REVUE FRANÇAISE D'AUTOMATIQUE INFORMATIQUE
RECHERCHE OPÉRATIONNELLE. INFORMATIQUE THÉORIQUE

M. P. SCHÜTZENBERGER

Sur les monoïdes finis dont les groupes sont commutatifs

Revue française d'automatique informatique recherche opérationnelle. Informatique théorique, tome 8, n° R1 (1974), p. 55-61

<http://www.numdam.org/item?id=ITA_1974__8_1_55_0>

R.A.I.R.O.
(8e année, R-1, 1974, p. 55-61)

SUR LES MONOIDES FINIS DONT LES GROUPES SONT COMMUTATIFS

par M. P. SCHÜTZENBERGER (1)

Communiqué par J.-F. PERROT

Résumés. — *On examine une caractérisation des parties reconnaissables dont les groupes dans le monoïde syntactique sont commutatifs.*

1. INTRODUCTION

Appelons *groupe latent* $Gp(F)$ d'une partie quelconque F d'un semi-groupe S, le plus petit groupe que divisent tous les groupes dans le semi-groupe syntactique de F. Nous nous proposons ici d'appliquer la théorie de la décomposition de Krohn et Rhodes pour examiner la famille $\mathbf{C}$ des parties reconnaissables au sens de Eilenberg ([1], c'est-à-dire des parties dont le monoïde syntactique est fini) d'un monoïde libre A^* dont le groupe latent appartient à une variété donnée Γ de groupes commutatifs finis.

Le cas où A^* est remplacé par un monoïde commutatif libre a été traité par J.-F. Perrot en 1965 ([2]).

La famille $\mathbf{C}$ elle-même a été introduite par McNaughton en 1960 ([3]) à l'intérieur d'une problématique différente.

Pour alléger les énoncés, nous dirons qu'une famille $\mathcal{F}$ de parties est *fermée par les opérations polynomiales* ssi elle contient la partie vide, la partie $\{1\}$ et toutes les parties de l'alphabet et si elle contient en outre l'union et le produit de deux quelconques de ses membres.

Étant donnés deux alphabets B et A et une famille $\mathcal{F}$ de parties de A^*, une *$\mathcal{F}$-substitution élémentaire* α est un morphisme de B^* dans le monoïde des parties de A^* qui satisfait les conditions suivantes :

(1) α est injectif en ce sens que les images par α de deux lettres distinctes de B sont disjointes ;

(1) Groupe d'Informatique théorique, UER de Mathématiques Université de Paris-VII.

(2) Il existe une partie A' de A telle que $B\alpha$ soit contenue dans $(A/A')^* A'$;

(3) $b\alpha$ appartient à la famille $\mathcal{F}$ pour chaque lettre b.

Pour simplifier, nous supposerons toujours A infini dénombrable et nous nous limiterons aux substitutions élémentaires $\alpha : A^* \to A^*$ telles que $a\alpha$ soit non vide pour un nombre *fini* de lettres de A.

Enfin, une partie préfixe ayant un délai de synchronisation fini est une partie P du semi-groupe libre $A^+ = AA^*$ qui est préfixe ($P \cap PA^+ = \varnothing$) et pour laquelle il existe un entier fini d (le délai de synchronisation) tel que l'on ait la relation :

$$a \in A^*, p \in P^d, apA^* \cap P^* \neq \varnothing \Rightarrow ap \in P^*.$$

Cette notion, sous une forme ou une autre a été souvent considérée (cf. [4], [5], [6]). En particulier quand α est une substitution élémentaire, l'image par α de l'alphabet A est préfixe avec un délai de synchronisation un (ou zéro si $A\alpha = A$).

Première caractérisation. La famille $\mathcal{C}$ est la plus petite famille $\mathcal{C}'$ fermée par les opérations polynomiales et les $\mathcal{C}'$-substitutions élémentaires qui contienne tous les noyaux $1\gamma^{-1}$ où γ est un morphisme dans un groupe de Γ d'un monoïde libre $A'^*(A' \subset A$, fini).

Notons $N(F)$ l'ensemble des entiers qui sont l'ordre d'au moins un groupe cyclique appartenant à Γ. Comme Γ est une variété de groupes commutatifs finis, elle est définie de façon unique par la donnée de $N(\Gamma)$.

