RAIRO. INFORMATIQUE THÉORIQUE ET APPLICATIONS

VÉRONIQUE BRUYÈRE

Factorisation des ensembles préfixiels

RAIRO. Informatique théorique et applications, tome 23, n° 3 (1989), p. 295-315

<http://www.numdam.org/item?id=ITA_1989__23_3_295_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 23, n° 3, 1989, p. 295 à 315)

FACTORISATION DES ENSEMBLES PRÉFIXIELS (*)

par Véronique BRUYÈRE (1)

Communiqué par J. BERSTEL

Résumé. – *Dans cet article, nous étudions les factorisations non ambiguës des ensembles de mots, préfixiels et finis. Une description complète de telles factorisations est obtenue pour deux familles particulières. L'une d'elles contient les exemples envisagés par D. Perrin dans la construction de codes asynchrones.*

Abstract. – *This paper deals with unambiguous factorizations of finite prefix-closed subsets of words. Such factorizations are completely described for two particular families. One of them includes the examples treated by D. Perrin in the construction of asynchronous codes.*

1. INTRODUCTION

Un ensemble de mots est dit *préfixiel* s'il contient les débuts de chacun de ses mots. Nous étudions dans cet article les factorisations des ensembles préfixiels finis P en deux ensembles R et S

$$P=RS$$

où le produit de R et de S est *non ambigu*, c'est-à-dire où tout mot w de P s'écrit d'une et une seule manière $w=rs$ avec r un mot de R et s un mot de S.

De telles factorisations apparaissent dans les exemples de codes asynchrones construits par D. Perrin dans [5]. Nous tentons ici de décrire complètement la structure des facteurs R et S dont le produit est non ambigu et forme un ensemble RS préfixiel et fini, ou tout au moins de généraliser les exemples connus de ces produits.

(*) Reçu avril 1987, version finale janvier 1988.
(1) Faculté des Sciences, 15, avenue Maistriau, B-7000 Mons, Belgique.

Informatique théorique et Applications/Theoretical Informatics and Applications
0296-1598 89/03 295 21/$4.10/

Nous présentons deux familles particulières de ces factorisations, dont nous donnons la caractérisation complète. La première famille est assez simple à construire. La seconde famille contient en particulier tous les exemples proposés dans [5]. Nous avons pu caractériser complètement les factorisations qui y apparaissent grâce à un résultat de Krasner [3] qui décrit toutes les décompositions possibles du polynôme

$$1+x+x^2+\ldots+x^n$$

en un produit de deux polynômes à coefficients 0 et 1.

Les factorisations présentées dans la seconde famille peuvent être considérées comme une généralisation en variables non commutatives du résultat de Krasner.

Les résultats obtenus sont un premier pas vers la description de toutes les factorisations non ambiguës d'ensembles préfixiels finis. Cependant, la forme des factorisations quelconques ne semble pas être connue.

Nous allons maintenant décrire plus en détail les résultats de cet article.

Une première méthode pour construire des factorisations non ambiguës d'ensembles préfixiels finis utilise des résultats connus en théorie des codes à longueur variable [1]. Rappelons qu'un *code* X est un ensemble de mots tel qu'un message dont les lettres sont codées par des mots de X, est déchiffrable d'une et une seule façon. Les *codes préfixes* sont les codes les plus simples à construire : ce sont les ensembles X de mots dont aucun mot n'apparaît au début d'un autre mot de X. On dit aussi que X est un *code préfixe maximal* s'il est préfixe et n'est inclus strictement dans aucun autre code préfixe. Il existe une *correspondance biunivoque* entre l'ensemble des codes préfixes maximaux finis et celui des ensembles préfixiels finis [1]. On peut penser qu'une certaine opération définie sur les codes préfixes, se traduise, *via* cette bijection, en une opération sur les ensembles préfixiels correspondants, qui fournisse une factorisation non ambiguë. Cette opération existe : étant donnés deux codes Y et Z préfixes maximaux finis, on peut les *composer* pour obtenir un nouveau code X préfixe maximal fini. Si on passe aux ensembles préfixiels associés, celui qui correspond à X se factorise en un produit non ambigu RS où S est l'ensemble préfixiel placé en bijection avec le code Z (proposition 6).

La réciproque de ce résultat est en général fausse : toute factorisation non ambiguë d'un ensemble préfixiel fini ne s'obtient pas nécessairement grâce à l'opération de composition des codes préfixes, sauf si le second facteur S est lui-même un ensemble préfixiel (proposition 7).

Nous présentons ensuite une deuxième famille de factorisations, pour laquelle les exemples envisagés par D. Perrin dans [5] forment une sous-famille.

Le facteur S qui apparaît dans les ensembles RS de cette famille, est supposé non préfixiel; on suppose aussi que les mots qui manquent à S pour être préfixiel sont tous début d'une puissance d'un mot t, t étant le plus petit de ces mots. Ces ensembles RS sont appelés *t-préfixiels*. Dans un premier temps, nous étudions les ensembles R et S restreints à l'ensemble T des puissances du mot t. On montre dans le théorème 10 que ces ensembles, écrits comme des polynômes en t à coefficients 0 et 1, factorisent un polynôme du type

$$1+t+t^2+\ldots+t^n.$$

Le résultat de Krasner permet donc de décrire leur forme. Ensuite, nous étendons cette description aux ensembles R et S entiers et donnons une caractérisation complète des ensembles t-préfixiels finis se factorisant de manière non ambiguë (théorème 11). Lorsqu'on passe des ensembles $R \cap T$ et $S \cap T$ aux ensembles R et S entiers, on perd la factorisation obtenue dans le théorème 10 à cause des mots de S situés en dehors T.

Les deux familles décrites présentent une intersection commune : par exemple, certains ensembles RS t-préfixiels de la seconde famille peuvent se factoriser en deux ensembles P et Q, où Q est préfixiel, et appartenir ainsi à la première famille.

