $\mathcal{L}^i(2FA)$ [$\mathcal{L}^i(2DFA)$] la famille des langages de A^* reconnus par les automates boustrophédon [déterministes] selon le mode i ($i = 1, 2, 3$).

Rappelons d'abord le théorème de Shepherdson (1959) qui prouve l'identité entre les familles $\mathcal{L}^1(2FA)$, $\mathcal{L}^1(2DFA)$ et $\text{Rat}(A^*)$:

THÉORÈME 3.1 : *Pour tout automate boustrophédon \mathcal{A} on peut construire un automate fini \mathcal{A}' tel que $\mathcal{L}^1(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$.*

On trouvera une preuve de ce résultat dans [4].

Nous nous proposons d'étendre ce résultat aux deux autres modes de reconnaissance. La technique utilisée s'inspire de celle décrite dans [4] et repose sur les deux lemmes suivants.

Soit $\mathcal{A} = (Q, Q_-, Q_+, F)$ un automate boustrophédon à $n = |Q|$ états et $w \in A^*$ un mot. On a :

LEMME 3.2 : Si w admet un calcul réussi et couvrant dans \mathcal{A} , il en admet un de longueur au plus égale à $3n(|w| + 1)$.

Preuve : Soit $(w_i q_i w'_i)_{i=0 \dots m}$ un calcul réussi et couvrant de w dans \mathcal{A} de longueur $m+1 > 3n(|w| + 1)$. Il existe alors $0 \leq i_1 < i_2 < i_3 < i_4 \leq m$ tels que $w_{i_k} q_{i_k} w'_{i_k} = uqv$ pour $k = 1 \dots 4$. Chacune des trois suites :

$$C_k = (w_i q_i w'_i)_{i=1, \dots, i_k, i_{k+1}+1, \dots, m} \quad (k=1, 2, 3)$$

fournit donc un calcul réussi de w dans \mathcal{A} de longueur au plus m . De plus un au-moins de ces calculs est couvrant. En effet parmi les trois suites $(w_i q_i w'_i)_{i=i_k \dots i_{k+1}}$ ($k=1, 2, 3$) on peut en choisir une pour laquelle la valeur $\min \{ |w_i|/i = i_k \dots i_{k+1} \}$ est minimale et une pour laquelle $\min \{ |w'_i|/i = i_k \dots i_{k+1} \}$ est minimal. Alors si k est l'indice d'une suite non choisie, le calcul C_k couvre W . ■

Soit $\mathcal{A} = (Q, i, Q_+, F)$ un automate boustrophédon. Pour tout mot $w \in A^*$ considérons les ensembles de chemins suivants :

$$id(w) = \{ c/\exists u, v \in A^*, \exists q \in Q, uiv \underset{c}{\vdash}^* wq \},$$

$$ig(w) = \{ c/\exists u, v \in A^*, \exists q \in Q, uiv \underset{c}{\vdash}^* qw \},$$

$$gd(w) = \{ c/\exists p, q \in Q, pw \underset{c}{\vdash}^* wq \},$$

$$dg(w) = \{ c/\exists p, q \in Q, wp \underset{c}{\vdash}^* qw \},$$

$$dd(w) = \{ c/\exists p, q \in Q, wp \underset{c}{\vdash}^* wq \},$$

$$gg(w) = \{ c/\exists p, q \in Q, pw \underset{c}{\vdash}^* qw \},$$

$$gt(w) = \{ c/\exists u, v \in A^*, \exists p \in Q, \exists t \in Q_+, pw \underset{c}{\vdash}^* utv \},$$

$$dt(w) = \{ c/\exists u, v \in A^*, \exists p \in Q, \exists t \in Q_+, wp \underset{c}{\vdash}^* utv \},$$

$$c(w) = id(w) \cup ig(w) \cup gd(w) \cup dg(w) \cup dd(w) \cup gg(w) \cup gt(w) \cup dt(w).$$

Avec ces notations, on a le :

LEMME 3.3 : *Tout chemin c associé à un calcul réussi et couvrant du mot $z = wu$ ($w, u \in A^*$) admet une factorisation dans l'un des deux ensembles :*

$$E = E(z) = (I \cup I')(T \cup T' \cup T'')$$

ou

$$E' = E'(z) = I''(T' \cup T''),$$

avec :

$$\begin{aligned} I &= I(z) = id(w)gg(u)(dd(w)gg(u))^*dg(w); \\ I' &= I'(z) = ig(u)(dd(w)gg(u))^*dg(w), \\ I'' &= I''(z) = ig(w); \quad T = T(z) = gt(w); \\ T' &= T'(z) = gd(w)gg(u)(dd(w)gg(u))^*dt(w); \\ T'' &= T''(z) = gd(w)(gg(u)dd(w))^*gt(u). \end{aligned}$$

De plus, dans une telle factorisation, l'un des facteurs dans $c(u)$ couvre le mot u .

Réciproquement, tout chemin de $E \cup E'$ dont l'un des facteurs dans $c(u)$ couvre le mot u est associé à un calcul réussi et couvrant de wu .

Preuve : Les ensembles I, I' et I'' sont formés des chemins associés aux calculs du type $\alpha i \beta \vdash^* qwu$ et qui respectivement commencent dans w et couvrent w , commencent dans u , commencent dans w et ne passent pas dans u . De même T, T' et T'' sont formés des chemins associés aux calculs du type $qwu \vdash^* \alpha' t \beta'$ et qui respectivement se terminent dans w sans passer dans u , se terminent dans w et couvrent w , se terminent dans u . Ces ensembles sont illustrés par le schéma 3.4.

Soit maintenant $\alpha i \beta \vdash^* \alpha' t \beta'$ le calcul réussi et couvrant de wu auquel est associé c . Soit $c = \sigma\tau$ la factorisation de c dans laquelle σ est le plus petit facteur gauche de c vérifiant $\alpha i \beta \vdash^* qwu$ pour un $q \in Q$. On a alors $\sigma \in I \cup I' \cup I'', \tau \in T \cup T' \cup T''$ et il faut montrer que $c \in E \cup E'$.

Si $\sigma \in I \cup I'$ alors $c \in E$. Si $\sigma \in I''$ couvre w et se factorise en $\sigma = \sigma_1 \sigma_2$ avec $uiv \vdash^*_{\sigma_1} wp \vdash^*_{\sigma_2} qw$ alors, si ε désigne le chemin vide, on a

$\sigma = \sigma_1 \varepsilon \sigma_2 \in id(w)gg(u)dg(w) \subset I$ d'où $c \in E$. Enfin si $\sigma \in I' - I$, puisque σ ne couvre pas w , τ admet un facteur gauche dans $gd(w)$. On a donc $\tau \in T' \cup T''$ d'où $c \in E'$. On a donc bien dans tous les cas $c \in E \cup E'$. De plus, puisque c couvre wu , dans toute factorisation de c dans E ou E' l'un des facteurs dans $c(u)$ couvre u .

La réciproque consiste en une simple vérification laissée au soin du lecteur. ■

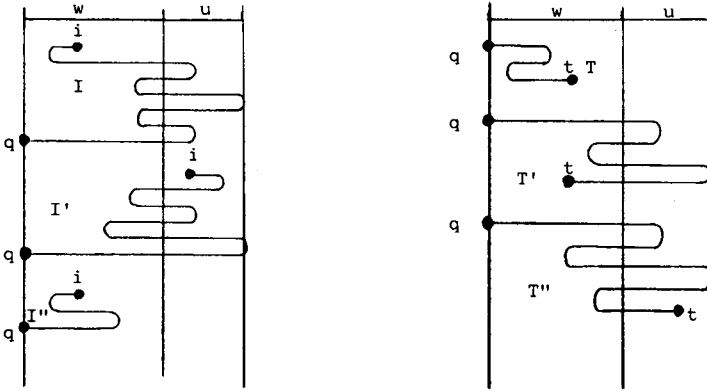


Schéma 3.4

Nous sommes maintenant en mesure de prouver le :

THÉORÈME 3.5 : *Pour tout automate boustrophedon \mathcal{A} on peut construire un automate fini \mathcal{A}' tel que $\mathcal{L}^2(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$.*

Preuve : On peut sans perte de généralité supposer que $\mathcal{A} = (Q, i, Q_+, F)$ contient un seul état initial. Posons $L = \mathcal{L}^2(\mathcal{A})$ et pour $w \in A^*$ posons $w^{-1}L = \{u \in A^* / wu \in L\}$.

Le lemme 3.2 montre que l'on peut décider si un mot w appartient ou non à L . Pour pouvoir construire \mathcal{A}' il suffit donc (cf. [4], p. 57) de montrer que la famille $\{w^{-1}L / w \in A^*\}$ est finie et de borner son cardinal.

Pour cela on introduit l'application $\theta : A^* \rightarrow (2^Q)^2 \times (2^{Q \times Q})^5$ qui à w associe $(\theta_1(w), \dots, \theta_7(w))$ défini de la façon suivante :

$$\theta_1(w) = \{q / \exists u, v, uiv \vdash^* wq\},$$

$$\theta_2(w) = \{q / \exists u, u, v, uiv \vdash^* qw\},$$

$$\theta_3(w) = \{(p, q) / pw \vdash^* wq\},$$

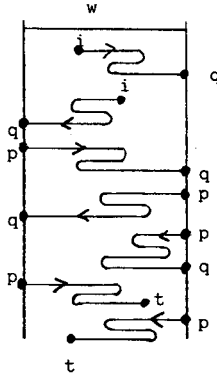
$$\theta_4(w) = \{(p, q) / wp \vdash^* qw\},$$

$$\theta_5(w) = \{(p, q) / wp \vdash^* wq\},$$

$$\theta_6(w) = \{(p, t) / \exists u, v, \exists t \in Q^+, pw \vdash^* utv\},$$

$$\theta_7(w) = \{(p, t) / \exists u, v, \exists t \in Q^+, wp \vdash^* utv\}.$$

La définition de θ est illustrée par le schéma ci-dessous :



Puisque θ est d'image finie il suffit de montrer que $\theta(w) = \theta(w')$ implique $w^{-1}L = w'^{-1}L$, et pour cela il suffit, par symétrie, de prouver que $wu \in L$ implique $w'u \in L$.

Soient donc $w, w', u \in A^*$ avec $\theta(w) = \theta(w')$, $wu \in L$ et c un chemin associé à un calcul réussi et couvrant de wu dans \mathcal{A} .

