PUBLICATIONS DU DÉPARTEMENT DE MATHÉMATIQUES DE LYON

J. E. PIN

Introduction aux langages reconnaissables

Publications du Département de Mathématiques de Lyon, 1984, fascicule 6B
« Théorie des langages et complexité des algorithmes », , p. 13-38
<http://www.numdam.org/item?id=PDML_1984___6B_A2_0>

INTRODUCTION AUX LANGAGES RECONNAISSABLES

par J.E. PIN

1. Introduction

Le concept d'automate fini intervient fréquemment en informatique, que ce soit pour modéliser telle ou telle machine, pour concevoir des protocoles de communication ou pour décrire des circuits logiques. On pourra se reporter à l'article "automates finis" de D. Perrin et J. Sakarovitch (Journées d'Avignon 1982) pour une présentation de ce sujet. La notion de langage reconnaissable est tout aussi fondamentale : elle permet de rendre compte de l'enchainement des calculs dans un programme, d'exprimer le comportement d'un processus, de décrire certaines opérations d'éditeur de texte et plus généralement tout algorithme itératif. Les langages reconnaissables (appelés aussi langages réguliers) constituent par ailleurs le premier chainon d'une hiérarchie de langages de plus en plus complexes, la hiérarchie de Chomsky. Cette hiérarchie est présentée dans l'article "Introduction à la théorie des langages" de S. Grigorieff (Journées d'Avignon 1982) et les langages algébriques, qui constituent le deuxième chainon de la hiérarchie, sont présentés dans l'article de L. Boasson (Journées d'Avignon, 1983).

Les langages reconnaissables ont été étudiés très tôt dans le passé : le théorème de Kleene, sur lequel se fonde toute la théorie, date de 1956. Par la suite, les travaux de Schützenberger, Mc Naughton, Brzozowski, Simon, etc., ont mis en relief les liens profonds qui existent entre automates finis, langages reconnaissables et semigroupes finis. Le concept de variétés de langages, introduit par S. Eilenberg en 1976 a permis de formaliser cette triple approche - automates, langages et semigroupes - et a fourni un cadre cohérent et unifié à la théorie. Il s'agit maintenant d'une théorie mathématique vigoureuse et féconde, riche en problèmes ouverts.

Cet article, qui ne suppose aucune connaissance préalable sur les langages ou les automates, se présente comme une introduction à cette théorie.

On ne trouvera pas de démonstrations, sauf dans la première partie de l'article, qui présente en détail les définitions et propriétés de base : semigroupes, automates finis, langages reconnaissables, semigroupes syntactiques. Cette première partie est illustrée par un exemple de calcul détaillé d'un semigroupe syntactique. Le théorème des variétés est présenté dans la quatrième section et est illustré de quelques exemples classiques. La cinquième section est consacrée à l'opération étoile. On verra que certains des problèmes liés à l'opération étoile (la hauteur d'étoile restreinte, par exemple) ne sont pas, stricto sensu, des problèmes sur les langages reconnaissables. La dernière section présente quelques résultats et problèmes ouverts liés au produit de concaténation, notamment le lien entre les hiérarchies de langages et les hiérarchies en logique du premier ordre.

2. Automates, semigroupes et langages

Un semigroupe est un ensemble muni d'une loi associative (généralement notée multiplicativement). Un idempotent d'un semigroupe S est un élément e de S tel que $e = e^2$ (e^2 est une abréviation pour ee). Un monoïde M est un semigroupe muni d'un élément neutre 1. (C'est-à-dire un élément tel que pour tout $x \in M$, $1x = x1 = x$).

Etant donnés deux semigroupes S et T, un morphisme de semigroupes $\varphi : S \to T$ est une application φ de S dans T telle que pour tout $x,y \in S$ $(xy)\varphi = (x\varphi)(y\varphi)$. Etant donnés deux monoïdes M et N, un morphisme de monoïdes $\varphi : M \to N$ est un morphisme de semigroupes tel que $1\varphi = 1$.

Soit S un semigroupe (resp. monoïde). Un sous-semigroupe (sous-monoïde) T de S est un sous-ensemble de S tel que $t,t' \in T$ entraine $tt' \in T$ (et $1 \in T$). On dit que T est un semigroupe (monoïde) quotient de S s'il existe un morphisme de semigroupes (monoïdes) surjectif $\varphi : S \to T$. Enfin S divise T (noté $S < T$) si S est quotient d'un sous-semigroupe de T. On vérifie facilement que la relation "divise" est transitive.

Soit S un semigroupe. On note $\mathcal{P}(S)$ le <u>semigroupe des parties de S</u> muni du "produit des parties" : si $X,Y \in \mathcal{P}(S)$,

$$XY = \{ xy \mid x \in X,\ y \in Y \}$$

On définit également sur $\mathcal{P}(S)$ les opérations booléennes habituelles : $X \cup Y$, $X \cap Y$, $\bar{X} = S \setminus X$. On note également X^+ le sous-semigroupe de S engendré par X.

$$X^+ = \bigcup_{n \geq 1} X^n$$

Si S est un monoïde, on note X^* le sous-monoïde engendré par X :

$$X^* = \bigcup_{n \geq 0} X^n$$

où par convention, $X^0 = \{1\}$ pour tout $X \in \mathcal{P}(S)$

Soit $(S_i)_{i \in I}$ une famille de semigroupes. Le <u>produit direct</u> $\prod_{i \in I} S_i$ est le semigroupe dont le support est le produit cartésien $\prod_{i \in I} S_i$ et dont la loi est donnée par

$$(s_i)_{i \in I} \cdot (t_i)_{i \in I} = (s_i t_i)_{i \in I}$$

Par convention $\prod_{i \in \emptyset} S_i$ est le semigroupe 1 contenant un seul élément.

Soit A un ensemble fini appelé alphabet, dont les éléments sont appelés des lettres. On notera A^+ (resp A^*) le semigroupe (monoïde) libre sur A. Pratiquement, A^+ est construit de la façon suivante : les éléments de A^+ sont des suites finies non vides de lettres, appelées mots. Pour simplifier, on les note par simple juxtaposition : $a_1 \ldots a_n$. Maintenant le produit de deux mots $a_1 \ldots a_p$ et $b_1 \ldots b_q$ est le mot $a_1 \ldots a_p b_1 \ldots b_q$. Dans le cas du monoïde libre il faut adjoindre le mot vide (noté 1) qui correspond à la suite vide. On note $|u|$ la longueur du mot u, c'est-à-dire le nombre de lettres du mot u. On appelle automate (resp. automate fini) un triplet $\mathcal{A} = (Q, A, .)$ où Q est un ensemble (resp. un ensemble fini) appelé ensemble des états, A est un alphabet fini et $.$ désigne une action $Q \times A \to Q$. On étend par associativité l'action à A^* à partir des règles suivantes, où $q \in Q$, $w \in A^*$ et $a \in A$:

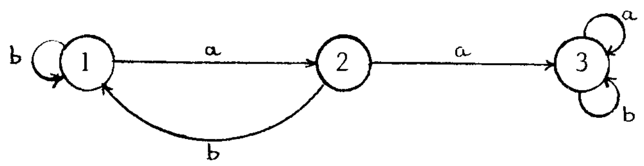
$q.1 = q$

$q.(wa) = (q.w).a$

<u>Exemple 1</u> $Q = \{1,2,3\}$ $A = \{a,b\}$

$1.a = 2$ $2.a = 3$ $3.a = 3$

$1.b = 1$ $2.b = 1$ $3.b = 3$



$2.babaab = 3$

Chaque mot w de A^* définit ainsi une application de Q dans Q. Le monoïde engendré par toutes les applications ainsi définies (w variant dans A^*) est un sous-monoïde de $\mathcal{T}(Q)$: c'est le "<u>monoïde de transition</u>" $M(\mathcal{A})$ de l'automate $\mathcal{A}$. Bien entendu $M(\mathcal{A})$ est engendré par les applications définies par les lettres de l'alphabet et on a un morphisme canonique $A^* \longrightarrow M(\mathcal{A})$

<u>Exemple 1</u> (suite)

		1	2	3
générateurs	a	2	3	3
	b	1	1	3
	a^2	3	3	3
	ab	1	3	3
	ba	2	2	3

La suite du calcul montre que b, b^2 et 1 définissent la même application de Q dans Q, de même que a^2, a^3, a^2b et ba^2 d'une part et a, aba d'autre part. Le monoïde de transition $M(\mathcal{A})$ contient donc 6 éléments correspondant aux mots $1, a, b, a^2, ab$ et ba.

Soit A un alphabet fini. On appelle <u>langage</u> (sur l'alphabet A) une partie du monoïde libre A . Par abus de notation, on identifiera éventuellement le mot $u \in A^*$ et le langage $\{u\}$. On peut définir sur les langages une grande quantité d'opérations. Outre les opérations booléennes, nous utiliserons principalement le produit, l'étoile et le quotient. Par définition si K et L sont deux langages de A^* , le <u>quotient</u> - ou <u>résiduel</u> - à gauche (resp. à droite de L par K est le langage $K^{-1}L$ (resp. LK^{-1}) défini par

$K^{-1}L = \{v \in A^* \mid Kv \cap L \neq \emptyset\} = \{v \in A^* \mid$ il existe $u \in K$ tel que $uv \in L\}$
$LK^{-1} = \{v \in A^* \mid vK \cap L \neq \emptyset\} = \{v \in A^* \mid$ il existe $u \in K$ tel que $vu \in L\}$

Les opérations ci-dessus permettent d'exprimer simplement certains langages définis par des propriétés combinatoires. Par exemple, si $A = \{a,b\}$, l'ensemble des mots commençant par la lettre a et n'ayant aucun facteur égal à aba est le langage $L = aA^* \setminus A^*abaA^*$. Un moment de réflexion convaincra le lecteur que le même langage peut s'exprimer sous la forme $L = a(a \cup bbb^*a)^*b^*$. On va maintenant délimiter de façon précise la puissance des opérations que nous avons introduites. Commençons par quelques définitions

Définition 2.1 *Soit A un alphabet fini. L'ensemble des langages rationnels de A^* est le plus petit ensemble, noté A^*Rat, de langages de A^* tel que*

(a) Pour tout mot $u \in A^*$, $\{u\} \in A^*Rat$

(b) A^*Rat est fermé par union finie, produit et étoile.

Précisons que l'ensemble vide est un langage rationnel puisque $\emptyset = \bigcup_{i \in \emptyset} L_i$

Définition 2.2. *Soit A un alphabet fini, $\eta : A^* \to M$ un morphisme de monoïdes et $L \subset A^*$ un langage. On dit que η reconnaît L s'il existe une partie P de M telle que $L = P\eta^{-1}$. Par extension, on dit également dans ce cas que M reconnaît L.*

Définition 2.3 *Un langage est dit reconnaissable s'il est reconnu par un monoïde fini.*

Les deux dernières définitions se rattachent à la notion d'automate de la façon suivante. Soit A un alphabet fini et $\mathcal{A} = (Q,A,.)$ un automate. On dit qu'un langage L est *reconnu par l'automate* $\mathcal{A}$ s'il existe un état $q_0 \in Q$ (appelé état *initial*) et un ensemble d'états F (l'ensemble des états *finaux*) tels que $u \in L$ si et seulement si $q_0.u \in F$.

Exemple 2 En prenant $q_0 = 1$ et $F = \{3\}$, on voit que l'automate de l'exemple 1 reconnaît le langage $L = A^*aaA^*$.

Le lien entre monoïde et automate est assuré par la proposition suivante

<u>Proposition 2.1</u> *Si* $L \subset A^*$ *est reconnu par un automate, il est reconnu par le monoïde de transition de cet automate. De plus* L *est reconnu par un automate fini si et seulement si* L *est reconnaissable.*

<u>Preuve</u> Soit $\mathcal{A}$ un automate reconnaissant L, muni de l'état initial q_0 et de l'ensemble des états finaux F. On note $\eta : A^* \to M(\mathcal{A})$ le morphisme canonique, et on pose $P = L\eta$. On va montrer que $P\eta^{-1} = L$. Soit $u \in P\eta^{-1}$: alors $u\eta \in P = L\eta$ par définition et donc $u\eta = v\eta$ pour un certain $v \in L$. Cela signifie que u et v définissent la même application de l'ensemble des états dans lui-même. En particulier $q_0.u = q_0.v$ et comme $v \in L$, $q_0.v \in F$. Donc $q_0.u \in F$ et on en déduit $u \in L$. On vient ainsi d'établir l'inclusion $L\eta\eta^{-1} \subset L$. L'inclusion opposée est évidente et donc $L = P\eta^{-1}$ comme annoncé. Donc $M(\mathcal{A})$ reconnaît L.

En particulier si $\mathcal{A}$ est un automate fini, L est reconnu par le monoïde fini $M(\mathcal{A})$ et L est reconnaissable. Réciproquement si L est reconnaissable il existe un monoïde fini M, un morphisme $\eta : A^* \to M$ et une partie de M telle que $P\eta^{-1} = L$. Soit $\mathcal{A}_M = (M, A, .)$ l'automate défini par l'action $m.a = m(a\eta)$. Si on prend 1 comme état initial et P comme ensemble d'états finaux, l'automate $\mathcal{A}_M$ reconnaît L puisque

$$1.u \in P \Leftrightarrow 1(u\eta) \in P \Leftrightarrow u \in P\eta^{-1}.$$

Donc L est reconnu par un automate fini. $\square$

Les énoncés qui suivent traduisent en termes de monoïdes quelques résultats très classiques de théorie des automates.