Nous ne sommes pas parvenus à décrire complètement l'intersection. Cependant, nous établissons dans la proposition 13 des hypothèses sous lesquelles un ensemble t-préfixiel fini qui s'écrit comme un produit RS non ambigu, se factorise de façon non ambiguë en un produit PQ où Q est un ensemble préfixiel.

2. PRÉLIMINAIRES

Les notations, définitions et résultats de ce paragraphe sont empruntés au livre *Theory of codes* [1].

A désigne un alphabet (que nous supposerons toujours fini). 1 dénote le mot vide, A^* le monoïde libre engendré par A et A^+ le semi-groupe libre $A^* \setminus \{1\}$. On note $|w|$ la longueur du mot w. Les opérations usuelles définies sur les sous-ensembles de A^* sont notées $+$ pour l'union, . pour le produit

de concaténation, * pour l'étoile; on note également

$$X-Y \text{ l'ensemble } \{ w \in A^* \mid w \in X \setminus Y \}$$
$$XY^{-1} \text{ l'ensemble } \{ w \in A^* \mid \exists\, x \in X, \exists\, y \in Y : x = wy \}.$$

Parfois, l'ensemble $\{1\}$ sera noté 1.

Si u, v sont deux mots de A^*, u est dit *facteur gauche (propre)* de v si

$$\exists\, w \in A^* (\in A^+) : \quad uw = v.$$

Cette relation définit un ordre partiel sur A^*. On note

$$u \leqq (<) v.$$

XA^- désigne l'ensemble de tous les mots qui sont facteur gauche propre d'un mot de $X \subset A^*$.

Un ensemble $X \subset A^*$ est dit *préfixiel* si

$$\forall x \in X, \quad \forall w \in A^* : \qquad w < x \quad \Rightarrow \quad w \in X.$$

Par contre, X est dit *préfixe* si

$$\forall x, x' \in X : \qquad x' \leqq x \quad \Rightarrow \quad x' = x.$$

X est *préfixe maximal* s'il est préfixe et n'est contenu strictement dans aucun autre ensemble préfixe sur le même alphabet A.

Rappelons qu'un ensemble $X \subset A^*$ préfixe différent du mot vide est un exemple particulier de code, appelé *code préfixe.* Rappelons aussi qu'un *code* $X \subset A^*$ est la base d'un sous-monoïde libre de A^*. En d'autres termes, X est un code si

$$\forall n, m > 0, \quad \forall x_1, \ldots, x_n \in X, \quad \forall y_1, \ldots, y_m \in X$$
$$x_1 x_2 \ldots x_n = y_1 y_2 \ldots y_m \quad \Rightarrow \quad n = m \qquad \text{et} \qquad x_i = y_i, \qquad \forall i.$$

Comme la notion d'ensemble préfixe maximal, on définit celle de code maximal : un code $X \subset A^*$ est dit *maximal* s'il n'est contenu strictement dans aucun autre code sur A.

A présent, citons des résultats classiques en théorie des codes, dont nous aurons besoin plus loin dans les preuves.

Il existe une correspondance biunivoque entre l'ensemble des codes préfixes maximaux finis et celui des ensembles préfixiels finis. Dans la proposition

qui suit, $\underline{X}$ désigne la série caractéristique de l'ensemble X :

$$\underline{X} = \sum_{w \in A^*} c_w w \qquad \text{où} \quad c_w = 1 \text{ si } w \in X \quad \text{et} \quad c_w = 0 \text{ sinon.}$$

PROPOSITION 1 : *Si $X \subset A^*$ est un code préfixe maximal fini, alors l'ensemble $P = XA^-$ est préfixiel fini non vide et*

$$\underline{X} - 1 = \underline{P}(\underline{A} - 1).$$

Réciproquement, si $P \subset A^*$ est un ensemble préfixiel fini non vide, alors $X = PA - P$ est un code préfixe maximal fini, et $P = XA^-$. ■

Il existe une opération de *composition* sur les codes qui permet de construire un nouveau code à partir de deux codes. Plus précisément, soit A, B deux alphabets et Y, Z deux codes respectivement sur B et A tels que les mots de Z soient mis en bijection β avec les lettres apparaissant dans les mots de Y. On suppose que toutes les lettres de B figurent dans les mots de Y. Y et Z sont alors appelés *codes composables.* L'ensemble $X = \beta(Y)$ obtenu en remplaçant, *via* β, les lettres des mots de Y par les mots de Z, est un code sur A, noté $Y \circ_\beta Z$. X est dit *décomposable* sur Z. Dans les cas triviaux où X est décomposable uniquement sur A et X, on dit que X est *indécomposable.*

L'opération de composition des codes préserve certaines propriétés comme par exemple celle d'ensemble préfixe maximal fini :

PROPOSITION 2 : *Soit Y, Z deux codes composables et $X = Y \circ_\beta Z$.*

Alors X est préfixe maximal fini ssi Y et Z sont préfixes maximaux finis. ■

Il est facile de tester si un code X est décomposable sur un code Z, dans le cas où X est un code maximal :

PROPOSITION 3 : *Soit X et Z deux codes sur un alphabet A, où X est un code maximal. X se décompose sur Z ssi $X \subset Z^*$.* ■

3. UNE CONSTRUCTION SIMPLE

Nous essayons de décrire, dans ce paragraphe et le paragraphe suivant, les ensembles P préfixiels finis qui se factorisent de manière non ambiguë en deux ensembles R et S :

$$P = RS$$

et

$$\forall r, r' \in R, \quad \forall s, s' \in S : \qquad rs = r's' \Rightarrow r = r' \qquad \text{et} \qquad s = s'.$$

La non-ambiguïté du produit signifie que tout mot w de l'ensemble RS s'écrit de manière unique $w = rs$ avec $r \in R$ et $s \in S$. En termes de séries caractéristiques, elle se traduit par

$$\underline{P} = \underline{R}\,\underline{S}.$$

Nous ne considérerons pas les factorisations triviales où l'un des ensembles R, S est égal à 1. Nous appellerons *produits non ambigus préfixiels finis* les ensembles préfixiels et finis qui se factorisent de manière non ambiguë en deux ensembles différents du mot vide.