D'après le lemme 3.3 c admet une décomposition dans $E(wu) \cup E'(wu)$. Mais d'après l'égalité $\theta(w) = \theta(w')$ on peut, à tout chemin de p à q apparaissant dans la décomposition de c et appartenant respectivement à $id(w), ig(w), gd(w), dg(w), dd(w), gt(w), dt(w)$, substituer un chemin de p à q appartenant respectivement à $id(w'), ig(w'), gd(w'), dg(w'), dd(w'), gt(w'), dt(w')$. On obtient ainsi un chemin c' de $E(w'u) \cup E'(w'u)$ dont l'un des facteurs dans $c(u)$ couvre le mot u . Il est donc, d'après le lemme 3.3 associé à un calcul réussi et couvrant de $w'u$, ce qui montre que $w'u \in L$ et achève la preuve. ■

Puisqu'un rationnel est reconnu par un automate fini déterministe, on a d'après la remarque 2.4 les inclusions $\text{Rat}(A^*) \subseteq \mathcal{L}^2(2DFA) \subseteq \mathcal{L}^2(2FA)$. Le théorème précédent montre qu'en fait ces trois familles coïncident.

Il n'en est plus de même pour le troisième mode de reconnaissance. En effet un mot $w = a_1 \dots a_n \in A^+$ ($n \geq 1$) ne peut-être reconnu par un automate déterministe selon le mode 3 puisque celui-ci devrait contenir à la fois une flèche du type (i, a_1, p) et une flèche du type (i, \bar{a}_n, q) , ce qui est exclu. On a

donc $\mathcal{L}^3(2DFA) = \{1\}$. Les deux résultats qui suivent établissent l'égalité $\mathcal{L}^3(2FA) = \text{Rat}(A^*)$.

PROPOSITION 3.6 : *Pour tout automate fini \mathcal{A} on peut construire un automate boustrophedon \mathcal{A}' tel que $\mathcal{L}(\mathcal{A}) = \mathcal{L}^3(\mathcal{A}')$.*

Preuve : Soit $\mathcal{A} = (Q, Q_-, Q_+, F)$. On construit \mathcal{A}' par adjonction à \mathcal{A} d'un nouvel état initial i et d'un nouvel état final t de la façon suivante.

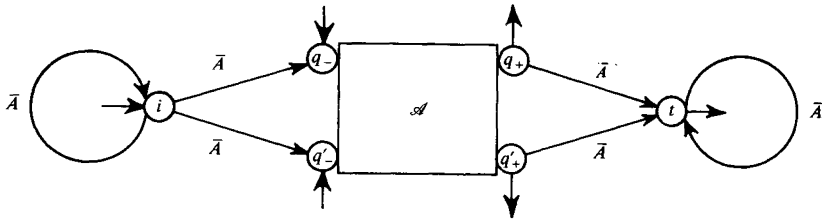
On pose :

$$\mathcal{A}' = (Q \cup \{i, t\}, Q_- \cup \{i\}, Q_+ \cup \{t\}, F \cup F'),$$

avec :

$$F' = \{(i, \bar{a}, p) / p \in Q_- \cup \{i\}, \bar{a} \in \bar{A}\} \cup \{(p, \bar{a}, t) / p \in Q_+ \cup \{t\}, \bar{a} \in \bar{A}\}.$$

La construction est illustrée par le schéma suivant.



On a de façon évidente $\mathcal{L}^3(\mathcal{A}') \subseteq \mathcal{L}^1(\mathcal{A}') = \mathcal{L}^1(\mathcal{A}) = \mathcal{L}(\mathcal{A})$.

Réciproquement si $w \in \mathcal{L}(\mathcal{A})$ est reconnu par un calcul $q_- \overset{*}{w} \vdash_c wq_+$ dans \mathcal{A} , on obtient facilement un calcul réussi et couvrant de w dans \mathcal{A}' de section initiale (u, v) et finale (u', v') de la forme $uiv \overset{*}{w} \vdash_c q_- \overset{*}{w} \vdash_c wq_+ \overset{*}{w} \vdash_c u'v'$, ce qui montre que $w \in \mathcal{L}^3(\mathcal{A}')$ et achève la preuve. ■

THÉORÈME 3.7 : *Pour tout automate boustrophedon \mathcal{A} on peut construire un automate fini \mathcal{A}' tel que $\mathcal{L}^3(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$.*

Preuve : Posons $L = \mathcal{L}^3(\mathcal{A})$ où $\mathcal{A} = (Q, i, Q_+, F)$ est un automate boustrophédon. En reprenant les arguments utilisés dans la preuve du théorème 3.5, il nous suffit de borner le cardinal de la famille $\{w^{-1}L/w \in A^*\}$.

Pour cela introduisons l'application :

$$\theta : A^* \rightarrow 2^{2^Q \times 2^Q} \times (2^Q \times Q)^3 \times 2^{2^Q \times 2^Q},$$

qui à $w \in A^*$ associe $(\theta_1(w), \dots, \theta_5(w))$ défini de la façon suivante :

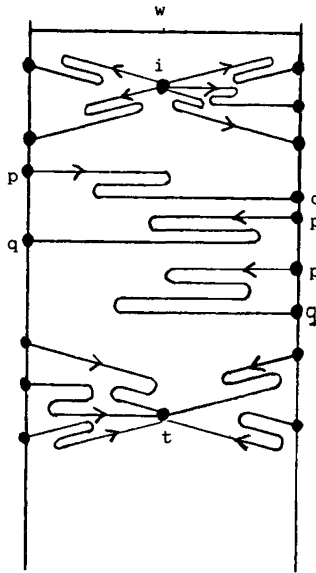
$$\theta_1(w) = \{ (\{ q/uv \vdash^* qw \}, \{ q/uiv \vdash^* wq \}) / uv = w \},$$

$$\theta_2(w) = \{ (p, q) / pw \vdash^* wq \},$$

$$\theta_3(w) = \{ (p, q) / wp \vdash^* qw \},$$

$$\theta_4(w) = \{ (p, q) / wp \vdash^* wq \},$$

$$\theta_5(w) = \{ (\{ q/\exists t \in Q_+, qw \vdash^* utv \}, \{ q/\exists t \in Q_+, wq \vdash^* utv \}) / uv = w \}.$$



Comme dans la preuve du théorème 3.5 il suffit de prouver que si $\theta(w) = \theta(w')$ alors $wu \in L$ implique $w'u \in L$.

Soient donc $w, w', u \in A^*$ avec $\theta(w) = \theta(w')$ et $v = wu \in L$. Considérons deux sections (v'_1, v'_2) et (v'_3, v'_4) de $v' = w'u$ et montrons qu'il existe un calcul réussi et couvrant de $w'u$ de section initiale (v'_1, v'_2) et de section finale (v'_3, v'_4) .

Nous devons distinguer quatre cas selon que le calcul commence ou se termine dans u ou w' .

Lorsque le calcul commence et se termine dans u c'est-à-dire lorsque $(v'_1, v'_2) = (w'u_1, u_2)$ et $(v'_3, v'_4) = (w'u_3, u_4)$, on peut choisir un chemin c dans \mathcal{A} associé à un calcul réussi et couvrant $wu_1 iu_2 \vdash^* wu_3 tu_4$ de $v = wu$. D'après

le lemme 3.3 *c* se factorise dans $I'(v)T''(v)$. D'après les égalités $\theta_i(w) = \theta_i(w')$ ($i=2, 3, 4$) on peut dans cette factorisation de c substituer aux facteurs dans $gd(w)$, $dg(w)$, $dd(w)$ des chemins dans $gd(w')$, $dg(w')$, $dd(w')$ qui ont mêmes origines et extrémités. On obtient ainsi un chemin c' dans $I'(v)T''(v)$ dont un des facteurs dans $c(u)$ couvre u et qui est associé à un calcul réussi et couvrant de $v' = w'u$ de section initiale (v'_1, v'_2) et finale (v'_3, v'_4) .

Lorsque le calcul commence et se termine dans w' , c'est-à-dire lorsque $(v'_1, v'_2) = (w'_1, w'_2 u)$ et $(v'_3, v'_4) = (w'_3, w'_4 u)$, puisque $\theta_i(w) = \theta_i(w')$ ($i=1, 5$) on peut trouver deux sections (w_1, w_2) et (w_3, w_4) de w vérifiant les conditions suivantes :

$$\begin{aligned} \{q/w_1 iw_2 \vdash^* qw\} &= \{q/w'_1 iw'_2 \vdash^* qw'\}, \\ \{q/w_1 iw_2 \vdash^* wq\} &= \{q/w'_1 iw'_2 \vdash^* w'q\}, \\ \{q/qw \vdash^* w_3 tw_4, t \in Q_+\} &= \{q/qw' \vdash^* w'_3 tw'_4, t \in Q_+\}, \\ \{q/wq \vdash^* w_3 tw_4, t \in Q_+\} &= \{q/w'q \vdash^* w'_3 tw'_4, t \in Q_+\}. \end{aligned} \tag{3.7.1}$$

Soit c un chemin dans \mathcal{A} associé à un calcul réussi et couvrant de $v = wu$ de section initiale $(w_1, w_2 u)$ et de section finale $(w_3, w_4 u)$. D'après le lemme 3.3, c admet une décomposition dans $I(v)(T(v) \cup T'(v))$ ou $I''(v)T'(v)$. D'après les relations (3.7.1) on peut remplacer le premier facteur, élément de $id(w)$ ou $ig(w)$, par un élément de $id(w')$ ou $ig(w')$ respectivement, se terminant sur le même état et associé à un calcul de w' de section initiale (w'_1, w'_2) . On peut de même remplacer le dernier facteur, élément de $dt(w)$ ou $gt(w)$, par un élément de $dt(w')$ ou $gt(w')$ se terminant sur le même état et associé à un calcul de w' de section finale (w'_3, w'_4) . Enfin, d'après les égalités $\theta_i(w) = \theta_i(w')$ ($i=2, 3, 4$) on peut remplacer les facteurs de c appartenant à $gd(w)$, $dg(w)$ ou $dd(w)$ par des chemins dans $gd(w')$, $dg(w')$ ou $dd(w')$ de mêmes origines et extrémités. On obtient ainsi un chemin c' élément de $E(v) \cup E'(v)$ associé, d'après le lemme 3.3, à un calcul réussi et couvrant de $v' = w'u$ de section initiale (v'_1, v'_2) et finale (v'_3, v'_4) .

Les deux cas restant se traitent comme les deux premiers et sont laissés au soin du lecteur.