<u>Proposition 2.2</u> *Soit* L *un langage de* A^*. *Si* M *reconnaît* L, M *reconnaît* $A^* \backslash L$.

<u>Preuve</u>

Soit $\eta : A^* \to M$ et $P \subset M$ tels que $L = P\eta^{-1}$. Alors $(M\backslash P)\eta^{-1} = A^*\backslash L$. $\square$

Proposition 2.3 *Soient* L_1 *et* L_2 *deux langages de* A^* *reconnus respectivement par des monoïdes* M_1 *et* M_2. *Alors* $L_1 \cap L_2$ *et* $L_1 \cup L_2$ *sont reconnus par* $M_1 \times M_2$.

Preuve Soient $\eta_1 : A^* \to M_1$, $\eta_2 : A^* \to M_2$, $P_1 \subset M_1$, $P_2 \subset M_2$ tels que $L_1 = P_1\eta_1^{-1}$, $L_2 = P_2\eta_2^{-1}$. Soit $\eta : A^* \to M_1 \times M_2$ le morphisme défini par $u\eta = (u\eta_1, u\eta_2)$. On a alors les formules $(P_1 \times P_2)\eta^{-1} = L_1 \cap L_2$ et $((P_1 \times M_2) \cup (M_1 \times P_2))\eta^{-1} = L_1 \cup L_2$. □

Proposition 2.4 *Soit* $\varphi : A^* \to B^*$ *un morphisme de monoïdes libres et* $L \subset B^*$ *un langage reconnu par un monoïde* M. *Alors* M *reconnaît aussi* $L\varphi^{-1}$.

Preuve Soient $\eta : B^* \to M$ et $P \subset M$ tels que $L = P\eta^{-1}$. Alors $L\varphi^{-1} = (P\eta^{-1})\varphi^{-1} = P(\varphi\eta)^{-1}$, ce qui montre que $L\varphi^{-1}$ est reconnu par le morphisme $\varphi\eta : A^* \to M$. □

Proposition 2.5 *Soient* K *et* L *deux langages de* A^*. *Si* M *reconnaît* L, M *reconnaît* $K^{-1}L$ *et* LK^{-1}.

Preuve Soit $\eta : A^* \to M$ et $P \subset M$ tels que $P\eta^{-1} = L$. Posons

$$Q = \{m \in M \mid \exists\, u \in K \ \ (u\eta)m \in P\}.$$ Il vient

$$\begin{aligned} Q\eta^{-1} &= \{v \in A^* \mid v\eta \in Q\} = \{v \in A^* \mid \exists\, u \in K\ (u\eta)(v\eta) \in P\} \\ &= \{v \in A^* \mid \exists\, u \in K \ \ (uv)\eta \in P\} = \{v \in A^* \mid \exists\, u \in K \ \ uv \in L\} \\ &= K^{-1}L \end{aligned}$$

Donc $K^{-1}L$ est reconnu par M. La preuve pour LK^{-1} est similaire. □

On note $A^*\mathrm{Rec}$ l'ensemble des langages reconnaissables de A^*. On a alors

Proposition 2.6 (1) *Pour tout alphabet* A, A^*Rec *est une algèbre de Boole.*

(2) *Si* $L \in A^*Rec$ *et* $K \subset A^*$ *alors* $K^{-1}L$, $LK^{-1} \in A^*Rec$.

(3) *Si* $\varphi : A^* \to B^*$ *est un morphisme de monoïdes libres et si* $L \in B^*Rec$, *alors* $L\varphi^{-1} \in A^*Rec$.

Preuve : résulte des propositions 2.2 à 2.5 □

Nous pouvons maintenant énoncer un résultat fondamental de la théorie des langages, le théorème de Kleene. Nous admettrons ce théorème dont la démonstration figure dans la plupart des livres consacrés à la théorie des automates ou des langages (cf. par exemple Eilenberg, Vol. 1, chap. 7 ou Lallement, chap. 6)

Théorème (Kleene) *Soit A un alphabet fini. Un langage $L \subset A^*$ est rationnel si et seulement si il est reconnaissable.*

Il résulte en particulier du théorème de Kleene que l'ensemble des langages rationnels de A^* est fermé pour les opérations booléennes finies et pour le quotient à gauche (resp. à droite) par un langage quelconque.

2.4 Monoïdes syntactiques

Soit L un langage de A^*. Conformément à la définition générale, on appelle congruence syntactique de L la congruence $\sim_L$ définie sur A^* par $u \sim_L v$ si et seulement si pour tout $x,y \in A^*$, $xuy \in L \Leftrightarrow xvy \in L$. Le *monoïde syntactique* de L est le monoïde quotient $M(L) = A^*/\sim_L$. La proposition qui suit montre que le monoïde syntactique d'un langage L est le plus petit monoïde reconnaissant L, "plus petit" étant pris au sens de la relation de division.

Proposition 2.7 *Soit L un langage de A^*.*

(1) M reconnaît L si et seulement si $M(L)$ divise M.

(2) Si M reconnaît L et si M divise N, N reconnaît L.

Preuve (a) Tout d'abord $M(L)$ reconnaît L. Notons en effet $\eta : A^* \to A^*/\sim_L = M(L)$ le morphisme canonique. On va montrer que $L = L\eta\eta^{-1}$. L'inclusion de gauche à droite est évidente. Réciproquement soit $u \in L\eta\eta^{-1}$. Alors $u\eta \in L\eta$ et il existe $v \in L$ tel que $u\eta = v\eta$ i.e. $u \sim_L v$. Comme $v \in L$ on a aussi $u \in L$ - en prenant $x = y = 1$ dans la définition de $\sim_L$ -. En posant $P = L\eta$, on trouve $P\eta^{-1} = L$ et donc $M(L)$ reconnaît L.

(b) Supposons que $M(L)$ divise M. Il existe alors un monoïde N et des morphismes $\alpha : N \to M$ injectif et $\beta : N \to M(L)$ surjectif. D'après la proposition 1.9, il existe un morphisme $\varphi : A^* \to N$ tel que $\eta = \varphi\beta$. Posons $P = L\eta\beta^{-1}\alpha \subset M$. Il vient

$$\begin{tikzcd} A^* \arrow[d, "\eta"'] \arrow[dr, dashed, "\varphi"] & & \\ M(L) & N \arrow[l, "\beta"] \arrow[r, "\alpha"] & M \end{tikzcd}$$

$$P(\varphi\alpha)^{-1} = P\alpha^{-1}\varphi^{-1} = L\eta\beta^{-1}\alpha\alpha^{-1}\varphi^{-1}$$
$$= L\eta\beta^{-1}\varphi^{-1} = L\eta\eta^{-1} = L$$

Donc M reconnaît L.

