Sur le monoide syntactique de L^* lorsque L est un langage fini

Jean-Éric Pin[1]

Résumé

We prove that for every finite monoid M, there exists a finite language L such that M divides the syntactic monoid of L^*. Moreover, one can choose for L a full finite prefix code. The same result for finite group has already been proved by Schützenberger. This result is crucial in the proof of the J.-F. Perrot's theorem that the only variety closed under star operation is the variety of all rational languages.

1 Introduction

Une série de recherches se sont attachées à l'étude des propriétés syntactiques de l'opération étoile. En particulier, les travaux de Restivo [3], de Schützenberger [5] et de Perrin et Schützenberger [1] ont montré l'intérêt de l'étude des groupes contenus dans le monoïde syntactique M de L^*. On sait par exemple que si L est fini, le groupe des éléments inversibles de M est cyclique [1]. En outre, Schützenberger [4, 6] a montré que tout groupe divisait le monoïde syntactique d'un langage L^*, L étant lui-même un langage fini. Nous montrons ici que cette propriété est vraie non seulement pour les groupes finis, mais aussi pour tous les monoïdes finis. On peut en outre choisir pour L un code préfixe et même un code préfixe complet fini. Ce résultat intervient dans la démonstration du fait que la seule variété fermée par étoile est la variété des langages rationnels. Sur ce sujet, et pour de plus amples commentaires, nous renvoyons le lecteur à l'article de Perrot [2] qui précède.

Theorem 1.1 *Pour tout monoïde fini M, il existe un langage fini L tel que M divise le monoïde syntactique de L. En outre, on peut choisir pour L un code préfixe complet fini.*

Ce théorème repose sur une proposition d'aspect plus technique, mais qui admet plusieurs corollaires intéressants. Avant de l'énoncer, introduisons quelques notations : E_n désignera l'ensemble $\{0, 1, \ldots, n-1\}$ et $\mathscr{T}_n$ le monoïde des applications de E_n dans E_n. Si Σ est un ensemble d'applications de E_n dans E_n, nous noterons $T(\Sigma)$ le sous-monoïde de $\mathscr{T}_n$ engendré par Σ. On sait que tout monoïde fini est isomorphe à un tel $T(\Sigma)$. Enfin $X_\sigma = \{x\} \cup \{x_\sigma\}_{\sigma\in\Sigma}$ designera un alphabet de $(1 + \operatorname{Card}\Sigma)$ lettres. À chaque ensemble Σ, on associe un code

1. LIAFA, CNRS-Informatique Théorique et Programmation, Tour 55, 4 Place Jussieu, 75230 Cedex, France.

prefixe fini $A(\Sigma)$ sur l'alphabet X_σ en posant :

$$A(\Sigma) = \{x^n\} \cup \Big(\bigcup_{\sigma \in \Sigma} \{x^i x_\sigma x^{n-\sigma(i)} \mid i \in E_n\} \Big)$$

On peut alors énoncer la

Proposition 1.2 *Soit n un entier quelconque et $\Sigma \subseteq \mathscr{T}_n$. Pour tout alphabet X tel que $X \supseteq X_\Sigma$ et pour tout langage fini L sur l'alphabet X vérifiant $L \cap (x^* X_\Sigma x^* \cup \{e\}) = A(\Sigma)$, le monoïde $T(\Sigma)$ divise le monoïde syntactique de L^*.*

Nous en déduirons le résultat de Schützenberger [4, 6] déjà cité, ainsi que les Corollaires 1.4 et 1.5 qui suivent.