On obtient ainsi un calcul réussi et couvrant de $w'u$ dans \mathcal{A}' pour toute section initiale (v'_1, v'_2) et toute section finale (v'_3, v'_4) , ce qui montre que $w'u \in L$ et achève la preuve. ■

5. RÉDUCTION DES LECTURES BOUSTROPHEDON

Nous montrons dans ce chapitre comment, à l'aide de réductions, on peut reconstituer un mot à partir d'une de ses lectures par un automate boustrophédon \mathcal{A} (prop. 4. 7) et calculer le langage $\mathcal{L}^1(\mathcal{A})$ reconnu par \mathcal{A} selon le mode 1 (le mode classique) à partir du langage $\mathcal{L}^0(\mathcal{A})$ reconnu par \mathcal{A} en tant qu'automate fini au sens habituel sur l'alphabet $A \cup \bar{A}$ (prop. 4. 12). Nous verrons que les formules obtenues ne permettent pas de rendre compte directement de la rationalité de $\mathcal{L}^1(\mathcal{A})$ et la rendent au contraire plutôt surprenante. Ces formules mettent également en évidence des propriétés inattendues des réductions utilisées.

Dans ce qui suit nous désignerons par Π_A et $\Pi_{\bar{A}}$ les projections de $(A \cup \bar{A})^*$ sur A^* et \bar{A} respectivement et par $w \mapsto \bar{w}$ l'anti-isomorphisme défini par $\bar{x} = \bar{a}$ si $x = a \in A$ et $\bar{x} = a$ si $x = \bar{a} \in \bar{A}$. Nous noterons σ_1, σ_2 et σ les congruences sur $(A \cup \bar{A})^*$ engendrées respectivement par les relations $\{(a\bar{a}, 1)/a \in A\}, \{(\bar{a}a, 1)/a \in A\}$ et $\{(u\bar{u}u, u)/u \in A^+\} \cup \{(\bar{u}\bar{u}\bar{u}, \bar{u})/u \in A^+\}$. Il est connu (cf. [2]) que pour chacune des congruences σ_1 et σ_2 tout mot $w \in (A \cup \bar{A})^*$ admet un unique mot réduit (c'est-à-dire un mot équivalent ne contenant respectivement aucun facteur du type $a\bar{a}$ ou $\bar{a}a$) que nous noterons respectivement $\rho_1(w)$ et $\rho_2(w)$. Il en est de même pour σ d'après le lemme suivant :

LEMME 4.1 : *Pour tout mot $w \in (A \cup \bar{A})^*$ il existe un et un seul mot $\rho(w) \in (A \cup \bar{A})^*$ qui soit congru à w modulo σ et qui ne contienne aucun facteur du type $u\bar{u}u$ ou $\bar{u}\bar{u}\bar{u}$ ($u \in A^+$).*

Preuve : Comme dans le cas des congruences σ_1 et σ_2 il suffit de montrer (cf. [2], p. 36-40) que l'on dispose de la propriété suivante :

$$\left. \begin{array}{l} \text{si } w \vDash w_1 \text{ et } w \vDash w_2 \text{ il existe } w' \text{ tel que :} \\ w_1^* \vDash w' \text{ et } w_2^* \vDash w' \quad [w, w_1, w_2, w' \in (A \cup \bar{A})^*], \end{array} \right\} \quad (4.1.1)$$

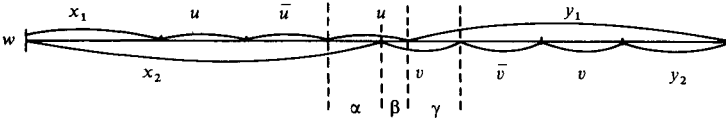
où \vDash est la relation binaire définie sur $(A \cup \bar{A})^*$ par $\alpha \vDash \beta$ ssi $(\alpha = \alpha_1 u\bar{u}u \alpha_2$ et $\beta = \alpha_1 u \alpha_2)$ ou $(\alpha = \alpha_1 \bar{u}\bar{u}\bar{u} \alpha_2$ et $\beta = \alpha_1 \bar{u} \alpha_2)$ ($u \in A^+$) et où \vDash^* est la fermeture réflexive et transitive de \vDash .

Or si $x \vDash w_1$ et $w \vDash w_2$ il est facile de voir qu'à une symétrie près les seuls cas non triviaux (c'est-à-dire pour lesquels les facteurs $u\bar{u}u$ ou $\bar{u}\bar{u}\bar{u}$ à simplifier se chevauchent) qui peuvent se présenter sont au nombre de deux.

Dans le premier cas on a $w = x_1 u\bar{u}u y_1 = x_2 \bar{v}\bar{v}v y_2$ avec $w_1 = x_1 u y_1,$
 $w_2 = x_2 v y_2,$

$$|x_1| + 2|u| \leq |x_2| \leq |x_1| + 3|u| \quad \text{et} \quad |y_2| + 2|v| \leq |y_1| \leq |y_2| + 3|v|$$

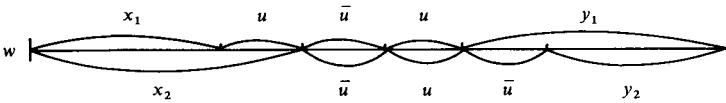
selon le schéma suivant :



On a alors $x_2 = x_1 u \bar{u} \alpha$, $x_2 \beta = x_1 u \bar{u} u$ et $y_1 = \gamma \bar{v} v y_2$.

On en déduit $u = \alpha \beta$, $v = \beta \gamma$ et où $w_1 \vDash w'$ et $w_2 \vDash w'$ avec $w' = x_1 \alpha \beta \gamma y_2$.

Dans le deuxième cas on a $w = x_1 u \bar{u} u y_1 = x_2 \bar{u} u \bar{u} y_2$ avec $w_1 = x_1 u y_1$, $w_2 = x_2 \bar{u} y_2$ et $x_2 = x_1 u$ selon le schéma suivant :



On a alors $w_1 \vDash^* w'$ et $w_2 \vDash^* w'$ avec $w' = w_1 = w_2 = x_1 u \bar{u} y_2$, ce qui achève de prouver (4.1.1). Le reste de la preuve se fait comme dans [2], p. 37-40. ■

Concernant les réductions des mots que nous venons d'introduire nous noterons le résultat suivant :

LEMME 4.2 : Pour tout $z \in (A \cup \bar{A})^*$ on a $\rho_1(z) = \rho_1(\rho(z))$, $\rho_2(z) = \rho_2(\rho(z))$, $\rho_1(\bar{z}) = \bar{\rho}_1(z)$, $\rho_2(\bar{z}) = \bar{\rho}_2(z)$, $\rho(\bar{z}) = \bar{\rho}(z)$.

Preuve : Immédiate à partir des définitions et des relations $\bar{a}\bar{a} = a\bar{a}$, $\bar{a}a = a\bar{a}$, $u\bar{u}u = \bar{u}u\bar{u}$ et $\bar{u}u\bar{u} = u\bar{u}u$. ■

Les notions de calcul ou de lecture d'un mot par un automate boustrophédon nous amènent à poser les définitions suivantes :

DÉFINITIONS 4.3 : On appelle *couverture* d'un mot $w \in A^*$ toute suite $\mathcal{C} = ((u_i, v_i))_{i=0, \dots, n}$ de sections de w vérifiant $|u_{i+1}| - |u_i| = \pm 1$ pour tout i .

On appelle *trace* de la couverture \mathcal{C} le mot $\theta = \theta_0 \dots \theta_{n-1} \in (A \cup \bar{A})^*$ défini par $\theta_i = a$ si $u_{i+1} = u_i a$ et $\theta_i = \bar{a}$ si $u_i = u_{i+1} a$ ($a \in A$, $i = 0, \dots, n-1$).

La couverture \mathcal{C} est dite *couvrante* si elle contient les sections $(1, w)$ et $(w, 1)$.

Ainsi, à tout calcul $u_0 q_0 v_0 \vdash u_1 q_1 v_1 \vdash \dots \vdash u_n q_n v_n$ d'un mot w par un automate boustrophédon on peut associer la couverture $((u_i, v_i))_{i=0, \dots, n}$ dont la trace est la lecture associée au calcul. Le calcul est couvrant ssi la couverture l'est. Réciproquement, si la trace θ d'une couverture $((u_i, v_i))_{i=0, \dots, n}$ d'un mot $w \in A^*$ est l'étiquette d'un chemin $c = (q_0, \theta_0, q_1) \dots (q_{n-1}, \theta_{n-1}, q_n)$ dans un automate boustrophédon, on a un calcul de w dans cet automate donné

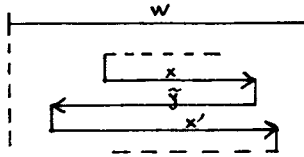
par $u_0 q_0 v_0 \vdots \dots \vdots u_n q_n v_n$. Le calcul est couvrant si la couverture l'est et réussi si le chemin l'est.

La trace d'une couverture ne peut pas être un mot quelconque. En effet tout mot $\theta \in (A \cup \bar{A})^*$ admet une unique factorisation du type :

$$\left. \begin{aligned} \theta &= x_1 \bar{y}_1 x_2 \bar{y}_2 \dots x_n \bar{y}_n x_{n+1} \quad (n \geq 1), \\ \text{avec } x_1, x_{n+1} &\in A, y_i \in A^+ (i=1, \dots, n), x_i \in A^+ (i=2, \dots, n). \end{aligned} \right\} (4.3.1)$$

Or si θ est la trace d'une couverture d'un mot $w \in A^*$ cette factorisation doit vérifier la propriété :

$$x_i \leq_s y_i \leq_p x_{i+1} \quad (i=1 \dots n). \tag{4.3.2}$$



où $\leq_p \leq_s$] dénote l'ordre préfixe défini par $u \leq_p v$ ssi u facteur gauche de v [suffixe défini par $u \leq_s v$ ssi u facteur droit de v] et $\leq_p [\leq_s]$ la comparabilité pour ces ordres.

Supposons en effet que θ se factorise en $\theta = z_1 x \bar{y} x' z_2$ avec $z_1, z_2 \in (A \cup \bar{A})^*$ et $x, y, x' \in A^*$. Si θ est la trace de la couverture $\mathcal{C} = ((u_i, v_i))_{i=0, \dots, m}$ de w , on peut factoriser \mathcal{C} en cinq couvertures :

$$\mathcal{C}_k = ((u_i, v_i))_{i=i_k, \dots, i_{k+1}} \quad (k=1, \dots, 5, 0=i_1 \leq i_2 \leq \dots \leq i_6 = m)$$

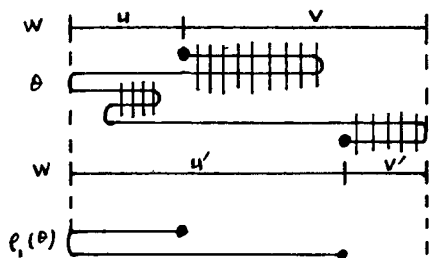
de w de traces respectives z_1, x, \bar{y}, x' et z_2 . On aura alors $(u_3, v_3) = (u_2 x, v_3)$, $(u_4, v_4) = (u_4, y v_3)$ et $(u_5, v_5) = (u_4 x', v_5)$.