Deuxième caractérisation. La famille $\mathcal{C}$ est la plus petite famille $\bar{\mathcal{C}}$ fermée par les opérations polynomiales qui contienne toutes les intersections finies $\cap \{ (P_i^{n_i})^* : 1 \leqslant i \leqslant m \}$ où les n_i appartiennent à $N(\Gamma)$ et où les P_i sont des parties préfixes ayant un délai de synchronisation fini et appartenant à $\bar{\mathcal{C}}$.

II. DELAI DE SYNCHRONISATION FINI

Observons d'abord que si P est une partie préfixe ayant un délai de synchronisation d, toute relation $apa' \in P^*$, où $a, a' \in A^*$ et $p \in P^d P^*$ implique $ap, a' \in P^*$. En effet, on peut écrire $p = p'p''$ où $p' \in P^d$ et $p'' \in P^*$ et la relation $ap'p''a' \in P^*$ entraîne d'abord $ap' \in P^*$, d'après l'hypothèse de synchronisation, ensuite $p''a', a' \in P^*$ d'après l'hypothèse que P est préfixe.

Rappelons maintenant l'énoncé suivant dans lequel B est un alphabet quelconque, Q une partie préfixe dans B^+ ayant un délai de synchronisation fini d' et α un morphisme injectif de B^* dans A^* tel que $B\alpha$ soit contenue dans une partie préfixe P ayant un délai de synchronisation fini d.

II.1. La partie $R = Q\alpha$ est préfixe et a un délai de synchronisation $d + d'$.

Preuve : Il est trivial que R soit préfixe. Considérons un élément de $R^{d+d'}$. On peut l'écrire comme un produit rr' où $r \in R^d$ et $r' \in R^{d'}$. Supposons que $a, a' \in A^*$ satisfassent $arr'a' \in R^*$. Comme R est contenu dans P^+ ceci entraîne que $arr'a'$ soit dans P^* et que r soit dans P^dP^*, donc, comme on l'a vu ci-dessus, que ar, r' et a' appartiennent tous à P^*. Leurs images inverses par α sont donc des mots b, q' et b' de B^* où $q' = r'\alpha^{-1}$ est dans $Q^{d'}$ et l'on a $bq'b' \in Q^*$. D'après notre hypothèse sur Q, on en conclut que bq' et b' sont dans Q^* et enfin que $arr' = bq'\alpha$ et $a' = b'\alpha$ sont dans R^*.

Q.E.D.

Corollaire II.2. : Soient $\{\alpha_i : 1 \leqslant i \leqslant \kappa\}$ un ensemble fini de substitutions élémentaires $\alpha_i : A^* \to A^*$. L'image P de A par la substitution produit $\alpha_1\alpha_2 \ldots \alpha_k$ est préfixe et a un délai de synchronisation κ.

Preuve : Ceci résulte immédiatement de l'énoncé précédent et de ce que pour toute partie A' de A, l'ensemble $(A/A')^*A'$ est une partie préfixe ayant délai un.

Q.E.D.

REMARQUE : Les mêmes techniques permettent sans peine de vérifier que P est contenue dans une partie préfixe Q ayant un délai de synchronisation fini et satisfaisant les conditions supplémentaires suivantes :

(1) $a \in A^* \Rightarrow aA^* \cap A^* \neq \varnothing$;

(2) Il existe un entier fini κ tel que :

$$a_1, a_2, \ldots, a_k \in A^*,$$
$$a_1a_2, a_2a_3, \ldots, a_ja_{j+1}, \ldots, a_{k-1}a_k \in Q^* \Rightarrow a_k \in Q^*,$$

(3) Si X est une partie de A^* non contenue dans Q^*, il existe au moins un $x \in X$ tel que $X^*x \cap Q^* = \varnothing$.

Nous considérons maintenant une partie préfixe P ayant un délai de synchronisation fini d et une application γ de P dans un groupe G. Celle-ci se prolonge en un morphisme de P^* dans G. On sait que le noyau $1\gamma^{-1}$ est un sous-monoïde de P^* engendré par une partie préfixe R de P^+. Nous nous proposons d'examiner le groupe latent de R^* au moyen de la méthode des produits en couronne de Krohn et Rhodes.