De tels produits apparaissent dans les exemples de codes préfixes asynchrones proposés par D. Perrin dans [5]. L'exemple suivant est issu de ce papier.

Exemple 4 : Sur l'alphabet $\{a, b\}$, le produit des ensembles :

$R = \{1, ba\}$

et

$S = \{1, a, aa, aab, aaba, aabab, b, baba, babab, bb, bba, bbab\}$,

égal à

$RS = \{1, a, aa, aab, aaba, aabab, b, ba, baa, baaa, baaab, baaaba, baaabab, bab, baba, babab, bababa, bababab, babb, babba, babbab, bb, bba, bbab\}$

est non ambigu, préfixiel et fini.

Voici un autre exemple plus simple.

Exemple 5 : $RS = \{1, b, ba, bab\}$, $R = \{1, ba\}$ et $S = \{1, b\}$.

Nous proposons ici une première méthode — assez simple — pour construire un produit non ambigu préfixiel fini. La famille composée de ces produits contiendra le second exemple cité, mais pas le premier.

On sait (proposition 1) qu'à chaque code préfixe maximal fini est associé un et un seul ensemble préfixiel fini. On sait aussi que la composition de deux codes préfixes maximaux finis redonne un code X préfixe maximal fini (proposition 2). L'ensemble préfixiel fini associé à ce code X est en fait un produit non ambigu.

PROPOSITION 6 : *Soit $X \subset A^*$ un code préfixe maximal fini décomposable sur un code $Z \subset A^*$. On note $Y \subset B^*$ le code et β la bijection tels que $X = Y \circ_\beta Z$.*

Alors, l'ensemble XA^- est préfixiel, fini et se factorise de manière non ambiguë en

$$XA^- = \beta(YB^-).ZA^-.$$

Preuve : Posons $P_X = XA^-$, $P_Y = YB^-$ et $P_Z = ZA^-$. On a par la proposition 1

$$\beta(Y) = \beta(P_Y B - P_Y) = \beta(P_Y)\beta(B) - \beta(P_Y).$$

Le produit $\beta(P_Y)\beta(B)$ est non ambigu car β étendu à B^* est injectif. Comme $\beta(B) = Z$ et $\beta(Y) = X$, on a en passant aux séries caractéristiques

$$\underline{X} - 1 = \underline{\beta(P_Y)}(\underline{Z} - 1).$$

Par la proposition 1, $\underline{Z} - 1 = \underline{P_Z}(\underline{A} - 1)$. Donc,

$$\underline{X} - 1 = \underline{\beta(P_Y)\,P_Z}(\underline{A} - 1).$$

Comme $\underline{X} - 1 = \underline{P_X}(A - 1)$, P_X est un ensemble préfixiel fini, égal au produit non ambigu $\beta(P_Y)P_Z$ (proposition 1). ■

L'exemple 5 précédent est obtenu par la méthode décrite dans la proposition. Le code préfixe maximal fini

$$X = RSA - RS = \{ a, baa, baba, babb, bb \}$$

associé à RS est obtenu par composition des codes

$$Y = \{ c, dc, dd, de, e \} \text{ sur l'alphabet } \{ c, d, e \}$$
$$Z = \{ a, ba, bb \} \text{ sur l'alphabet } \{ a, b \}$$

en utilisant la bijection β qui envoie respectivement les lettres c, d et e sur les mots a, ba et bb de Z.

Dans le cas des compositions triviales où $X = X \circ_\beta A$, $X = B \circ_\beta X$, on obtient, selon la méthode décrite dans la proposition, les factorisations triviales $XA^- = XA^-.1$ et $XA^- = 1.XA^-$, que nous avons décidé de ne pas considérer en début de ce paragraphe. Par conséquent, les ensembles préfixiels correspondant aux codes préfixes maximaux finis qui sont indécomposables ne peuvent se factoriser selon cette méthode. C'est le cas de tous les ensembles préfixiels apparaissant dans [5], les codes préfixes associés étant indécomposables.

Par conséquent, la réciproque de la proposition 6 est fausse. Remarquons cependant que le second facteur du produit $\beta(YB^-).ZA^-$ construit dans la proposition, est préfixiel. Sous cette hypothèse supplémentaire, la proposition

admet alors une réciproque :

PROPOSITION 7 : *Si $RS \subset A^*$ est un produit non ambigu préfixiel fini non vide où S est aussi préfixiel, alors $X = RSA - RS$ est un code préfixe maximal fini qui se décompose sur le code $Z = SA - S$.*

Preuve : Par la proposition 1,

$$\underline{X} - 1 = \underline{R}\,\underline{S}\,(\underline{A} - 1).$$

Soit Z le code préfixe maximal fini $SA - S$. On a

$$\underline{X} - 1 = \underline{R}\,(\underline{Z} - 1)$$

$$\underline{Z}^* = \underline{X}^*\,\underline{R}$$

car $(\underline{X} - 1).\,\underline{X}^* = 1$ et $(\underline{Z} - 1).\,\underline{Z}^* = 1$ (cf. [1]).

Puisque l'ensemble RS est préfixiel et non vide, $1 \in RS$ et donc $1 \in R$. Par conséquent, $X^* \subset Z^*$ et X se décompose sur Z par la proposition 3. ■

4. ENSEMBLES *t*-PRÉFIXIELS

Nous construisons dans ce paragraphe une autre famille particulière de produits RS non ambigus préfixiels finis, dont le second facteur S n'est pas préfixiel. Cette famille rassemble toutes les factorisations non ambiguës d'ensembles préfixiels apparaissant dans [5].

Reprenons l'exemple 4 :

$R = \{1, ba\}$

$S = \{1, a, aa, aab, aaba, aabab, b, baba, babab, bb, bba, bbab\}$.