On en déduit $w = u_2 x v_3 = u_4 y v_3 = u_4 x' v_5$ et donc $x \leq_s y$ et $y \leq_p x'$, ce qui prouve (4.3.2).

C'est cette remarque que nous allons utiliser dans ce qui suit pour reconstituer le mot w à partir de la trace d'une de ses couverture couvrantes.

Nous commencerons par quelques lemmes.

LEMME 4.4 : Si θ est la trace d'une couverture couvrante de w de section initiale (u, v) et de section finale (u', v') on a $\rho_1(\theta) = \bar{u}u'$ et $\rho_2(\theta) = \bar{v}v'$.



Preuve : Nous ne traiterons que le cas de la réduction ρ_1 , celui de ρ_2 lui étant symétrique. Posons $\mathcal{C} = ((u_i, v_i))_{i=0, \dots, n}$, $\theta = \theta_0 \dots \theta_{n-1}$ et supposons que θ contienne le facteur $\theta_{i-1} \theta_i = a\bar{a}$ ($a \in A$). On a alors :

$$0 < i < n, \quad (u_i, v_i) = (u_{i-1} a, v_i) \neq (1, w)$$

et

$$(u_{i-1}, v_{i-1}) = (u_{i+1}, v_{i+1}).$$

Par conséquent la suite $((u_i, v_i))_{i=0, \dots, i-1, i+2, \dots, n}$ est encore une couverture de w contenant $(1, w)$, de sections initiale et finale (u, v) et (u', v') et de trace $\theta_0 \dots \theta_{i-2} \theta_{i+1} \dots \theta_{n-1}$. Par récurrence sur la longueur d'une réduction de θ en $\rho_1(\theta)$ on montre ainsi que l'on peut associer à \mathcal{C} une couverture \mathcal{C}' de w , contenant $(1, w)$ de sections initiales et finales (u, v) et (u', v') et de trace $\theta' = \rho_1(\theta)$. Or d'après (4.3.2) la trace d'une couverture de w ne peut contenir un facteur du type $a\bar{b}$ avec $a, b \in A, a \neq b$.

On en déduit que $\rho_1(\theta) \in \bar{A} A^*$. Il est alors clair que si $\rho_1(\theta) = \bar{x}y$ ($x, y \in A^*$) on a $x = u$ et $y = u'$ et par conséquent $\rho_1(\theta) = \bar{u}u'$. ■

LEMME 4.5 : Si la factorisation (4.3.1) d'un mot $\theta \in (A \cup \bar{A})^*$ vérifie (4.3.2) et si $\theta = \rho(\theta)$ est réduit modulo ρ , alors θ est un mot de Birget, c'est-à-dire vérifie l'une des quatre propriétés suivantes :

$$x_1 \langle_s y_1 \rangle_p \dots \langle_s y_{k-1} \rangle_p x_k \langle_s y_k \rangle_p \dots \langle_s y_n \rangle_p x_{n+1} \left. \vphantom{x_1} \right\} \quad (4.5.1 a)$$

$$(1 \leq k \leq n),$$

$$x_1 \langle_s y_1 \rangle_p \dots \langle_s y_{k-1} = x_k \rangle_s y_k \langle_p \dots \rangle_s y_n \rangle_p x_{n+1} \left. \vphantom{x_1} \right\} \quad (4.5.1 b)$$

$$(1 \leq k \leq n+1),$$

$$x_1 \langle_s y_1 \rangle_p \dots \langle_s y_{k-1} \rangle_p x_k = y_k \langle_p \dots \rangle_s y_n \rangle_p x_{n+1} \left. \vphantom{x_1} \right\} \quad (4.5.1 c)$$

$$(1 \leq k \leq n+1),$$

$$x_1 \langle_s y_1 \rangle_p \dots \langle_p x_k \rangle_s y_k \langle_p x_{k+1} \rangle_s \dots \langle_s y_n \rangle_p x_{n+1} \left. \vphantom{x_1} \right\} \quad (4.5.1 d)$$

$$(1 \leq k \leq n).$$

Réciproquement, on pourra remarquer que tout mot de Birget est réduit modulo ρ et vérifie (4.3.2). Selon la terminologie de Birget (cf. [3]) le mot x_k dans (4.5.1 a, b, c) [respectivement y_k dans (4.5.1 d)] est appelé le *centre* de θ et sera noté $c(\theta)$.

Les quatre types de mots de Birget sont illustrés dans le schéma (4.5.2) où les centres sont fléchés.

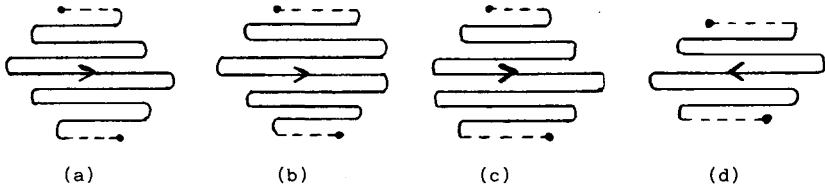


Schéma (4.5.2).

Preuve : Soit $\theta = \rho(\theta) = x_1 \bar{y}_1 \dots x_n \bar{y}_n x_{n+1}$ ($n \geq 1$) avec $x_1, x_{n+1} \in A^*$, $y_i \in A^+$ ($i = 1, \dots, n$), $x_i \in A^+$ ($i = 2, \dots, n$) avec $x_i \leq_s y_i \leq_p x_{i+1}$ ($i = 1, \dots, n$). On ne peut avoir la configuration $x_i = y_i \leq_p x_{i+1}$ ($1 \leq i \leq n$), sinon θ contiendrait le facteur $x_i \bar{x}_i x_i$ ($x_i \in A^+$) et ne serait pas réduit, on montre de la même façon qu'on ne peut avoir aucune des configurations $x_i \geq_s y_i = x_{i+1}$, $y_i = x_{i+1} \leq_s y_{i+1}$ ou $y_i \geq_p x_{i+1} = y_{i+1}$. On ne peut pas avoir la configuration $x_i \geq_s y_i \leq_p x_{i+1}$, sinon θ contiendrait le facteur $y_i \bar{y}_i y_i$ ($y_i \in A^+$) et ne serait pas réduit. On montre de même que l'on ne peut pas avoir $y_i \geq_p x_{i+1} \leq_s y_{i+1}$.

Les seules configurations permises pour trois composants successifs sont donc $x_i <_s y_i <_p x_{i+1}$, $x_i \leq_s y_i \geq_p x_{i+1}$ (une des deux inégalités étant stricte),

$$x_i >_s y_i >_p x_{i+1}, y_i <_p x_{i+1} <_s y_{i+1}, y_i \leq_p x_{i+1} \geq_s y_{i+1}$$

(une des deux inégalités étant stricte) ou $y_i >_p x_{i+1} >_s y_{i+1}$.

Enfin on ne peut avoir une inégalité $x_i >_s y_i$ ou $y_i >_p x_i$ précédant une inégalité $x_j <_s y_j$ ou $y_j <_p x_{j+1}$, sinon on aurait quelquepart $x_k \geq_s y_k \leq_p x_{k+1}$ ou $y_k \geq_p x_{k+1} \leq_s y_{k+1}$, ce qui est exclu.

On a donc d'abord des $<$ puis des $>$ et on obtient une des formes (4.5.1) selon que la transition s'effectue sur un x_i ou y_i . ■

LEMME 4.6 : Si θ est la trace d'une couverture couvrante de w alors $w = c(\rho(\theta))$ est le centre du mot de Birget $\rho(\theta)$. De plus la couverture débute sur la section (1, w) et se termine sur la section (w , 1) ssi $\rho(\theta) \in A^*$.

Preuve : Montrons d'abord que l'on peut supposer que $\theta = \rho(\theta)$ est un mot de Birget.

Supposons que $\theta = \theta_0 \dots \theta_{m-1}$ soit la trace de la couverture couvrante de w donnée par $\mathcal{C} = ((u_i, v_i))_{i=0, \dots, m}$ et supposons que $\theta = yx\bar{x}xz$ contienne le facteur $x\bar{x}x = \theta_{i_1} \dots \theta_{i_4-1}$ ($x \in A^+$, $y, z \in A^*$) avec :

$$x = \theta_{i_1} \dots \theta_{i_2-1} = \overline{\theta_{i_2} \dots \theta_{i_3-1}} = \theta_{i_3} \dots \theta_{i_4-1} \quad (0 \leq i_1 < i_2 < i_3 < i_4 \leq m).$$

On aura alors :

$$\begin{aligned} (u_{i_2}, v_{i_2}) &= (u_{i_1} x, v_{i_2}), & (u_{i_3}, v_{i_3}) &= (u_{i_1}, v_{i_1}), \\ (u_{i_4}, v_{i_4}) &= (u_{i_3} x, v_{i_4}) = (u_{i_2}, v_{i_2}). \end{aligned}$$

Par conséquent la suite $\mathcal{C}' = ((u_i, v_i))_{i=0, \dots, i_2-1, i_4, \dots, m}$ est encore une couverture couvrante de w de trace $\theta' = \theta_0 \dots \theta_{i_2-1} \theta_{i_4} \dots \theta_{m-1} = yxz$ et de mêmes sections initiales et finales que \mathcal{C} . De la même façon, on montre que si $\theta = y\bar{x}x\bar{x}z$ ($x \in A^+$; $y, z \in A^*$) on peut extraire de \mathcal{C} une couverture couvrante de w de trace $y\bar{x}z$ et de mêmes sections initiales et finales que \mathcal{C} . Une récurrence sur la longueur d'une réduction de θ en $\rho(\theta)$ montre donc que l'on peut extraire de \mathcal{C} une couverture couvrante de w de trace $\rho(\theta)$ et de mêmes sections initiales et finales que \mathcal{C} .

Nous supposons donc désormais que $\theta = \rho(\theta)$ est réduit. D'après le lemme 4.5 c'est donc un mot de Birget admettant une factorisation (4.3.1) vérifiant (4.5.1 a, b, c ou d). A cette factorisation de θ correspond une « factorisation » de la couverture :

$$\mathcal{C} = ((u_i, v_i))_{i=0, \dots, m}$$

en couvertures :

$$\mathcal{C}_r = ((u_i, v_i))_{i=i_r \dots i_{r+1}} \quad (r=0 \dots 2n; 0=i_0 \leq i_1 \leq \dots \leq i_{2n+1}=m),$$

de traces θ_r , définies par :

$$\theta_{2p} = x_{p+1} \quad (p=0 \dots n) \quad \text{et} \quad \theta_{2p+1} = \bar{y}_{p+1} \quad (p=0 \dots n-1).$$

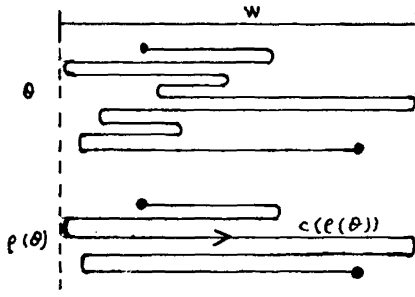
Chaque \mathcal{C}_r est une couverture de w et puisque \mathcal{C} est une couverture couvrante de w , une des \mathcal{C}_r contient $(1, w)$ et une des \mathcal{C}_r contient $(w, 1)$.