Pour cela, nous considérons d'abord l'automate minimal reconnaissant les parties $P_g = P \cap g^{-1}\gamma (g \in G)$. Soient Q son ensemble d'états, q_1 son état initial et Q_+ l'ensemble de ses états finaux.

Comme P est préfixe, l'état q_1 n'est pas contenu dans Q_+ et l'on a $Q_+a = \varnothing$ pour tout mot $a \neq 1$. De plus, il existe une bijection entre les parties $P_g(g \in G)$ et les états finaux qui permet de définir pour chaque $q \in Q_+$, l'élément $q\gamma$ de G égal à l'image par γ des mots de $q_1^{-1}q$ $(= \{ a \in A^* : q_1 a = q \})$.

Posons $\bar{Q} = (G \times Q/Q_+)$ et définissons une action de A^* sur $\bar{Q}$ en posant pour chaque lettre a et chaque état $(g, q) \in \bar{Q}$:

$$\begin{aligned} (g, q)a &= (g \cdot q\gamma, q_1) && \text{si} \quad qa \in Q_+ \\ &= (g, qa) && \text{si} \quad qa \in Q/Q_+ \\ &= \varnothing && \text{si} \quad qa = \varnothing. \end{aligned}$$

Il est facile de voir que cette définition entraîne que pour tout mot a, on ait :

$(g, q)a = (g \cdot p\gamma, q_1)a''$ si $a = a'pa''$ où $a' \in q^{-1}Q_+$, p est le plus long facteur dans P^* de $a'^{-1}a$ et $a'' = (a'p)^{-1}a$.

$(g, q)a = (g, qa)$ ou $\varnothing$ si a n'a aucun facteur gauche dans $q^{-1}Q_+$.

En particulier, quels que soient $g \in G$ et $a \in A^*$, l'état $(g, q_1)a$ appartient à (G, q_1) ssi a est dans P^*. On en conclut que l'automate ainsi défini (avec $(1, q_1)$ comme état initial et final) reconnaît R^* et on vérifie facilement qu'il est minimal.

II.3. Tout groupe maximal dans le monoïde syntactique $A^*\sigma$ de R est isomorphe à G ou à un groupe dans le monoïde syntactique M dans $P_g(g \in G)$.

Preuve : Soient $H \subset A^*$ l'image inverse d'un groupe maximal $H\sigma$ dans $A^*\sigma$ et $\bar{Q}_H = Q(H \cap A^+)$.

Supposons d'abord que pour tout $(g, q) \in \bar{Q}_H$ et tout $h \in H$ on ait $qh \neq \varnothing$. D'après notre définition de l'action de A^* sur $\bar{Q}$ ceci implique que l'on ait $(g, q)h = (g, qh)$ identiquement pour tout $h \in H$ et $(g, q) \in \bar{Q}_H$ et par conséquent que $H\sigma$ soit isomorphe à un groupe dans le monoïde M.

Dans le cas contraire il existe un facteur gauche $a \neq 1$ d'un mot $ab \in H$ et un état (g, q) de $\bar{Q}_H$ tels que $qa = q_+$. Prenons un mot $c \in H$ tel que $(abc)\sigma$ soit idempotent. Les relations

$$(g, q)a = (g', q_1) = (g, q)abca = (g', q_1)bca$$

montrent que $bca \in P^+$.

Il existe donc des mots f, g tels que $gf \in P^dP^*$, $fg \in H$, et que les images par σ de fg et gf soient idempotents. Comme $(fgHfg)\sigma = H\sigma$ l'image par σ du monoïde $H' = gHf$ est un groupe isomorphe à $H\sigma$ et il existe un $p \in P^dP^*$

tel que son image par σ soit l'idempotent de ce groupe. Nous pouvons donc supposer pour simplifier que $H' = H$, c'est-à-dire que $H \cap P^dP^*$ contient un mot p tel que $p\sigma = pp\sigma$.