L'ensemble S n'est pas préfixiel. Les mots qui lui manquent pour être préfixiel sont $\{ba, bab\}$. Remarquons que ces mots sont facteur gauche d'une puissance de ba.

Les ensembles RS préfixiels, envisagés ici, sont de cette forme : si t est un mot de plus petite longueur manquant à S pour être préfixiel, alors tous les autres mots manquants sont facteur gauche d'une puissance de t. Ces ensembles sont appelés *t-préfixiels.*

Nous avons pu caractériser complètement de tels produits RS t-préfixiels, non ambigus et finis, grâce à deux résultats, l'un de Fine et Wilf [2], l'autre de Krasner [3].

THÉORÈME 8 (Fine et Wilf) : *Soit $x, y \in A^*$, $n = |x|$, $m = |y|$ et d le plus grand commun diviseur de n et m.*

S'il existe un mot w préfixe de x^p et y^q tel que $|w| \geq n+m-d$, alors x et y sont puissance d'un même mot. ■

Le théorème de Krasner décrit les décompositions du polynôme

$$1+x^a+x^{2a}+\ldots+x^{na}$$

en deux polynômes $q(x)$ et $r(x)$ à coefficients réels positifs.

Notons $p_{n,a}(x)$ le polynôme $(x^n-1)/(x^a-1)$ où le naturel a divise n. On appelle *chaîne de diviseurs* de (n, a), toute suite de naturels deux à deux distincts $(n_0, n_1, \ldots, n_k)$ telle que

$$n_0=a, \qquad n_k=n \qquad \text{et} \qquad n_i \mid n_{i+1}, \qquad 0 \leq i<k.$$

On peut alors décomposer $p_{n,a}(x)$ en le produit

$$P_{n,a}(x)=\frac{x^{n_k}-1}{x^{n_{k-1}}-1}\cdot\ldots\cdot\frac{x^{n_2}-1}{x^{n_1}-1}\cdot\frac{x^{n_1}-1}{x^{n_0}-1}$$

$$=\prod_{i=0}^{k-1} p_{n_{i+1},n_i}(x).$$

Krasner a montré que les seules factorisations possibles de $p_{n,a}(x)$ en deux polynômes à coefficients réels positifs s'obtenaient en regroupant en deux paquets les polynômes $p_{n_{i+1},n_i}(x)$, $0 \leq i \leq k-1$.

THÉORÈME 9 (Krasner) : *Si le polynôme $p_{n,a}(x)$ se décompose en un produit $q(x)r(x)$ de deux polynômes à coefficients réels positifs, alors il existe une chaîne $(n_0, n_1, \ldots, n_k)$ de diviseurs de (n, a) et une partition $I \cup J$ de $\{0, 1, \ldots, k-1\}$ telles que*

$$q(x)=\prod_{i\in I} p_{n_{i+1},n_i}(x) \qquad \text{et} \qquad r(x)=\prod_{j\in J} p_{n_{j+1},n_j}(x). \quad ■$$

Nous dirons que $q(x)r(x)$ est une *factorisation de Krasner* du polynôme $p_{n,a}(x)$.

A présent, passons à l'étude des ensembles *RS t-préfixiels*, c'est-à-dire, comme nous l'avons dit précédemment, les ensembles *RS* préfixiels, où S n'est pas préfixiel, et où tout mot $w \in S(A^*)^{-1}-S$ est facteur gauche d'une puissance de t, t étant un mot de longueur minimale dans $S(A^*)^{-1}-S$.

Dans un premier temps, nous nous intéressons à décrire la répartition des mots de R et de S (dont le produit RS est non ambigu, t-préfixiel et fini), uniquement dans l'ensemble t^*A^- des mots facteur gauche d'une puissance

de t. A cause de la non-ambiguïté du produit RS, la disposition de mots est très régulière : les séries caractéristiques des ensembles R et S restreints à t^*, donnent une factorisation de Krasner de la série caractéristique de $RS \cap t^*$.

Ce résultat peut s'observer sur l'exemple 4. Le mot t est égal à ba et

$$\underline{R \cap t^*} = 1 + t, \qquad \underline{S \cap t^*} = 1 + t^2$$

$$\underline{RS \cap t^*} = 1 + t + t^2 + t^3$$

$$= (1 + t)(1 + t^2).$$

Dans le théorème qui suit, on note respectivement $(RS)_t$, R_t, S_t les ensembles $RS \cap t^* A^-$, $R \cap t^* A^-$, $S \cap t^* A^-$. Si $s_{\max}$ est le mot le plus long de S_t, $R_{t,S}$ désigne l'ensemble $\{ w \in R_t \mid w < s_{\max} \}$.

THÉORÈME 10 : *Soit $RS \subset A^*$ un produit non ambigu, t-préfixiel, fini et non vide. Alors, il existe un ensemble $T \subset t^*$, un naturel $n > 3$ tels que*

$$\underline{RS \cap t^*} = p_{n,1}(t)$$

$$\underline{S_t} = \underline{T} . \underline{t A^-},$$

$\underline{R_t}$ $\underline{T}$ est une factorisation de Krasner de $p_{n,1}(t)$ et $\underline{R_t}$ est divisible par $p_{m,1}(t)$, où m divise n.

Preuve : La non-ambiguïté du produit RS conduit aux égalités suivantes, qui seront plusieurs fois utilisées dans la preuve

$$R \cap (S - 1) = \varnothing \qquad \text{et} \qquad (R - 1) \cap S = \varnothing$$

$$R \cap (R - 1)(S - 1) = \varnothing \qquad \text{et} \qquad S \cap (R - 1)(S - 1) = \varnothing.$$

Pour la première égalité, puisque RS est préfixiel et non vide, $1 \in RS$ et donc $1 \in R$, $1 \in S$. Supposons qu'il existe un mot w non vide dans $R \cap S$. Alors, $1 . w = w . 1$ appartient à RS, en contradiction avec la non-ambiguïté du produit RS. Les autres cas se démontrent de la même façon.