Mais les relations sur les longueurs des mots u_i et v_i déduites des relations (4.5.1) montrent que les seules couvertures \mathcal{C}_r pouvant contenir $(1, w)$ au

$(w, 1)$ contiennent nécessairement les deux sections à la fois et sont respectivement les suivantes :

- (a) \mathcal{C}_{2k-2} de trace $x_k = c(\theta)$;
- (b) \mathcal{C}_{2k-3} de trace $\bar{y}_{k-1} = \overline{c(\theta)}$ et \mathcal{C}_{2k-2} de trace $x_k = c(\theta)$;
- (c) \mathcal{C}_{2k-2} de trace $x_k = c(\theta)$ et \mathcal{C}_{2k-1} de trace $\bar{y}_k = \overline{c(\theta)}$;
- (d) \mathcal{C}_{2k-1} de trace $\bar{y}_k = \overline{c(\theta)}$.

Chacune de ces couvertures est donc couvrante et l'on a d'après le lemme 4.4, $w = c(\theta)$ dans chacun des quatre cas considérés, ce qui prouve la première partie du lemme.



Pour la seconde partie, on remarque d'abord que le seul cas dans lequel \mathcal{C} débute sur $(1, w)$ et se termine sur $(w, 1)$ est le cas (a) avec $k = n + 1 = 1$ pour lequel $\theta = c(\theta) \in A^*$. Réciproquement c'est le seul cas pour lequel $\theta \in A^*$, ce qui achève la preuve du lemme. ■

Nous sommes maintenant en mesure de reconstituer un mot w à partir d'une de ses lectures par un automate boustrophédon.

PROPOSITION 4.7 : Soit \mathcal{A} un automate boustrophédon, $\theta \in (A \cup \bar{A})^*$ et $w \in A^*$.

(1) Si θ est la lecture de w dans \mathcal{A} associée à un calcul couvrant de section initiale (w_1, w_2) et de section finale (w_3, w_4) on a $\rho_1(\theta) = \bar{w}_1 w_3$, $\rho_2(\theta) = w_2 \bar{w}_4$, $c(\rho(\theta)) = w$.

(2) Si θ est une lecture couvrante de w dans \mathcal{A} selon le mode 1 on a $w = \rho_1(\theta) = \rho_2(\theta) = \rho(\theta)$.

(3) Si θ est une lecture couvrante de w dans \mathcal{A} selon le mode 2 (ou 3) on a $w = \pi_A(\rho_1(\theta) \rho_2(\bar{\theta})) = c(\rho(\theta))$.

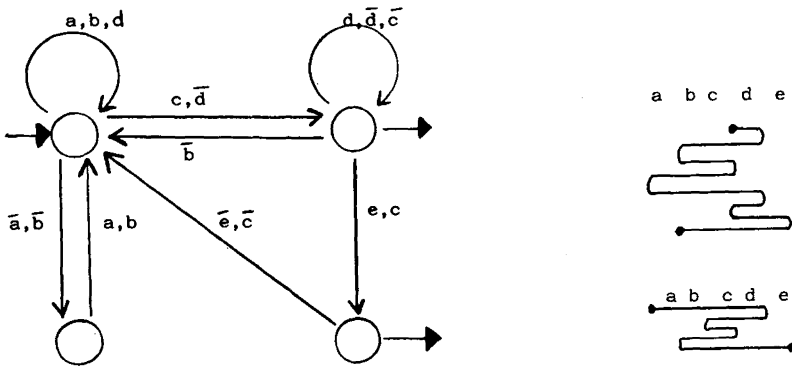
Preuve : Le (1) se déduit immédiatement des lemmes 4.4 et 4.6 en considérant la couverture couvrante de w , de trace θ et de sections initiales et finales (w_1, w_2) et (w_3, w_4) associée au calcul de w .

Le (2) se déduit du (1) en remarquant que l'on a dans ce cas $(w_1, w_2) = (1, w)$ et $(w_3, w_4) = (w, 1)$ d'où $w = \rho_1(\theta) = \rho_2(\theta)$. De plus on a dans ce cas $\rho(\theta) \in A^*$ d'après le lemme 4.6, d'où $w = c(\rho(\theta)) = \rho(\theta)$ puisque $\rho(\theta)$ est réduit à son centre.

Enfin le (3) se déduit également du (1) puisque l'on a :

$$\pi_A(\rho_1(\theta) \rho_2(\bar{\theta})) = \pi_A(\bar{w}_1 w_3 w_4 \bar{w}_2) = w_3 w_4 = w = c(\rho(\theta)). \quad \blacksquare$$

Exemple 4.8 :



Dans l'automate boustrophédon \mathcal{A} représenté ci-dessus le mot $w = abcde$ admet la lecture couvrante $z = d\bar{d}c\bar{b}bc\bar{c}b\bar{a}abc\bar{d}d\bar{d}e\bar{e}d\bar{c}$ selon le mode 2 et la lecture couvrante $z' = abc\bar{d}\bar{d}c\bar{c}b\bar{b}cde$ selon le mode 1.

On a :

$$\pi_A((\rho_1 z) \overline{(\rho_2 z)}) = \pi_A(\bar{c}\bar{b}\bar{a}ab) \pi_A(\overline{d\bar{e}\bar{e}\bar{d}\bar{c}}) = abcde = \rho_1(z') = \rho_2(z'),$$

$$\rho(z) = \overline{d\bar{a}b\bar{c}d\bar{a}b\bar{c}d\bar{e}c\bar{d}e} \text{ d'où } c(\rho(z)) = abcde \text{ et } \rho(z') = abcde.$$

Nous nous proposons maintenant d'établir des formules permettant de calculer le langage $\mathcal{L}^1(\mathcal{A})$ reconnu par un automate boustrophédon \mathcal{A} selon le mode 1 à partir du langage $\mathcal{L}^0(\mathcal{A})$ reconnu par \mathcal{A} en tant qu'automate fini au sens habituel sur l'alphabet $A \cup \bar{A}$ (prop. 4.12). L'obtention de ces formules repose sur une caractérisation des mots de $(A \cup \bar{A})^*$ qui correspondent à des lectures selon le mode 1, c'est-à-dire qui sont les traces des couvertures de section initiale $(1, w)$ et finale $(w, 1)$ (prop. 4.11). Cette caractérisation est obtenue à l'aide d'un procédé permettant de reconstituer une couverture à partir de sa trace et d'éliminer les mots qui ne sont la trace d'aucune couverture.

Passons maintenant à la description de ce procédé.

Soit $A^* \cup 0$ le semi-groupe obtenu à partir de A^* par l'adjonction d'un zéro 0. On définit sur $A^* \cup 0$ les opérations $\cdot^{\#} a$, et $a^{\#}$. ($a \in A$) par

$$1. \cdot^{\#} a = a^{\#}. 1 = 1, wa. \cdot^{\#} a = a^{\#}. aw = w (w \in A^*), wb. \cdot^{\#} a = a^{\#}. bw = 0 \\ (b \in A, b \neq a, w \in A^*), 0. \cdot^{\#} a = a^{\#}. 0 = 0.$$

Soit :

$$f : (A \cup \bar{A})^* \rightarrow (A^* \cup 0)^2$$

l'application :

$$z \mapsto (f_1(z), f_2(z)),$$

définie par les formules :

$$f(1) = (1, 1), \\ f(za) = (f_1(z)a, a^{\#}.f_2(z)), \\ f(z\bar{a}) = (f_1(z).^{\#}a, af_2(z)). \tag{4.9}$$

Soit $r : (A \cup \bar{A})^* \rightarrow A^* \cup 0$ l'application définie par $r(z) = f_1(z) f_2(z)$.

Les deux fonctions vérifient les propriétés suivantes :

LEMME 4.10 : (1) Si $r(\theta) \neq 0$ alors la factorisation (4.3.1) de θ vérifie (4.3.2).

(2) Si θ est la trace d'une couverture couvrante de w de section initiale (w_1, w_2) et finale (w_3, w_4) alors $f(\theta) = (w_3, w_4)$, $f(\bar{\theta}) = (w_1, w_2)$ et $r(\theta) = w$.

(3) Si $r(\theta) \neq 0$ alors θ est la trace d'une couverture couvrante de $r(\theta)$ de section initiale $f(\bar{\theta})$ et de section finale $f(\theta)$.

Preuve : (1) Supposons que la factorisation (4.3.1) de θ ne vérifie pas (4.3.2). On a alors $\theta = z_1 x \bar{y} z_2$ avec $x_s y$ ou bien $z = z_1 \bar{y} x z_2$ avec $x_p y$ ($z_1, z_2 \in (A \cup \bar{A})^*$; $x, y \in A^+$). Les deux cas étant symétriques, nous ne traiterons que le premier. On a donc $\theta = z_1 a x \bar{x} b z_2$ avec $z_1, z_2 \in (A \cup \bar{A})^*$, $x \in A^*$, $a, b \in A$ et $a \neq b$. On obtient alors :

$$f(z_1) = f(z_1 a x \bar{x}) = (f_1(z_1) a, f_2(z_1) a),$$

d'où :

$$f_1(z_1 a x \bar{x} b) = f_1(z_1) a^{\#}. b = 0 = f_1(\theta)$$

et finalement $r(\theta) = 0$, ce qui achève la preuve du (1).

(2) On a de façon évidente $f(\theta) = f(\rho(\theta))$, $f(\bar{\theta}) = f(\rho(\bar{\theta})) = f(\overline{\rho(\bar{\theta})})$ et $r(\theta) = r(\rho(\theta))$. On peut donc, quitte à extraire de la couverture \mathcal{C} une couverture couvrante de w de trace $\rho(\theta)$ de mêmes sections initiales et finales,

supposer comme dans la preuve du lemme 4.6 que $\theta = \rho(\theta)$ est un mot de Birget. Reprenons les mêmes notations et supposons, pour fixer les idées, que nous sommes dans le cas (4.5.1a). On a alors $x_k = c(\theta) = w$ d'après ce même lemme et on obtient facilement $f(x_1) = (x_1, 1)$, $f(x_1 \bar{y}_1) = (1, y_1)$ et par récurrence :

$$f(x_1 \bar{y}_1 \dots \bar{y}_{k-1} x_k) = (x_k, 1) = (w, 1) = (u_{i_{2k-1}}, v_{i_{2k-1}}).$$

On en déduit alors :

$$f(x_1 \bar{y}_1 \dots x_k \bar{y}_k) = (u_{i_{2k}}, v_{i_{2k}})$$

et par récurrence :

$$f(x_1 \bar{y}_1 \dots \bar{y}_n x_{n+1}) = (u_{i_{2n+1}}, v_{i_{2n+1}})$$

c'est-à-dire $f(\theta) = (w_3, w_4)$. Mais il est clair que $\bar{\theta}$ est également réduit modulo ρ et est la trace d'une couverture couvrante de w de section initiale (w_3, w_4) et finale (w_1, w_2) . En lui appliquant le résultat que nous venons d'obtenir il vient donc $f(\bar{\theta}) = (w_1, w_2)$ et finalement $r(\theta) = w_3 w_4 = w_1 w_2 = w$. Les cas (4.5.1b) à (4.5.1d) se traitent de la même façon.