De nouveau, $H\sigma = Hp\sigma$ et par conséquent $\bar{Q}_H$ est contenu dans $\bar{Q}_p$. Soit $(g, q) \in \bar{Q}_H$. Il existe des mots a, a' tels que $q_1 a = q$, $qa' = q_+$. D'après $(g, q)p = (g, q)$ on a $apa' \in P^*$ et comme $p \in P^dP^*$ où d est le délai de synchronisation de P, on en conclut que $ap \in P^*$ c'est-à-dire que $q = q_1$. Donc $\bar{Q}_H$ est contenu dans (G, q_1), ce qui entraîne que H soit contenu dans P^*.

Q.E.D.

Corollaire II.4. : Si P préfixe a un délai de synchronisation borné, le groupe latent de $(P^r)^*$ divise le produit direct du groupe latent de P par le groupe cyclique $\mathbf{Z}_{(r)}$.

Preuve: Immédiate.

Corollaire II.5. : Les familles $\mathcal{C}'$ et $\bar{\mathcal{C}}$ de l'Introduction sont contenues dans $\mathcal{C}$.

Preuve : Compte tenu de II.2 et de II.4, ceci résulte immédiatement de la fermeture de $\mathcal{C}$ par rapports aux opérations polynomiales et aux opérations booléennes.

Q.E.D.

III. FIN DE LA VERIFICATION

Pour achever la preuve, il nous suffit de considérer une partie $F \in \mathcal{C}$ et de montrer qu'elle appartient aux familles $\mathcal{C}'$ et $\bar{\mathcal{C}}$. Nous notons σ le morphisme syntactique de F et comme les familles $\mathcal{C}$, $\mathcal{C}'$ et $\bar{\mathcal{C}}$ sont fermées par rapport à l'union, nous supposons que $F\sigma$ est un singleton.

III.1. Si $F\sigma$ est un groupe G, la partie F appartient à $\mathcal{C}'$ et $\bar{\mathcal{C}}$.

Preuve : Soit $K = 1^{-1}\sigma$ le noyau de σ. Tout mot b admet une factorisation $b = \kappa_1 a_1 \kappa_2 a_2 \dots \kappa_n a_n \kappa_{n+1}$ où $a_1, a_2, \dots, a_n \in A$; $\kappa_1, \kappa_2, \dots, \kappa_{n+1} \in K$ et où chaque κ_i est défini comme le plus long facteur dans K du mot b_i tel que $\kappa_1 a_1 \dots a_{i-1} b_i = b = b_1$.

Cette condition implique que pour chaque $m \leqslant n$ tous les produits $(a_i a_{i+1} \dots a_m)\sigma\ (1 \leqslant i \leqslant m)$ soient différents. Donc $n < \text{Card } G$ établissant que F est dans la fermeture polynomiale de K, et, par définition, l'inclusion $F \in \mathcal{C}'$. En ce qui concerne $\bar{\mathcal{C}}$, l'hypothèse que G est un groupe commutatif fini implique que G soit groupe quotient d'un groupe $G' \in \Gamma$ dont le noyau K'

est défini par une partition $\{A_i\}$ de A, des entiers $n_i \in N(\Gamma)$ et la condition qu'un mot a appartienne à K' ssi le nombre $|a|_i$ de lettres de A_i figurant dans a est pour chaque i un multiple de n_i. On peut donc supposer que $G = G'$ et $K = K'$ et posant $P_i = (A/A_i)^* A_i$, on obtient K comme l'intersection des monoïdes $(P_i^{n_i})^*$.

Corollaire III.2. : Si $1 \in F$ on a $F \in \mathcal{C}' \cap \bar{\mathcal{C}}$.

Preuve : Comme $F\sigma$ est un singleton, l'hypothèse $1 \in F\sigma$ entraîne $F\sigma = 1$. Il existe donc une partie A' de A telle que F soit contenue dans A'^* et que la restriction de σ à A'^* soit un morphisme dans un groupe.

Q.E.D.

Nous pouvons désormais supposer que F ne contient pas 1 ce qui entraîne que son monoïde syntactique soit l'union disjointe d'un élément neutre et d'un semi-groupe S image de A^+ par σ.

III.3. Si $A\sigma$ est un singleton ou si S est un semi-groupe $\mathfrak{L}$-simple on a $F \in \mathcal{C}' \cap \bar{\mathcal{C}}$.