Corollary 1.3 (Schützenberger) *Le groupe symétrique de degré n $\mathfrak{S}_\mathfrak{n}$ divise le monoïde syntactique de A_n^*, où A_n est le code préfixe fini :*

$$A_n = \{x^n, xy, yx^{n-1}\} \cup \{x^i y x^{n-i} \mid 2 \leqslant i \leqslant n-1\}.$$

Corollary 1.4 *Le monoïde $\mathscr{T}_n$ divise le monoïde syntactique de B_n^* où B_n est le code préfixe fini :*

$$B_n = \{x^n, xy, z, x^{n-1}z, yx^{n-1}\} \cup \{x^i y x^{n-i} \mid 2 \leqslant i \leqslant n-1\} \cup \{x^i z x^{n-i} \mid 1 \leqslant i \leqslant n-2\}.$$

Rappelons que si A est un code préfixe fini, l'ensemble P des préfixes de A est l'ensemble des mots u de X^* tels qu'il existe $v \in X^+ = X^* \setminus \{e\}$ vérifiant $uv \in A$. On vérifie alors sans peine que $C = PX - P$ est un code préfixe complet *fini* contenant A, appelé *complété* de A. (C'est en fait l'unique code préfixe complet contenant A et de cardinal minimal pour cette propriété.) On a alors le

Corollary 1.5 *Si $C(X)$ est le complété du code $A(\Sigma)$, le monoïde $T(\Sigma)$ divise le monoïde syntactique du langage $C(\Sigma)^*$.*

2 Démonstration de la proposition 1.2

Si u est un mot de X^*, nous noterons $u^{-1}L^*$ l'ensemble $u^{-1}L^* = \{m \in X^* \mid um \in L^*\}$. On sait alors que les états de l'automate minimal reconnaissant L^* s'identifient aux ensembles $u^{-1}L^*$, u parcourant X^*, les transitions de ce même automate minimal étant données par les formules :

$$(u^{-1}L^*) \cdot v = v^{-1}(u^{-1}L^*) = (uv)^{-1}L^*.$$

Puisque L est supposé fini, il existe un entier a tel que an majore la longueur de tous les mots de L. Les résiduels $(x^{an+i})^{-1}L^*$, pour $i \in E_n$, sont alors deux à deux distincts. En effet, on a $x^n \in L$ d'après la définition de L. Par conséquent $x^{an+i}x^{n-i} \in L^*$ et $x^{n-i} \in (x^{an+i})^{-1}L^*$, pour $i \in E_n = \{0, 1, \cdots, n-1\}$. Maintenant, si $x^{n-i} \in (x^{an+j})^{-1}L^*$ pour un $j \in E_n$, il vient $x^{an+j+n-i} \in L^* \cap x^* = (x^n)^*$ d'où $j \equiv i \bmod n$ et finalement $j = i$ puisque i et j sont éléments de E_n : cela démontre que $x^{n-i} \notin (x^{an+j})^{-1}L^*$ si $i \neq j$ $(i, j \in E_n)$.

La suite résultera du lemme suivant :

Lemma 2.1 *On a la formule :*

$$\forall i \in E_n, \quad \forall \sigma \in \Sigma \quad [(x^{an+i})^{-1}L^*] \cdot (x_\sigma x^{an}) = (x^{an+\sigma(i)})^{-1}L^*.$$

On a en effet l'équivalence :

$$m \in [(x^{an+i})^{-1}L^*] \cdot (x_\sigma x^{an}) \iff w = x^{an+i}x_\sigma x^{an}m \in L^*.$$

Mais puisque an majore la longueur des mots de L^*, le mot w se factorise en $w = uv$, avec $u = x^{an+i}x_\sigma x^p \in L^*$, $v = x^q m \in L^*$ et $p + q = an$. On a alors $u \in L^* \cap x^* X_\Sigma x^*$ et donc $u = u_1 \cdots u_r$, avec $u_i \in L \cap x^* X_\Sigma x^* \subseteq A(\Sigma)$. Autrement dit, $u \in A(\Sigma)^*$ et puisque $A(\Sigma)$ est un code préfixe, la décomposition de u ne peut être que $u = (x^n)^a(x_i x_\sigma x^{n-\sigma(i)})(x^n)^k$ avec $p = kn+(n-\sigma(i))$. Il en résulte, puisque $p+q = an$, $q = (a-k-1)n+\sigma(i)$. On en déduit successivement, en utilisant le fait que L contient $A(\Sigma)$:

$$\begin{aligned} w \in L^* &\implies v = x^q m \in L^* \implies x^{(k+1)n}x^q m \in L^* \\ &\implies x^{an+\sigma(i)}m \in L^* \implies m \in (x^{an+\sigma(i)})^{-1}L^*. \end{aligned}$$