(c) Finalement supposons que M reconnaisse L. Il existe alors un morphisme $\varphi : A^* \to M$ et une partie P de M telle que $L = P\varphi^{-1}$. Posons $N = A^*\varphi$: N est un sous-monoïde de M. Supposons $u\varphi = v\varphi$ et $xuy \in L$. Alors $(xuy)\varphi = (xvy)\varphi \in L\varphi = P$ et donc $xvy \in L$ puisque $P\varphi^{-1} = L$. Il en résulte que $u\varphi = v\varphi$ entraîne $u \sim_L v$. D'après la proposition 1.4, il existe un morphisme surjectif $\pi : N \to A^*/\sim_L = M(L)$. Donc $M(L)$ divise M.

(1) La partie (2) de l'énoncé résulte de la partie (1).

<u>Corollaire 2.8</u> *Soient L, L_1 et L_2 des langages reconnaissables de A^* et K un langage quelconque. Alors :*

(a) $M(A^*\backslash L) = M(L)$,

(b) $M(L_1 \cap L_2)$ *divise* $M(L_1) \times M(L_2)$,

(c) $M(L_1 \cup L_2)$ *divise* $M(L_1) \times M(L_2)$,

(d) $M(LK^{-1})$ *et* $M(K^{-1}L)$ *divisent* $M(L)$,

(e) si $\varphi : B^* \to A^*$ *est un morphisme de monoïdes libres,* $M(L\varphi^{-1})$ *divise* $M(L)$.

<u>Preuve</u> C'est une conséquence immédiate des propositions 2.2 à 2.6. □

Le lien entre monoïdes syntactiques et automates est le suivant. Appelons automate minimal d'un langage L l'automate $\mathcal{A} = (Q,A,.)$ ainsi défini : l'ensemble des états est $Q = \{u^{-1}L \mid u \in A^*\}$ et les transitions sont données par

$$(u^{-1}L).a = a^{-1}(u^{-1}L) = (ua)^{-1}L \quad \text{pour tout } a \in A.$$

On démontre alors que *le monoïde de transition de l'automate minimal de L est égal au monoïde syntactique de L.*

Il est facile de voir que l'automate minimal de L reconnaît L en prenant pour état initial $q_0 = L$ et pour ensemble des états finaux $F = \{u^{-1}L \mid u \in L\}$. De plus le terme"minimal" correspond à une propriété de théorie des automates. D'une part l'automate minimal est *accessible* ce qui signifie que, pour tout état q, il existe un mot $u \in A^*$ tel que $q_0.u = q$. D'autre part c'est le plus petit automate accessible reconnaissant L. De façon précise, si $\mathcal{A}' = (Q',A,.)$, muni de l'état initial q'_0 et de l'ensemble d'états finaux F', est un automate accessible reconnaissant L, il existe une application surjective unique $\varphi : Q' \rightarrow Q$ telle que $q_o\varphi = q'_o$ et telle que pour tout mot $u \in A^*$, $(q.u)\varphi = (q\varphi).u$.

2. Le cas des semigroupes libres

Tout ce qui précède s'adapte sans difficulté au cas où les langages sont des parties du semigroupe libre A^+. On définit de façon analogue les notions d'automate minimal, semigroupe syntactique, semigroupe de transition, langage reconnaissable. La règle est de remplacer partout le terme monoïde par semigroupe et "*" par "+". Une exception à cette règle : la définition de la congruence syntactique d'un langage L de A^+. En effet puisque $A^* = (A^+)^1$, cette définition est :

$$u \sim_L v \text{ si et seulement si pour tout } x,y \in A^* \ (xuy \in L \Leftrightarrow xvy \in L).$$

3. Calculs explicites

Lors des calculs de résiduels, on pourra utiliser les formules suivantes, dont la démonstration est laissée au lecteur. Dans ces formules a désigne une lettre, u et v des mots et L, L_1, L_2 des langages.

$$u^{-1}(L_1 \cup L_2) = u^{-1}L_1 \cup u^{-1}L_2$$

$$u^{-1}(L_1 \setminus L_2) = u^{-1}L_1 \setminus u^{-1}L_2$$

$$u^{-1}(L_1 \cap L_2) = u^{-1}L_1 \cap u^{-1}L_2$$

$$\begin{aligned} a^{-1}(L_1L_2) &= (a^{-1}L_1)L_2 \quad \text{si} \quad 1 \notin L_1 \\ &= (a^{-1}L_1)L_2 \cup a^{-1}L_2 \quad \text{si} \quad 1 \in L_1 \end{aligned}$$

$$a^{-1}L^* = (a^{-1}L)L^*$$

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

Semigroupe syntactique de $L = A^*abaA^*$ sur l'alphabet $A = \{a,b\}$

On va calculer l'automate minimal par la méthode des résiduels

$$a^{-1}L = L \cup baA^*$$

$$b^{-1}L = L$$

$$a^{-1}(L \cup baA^*) = a^{-1}L = L \cup baA^*$$

$$b^{-1}(L \cup baA^*) = b^{-1}L \cup aA^* = L \cup aA^*$$

$$a^{-1}(L \cup aA^*) = a^{-1}L \cup A^* = A^*$$

$$b^{-1}(L \cup aA^*) = b^{-1}L = L$$

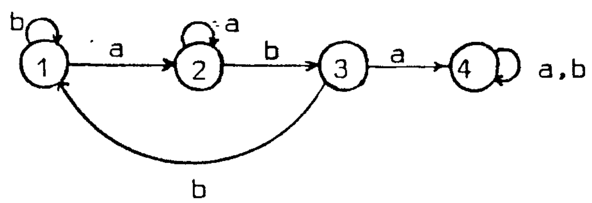
$$a^{-1}A^* = A^*$$

$$b^{-1}A^* = A^*$$

Donc l'automate minimal de L possède 4 états, correspondant aux résiduels $q_1 = L$ $q_2 = L \cup baA^*$ $q_3 = L \cup aA^*$ $q_4 = A^*$.

Les transitions de l'automate minimal sont données par le tableau suivant

Indice des états	1	2	3	4
a	2	2	4	4
b	1	3	1	4



Calcul du semigroupe syntactique de L

C'est le semigroupe de transition de l'automate minimal. Voici d'abord le principe du calcul

1) on ordonne l'alphabet $A : a_1 < a_2 < \ldots < a_n$. On va maintenant gérer

simultanément une table des transformations et une liste de relations.

2) On calcule successivement les transformations associées aux mots de longueur 1,2,3,...

3) Pour passer de la longueur n à la longueur $n+1$, on calcule les transformations associées aux mots de la forme ua_i $(1 \leq i \leq n)$, où u est un mot de longueur n figurant dans la table.

4) Soit $v = ua_i$. Si la transformation associée à v ne figure pas dans le tableau, on l'y ajoute. Si la transformation est déjà associée à un mot u, on inscrit $u = v$ dans la liste des relations.

Dans notre exemple, on prendra $a < b$.