A présent, nous allons montrer que pour tout $n \geqq 0$,

si $t^n \in RS$, alors,

soit $t^n . t A^- \subset S$,

soit $t^n \in (R - 1) S$ et $t^n . (t A^- - 1) \subset (R - 1)(S - 1)$.

Nous en déduirons ensuite le résultat final.

La preuve est basée sur une récurrence sur les puissances n-ièmes de t.

Pour $n=0$, il est clair que $tA^- \subset S$ par minimalité de $|t|$. Soit $n>0$ tel que $t^n \in RS$. Puisque RS préfixiel, pour tout $m<n$, $t^m \in RS$. Supposons le résultat vrai pour tout $m<n$.

Si $t^n \in (R-1)S$, par hypothèse de récurrence,

$$\exists r \in R \cap t^+, \quad \exists s \in S \cap t^*: \qquad t^n = rs.$$

Comme $|s| < |t^n|$, $s.tA^- \subset S$ et donc

$$t^n.(tA^- - 1) \subset (R-1)(S-1).$$

Si $t^n \in S$, prouvons par l'absurde que pour tout mot u tel que $t^n < u < t^{n+1}$, u appartient à S.

Comme $t \in R$, $t^{n+1} \in RS$ et donc

$$\forall u, t^n < u < t^{n+1}: \quad u \in RS.$$

Soit u de longueur minimale appartenant à $(R-1)S$:

$$\exists r \in R-1, \quad \exists s \in t^*.tA^-: \qquad u = rs.$$

Si $s \neq 1$, alors $r \in t^+$ (par non-ambiguïté du produit RS) et s s'écrit $t^m v$, avec $m<n$ et $1<v<t$. Par hypothèse de récurrence, $t^m \in S$ et $1.t^n$, $r.t^m$ sont deux factorisations distinctes d'un même mot de RS. Par conséquent, $s=1$ et $u \in R$. Soit $\sqrt{t}$ le plus petit mot tel que $t=(\sqrt{t})^k$ avec $k \geqq 1$.

Si

$$u = t^n(\sqrt{t})^{k'}, \qquad 0<k'<k,$$

alors $t^{k-k'} \in tA^- \subset S$, et $t.t^n = u.t^{k-k'}$. C'est impossible, par non-ambiguïté du produit RS. Donc, u n'appartient pas à $(\sqrt{t})^*$.

Considérons l'ensemble de mots

$$W = \{ w \in A^* \mid w \in u.t^*A^- \}.$$

W ne peut comprendre aucun mot de S. Sinon, soit un mot s appartenant à $S \cap W$. Comme $u.tA^- \subset (R-1)S$, $ut \leqq s$ par non-ambiguïté de RS. Soit v tel que $va=ut$, où a est une lettre de A. Si $v \in t^*A^-$, le théorème de Fine et Wilf avec $x=t$ et $y=tu'$ où u' est défini par $u=t^nu'$, montre que t et tu' sont puissances d'un même mot, en particulier du mot $\sqrt{t}$. C'est impossible car $u' \notin (\sqrt{t})^*$. Donc, $v \notin t^*A^-$, $v \notin S$ et $v<s$. L'existence d'un tel mot v est contradictoire avec les hypothèses faites sur S.

Par conséquent, RS étant préfixiel, la répartition des mots de R dans l'ensemble $u.u(A^*)^{-1}$ est identique à celle dans $u(A^*)^{-1}$; u joue le rôle du mot vide. On conclut que u^2 appartient à R.

En suivant le même raisonnement, on montre que l'ensemble $u^2.u(A^*)^{-1}$ se comporte lui aussi comme $u(A^*)^{-1}$, de même pour u^3, u^4 ...

Donc, $u^* \subset R$, ce qui est impossible car RS est fini.

Ceci achève la démonstration par récurrence sur n.

Le résultat précédent montre que $R_t \subset t^*$ et que S_t s'écrit $T.tA^-$ avec $T \subset t^*$. Donc,

$$(RS)_t = R_t.S_t = R_t.T.tA^-.$$

Soit $\underline{R}_t$ et $\underline{T}$ les séries caractéristiques respectivement des ensembles R_t et T. RS étant préfixiel, $\underline{R}_t\,\underline{T}$ est un polynôme en t de la forme

$$p_{n,1}(t) = \frac{t^n - 1}{t-1} \qquad \text{où} \qquad n > 3.$$

Par le résultat de Krasner, il existe une chaîne $(n_0, n_1, \ldots, n_k)$ de diviseurs de $(n, 1)$, et une partition en deux ensembles I_R, I_S non vides de $\{0, 1, \ldots, k-1\}$ telles que

$$\underline{R}_t = \prod_{i \in I_R} p_{n_{i+1}, n_i}(t)$$

$$\underline{T} = \prod_{i \in I_S} p_{n_{i+1}, n_i}(t).$$

Comme 1, $t \in R$, on peut préciser que $p_{n_1, n_0}(t)$ divise $\underline{R}_t$. ■

Le résultat qui suit élargit la description de R et de S à tout le monoïde libre A^* et donne une caractérisation des produits non ambigus t-préfixiels finis. La factorisation de S_t en $T.tA^-$ est perdue lorsqu'on passe à S tout entier, car les mots en dehors de t^*A^- ne présentent pas nécessairement la même régularité que celle des éléments de S_t.

THÉORÈME 11 : *Soit R, S deux sous-ensembles finis non vides de A^*. Alors, le produit RS est non ambigu et t-préfixiel ssi*

(1) *Il existe $T \subset t^*$, $n \in \mathbb{N}_0$ et $m \neq 1$ un diviseur de n tels que*

$$\underline{RS \cap t^*} = p_{n,1}(t)$$

$$\underline{S}_t = \underline{T}\,\underline{t}\,\underline{A^-}$$

$\underline{R}_t\,\underline{T}$ *est une factorisation de Krasner de* $p_{n,1}(t)$ *et* $p_{m,1}(t)$ *divise* $\underline{R}_t$.