Réciproquement si $m \in (x^{an+\sigma(i)})^{-1}L^*$ on a $x^{an+\sigma(i)}m \in L^*$ et le même raisonnement permet de conclure que $x^{(a-1)n+\sigma(i)}m \in L^*$. On a alors :

$$\begin{aligned} x^{(a-1)n+\sigma(i)}m \in L^* &\implies x^{an}(x_i x_\sigma x^{n-\sigma(i)})x^{(a-1)n+\sigma(i)}m \in L^*, \\ &\implies w = x^{an+i}x_\sigma x^{an}m \in L^*, \end{aligned}$$

ce qui achève la preuve du lemme.

Revenons à la proposition. Soit M le monoïde syntactique de L^* et ξ le morphisme syntactique de $X^* \to M$. On sait que M est aussi le monoïde de transition de l'automate minimal. Autrement dit, M est représenté comme monoïde de transformation. Les éléments $m_\sigma = \xi(x_\sigma x^{an})$ $(\sigma \in \Sigma)$ engendrent un sous-monoïde N de M, dont la restriction à l'ensemble des états $q_i = (x^{an+i})^{-1}L^*$, $i \in E_n$ — états qui sont deux à deux distincts comme on l'a vu — est décrite par les formules du lemme. Compte tenu des notations que nous venons d'introduire, ces formules peuvent s'écrire :

$$\forall i \in E_n, \ \forall \sigma \in \Sigma \quad q_i m_\sigma = q_{\sigma(i)}.$$

Cela montre que l'ensemble $\{q_i \mid i \in F_n\}$ est stable sous N et que la relation d'équivalence $\sim$ définie par $a \sim b$ si et seulement si pour tout $i \in E_n$, $q_i a = q_i b$, définit sur N une congruence telle que le quotient $N/\sim$ soit isomorphe à $T(\Sigma)$. Il en résulte que $T(\Sigma)$ est isomorphe à un quotient d'un sous-monoïde de M, ce qui achève la démonstration de la proposition.

Démonstration des Corollaires 1.3 et 1.4. Si $n = 1$, les deux résultats sont évidents. Si $n \geqslant 2$, le groupe $\mathfrak{S}_n$ est engendré par la permutation circulaire $\sigma_0 = (0, 1, \ldots, n-1)$ et par la transposition $\tau = (0, 1)$ et le monoïde $\mathscr{T}_n$ est engendré par σ_0, τ et par l'application ρ définie par : pour tout $i \in E_n$, $\rho(i) = [i]_{n-1}$ où $[a]_b$ désigne le reste de la division de a par b. La fin de la preuve s'appuie

sur deux remarques : tout d'abord, on vérifie facilement que la Proposition 1.2 s'applique encore si on remplace $A(\Sigma)$ par le code $A'(\Sigma)$ suivant :

$$A'(\Sigma) = \{x^n\} \cup \Big(\bigcup_{\sigma\in\Sigma} \{x^i x_\sigma x^{[n-\sigma(i)]_n} \mid i \in E_n\} \Big)$$

Ensuite, en reprenant les notations de la proposition, on constate que pour tout $i \in E_n$, $q_i \cdot \xi(x) = q_{[i+1]_n}$ et donc $\xi(x)$ a la même action que m_{σ_0}. Par conséquent, si l'ensemble Σ contient σ_0, la Proposition 1.2 s'appliquera encore si on remplace $A(\Sigma)$ par le code $A(\Sigma - \{\sigma_0\})$ ou par $A'(\Sigma - \{\sigma_0\})$.