(3) Soit θ tel que $r(\theta) \neq 0$. D'après le (2) il suffit de trouver un mot w dont θ soit la trace d'une couverture couvrante. Supposons pour l'instant que $\theta = \rho(\theta)$ est réduit modulo ρ . D'après le (1) et le lemme 4.5 θ est alors un mot de Birget dont la factorisation (4.3.2) vérifie (4.5.1a, b, c ou d). Supposons, par exemple, qu'elle vérifie (4.5.1a). Posons $m = |\theta|$ et soient m_1, m_2 tels que :

$$\begin{aligned} x_1 \bar{y}_1 \dots \bar{y}_{k-1} &= \theta_1 \dots \theta_{m_1}, \\ x_k &= \theta_{m_1+1} \dots \theta_{m_2}, \quad \bar{y}_k \dots \bar{y}_n x_{n+1} = \theta_{m_2+1} \dots \theta_m. \end{aligned}$$

Les relations $x_k >_s y_k >_p \dots >_s y_n >_p x_{n+1}$ montrent que la suite $(u_i, v_i)_{i=m_2 \dots m}$ définie par $(u_i, v_i) = f(x_k \theta_{m_2+1} \dots \theta_i)$ est une couverture de x_k de section initiale $(x_k, 1)$ et de trace $\theta_{m_2+1} \dots \theta_m$.

De même les relations $x_k >_p y_{k-1} >_s \dots >_p y_1 >_s x_1$ montrent que la suite $((u'_i, v'_i))_{i=0 \dots m_1}$ définie par :

$$(u'_i, v'_i) = f(x_k \bar{\theta}_{m_i} \bar{\theta}_{m_1-1} \dots \bar{\theta}_{m_1+1-i})$$

est une couverture de x_k de section initiale $(1, x_k)$ et de trace $\bar{\theta}_1 \dots \bar{\theta}_{m_1}$. Par conséquent la suite $((u_i, v_i))_{i=0 \dots m_1}$ définie par $(u_i, v_i) = (u'_{m_1-i}, v'_{m_1-i})$ est une couverture de x_k de section finale $(1, x_k)$ et de trace $\theta_1 \dots \theta_{m_1}$. Enfin il est

clair que la suite $((u_i, v_i))_{i=m_1 \dots m_2}$ définie par $(u_i, v_i) = (\theta_{m_1+1} \dots \theta_i, \theta_{i+1} \dots \theta_{m_2})$ est une couverture de x_k de section initiale $(1, x_k)$ et finale $(x_k, 1)$ et de trace x_k . On en déduit donc que la suite $((u_i, v_i))_{i=1 \dots m}$ est une couverture couvrante de $x_k = c(\theta)$ de trace θ .

On construit de la même façon une couverture de $c(\theta)$ dans les cas (4.5.1 b) à (4.5.1 d).

Enfin dans le cas général où θ n'est pas réduit, puisque $r(\rho(\theta)) = r(\theta) \neq 0$, on peut construire comme plus haut une couverture couvrante de $w = c(\rho(\theta))$ de trace $\rho(\theta)$ que l'on peut ensuite étendre de façon évidente en une couverture de w de trace θ . ■

Nous pouvons maintenant caractériser les mots qui sont des lectures boustrophédon selon le mode 1. Ce résultat met en évidence des propriétés étonnantes des réductions ρ , ρ_1 et ρ_2 .

PROPOSITION 4.11 : On a l'égalité $\rho^{-1}(A^*) = \rho_1^{-1}(A^*) \cap \rho_2^{-1}(A^*)$.

Ce langage est constitué des mots de $(A \cup \bar{A})^*$ qui sont la trace d'une couverture d'un mot w de A^* de section initiale $(1, w)$ et finale $(w, 1)$. Tout mot z appartenant à ce langage vérifie $\rho(z) = \rho_1(z) = \rho_2(z)$.

Preuve : Notons L le langage des mots de $(A \cup \bar{A})^*$ qui sont la trace d'une couverture d'un mot de A^* de section initiale $(1, w)$ et finale $(w, 1)$.

On a $L \subseteq \rho^{-1}(A^*)$ d'après le lemme 4.6. Réciproquement supposons $\rho(z) \in A^*$. Alors $f(z) = f(\rho(z)) = (\rho(z), 1)$ et donc $r(z) = r(\rho(z)) = \rho(z) \neq 0$.

On en déduit d'après de (3) du lemme 4.10 que z est la trace d'une couverture de $r(z) = \rho(z)$ de section initiale $f(\bar{z}) = f(\overline{\rho(z)}) = (1, \rho(z))$ et de section finale $f(z) = (\rho(z), 1)$. On a donc $z \in L$, d'où l'égalité $L = \rho^{-1}(A^*)$.

On a $L \subseteq \rho_1^{-1}(A^*) \cap \rho_2^{-1}(A^*)$ d'après le lemme 4.4 puisque d'après ce lemme, si z est la trace d'une couverture de w de section initiale $(1, w)$ et finale $(w, 1)$ on a $\rho_1(z) = \rho_2(z) = w \in A^*$.

Réciproquement, supposons $\rho_1(z) = z_1 \in A^*$ et $\rho_2(z) = z_2 \in A^*$. On en déduit alors $r(z) = f_1(z_1) f_2(z_2) = z_1 \neq 0$ et d'après le (3) du lemme 4.10 z est la trace d'une couverture de $r(z) = z_1$ de section initiale $f(\bar{z}) = (f_1(\bar{z}_1), f_2(\bar{z}_2)) = (1, z_2)$ et de section finale $f(z) = (f_1(z_1), f_2(z_2)) = (z_1, 1)$. On a donc $z_1 = z_2$ et $z \in L$ d'où la seconde égalité $L = \rho_1^{-1}(A^*) \cap \rho_2^{-1}(A^*)$.

Enfin si $z \in L$ est la trace d'une couverture de $w \in A^*$ de section initiale $(1, w)$ et finale $(w, 1)$ on a $\rho_1(z) = \rho_2(z) = w$ d'après le lemme 4.4 et puisque $\rho(z) \in A^*$ d'après ce qui précède, on a également $w = c(\rho(z)) = \rho(z)$ d'après le lemme 4.6, d'où l'égalité $\rho(z) = \rho_1(z) = \rho_2(z)$ qui achève la preuve. ■

Ce qui précède va nous fournir des formules permettant de reconstituer le langage reconnu par un boustrophédon selon le mode 1 à partir du langage qu'il reconnaît en tant qu'automate fini sur $A \cup \bar{A}$.

PROPOSITION 4. 12 : Pour tout automate boustrophedon \mathcal{A} on a :

$$\begin{aligned}\mathcal{L}^1(\mathcal{A}) &= \rho[\mathcal{L}^0(\mathcal{A})] \cap A^* \\ &= \rho_1[\mathcal{L}^0(\mathcal{A}) \cap \rho_2^{-1}(A^*)] \cap A^* = \rho_2[\mathcal{L}^0(\mathcal{A}) \cap \rho_1^{-1}(A^*)] \cap A^*.\end{aligned}$$

Preuve : Notons L le langage des mots de $(A \cup \bar{A})^*$ qui sont la trace d'une couverture d'un mot de A^* de section initiale $(1, w)$ et finale $(w, 1)$. D'après la proposition 4. 11 on a $L = \rho^{-1}(A^*) = \rho_1^{-1}(A^*) \cap \rho_2^{-1}(A^*)$ et les égalités à prouver peuvent encore s'écrire sous la forme :

$$\mathcal{L}^1(\mathcal{A}) = \rho[\mathcal{L}^0(\mathcal{A}) \cap L] = \rho_1[\mathcal{L}^0(\mathcal{A}) \cap L] = \hat{\rho}_2[\mathcal{L}^0(\mathcal{A}) \cap L].$$

Or si $w \in \mathcal{L}^1(\mathcal{A})$, w admet une lecture z associée à un calcul de w selon le mode 1. z est donc l'étiquette d'un chemin réussi dans \mathcal{A} et aussi la trace d'une couverture de w de section initiale $(1, w)$ et finale $(w, 1)$ d'où $z \in \mathcal{L}^0(\mathcal{A}) \cap L$. Mais d'après la proposition 4. 7 on a $w = \rho(z) = \rho_1(z) = \rho_2(z)$ d'où les inclusions dans le sens direct.

Réciproquement si $z \in \mathcal{L}^0(\mathcal{A}) \cap L$, z est l'étiquette d'un chemin réussi dans \mathcal{A} et la trace d'une couverture d'un mot $w \in A^*$ de section initiale $(1, w)$ et finale $(w, 1)$. z est donc une lecture de w dans \mathcal{A} selon le mode 1 et d'après la proposition 4. 7 on a $\rho(z) = \rho_1(z) = \rho_2(z) = w \in \mathcal{L}^1(\mathcal{A})$ d'où les inclusions dans le sens réciproque. ■

On pourrait espérer déduire de ces formules une nouvelle preuve de la rationalité de $\mathcal{L}^1(\mathcal{A})$. On sait en effet que les applications ρ_1 et ρ_2 conservent les rationnels (cf. [2]). Mais ceci n'est d'aucun secours puisque L n'étant pas rationnel le langage $\mathcal{L}^0(\mathcal{A}) \cap L$ ne l'est en général pas non plus. L n'est en fait même pas algébrique comme le montre l'intersection :

$$L \cap ab^* c \bar{c} \bar{b} \bar{a} ab^* c = \{ ab^n c \bar{c} \bar{b}^n \bar{a} b^n c / n \geq 0 \}.$$

De la même façon le langage $\rho(\mathcal{L}^0(\mathcal{A}))$ n'est en général pas rationnel comme on peut le voir à l'aide de l'égalité :

$$\rho(\{ a, \bar{a} \}^*) \cap a^* \bar{a}^* a^* = \{ a^p \bar{a}^q a^r / p < q \text{ ou } r < q \leq p \}.$$

Ainsi les formules obtenues plus haut, loin de donner une nouvelle preuve de la rationalité de $\mathcal{L}^1(\mathcal{A})$ la rendent plutôt surprenante.