Longueur 1 les transformations associées aux mots a et b sont données

	1	2	3	4
a	2	2	4	4
b	1	3	1	4

Longueur 2 On calcule successivement (1)aa (2)ab (3)ba (4)bb

Le résultat de ce calcul est reporté dans le tableau ci-dessous

		1	2	3	4
	a	2	2	4	4
	b	1	3	1	4
(2)	ab	3	3	4	4
(3)	ba	2	4	2	4
(4)	bb	1	1	1	4

Relations

(1) $a^2 = a$

Longueur 3 On calcule successivement (5)aba (6)abb (7)baa (8)bab (9)bba (10)bbb

(On ne doit calculer ni aaa, ni aab puisque aa ne figure pas dans la table de gauche). On observe déjà que baa = ba puisque aa = a. Il sera donc inutile de reporter cette relation dans la liste des relations. La suite du calcul donne alors

		1	2	3	4
	a	2	2	4	4
	b	1	3	1	4
	ab	3	3	4	4
	ba	2	4	2	4
	bb	1	1	1	4
(5)	aba	4	4	4	4
(6)	abb	1	1	4	4
(8)	bab	3	4	3	4
(9)	bba	2	2	2	4

Relations

$a^2 = a$

(10) $b^3 = b^2$

Longueur 4 On calcule successivement (11)abaa (12)abab (13)abba (14)abbb (15)baba (16)babb (17)bbaa (18)bbab

Les relations déjà connues permettent d'éviter le calcul de (11) (14) et (17) puisque abaa = aba, abbb = abb et bbaa = bba. D'autre part le mot aba est un zéro du semigroupe de transition (cf. problème 2.7). On reporte directement cette remarque sous la forme aba = 0 dans la liste des relations. Il est donc inutile de calculer (12) et (15). On obtient alors

	1	2	3	4
a	2	2	4	4
b	1	3	1	4
ab	3	3	4	4
ba	2	4	2	4
bb	1	1	1	4
aba	4	4	4	4
abb	1	1	4	4
bab	3	4	3	4
bba	2	2	2	4
bbab	3	3	3	4
babb	1	4	1	4

Relations

(a) $a^2 = a$

(b) $b^3 = b^2$

(c) aba = 0

(d) abba = a

Longueur 5 On calcule successivement babba = ba, babbb = babb, bbaba = 0, bbabb = b

Conclusion Le semigroupe syntactique de L est donné soit par la table précédente (comme semigroupe de transformation), soit par "générateurs et relations", les relations étant $a^2 = a$, $b^3 = b^2$, aba = 0, abba = a, bbabb = bb.

4. Variétés

4.1. Le théorème des variétés

La notion de monoïde syntactique suggère de classer les langages reconnaissables à partir des propriétés algébriques des monoïdes qui les reconnaissent. Cette idée naturelle s'est révélée extrêmement féconde au fil des années. Comme la plupart des "bonnes" propriétés algébriques d'un monoïde sont conservées par passage au sous-monoïde, au monoïde quotient et au produit direct, on est amené à introduire la définition suivante

Définition 4.1 Une variété de monoïdes finis est une classe $\underline{V}$ de monoïdes finis telle que

(1) Si $M \in \underline{V}$ et si N est un sous-monoïde de M, alors $N \in \underline{V}$

(2) Si $M \in \underline{V}$ et si N est un quotient de M, alors $N \in \underline{V}$

(3) Si $(M_i)_{i \in I}$ est une famille finie d'éléments de $\underline{V}$, alors $\prod_{i \in I} M_i \in \underline{V}$

On définit de manière analogue les variétés de semigroupes finis.

Etant donné un alphabet A, on associe maintenant à chaque variété de monoïdes $\underline{V}$, l'ensemble $A^*\mathcal{V}$ des langages reconnaissables de A^* qui sont reconnus par un monoïde de $\underline{V}$. D'après la proposition 2.7, un langage L est dans $A^*\mathcal{V}$ si et seulement si son monoïde syntactique est élément de $\underline{V}$. On obtient ainsi une correspondance $\underline{V} \longrightarrow \mathcal{V}$ associant à une variété de monoïdes finis un objet baptisé "variété de langages" et dont voici la définition précise.

Définition 4.2 Une variété de langages $\mathcal{V}$ associe à chaque alphabet A un ensemble $A^*\mathcal{V}$ de langages reconnaissables de A^* tels que

(1) Pour tout alphabet A, $A^*\mathcal{V}$ est une algèbre de Boole

(2) Pour tout alphabet A, si $a \in A$ et $L \in A^*\mathcal{V}$, alors

$$a^{-1}L \text{ et } La^{-1} \in A^*\mathcal{V}$$

(3) Pour tout morphisme de monoïdes $\varphi : A^* \longrightarrow B^*$, $L \in B^*\mathcal{V}$ entraine $L\varphi^{-1} \in A^*\mathcal{V}$.

Le théorème des variétés, dû à S. Eilenberg, s'énonce alors ainsi :

Théorème 4.1 La correspondance $\underline{V} \longrightarrow \mathcal{V}$ induit une bijection entre les variétés de monoïdes finis et les variétés de langages.

On a un théorème analogue pour les variétés de semigroupes finis à condition de considérer les langages comme des parties du semigroupe libre A^+. Autrement dit, dans la définition d'une variété de langages il faut remplacer le mot "monoïde" par "semigroupe" et le symbole "$*$" par le symbole "+". Dans ce cas, on parle de +-variété de langages.

4.2 <u>Exemples de variétés</u>

Le premier exemple découle directement du théorème de Kleene. Soit $\underline{M}$ la variété de tous les monoïdes finis. La variété de langages qui lui correspond est la variété des langages rationnels.

Autre exemple, la variété triviale $\underline{I}$ constituée du seul monoïde à un élément. Pour tout alphabet A, la variété $\mathcal{V}$ correspondant à I est définie par $A^*\mathcal{V} = \{\emptyset, A^*\}$

Appelons <u>semigroupe nilpotent</u> un semigroupe fini tel que pour tout idempotent $e \in S$ et pour tout $s \in S$, on ait $es = e = se$. Un petit raisonnement montre que si S n'est pas vide, S contient en fait un seul idempotent, noté 0, qui vérifie donc $0s = 0 = s0$ pour tout $s \in S$. De plus tout produit d'au moins Card(S) éléments de S est égal à 0.

On démontre que les semigroupes nilpotents forment une variété de semigroupes finis notée <u>Nil</u>. La +variété de langages correspondante, notée $\mathcal{N}il$, est définie par

$$A^+\mathcal{N}il = \{ \text{langages finis ou cofinis de } A^+ \}$$

(rappelons qu'un langage est cofini si son complémentaire est fini).

Cet exemple montre la nécessité de distinguer les variétés et les +-variétés. En effet les langages finis-cofinis de A^* ne définissent pas une variété de langages car ils ne vérifient pas la condition (3) de la définition 4.2. Il suffit de prendre $A = B = \{a,b\}$, $a\varphi = a$, $b\varphi = 1$ et $L = \{a\}$: L est fini mais $L\varphi^{-1} = b^*ab^*$ n'est ni fini, ni cofini...