(2) *Pour tout* $s \in S - S_t$, *soit* $s = uav$ *où* $u \in t^* A^-$ *est de longueur maximale,* $a \in A$ *et* $v \in A^*$.

Alors $u \in S_t (A^*)^{-1}$,

$$\forall v' \in A^*,\ v' < v: \quad uav' \in S$$

et

$$\forall r_1, r_2 \in R_t, r_1 \neq r_2: \quad r_1 s \notin r_2 (S - S_t).$$

(3) *Il existe* $R' \subset A^*$ *comprenant le mot vide tel que*

$$\underline{R} = \underline{R}'\,\underline{R}_{t,S}$$

$$\forall r' \in R' - 1: \quad r' \in R_{<r'} R_{t,S} SA - R_{<r'} R_{t,S} S$$

où $R_{<r'} = \{ r \in R' \mid r < r' \}$.

Les conditions 1-3 du théorème ci-dessus signifient respectivement que

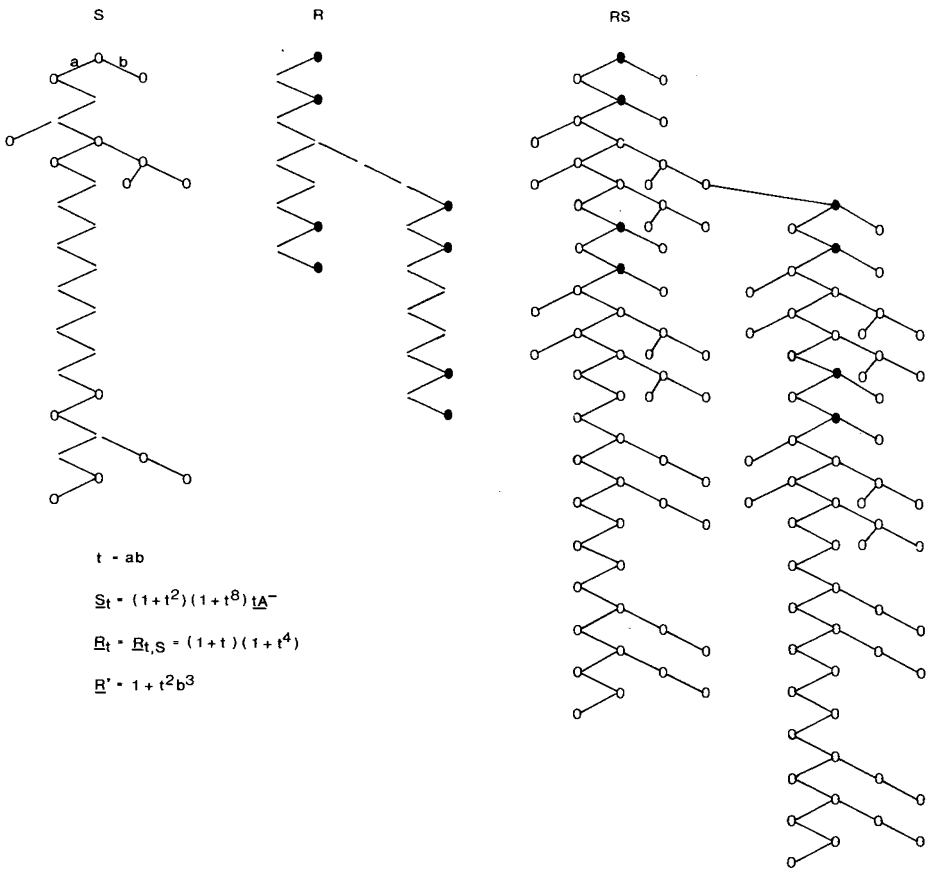
(1) Les séries caractéristiques des ensembles R et S restreints à t^* forment une factorisation de Krasner du polynôme $p_{n,1}(t)$ égal à $\underline{RS \cap t^*}$ (théorème 10).

(2) L'ensemble S en dehors de $t^* A^-$ est préfixiel et forme avec R un produit non ambigu. De plus, pour les mots $s = uw$ de S où u représente le début du mot s situé dans $t^* A^-$, u ne peut être plus long que le plus long mot de S_t.

(3) L'ensemble R se décompose en une ou plusieurs copies C de la structure $R_{t,S}$. Ces différentes copies s'agencent de telle sorte que les ensembles CS « collent » les uns aux autres sans superposition.

L'exemple qui suit, ainsi que l'exemple déjà cité, illustrent ces différentes propriétés.

Exemple 12 :



Preuve du théorème 11 :

(A) Avant de prouver le théorème, montrons que pour des ensembles R, $S \subset A^*$ finis non vides, les conditions rassemblées dans les points (1) et (2) impliquent que les ensembles $R_{t,S}S$ et R_tS sont préfixiels.

L'hypothèse (1) indique que R_tS_t est préfixiel.

Soit un mot w appartenant à $R_{t,S}S$. On factorise w en uv où $u \in t^*A^-$ est de longueur maximale. On considère aussi une factorisation rs de w dans $R_{t,S}S$. On a $r \leqq u \leqq rs$.

Soit $u_1 \in A^*$ tel que $u = ru_1$ et $s = u_1 v$. Le mot u appartient à $R_{t,S}S_t$, car sinon $u_1 \notin S_t(A^*)^{-1}$, ce qui contredit l'hypothèse (2).

Soit w' un mot facteur gauche de w. Si $w' \leqq u$, alors $w' \in R_{t,S}S$ car R_tS_t est préfixiel. Si $u < w' < w$, alors, par définition de u, $w' \notin t^*A^-$ et $w' \in R_{t,S}S$ par

l'hypothèse (2). Il est facile de prouver que $R_t S$ est préfixiel : R_t égale $R_{t,S}$ ssi S_t possède des puissances de t supérieures à celles de R_t, c'est-à-dire ssi il existe k diviseur de n tel que $p_{n,k}(t)$ divise S_t. Donc, si $R_t \neq R_{t,S}$, on a

$$\underline{R}_t = p_{n,k}(t)\,\underline{R}_{t,S}$$

et donc $R_t S$ est préfixiel.