5. PARTIES RATIONNELLES DU SEMI-GROUPE DE BIRGET ET DU MONOÏDE INVERSIF LIBRE

Ce chapitre est destiné à donner un exemple d'application des théorèmes boustrophédon. Nous y montrons que les résultats précédents peuvent être utilisés pour obtenir des propriétés des parties rationnelles du semi-groupe de Birget ou du monoïde inversif libre.

Nous montrons que les parties rationnelles du semi-groupe de Birget sont stables par restriction à A^* et par passage au centre, et que celles du monoïde inversif libre sont stables par restriction à A^* et par les opérations consistant à conserver les supports de toutes les chaînes contenues dans une partie X , ou seulement de celles pour lesquelles toutes les chaînes déduites par changement du nœud d'entrée ou de sortie sont contenues dans X .

Nous commencerons avec le semi-groupe de Birget. Soit $S = (A \cup \bar{A})^* \cup 0$ le semi-groupe obtenu à partir du monoïde libre $(A \cup \bar{A})^*$ par l'adjonction d'un zéro et soit ε la congruence engendrée dans S par la relation :

$$\{(u\bar{u}u, u)/u \in A^+\} \cup \{(\bar{u}u\bar{u}, \bar{u})/u \in A^+\} \cup \{(u\bar{v}, 0)/u, v \in A^+, \\ u_s v\} \cup \{(\bar{v}u, 0)/u, v \in A^+, u_p v\}.$$

Birget a montré que le semi-groupe quotient $(A^*)_{\text{reg}} = S/\varepsilon$ est régulier.

Si $\mu : S \rightarrow (A^*)_{\text{reg}}$ désigne l'homomorphisme surjectif canonique, il est clair que pour tout mot $\theta \in (A \cup \bar{A})^*$ vérifiant $\mu(\theta) \neq 0$, la classe de θ modulo ε coïncide avec celle de θ modulo σ et que la factorisation (4.3.1) de θ vérifie (4.3.2). Puisque chaque σ -classe contient un seul mot réduit modulo ρ on en déduit, d'après le lemme 4.5, que chaque élément non nul $\xi \in (A^*)_{\text{reg}}$ admet un unique représentant $w = \rho(w) \in (A \cup \bar{A})^*$ qui soit un mot de Birget.

Par abus de notation nous noterons $c(\xi)$ et appellerons centre de ξ le centre du mot de Birget w . Pour une partie $X \subset (A^*)_{\text{reg}}$ nous noterons $c(X) = \{c(\xi)/\xi \in X - 0\}$.

On remarquera également que μ est injectif lorsqu'on le restreint à A^* ou à \bar{A}^* , ce qui nous permettra dans la suite d'identifier les éléments de $A^* \cup \bar{A}^*$ avec leurs images dans $(A^*)_{\text{reg}}$.

Les relations entre les lectures boustrophédon et le semi-groupe de Birget $(A^*)_{\text{reg}}$ reposent sur les deux lemmes suivants.

LEMME 5.1 : *Un mot $\theta \in (A \cup \bar{A})^*$ est la trace d'une couverture d'un mot de A^* ssi son image $\xi = \mu(\theta)$ dans $(A^*)_{\text{reg}}$ est non nulle. θ est alors la trace d'une couverture couvrante de $c(\xi)$.*

Preuve : La première partie se déduit du lemme 4. 10 en remarquant que $\mu(\theta) \neq 0$ ssi $r(\theta) \neq 0$. La seconde partie est une conséquence immédiate du lemme 4. 6. ■

LEMME 5. 2 : Si $R \in \text{Rat}(A^*)_{\text{reg}}$ il existe un automate boustrophédon \mathcal{A} vérifiant :

$$\mathcal{L}^1(\mathcal{A}) = \mu[\mathcal{L}^1(\mathcal{A})] = R \cap A^* \quad \text{et} \quad \mathcal{L}^2(\mathcal{A}) = \mu[\mathcal{L}^2(\mathcal{A})] = c(R).$$

Preuve : Puisque μ est surjectif les parties rationnelles de $(A^*)_{\text{reg}}$ s'identifient aux images par μ des parties rationnelles de S (cf. [2], p. 56). Soit donc $R' \in \text{Rat}(S)$ tel que $\mu(R') = R$.

Alors $R'' = R' - 0$ est une partie rationnelle de $(A \cup \bar{A})^*$ reconnue au sens habituel du terme par un automate fini \mathcal{A} sur l'alphabet $A \cup \bar{A}$ et vérifiant les relations :

$$\mu(R'') \cap A^* = R \cap A^* \quad \text{et} \quad c(\mu(R'')) = c(R).$$

Considérons maintenant \mathcal{A} en tant qu'automate boustrophédon sur l'alphabet A .

D'après la proposition 4. 12 on a :

$$\mathcal{L}_1(\mathcal{A}) = \rho(\mathcal{L}^0(\mathcal{A})) \cap A^* = \rho(R'') \cap A^*.$$

μ étant injectif sur A^* on en déduit :

$$\mathcal{L}_1(\mathcal{A}) = \mu(\mathcal{L}_1(\mathcal{A})) = \mu[\rho(R'')] \cap A^* = \mu(R'') \cap A^* = R \cap A^*,$$

ce qui fournit la première formule.

Maintenant si $w \in \mathcal{L}^2(\mathcal{A})$, soit $\theta \in R'' = \mathcal{L}^0(\mathcal{A})$ une lecture de w dans \mathcal{A} selon le mode 2. D'après la proposition 4. 7 on a $w = \mu(w) = c(\rho(\theta)) = c(\mu(\theta)) \in c(R)$, d'où l'inclusion $\mathcal{L}^2(\mathcal{A}) \subseteq c(R)$.

Réciproquement, soit $w \in c(R)$, $\xi \in R - 0$ tel que $w = c(\xi)$ et $\theta \in R'' = \mathcal{L}^0(\mathcal{A})$ tel que $\mu(\theta) = \xi$. D'après le lemme 5. 1, θ est la trace d'une couverture couvrante de $c(\rho(\theta)) = w$ et donc $w \in \mathcal{L}^2(\mathcal{A})$, d'où l'égalité $\mathcal{L}^2(\mathcal{A}) = c(R)$. ■

Voyons maintenant comment les théorèmes boustrophédon se transcrivent en terme de propriétés des parties rationnelles du semi-groupe de Birget.

Tout d'abord le théorème de Shepherdson (th. 3. 1) permet de déduire immédiatement du lemme 5. 2 le résultat suivant :

THÉORÈME 5. 3 : Si $R \in \text{Rat}(A^*)_{\text{reg}}$ il en est de même de $R \cap A^*$.

On remarquera que ce résultat n'est pas trivial puisque A^* est une partie rationnelle mais pas une partie reconnaissable de $(A^*)_{\text{reg}}$. Supposons

en effet qu'il existe un semi-groupe fini F et un morphisme surjectif $\alpha : (A^*)_{reg} \rightarrow F$ vérifiant $A^* = \alpha^{-1} \circ \alpha(A^*)$. Si $a \in A$ on aura $\alpha(a^n) = \alpha(a^n \bar{a}^n a^n) = \alpha(a^n) \alpha(\bar{a})^n \alpha(a^n)$ pour tout $n \geq 1$. En choisissant n de telle sorte que $\alpha(\bar{a})^n$ soit idempotent on obtient $\alpha(a^n \bar{a}^{2n} a^n) = \alpha(a^n) \in \alpha(A^*)$ d'où $a^n \bar{a}^{2n} a^n \in A^*$, contradiction.

Enfin le théorème 3.5 et le lemme 5.2 fournissent immédiatement le :

THÉORÈME 5.4 : *Si $R \in \text{Rat}(A^*)_{reg}$ il en est de même de $c(R)$.*

Passons maintenant au monoïde inversif libre.

Soit τ la congruence engendrée sur $(A \cup \bar{A})^*$ par la relation :

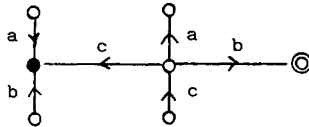
$$\{ (u\bar{u}u, u) / u \in (A \cup \bar{A})^+ \} \cup \{ (u\bar{u}\bar{v}\bar{v}, v\bar{v}u\bar{u}) / u, v \in (A \cup \bar{A})^+ \}.$$

Le monoïde quotient $FI_A = (A \cup \bar{A})^* / \tau$ est le monoïde inversif libre sur A (cf. [6]). Nous noterons $\mu : (A \cup \bar{A})^* \rightarrow FI_A$ l'homomorphisme surjectif canonique.

On peut identifier FI_A avec l'ensemble des arbres orientés non vides bipoints étiquetés sur A ne contenant pas de sous-graphe du type $0 \xrightarrow{a} 0 \xleftarrow{a} 0$ ou $0 \xleftarrow{a} 0 \xrightarrow{a} 0$ ($a \in A$) (cf. [5]). De tels arbres seront appelés *arbres-mots*.

Exemple : L'arbre-mot représenté ci-contre est associé à la classe :

$$\mu(\bar{a}ab\bar{b}\bar{c}\bar{a}\bar{a}\bar{c}cb) = \mu(\bar{c}\bar{b}\bar{c}\bar{c}a\bar{a}\bar{c}\bar{b}\bar{b}\bar{a}\bar{a}\bar{c}\bar{b}).$$



On obtient les mots de la τ -classe associée à un arbre-mot donné comme les étiquettes des chemins couvrants d'origine le nœud d'entrée (symbolisé ici par ●) et d'extrémité le nœud de sortie (symbolisé par ⊙), une arête $0 \xrightarrow{a} 0$ étant lue a lorsqu'elle est parcourue dans le sens de la flèche, et \bar{a} lorsqu'elle est parcourue dans le sens opposé.

On note $(T(\xi), \alpha(\xi), \beta(\xi)) = (T(w), \alpha(w), \beta(w))$ l'arbre-mot associé à l'élément $\xi = \mu(w) \in FI_A$ ($w \in (A \cup \bar{A})^*$) où T représente l'arbre orienté étiqueté, α et β les nœuds d'entrée et de sortie.