La variété de langages la plus importante est certainement la variété des langages sans étoile. Par définition, l'ensemble $A^*\mathcal{S}$ des langages sans étoile de A^* est le plus petit ensemble de langages de A^* tel que

(1) $\{1\} \in A^*\mathcal{S}$, $\{a\} \in A^*\mathcal{S}$ pour tout $a \in A$

(2) $A^*\mathcal{S}$ est fermée pour les opérations booléennes finies (y compris la complémentation) et le produit de concaténation.

Notons que $\emptyset = \bigcup_{i \in \emptyset} L_i$ et A^*, complémentaire de $\emptyset$, sont des langages sans étoile.

De même si $A = \{a,b\}$

$b^* = A^* \setminus A^* a A^*$ est sans étoile... Plus compliqué :

$(ab)^+ = (aA^* \cap A^* b) \setminus (A^* aaA^* \cup A^* bbA^*)$ et $(ab)^* = (ab)^+ \cup \{1\}$ sont des langages sans étoile. En revanche $(a^2)^*$ n'est pas sans étoile comme on va le voir.

Un monoïde fini M est <u>apériodique</u> si la suite des exposants successifs d'un élément quelconque $x \in M$ ne contient pas de cycle de longueur $\geqslant 2$. Plus formellement M est apériodique si pour tout $m \in M$, il existe $n > 0$ tel que $x^n = x^{n+1}$

$$1 \quad x \quad x^2 \quad x^3 \quad \cdots \quad x^{n-1} \quad x^n = x^{n+1}$$

Il est équivalent de dire que tout groupe qui divise M est trivial. Autrement dit les monoïdes apériodiques sont "à l'opposé des groupes" en théorie des semigroupes. Le théorème de Schützenberger s'énonce ainsi

<u>Théorème 4.2</u> Un langage reconnaissable est sans étoile si et seulement si son monoïde syntactique est apériodique.

Puisque le monoïde syntactique de $(a^2)^*$ est le groupe à deux éléments, $(a^2)^*$ n'est pas un langage sans étoile. En termes de variétés, le théorème 4.2 s'énonce ainsi

<u>Corollaire 4.2</u> La variété $\underline{A}$ des monoïdes apériodiques correspond à la variété des langages sans étoile.

5. L'opération étoile.

C'est une opération qui reste relativement mystérieuse malgré les recherches dont elle a fait l'objet. En particulier les problèmes de hauteur d'étoile ne sont toujours pas résolus, bien que des progrès aient été accomplis récemment. Rappelons les énoncés de ces problèmes qui reposent sur la notion d'expression rationnelle.

Les expressions rationnelles sur l'alphabet A sont définies récursivement par :

(a) $\emptyset$, 1 et a (pour tout $a \in A$) sont des expressions rationnelles.

(b) Si E et F sont des expressions rationnelles, $(E \cup F)$, $(E \cap F)$, (EF), $(\overline{E})$ et (E^*) sont des expressions rationnelles.

La hauteur d'étoile $h(E)$ d'une expression rationnelle E est définie récursivement par :

(a) $h(\emptyset) = 0$, $h(1) = 0$, et pour tout $a \in A$, $h(a) = 0$.

(b) Si E et F sont des expressions rationnelles,

$$h(E \cup F) = h(E \cap F) = h(EF) = \max(h(E), h(F))$$

$$h(\overline{E}) = h(E) \qquad h(E^*) = h(E)+1 .$$

La valeur $v(E)$ d'une expression rationnelle E est le langage de A^* représenté par E. Formellement, v est une application de l'ensemble des expressions rationnelles dans l'ensemble des langages rationnels de A^* définie par :

(a) $v(\emptyset) = \emptyset$, $v(1) = \{1\}$, $v(a) = \{a\}$ pour tout $a \in A$.

(b) Si E et F sont deux expressions rationnelles,

$$v(E \cup F) = v(E) \cup v(F) \qquad v(E \cap F) = v(E) \cap v(F)$$

$$v(EF) = v(E)v(F) \qquad v(\overline{E}) = A^* \setminus v(E) \qquad v(E^*) = (V(E))^* .$$

La hauteur d'étoile $h(L)$ d'un langage reconnaissable L est par définition le minimum de la hauteur d'étoile des expressions rationnelles qui le représentent. Autrement dit :

$$h(L) = \min \{h(E) \mid E \text{ est une expression rationnelle et } v(E) = L \}.$$

On voit que les langages de hauteur d'étoile 0 sont exactement les langages sans étoile définis au paragraphe précédent. Par conséquent, le théorème de Schützenberger permet d'énoncer

Théorème 5.1. Un langage est de hauteur d'étoile 0 si et seulement si son monoïde syntactique est apériodique.

En particulier, on voit qu'il existe un algorithme pour décider si un langage reconnaissable est de hauteur d'étoile 0. Si le langage est donné par exemple par un automate qui le reconnait, on calculera d'abord l'automate minimal, puis le monoïde syntactique $M(L)$ - qui est fini - suivant l'algorithme décrit dans les sections 2 et 3 et on vérifiera enfin que ce monoïde est apériodique.

Le théorème 5.1. suggère que les langages de hauteur d'étoile au plus n peuvent être caractérisés par une propriété de leur monoïde syntactique, autrement dit que les langages de hauteur d'étoile au plus n forment une variété de langages. Cette hypothèse est explicitement mentionnée dans la thèse d'Henneman mais est maintenant peu probable compte tenu du résultat suivant

Théorème 5.2. Pour tout monoïde fini M, il existe un langage fini F tel que M divise le monoïde syntactique de F^*.

Corollaire 5.3. Si les langages de hauteur d'étoile 0 ou 1 forment une variété de langages, tous les langages sont de hauteur d'étoile 0 ou 1.

Toutefois, on ne connait toujours pas d'exemple de langage de hauteur d'étoile > 1 ! Pour illustrer la difficulté du problème, voici un exemple célèbre dont on pensait qu'il était de hauteur d'étoile 2 avant qu'Henneman ne trouve une expression de hauteur 1 pour le représenter.

$$L = (a(b^*)a \cup b(a(b^*)a)^*b)^*$$

L'expression trouvée par Henneman est la suivante.

$$(E \cap F) \cup [(b \cup a(b^*)a)E \cap ((a^*)b(a^*)b)F]$$

où

$$E = ((b \cup a(b^*)a)(b \cup a(b^*)a))^*$$

et

$$F = (b \cup ((b^*)a(b^*)a(b^*)a(b^*)a)^* \quad .$$

Il faut également substituer à (b^*) l'expression sans étoile décrite dans la section précédente ...

Voici un autre candidat pour la hauteur d'étoile 2, proposé par Thérien.

$$T = [a(b^*)a \cup b(a^*)b(a(b^*)a)^*b(a^*)b\ ]^* \quad .$$

Les résultats positifs sont rares. Il faut toutefois mentionner les résultats obtenus par Henneman dans sa thèse.