(B) Supposons d'abord que le produit RS soit non ambigu et t-préfixiel et prouvons les points 1-3.

(1) Par le théorème 10, on connaît exactement la forme de R et de S dans l'ensemble $t^* A^- : (RS)_t$ se factorise en $\underline{R_t}\,\underline{TtA^-}$ où T est inclus dans t^*, $\underline{S_t} = \underline{TtA^-}$ et $\underline{R_t}\,\underline{T}$ est une factorisation de Krasner de $\underline{RS \cap t^*}$.

(2) Le comportement de S en dehors de $t^* A^-$ est facile à décrire.
Soit $s \in S - S_t$. Soit u de longueur maximale appartenant à $t^* A^-$, $a \in A$ et $v \in A^*$ tels que $s = uav$. Alors, $u \in S_t (A^*)^{-1}$.

En effet, soit $r_{\max}$ le plus long mot de R_t. Comme $r_{\max} s \in RS$, et que RS est préfixiel, $r_{\max} u \in (RS)_t = R_t S_t$. Donc, $u \in S_t (A^*)^{-1}$.

De plus, vu les hypothèses faites sur S, $uav' \in S$ quel que soit $v' < v$.

Enfin, soit $r_1, r_2 \in R_t$ distincts. Il est évident que $r_1 s \notin r_2 (S - S_t)$ car RS est un produit non ambigu.

Par le point (A) précédent, on peut conclure que $R_{t,S} S$ et $R_t S$ sont préfixiels.

(3) Maintenant, montrons que R est égal au produit non ambigu $R' R_{t,S}$ où R' satisfait aux conditions du théorème.

On sait déjà que R_t s'écrit sous cette forme (théorème 10). Notons $R_t = R_0$. Supposons que R comprenne d'autres éléments, en dehors de l'ensemble $t^* A^-$. Puisque RS, $R_t S$ sont préfixiels, et que le produit RS est non ambigu,

$$R \cap Y_0 \neq \varnothing \qquad \text{avec} \quad Y_0 = R_0 SA - R_0 S.$$

Soit $r \in R \cap Y_0$. Aucun mot de S ne possède r comme facteur gauche. Donc, $r R_{t,S} \subset R$ car RS est préfixiel.

On note R_1 l'ensemble $R_0 + \sum_{r \in R \cap Y_0} R_{t,S}$.

Supposons que $R - R_1$ soit encore non vide. Par le même raisonnement,

$$R \cap Y_1 \neq \varnothing \qquad \text{avec} \quad Y_1 = R_1 SA - R_1 S$$

et $r R_{t,S} \subset R$ pour tout mot r dans $R \cap Y_1$, etc.

Par conséquent, il existe un ensemble $R' \subset A^*$ comprenant 1 tel que

$$\underline{R} = \underline{R'}\, \underline{R}_{t,S}$$

et pour tout mot r' dans $R' - 1$,

$$r' \in R_{<r'} R_{t,S} SA - R_{<r'} R_{t,S} S$$

où $R_{<r'}$ est l'ensemble des mots de R' qui sont facteur gauche propre de r'.

(C) Pour la réciproque, on sait que $R_{t,S} S$ est préfixiel par le point (A).

Montrons par récurrence sur les mots r' de $R' - 1$ que RS est préfixiel. Soit $r' \in R - 1$. Par hypothèse de récurrence, supposons que $R_{<r'} R_{t,S} S$ soit préfixiel. Vu l'hypothèse (3) et comme $R_{t,S} S$ est préfixiel, il est clair que l'ensemble $(R_{<r'} + r') R_{t,S} S$ est lui aussi préfixiel.

L'ensemble RS est t-préfixiel. En effet, on vient de montrer que RS est préfixiel; l'hypothèse R_t divisible par $p_{m,1}(t)$, implique que S n'est pas préfixiel; les conditions (1) et (2) montrent que tout mot w de $S(A^*)^{-1} - S$ est facteur gauche d'une puissance de t où t est le plus petit mot de $S(A^*)^{-1} - S$.

Il reste à prouver que le produit RS est non ambigu. Par les hypothèses (1) et (2), on sait que $R_t S$ est non ambigu. Supposons qu'il existe $r_1, r_2 \in R$, $s_1, s_2 \in S$ tels que

$$r_1 s_1 = r_2 s_2, \qquad r_2 < r_1 \qquad \text{et} \qquad r_1 \notin R_t.$$

Soit $r'_1 \in R'$ tel que $r_1 \in r'_1 R_{t,S}$. Donc, $r'_1 \notin R_t$.

Si $r_2 < r'_1$, soit le mot s'_2 facteur gauche de s_2 tel que $r'_1 = r_2 s'_2$. Par l'hypothèse (2), $s'_2 \in S$. Donc, $r_2 \in R_{<r'_1} R_{t,S}$ et $r'_1 \in R_{<r'_1} R_{t,S} S$. C'est impossible vu l'hypothèse (3).

Si $r'_1 \leqq r_2$, alors r_2 appartient à $r'_1 R_{t,S}$. Si s'_2 désigne le mot tel que $r_1 = r_2 s'_2$, alors $s'_2 \in S$ vu l'hypothèse (2). C'est contradictoire avec la non-ambiguïté du produit $R_t S$ car

$$r_1 = r'_1 r''_1, \qquad r''_1 \in R_{t,S}$$
$$r_2 = r'_1 r''_2, \qquad r''_2 \in R_{t,S}$$

et

$$r''_1 = r''_2 s'_2.$$

Par conséquent, le produit RS est non ambigu. ■

Dans [5], comme nous l'avons déjà dit, D. Perrin construit des codes X préfixes maximaux finis, dont on peut vérifier que les ensembles préfixiels

XA^- correspondants se factorisent en un produit RS non ambigu t-préfixiel. Ces produits RS font donc partie de la famille qui vient d'être décrite dans ce paragraphe. Cependant, ils sont assez particuliers car leur facteur R est toujours égal à $\{1, t\}$.