Preuve : Si $A\sigma$ est un singleton, F est de la forme $A^k (k \in \mathbf{N})$ ou $A^k (A')^* (r \in \mathbf{N}(\Gamma))$ et le résultat est établi puisque A est une partie préfixe ayant un délai de synchronisation zéro.

Si S est $\mathfrak{L}$-simple, il existe un morphisme γ de A^+ dans un groupe G de Γ et une partition $\{A_i\}$ de A tels que F soit une union finie de termes de la forme $A_i(g\gamma^{-1})(g \in G)$.

Q.E.D.

Nous faisons maintenant intervenir le lemme classique de Krohn et Rhodes. Il affirme qu'en dehors des trois cas traités dans III.1, III.2 et III.3 il existe une partie A' de l'alphabet A ayant les propriétés suivantes :

(1) Posant $B = (A/A')^* A'$, on a $B^+\sigma \neq S$;

(2) Si $(A/A')\sigma$ n'est pas un singleton on a $(A/A')^+\sigma \neq S$.

Comme tout mot de A^* admet exactement une factorisation comme produit d'un mot de B^* par un mot de $(A/A')^*$, il en résulte que F est une union finie disjointe de produits non ambigus de la forme GF' où $G \subset B^*$, $F' \subset (A/A')^*$ et où $G\sigma$ et $F'\sigma$ sont des singletons.

Procédant par induction sur Card $(A^*\sigma)$, nous en concluons que compte tenu du premier cas de III.3, la partie F' appartient à $\mathcal{C}'$ et $\bar{\mathcal{C}}$.

Montrons qu'il en est de même pour G. Pour ce faire, prenons une partie A_1 de A en correspondance bi-univoque (par β) avec les paires dans

$((A/A')^*\sigma, A'\sigma)$. Pour chaque $a \in A_1$ nous définissons $a\alpha$ comme le produit de $s\sigma^{-1} \cap (A/A')^*$ par $s'\sigma^{-1} \cap A'$ si $a\beta = (s, s')$. Comme on vient de le voir, $a\alpha$ appartient à $\mathcal{C}' \cap \bar{\mathcal{C}}$ et par conséquent α peut être considérée comme une $\mathcal{C}' \cap \bar{\mathcal{C}}$-substitution élémentaire. Par définition, il existe une partie $G_1 \subset A_1^*$ telle que $G_1\alpha = G$. D'après l'hypothèse d'induction et $B^+\sigma \neq S$, on a $G_1 \in \mathcal{C}' \cap \bar{\mathcal{C}}$ ce qui établit directement $G \in \mathcal{C}'$ et qui donne $G \in \bar{\mathcal{C}}$ moyennant le corollaire II.2.

Q.E.D.

REMARQUE : La technique de factorisation de III.1 empêche (hormis certains cas particuliers) que les parties obtenues le soient de manière non ambiguë. Il serait évidemment possible d'obtenir ce résultat en admettant comme données dans la deuxième caractérisation toutes les intersections finies de parties de la forme

$$(P_i^{n_i})^* P_i^{m_i} (0 \leqslant m_i < n_i, n_i \in \mathbf{N}(\Gamma),$$

$P_i \in \bar{\mathcal{C}}$, préfixe ayant un délai de synchronisation fini). Inversement, on pourrait (en perdant la non-ambiguïté) remplacer les intersections par des « produits de mélange ». On notera que la deuxième caractérisation n'utilise pas le fait que l'alphabet A soit de cardinalité non bornée.

REFERENCES

[1] EILENBERG S., Livre sous presse, vol. 1.

[2] PERROT J.-F., *Sur quelques familles de parties des monoïdes abéliens libres*, C.R. Acad. Sc. Paris, 1965, *261*, 3008-3011.

[3] MCNAUGHTON R., *Symbolic Logic for automata.* Wright Air Dept. Div. Tech. Note. 60-244. Cincinnati. Ohio 1960.

[4] NEUMAN P. G., *On error limiting Codes.* IRE Trans. IT, 1963, *9*, 209-212.

[5] WINOGRAD S., *Input error limiting Automata.* IBM Research Report. RC 966, 1963.

[6] GOLOMB S. W. et GORDON B., *Codes with bounded synchronization delay.* Information and Control, 1965, *8*, 355-372.