Théorème 5.4. Si un langage L est reconnu par un groupe commutatif alors L est de hauteur d'étoile au plus 1.

Henneman a également obtenu des bornes pour les langages reconnus par des groupes résolubles ou superrésolubles et plusieurs autres résultats partiels plus techniques. En utilisant un théorème de Straubing sur les variétés de langages fermées par concaténation on peut améliorer le théorème 5.4. Nous dirons qu'un monoïde M est une extension apériodique d'un groupe commutatif s'il existe un groupe commutatif G est un morphisme surjectif de monoïdes $\varphi : M \dashrightarrow G$ tel que $1\varphi^{-1}$ soit un semi-groupe apériodique.

Théorème 5.5. Si un langage L est reconnu par un extension apériodique d'un groupe commutatif, alors L est de hauteur d'étoile au plus 1.

Il existe une variante de la définition de la hauteur d'étoile que nous appelerons hauteur d'étoile restreinte. Pour cela on supprime les opérations d'intersection et de complémentation dans la définition d'une expression rationnelle : on définit ainsi des expressions rationnelles restreintes. La définition de la valeur et de la hauteur d'une expression sont les mêmes que précédemment (compte tenu de la suppression des opérateurs d'intersection et de complémentation). On notera hr(E) la hauteur d'étoile restreinte d'une expression rationnelle restreinte E. La hauteur d'étoile restreinte d'un langage L est définie par

$hr(L) = \min \{ hr(E) \mid$ E est une expression rationnelle restreinte et $v(E) = L \}$.

Comme pour la hauteur d'étoile, on ne dispose que de résultats très partiels. cf. [1] pour plus de détails.

Théorème 5.6. Pour tout entier n, il existe un langage reconnaissable L tel que $hr(L) = n$.

Les travaux de Cohen et Mc Naughton notamment donnent des techniques pour calculer la hauteur d'étoile restreinte de certains langages. Mais le résultat la plus significatif a été obtenu récemment par Hashigushi :

Théorème 5.7. [3]. Il existe un algorithme pour décider si la hauteur d'étoile restreinte d'un langage reconnaissable est 1.

Bien entendu, on sait également décider si un langage reconnaissable est de hauteur d'étoile restreinte 0 : il faut et il suffit que ce langage soit fini. En revanche, on ne sait toujours pas s'il existe un algorithme pour décider si un langage reconnaissable est de hauteur d'étoile restreinte n, pour $n > 1$.

On a aussi étudié l'opération étoile sous l'angle des variétés de langages. Une variété de langages $\mathcal{V}$ est dite fermée par étoile si pour tout alphabet A, $L \in A^*\mathcal{V}$ entraine $L^* \in A^*\mathcal{V}$.

Théorème 5.8. 7 . La seule variété de langages fermée pour l'opération étoile est la variété de tous les langages rationnels.

Le théorème 5.8. indique que l'opération étoile est une opération très "puissante". Pour limiter cette "puissance" on peut imaginer diverses restrictions sur l'opération étoile. Nous n'en donnerons qu'un seul exemple. Un langage de la forme L^* est dit pur si pour tout $n > 0$, $u^n \in L^*$ entraine $u \in L^*$. On dit alors que l'opération $L \longrightarrow L^*$ est une étoile pure.

Théorème 5.9. La variété des langages sans étoile est la plus petite variété fermée par étoile pure.

On remarquera que le produit de concaténation n'est pas mentionné dans ce énoncé qui rend un peu curieuse la terminologie "sans étoile". En fait le produit de concaténation est relié à cette opération d'étoile pure par le résultat suivant qui découle des travaux de Straubing.

<u>Théorème 5.10.</u> Toute variété de langages fermée par produit de concaténation est fermée par étoile pure.

On ne sait pas si la réciproque du théorème 5.10 est vraie ou fausse.

6. Le produit de concaténation.

Cette opération est pratiquement aussi difficile à étudier que l'étoile, si on se réfère aux problèmes encore ouverts à ce jour. Comme pour l'étoile on peut définir des hiérarchies de langages en utilisant le produit de concaténation. La première hiérarchie de ce type a été introduite par Brzozowski et a servi de modèle pour les hiérarchies qui ont été proposées par la suite. Nous nous contenterons de présenter la plus simple de ces hiérarchies qui a été proposée récemment par Straubing. Soit A un alphabet fini. Chaque niveau de la hiérarchie sera constitué par une algèbre de Boole formée de langages reconnaissables. Le niveau 0 est constitué par l'algèbre de Boole triviale : $\{ \emptyset , A^*\}$. Le niveau n+1 est l'algèbre de Boole engendrée par les langages de la forme $L_0a_1L_1a_2\dots a_kL_k$ où $k \geq 0$, $a_1,a_2,\dots,a_k \in A$ et où les langages $L_0,L_1,\dots,L_k$ sont des langages de niveau n.

Commençons par l'examen du niveau 1. Il résulte immédiatement de la définition ci-dessus que le niveau 1 est l'algèbre de Boole engendrée par les langages de la forme $A^*a_1A^*a_2 \dots a_kA^*$ ou $k \geq 0$ et $a_1,a_2,\dots,a_k \in A$. Simon a obtenu une caractérisation syntactique de ces langages qui demeure l'un des plus beaux résultats de la théorie des variétés. Nous dirons qu'un monoïde fini M est J-trivial si deux éléments de M qui engendrent le même idéal sont égaux, i.e. si la condition MaM = MbM entraine a = b. Les monoïdes J-triviaux forment une variété de monoïdes, notée $\underline{J}$, et contenue (strictement) dans la variété $\underline{A}$ des monoïdes apériodiques. Le théorème de Simon s'énonce alors ainsi :

Théorème 6.1. Un langage est de niveau 1 dans la hiérarchie de Straubing si et seulement si son monoïde syntactique est J-trivial.

Ce théorème montre que les langages de niveau 1 forment une variété de langages. En fait le résultat est général.

Théorème 6.2. Pour tout entier positif n, les langages de niveau n dans la hiérarchie de Straubing constituent une variété de langages.

Le théorème de Simon fournit également un algorithme pour décider si un langage est de niveau 1 dans la hiérarchie de Straubing. La situation est à peu près la même que pour la hauteur d'étoile restreinte : le niveau 1 est décidable, mais la question est ouverte à partir du niveau 2. De même, on sait que la hiérarchie est infinie. Ce dernier résultat découle du théorème de Brzozowski et Knast qui établit le même résultat pour une hiérarchie parallèle à celle de Straubing.

Théorème 6.3. La hiérarchie de Straubing est stricte : pour tout entier positif n, il existe des langages de niveau n+1 qui ne sont pas de niveau n. La réunion de tous les niveaux constitue la classe des langages sans étoile.