Il reste à évaluer $A'(\Sigma - \{\sigma_0\})$ dans les deux cas qui nous intéressent. Or on a vu que pour le groupe symétrique $\mathfrak{S}_n$, $\Sigma = \{\sigma_0, \tau\}$ et pour le monoïde $\mathscr{T}_n$, $\Sigma = \{\sigma_0, \tau, \rho\}$. Il vient donc dans le premier cas $A'(\Sigma - \{\sigma_0\}) = A'(\{\tau\})$ et dans le second cas $A'(\Sigma - \{\sigma_0\}) = A'(\{\tau, \rho\})$. Or si on pose $x_r = y$ et $x_p = z$, on constate que les codes $A'(\{\tau\})$ et $A'(\{\tau, \rho\})$ sont égaux respectivement à A_n et B_n ce qui démontre les Corollaires 1.3 et 1.4.

Démonstration du Corollaire 1.5. Il suffit de vérifier que nous sommes dans les conditions d'application de la proposition, i.e. que :

$$C(\Sigma) \cap (x^* X_\Sigma x^* \cup \{e\}) = A(\Sigma).$$

Puisque $e \notin C(\Sigma)$ et que $A(\Sigma) \subseteq C(\Sigma) \cap x^* X_\Sigma x^*$, il suffit de prouver l'inclusion $C(\Sigma) \cap x^* X_\Sigma x^* \subseteq A(\Sigma)$. Or par définition $C(\Sigma) = PX - P$, où P est l'ensemble des préfixes de $A(\Sigma)$:

$$P = \{x^i \mid i \in E_n\} \cup \Big(\bigcup_{\sigma\in\Sigma} \{x^i x_\sigma x^j \mid 0 \leqslant j < n - \sigma(i), i \in E_n\} \Big).$$

D'où :

$$PX \cap x^* X_\Sigma x^* = \{x^i \mid 0 \leqslant i \leqslant n\} \cup \Big(\bigcup_{\sigma\in\Sigma} \{x^i x_\sigma x^j \mid 0 \leqslant j \leqslant n - \sigma(i), i \in E_n\} \Big)$$

et finalement

$$\begin{aligned} (PX - P) \cap x^* X_\Sigma x^* &= (PX \cap x^* X_\Sigma x^*) - P \\ &= \{x^n\} \cup \Big(\bigcup_{\sigma\in\Sigma} \{x^i x_\sigma x^{n-\sigma(i)} \mid i \in E_n\} \Big) \end{aligned}$$

soit encore $C(\Sigma) \cap x^* X_\Sigma x^* \subseteq A(\Sigma)$ comme annoncé.

Démonstration du Théorème 1.1. Elle résulte immédiatement du Corollaire 1.5.

Remerciements

Je tiens à remercier ici Dominique Perrin pour ses nombreuses suggestions lors de la rédaction de cet article.

Références

[1] D. Perrin and M.-P. Schützenberger, Codes et sous-monoïdes possédant des mots neutres, in *Theoretical computer science (Third GI Conf., Darmstadt, 1977)*, pp. 270–281, *Lect. Notes Comp. Sci.* vol. 48, Springer, Berlin, 1977.

[2] J.-F. Perrot, Variétés de langages et operations, *Theoret. Comput. Sci.* **7** (1978), 197–210.

[3] A. Restivo, Codes and aperiodic languages, in *Erste Fachtagung der Gesellschaft für Informatik über Automatentheorie und Formale Sprachen (Bonn, 1973)*, pp. 175–181, *Lect. Notes Comp. Sci.* vol. 2, Springer, Berlin, 1973.

[4] M.-P. Schützenberger, On finite monoids having only trivial subgroups, *Information and Control* **8** (1965), 190–194.

[5] M. P. Schützenberger, Sur certaines pseudo-variétés de monoïdes finis, in *Comptes Rendus des Journées Mathématiques de la Société Mathématique de France (Univ. Sci. Tech. Languedoc, Montpellier, 1974)*, pp. 317–327. Cahiers Math. Montpellier, No. 3, U.E.R. De Math., Univ. Sci. Tech. Languedoc, Montpellier, 1974.

[6] M.-P. Schützenberger, A property of finitely generated submonoids of free monoids, in *Algebraic theory of semigroups (Proc. Sixth Algebraic Conf., Szeged, 1976)*, pp. 545–576, *Colloq. Math. Soc. János Bolyai* vol. 20, North-Holland, Amsterdam-New York, 1979.