Nous dirons qu'un arbre orienté étiqueté T de nœuds x_1, \dots, x_n est une chaîne d'origine x_1 et d'extrémité x_n si l'ensemble de ses arêtes est formé des

couples (x_i, x_{i+1}) ($1 \leq i \leq n-1$). Nous poserons :

$$\text{Ch}(FI_A) = \{ \xi \in FI_A / T(\xi) \text{ est une chaîne} \}.$$

Puisque μ est injectif sur A^* nous identifierons les éléments de A^* avec leurs images dans FI_A . On se persuadera facilement de l'égalité :

$$A^* = \{ \xi \in FI_A / T(\xi) \text{ est une chaîne d'origine } \alpha(\xi)$$

$$\text{et d'extrémité } \beta(\xi) \}.$$

On peut ainsi à chaque $\xi \in \text{Ch}(FI_A)$ associer l'unique mot $w \in A^*$ tel que $T(\xi) = T(w)$. Nous dirons que w est le *support* de ξ et noterons $w = \text{supp}(\xi)$. Le support d'une partie X de $\text{Ch}(FI_A)$ sera l'ensemble $\text{supp}(X) = \{ \text{supp}(\xi) / \xi \in X \}$ et son *socle* l'ensemble

$$\text{socle}(X) = \{ w \in A^* / \text{supp}(\xi) = w \Rightarrow \xi \in X \}.$$

Exemple : L'élément $\xi = \mu(a\bar{a}b\bar{a}a) \in \text{Ch}(FI_{\{a, b\}})$ est représenté par l'arbre-mot $o \xrightarrow{a} \odot \xrightarrow{b} \bullet \xrightarrow{a} o$.

Son support $w = \text{supp}(\xi) = aba$ est représenté par l'arbre-mot $\bullet \xrightarrow{a} o \xrightarrow{b} o \xrightarrow{a} \odot$.

Les relations entre les automate boustrophédon et le monoïde inversif libre sont précisées dans les deux lemmes suivants.

LEMME 5.5 : (1) Si $\mu(w) \in \text{Ch}(FI_A)$ et si $T(w)$ est la chaîne constituée des arêtes (x_i, x_{i+1}) étiquetées par les lettres a_i ($1 \leq i \leq n$) alors w est la trace d'une couverture couvrante de $\text{supp}(\mu(w))$ de section initiale $(a_1 \dots a_{p-1}, a_p \dots a_n)$ et de section finale $(a_1 \dots a_{q-1}, a_q \dots a_n)$ où p et q sont donnés par $\alpha(w) = x_p$ et $\beta(w) = x_q$.

(2) Un mot $w \in (A \cup A)^*$ est la trace d'une couverture d'un mot de A^* ssi $\mu(w) \in \text{Ch}(FI_A)$.

Preuve : Le (1) est évident et fournit la réciproque du (2). D'autre part si w est la trace d'une couverture couvrante du mot $u \in A^*$, la construction de l'arbre-mot $(T(w), \alpha(w), \beta(w))$ associé à w décrite dans [5], p. 391 montre que l'arbre $T(w)$ obtenu est une chaîne de support u . ■

LEMME 5.6 : Si $R \in \text{Rat}(FI_A)$ il existe un automate boustrophédon \mathcal{A} vérifiant :

$$\mathcal{L}^1(\mathcal{A}) = R \cap A^*,$$

$$\mathcal{L}^2(\mathcal{A}) = \text{supp}(R \cap \text{Ch}(FI_A)),$$

$$\mathcal{L}^3(\mathcal{A}) = \text{socle}(R \cap \text{Ch}(FI_A)).$$

Preuve : Soit $R \in \text{Rat}(FI_A)$. μ étant surjectif, il existe un rationnel $R' \in \text{Rat}(A \cup \bar{A})^*$ tel que $R = \mu(R')$. Soit \mathcal{A} un automate fini sur $A \cup \bar{A}$ reconnaissant R' et considérons \mathcal{A} en tant qu'automate boustrophédon sur A .

Soit $w \in A^*$ un mot admettant dans \mathcal{A} une lecture $z \in (A \cup \bar{A})^*$ associée à un calcul réussi et couvrant de section initiale (u, v) et finale (u', v') . On a alors $z \in \mathcal{L}(\mathcal{A}) = \mathcal{L}^0(\mathcal{A}) = R'$ et d'après le lemme 5.5 on a $\mu(z) \in R \cap \text{Ch}(FI_A)$ et $w = \text{supp}(\mu(z))$, d'où l'inclusion $\mathcal{L}^2(\mathcal{A}) \subseteq \text{supp}(R \cap \text{Ch}(FI_A))$. Si $w \in \mathcal{L}^1(\mathcal{A})$ on peut supposer que $(u, v) = (1, w)$ et $(u', v') = (w, 1)$ et le (1) du lemme 5.5 montre qu'alors $\mu(z) \in A^*$, d'où l'inclusion $\mathcal{L}^1(\mathcal{A}) \subseteq R \cap A^*$. Enfin si $w \in \mathcal{L}^3(\mathcal{A})$, on a une lecture z pour toutes les sections (u, v) et (u', v') , ce qui montre que $R \cap \text{Ch}(FI_A)$ contient toutes les chaînes de support w , c'est-à-dire que $w \in \text{socle}(R \cap \text{Ch}(FI_A))$, d'où l'inclusion $\mathcal{L}^3(\mathcal{A}) \subseteq \text{socle}(R \cap \text{Ch}(FI_A))$.

Réciproquement, soit $(T, \alpha, \beta) \in R \cap \text{Ch}(FI_A)$ une chaîne de nœuds x_1, \dots, x_n , de support $w = a_1 \dots a_n$ d'entrée $\alpha = x_p$ et de sortie $\beta = x_q$. Posons $u = a_1 \dots a_{p-1}$, $v = a_p \dots a_n$, $u' = a_1 \dots a_{q-1}$, $v' = a_q \dots a_n$. Soit $z \in R' = \mathcal{L}^0(\mathcal{A})$ tel que $\mu(z) = w$. z est alors l'étiquette d'un chemin réussi dans \mathcal{A} et la trace d'une couverture couvrante de w de section initiale (u, v) et finale (u', v') , ce qui montre que w admet dans le boustrophédon \mathcal{A} un calcul réussi et couvrant de section initiale (u, v) et finale (u', v') . On en déduit immédiatement l'inclusion $\text{supp}(R \cap \text{Ch}(FI_A)) \subseteq \mathcal{L}^2(\mathcal{A})$. Si $(T, \alpha, \beta) = \mu(w) = w \in A^*$ on a $p=1$, $q=n$ et donc w admet dans \mathcal{A} un calcul réussi de section initiale $(1, w)$ et finale $(w, 1)$, d'où l'inclusion $R \cap A^* \subseteq \mathcal{L}^1(\mathcal{A})$. Enfin si $w \in \text{socle}(R \cap \text{Ch}(FI_A))$ on a $(T, \alpha, \beta) \in R \cap \text{Ch}(FI_A)$ pour toutes les valeurs de p et q . On en déduit que w admet un calcul réussi et couvrant dans \mathcal{A} pour toutes les sections initiales (u, v) et finales (u', v') d'où l'inclusion $\text{socle}(R \cap \text{Ch}(FI_A)) \subseteq \mathcal{L}^3(\mathcal{A})$. ■

Le lemme 5.6 permet d'interpréter chacun des théorèmes boustrophédon (th. 3.1, 3.5, 3.6) en une propriété des parties rationnelles du monoïde inversif libre. On obtient en effet de façon évidente le :

THÉORÈME 5.7 : *Si R est une partie rationnelle du monoïde inversif libre FI_A , il en est de même de $R \cap A^*$, $\text{supp}(R \cap \text{Ch}(FI_A))$ et $\text{socle}(R \cap \text{Ch}(FI_A))$.*

On pourra remarquer que, comme dans le cas du semi-groupe de Birget, A^* est une partie rationnelle non reconnaissable de FI_A , ce qui rend non triviale de la rationalité de $R \cap A^*$.

6. CONCLUSION

Nous avons défini trois modes de reconnaissance pour les automates boustrophédon. En ce qui concerne le premier, le théorème de Shepherdson montre que, pour tout boustrophédon \mathcal{A} , on peut construire un automate fini \mathcal{A}_1 vérifiant $\mathcal{L}^1(\mathcal{A}) = \mathcal{L}(\mathcal{A}_1)$. Nous avons étendu ce résultat aux deux autres modes en montrant comment construire des automates finis \mathcal{A}_i vérifiant $\mathcal{L}^i(\mathcal{A}) = \mathcal{L}(\mathcal{A}_i)$ ($i=2, 3$).

On pourrait encore définir deux autres modes de reconnaissance consistant à accepter les mots qui admettent dans \mathcal{A} un calcul réussi et couvrant de section initiale arbitraire (mode 4) ou de section finale arbitraire (mode 5). Le lecteur adaptera facilement les techniques utilisées dans les cas précédents pour construire des automates finis \mathcal{A}_i vérifiant $\mathcal{L}^i(\mathcal{A}) = \mathcal{L}(\mathcal{A}_i)$ ($i=4, 5$).

Lorsque le nombre d'états de l'automate boustrophédon \mathcal{A} est de l'ordre de n , celui des automates \mathcal{A}_i est de l'ordre de 2^{n^2} dans les deux premiers modes et de 2^{2n} dans les trois derniers. Il est facile, dans chacun des cas, de trouver des exemples pour lesquels l'automate \mathcal{A}_i doit contenir de l'ordre de 2^n états, mais nous ne connaissons aucun exemple permettant d'atteindre une des bornes. Un travail resterait donc à faire pour trouver de tels exemples ou optimiser les constructions.

Enfin signalons que le comportement infini des boustrophédon est identique à celui des automates finis et correspond aux parties rationnelles de mots infinis (cf. [8]).

REMERCIEMENTS

Je tiens à remercier R. Cori, S. Margolis, D. Perrin et G. Viennot pour leurs nombreuses suggestions. Je remercie également J. C. Birget qui a bien voulu référer cet article et m'a aidé à préciser quelques idées et quelques notations.

BIBLIOGRAPHIE

1. AHO, HOPCROFT et HULLMAN, *Time and Tape Complexity of Pushdown Automaton Languages*, Information and Control, vol. 13, n° 3, 1968, p. 186-206.
2. BERSTEL, *Transductions and Context-Free Languages*, Teubner, 1979.
3. BIRGET, Ph. D, Un. of California, Berkeley, 1983.
4. EILENBERG, *Automata, Languages and Machines*, Vol. A, Acad. Press, 1974.
5. MUNN, *Free Inverse Semigroup*, Proc. London Math. Soc., (3), vol. 29, 1974, p. 385-404.
6. SCHEIBLICH, *Free Inverse semigroups*, Proc. Amer. Math. Soc., vol. 38, 1973, p. 1-7.
7. SHEPHERDSON, *The Reduction of Two-Way Automata to One-Way Automata*, I.B.M. J. Res., vol. 3, n° 2, 1959, p. 198-200.
8. PÉCUCHET, *Automates Boustrophédon et mots infinis*, T.C.S. (à paraître).