On dispose cependant d'une description non triviale du niveau 2 de la hiérarchie, obtenue par Straubing et l'auteur. Si $\underline{V}$ est une variété de monoïdes on note $\underline{PV}$ la variété de monoïdes engendrée par les monoïdes de la forme P(M) où M varie dans $\underline{V}$. On a alors

Théorème 6.4. Soit L un langage reconnaissable de A*. Les conditions suivantes sont équivalentes :

(1) L est de niveau 2 dans la hiérarchie de Straubing.

(2) L est dans l'algèbre de Boole engendrée par les langages de la forme $A_0^* a_1 A_1^* a_2 \dots a_k A_k^*$ où $k \geqslant 0$, $A_i \in A$ pour $0 \leqslant i \leqslant k$ et $a_i \in A$ pour $1 \leqslant i \leqslant k$.

(3) Le monoïde syntactique de L est élément de $\underline{PJ}$.

Malheureusement, ce résultat ne fournit pas d'algorithme pour décider si un langage est de niveau 2 dans la hiérarchie de Straubing. On connait également une description algébrique de la hiérarchie de variétés de monoïdes correspondant à la hiérarchie de Straubing, mais cette description est trop technique pour être exposée ici.

L'un des résultats récents les plus remarquables est le lien entre la hiérarchie de Straubing et une hiérarchie classique de la logique. Ce lien a été mis en évidence par W. Thomas à propos d'une autre hiérarchie mais son résultat s'adapte sans difficulté au cas de la hiérarchie de Straubing.

Considérons le langage $L = A^*aA^*$ sur l'alphabet $A = \{a,b\}$. Le langage L est l'ensemble des mots de A^* dont une lettre est un a. En termes de logique, nous dirons que L est l'ensemble des mots u qui satisfont la formule

$$\exists x \quad R_a x$$

Intuitivement cette formule se traduit par "il existe une occurrence du mot u qui est un a". De même la formule

$$\varphi = \exists x \ (R_a x \wedge \forall y (y < x \dashrightarrow R_b y))$$

se traduit par "il existe une occurrence du mot qui est un a et telle que toute occurrence qui la précède soit un b". L'ensemble des mots qui satisfont φ est donc le langage b^*aA^*.

Il reste à formaliser ce qui précède. Nous supposerons connu le vocabulaire de base de la logique. Soit A un alphabet fini. On considère l'ensemble des symbole $S = \{<\} \cup \{R_a \mid a \in A\}$. De façon classique, on construit des S-formules à partir des symboles de S et des symboles logiques (égalité, connexions logiques $\vee$, $\wedge$, $\neg$, $\longrightarrow$, $\longleftrightarrow$, variables, quantificateurs et constantes vrai et faux). A chaque mot u de A^* on associe une S-structure

$$M(u) = (M, <, (R_a)_{a \in A}).$$

où $M = \{1, \ldots, |u|\}$, $<$ est la relation d'ordre habituelle sur les entiers et pour $a \in A$, R_a est le sous-ensemble de M défini par $R_a x$ si et seulement si la x-ième lettre de u est un a. On dit qu'un langage L de A^* est défini par un énoncé φ si L est l'ensemble des mots u de A^* tels que M(u) soit un modèle pour φ. On a alors les résultats suivants :

Théorème 6.5. (Buchi-Elgot). Un langage L de A^* est reconnaissable si et seulement si il peut être défini par un S-énoncé du second ordre monadique faible.

Théorème 6.6. (Mc Naughton; Ladner). Un langage L de A^* est sans étoile si et seulement si il peut être défini par un S-énoncé du premier ordre.

On sait que toute formule du premier ordre est équivalente à une formule sous forme normale prénexe, c'est-à-dire de la forme $\varphi = Q(x_1,\ldots,x_k)\psi$ où $Q(x_1,\ldots,x_n)$ est une suite de quantificateurs $\exists x_i$ ou $\forall x_i$ et où ψ est une formule sans quantificateurs. Si $Q(x_1,\ldots,x_k)$ est formée de n blocs de quantificateurs tels que le premier bloc ne contienne que des quantificateurs existentiels, le second bloc que des quantificateurs universels, etc. , on dit que φ est une Σ_n-formule. On note Σ_n l'ensemble des Σ_n-formules et $B\Sigma_n$ l'ensemble des combinaisons booléennes de Σ_n-formules. Le résultat de Thomas peut alors s'adapter sous la forme :

Théorème 6.7. Un langage de A^* est de niveau n dans la hiérarchie de Straubing si et seulement si il peut être défini par un S-énoncé de $B\Sigma_n$.

Comme pour l'opération étoile on sait caractériser les variétés de langages fermées pour le produit de concaténation. Nous avons déjà défini les extensions apériodiques d'un groupe commutatif. C'est en fait un cas particulier de la définition suivante. Un monoïde M est une extension apériodique d'un monoïde N s'il existe un morphisme surjectif $\varphi : M \longrightarrow N$ tel que pour tout idempotent e de N, $e\varphi^{-1}$ soit un semi-groupe apériodique. Si $\underline{V}$ est une variété de monoïdes, on dira que $\underline{V}$ est fermée par extension apériodique si toute extension apériodique d'un monoïde de $\underline{V}$ est encore dans $\underline{V}$.

Théorème 6.8. (Straubing) Une variété de langages est fermée par produit si et seulement si la variété de monoïdes correspondante est fermée par extension apériodique.

Comme pour l'opération étoile on peut imposer diverses conditions restrictives au produit de concaténation (produit non ambigu, déterministe,..) et on a des théorèmes analogues au théorème 6.8. pour ces restrictions. cf. [9].

BIBLIOGRAPHIE.

N.B. La référence [1] contient une bibliographie détaillée sur les problèmes de hauteur d'étoile (à l'exception des références [3] et [10]). De même la référence [9] contient une bibliographie détaillée sur les variétés de langages et de semi-groupes.

[1] J.A. BRZOZOWSKI, Open problems about regular languages, Formal language theory, perpectives and open problems (R.V. Book editeur) Academic Press (1980), 23-47.

[2] S. EILENBERG, Automata, Languages and Machines, Academic Press, Vol. A (1974), Vol. B (1976).

[3] K. HASHIGUSHI, Regular languages of star height one, Information and Control 53 (1982), 199-210.

[4] W.H. HENNEMAN, Algebraic Theory of Automata, Ph. D. Dissertation, MIT (1971).

[5] G. LALLEMENT, Semigroups and Combinatorial Applications, Wiley, New-York (1979).

[6] M. LOTHAIRE, Combinatorics on words, Addison Wesley, Encyclopedia of Mathematics 17 (1983).

[7] J.F. PERROT, Variétés de langages et opérations. Theoretical Computer Science 7 (1978), 197-210.

[8] J.E. PIN, Sur le monoïde syntactique de L* lorsque L est un langage fini, Theoretical Computer Science 7 (1978), 211-215.

[9] J.E. PIN, Variétés de langages formels, Masson, Paris, (1984).

[10] W. THOMAS, Remark on the Star-Height Problem.

==========