5. DÉCOMPOSITION

Dans les paragraphes 3 et 4, nous avons envisagé deux familles de produits RS non ambigus, préfixiels et finis. Les deux familles ont été entièrement caractérisées : Pour la première, la non-ambiguïté du produit RS a été obtenue grâce à l'opération de composition de codes préfixes; dans ce cas, S est toujours préfixiel. Pour la seconde famille, l'ensemble S n'est plus préfixiel mais les mots qui lui manquent pour être préfixiel sont tous facteur gauche d'une puissance d'un même mot t. Un résultat de Krasner [3] a permis de décrire la forme des facteurs R et S.

Ces familles ont une intersection commune. En effet, certains produits RS t-préfixiels peuvent se factoriser en un autre produit non ambigu PQ où Q est préfixiel. L'exemple 12 illustre cette situation : on peut écrire RS comme

$$\underline{R}\,\underline{S} = (1 + t^2 b^3)\,\underline{R_{t,S}}\,\underline{S}$$

où $R_{t,S}S$ est préfixiel. En d'autres termes, le code préfixe maximal fini RSA-RS associé à RS n'est pas indécomposable.

C'est le cas de tous les produits RS non ambigus t-préfixiels dont le facteur R s'écrit

$$R = R' R_{t,S} \qquad \text{avec} \quad R' \neq 1 \qquad (cf.\ \text{théorème 11}).$$

Comme précédemment, $\underline{R}\,\underline{S}$ se factorise en

$$\underline{R}\,\underline{S} = \underline{R'}\,(\underline{R_{t,S}}\,\underline{S})$$

où $R_{t,S}S$ est préfixiel.

Si $R = 1$, nous avons pu établir des hypothèses — liées à la forme de S uniquement —, sous lesquelles X est encore décomposable. Nous ne connaissons pas de caractérisation de l'intersection de deux familles.

Ces conditions indiquent que S se factorise de manière non ambiguë en $S'_1 S'_2$ de telle sorte que le produit RS soit formé de plusieurs copies juxtaposées d'un même ensemble Q préfixiel. Le code préfixe maximal correspondant à RS est alors décomposable sur le code préfixe $QA - Q$ associé à Q.

On reprend les notations utilisées dans la preuve du théorème 10 pour la factorisation de Krasner $\underline{R}_t\,\underline{T}$ du polynôme $\underline{RS \cap t^*} = p_{n,1}(t)$:

$$I_R \cup I_S = \{0, 1, \ldots, k-1\}$$

et

$$\underline{R}_t = \prod_{i \in I_R} p_{n_{i+1}, n_i}(t)$$

$$\underline{T} = \prod_{i \in I_S} p_{n_{i+1}, n_i}(t).$$

PROPOSITION 13 : *Soit $RS \subset A^*$ un produit non ambigu, t-préfixiel, fini, non vide tel que $R_t = R_{t,S}$.*

S'il existe S'_1, $S'_2 \subset A^*$ et $0 \leqq j \leqq k-2$ tels que

(1) $\underline{S} = \underline{S'_1}\,\underline{S'_2}$.

(2) *Soit $U = S'_1 - t^* A$. Il existe $U' \subset A^*$ tel que*

$$\underline{U} = \underline{U'}\,\underline{U''} \qquad \text{où} \quad \underline{U''} = \prod_{\substack{i \in I_R \\ i<j}} p_{n_{i+1}, n_i}(t).$$

(3) $\underline{S'_2 \cap t^* A^-} = \prod_{\substack{i \in I_S \\ i \leqq j}} p_{n_{i+1}, n_i}(t) . \underline{tA^-}$

(4) *Pour tout $s \in S'_2$ tel que $s = uav$ où $u \in t^* A^-$ est de longueur maximale, $a \in A$, $v \in A^*$, on a*

$$\forall v' \in A^*,\ v' < v : uav' \in S'_2$$

alors, le code $RSA - RS$ n'est pas indécomposable.

Preuve : On note :

$$\underline{R}_{<j} = \prod_{\substack{i \in I_R \\ i<j}} p_{n_{i+1}, n_i}(t) \qquad \text{et} \qquad \underline{R}_{\geqq j} = \prod_{\substack{i \in I_R \\ i \geqq j}} p_{n_{i+1}, n_i}(t)$$

$$\underline{S}_{\leqq j} = \prod_{\substack{i \in I_S \\ i \leqq j}} p_{n_{i+1}, n_i}(t) . \underline{tA^-} \qquad \text{et} \qquad \underline{S}_{>j} = \prod_{\substack{i \in I_S \\ i>j}} p_{n_{i+1}, n_i}(t)$$

Avec ces notations, $\underline{R}_t = \underline{R}_{\geqq j} . R_{<j}$ et $\underline{S}_t = \underline{S}_{>j} . \underline{S}_{\leqq j}$.

Posons $Q = R_{<j} . S'_2$ et montrons que le code $X = RSA - RS$ est décomposable sur $QA - Q$, en prouvant que $\underline{R}\,\underline{S}$ s'écrit $\underline{P}\,\underline{Q}$ où Q est préfixiel

(proposition 7). Nous montrerons aussi que P est différent de 1, donc que la décomposition du code X n'est pas triviale.

Comme $\underline{S}_t = \underline{S}_{>j}\underline{S}_{\leqq j}$, et que $\underline{S} = \underline{S}'_1\underline{S}'_2$ avec $\underline{S'_2 \cap t^* A^-} = \underline{S}_{\leqq j}$, on a

$$\underline{S'_1 \cap t^* A^-} = \underline{S}_{>j}.$$

Par conséquent

$$\begin{aligned}\underline{R}\,\underline{S} &= \underline{R}\,\underline{S}'_1\,\underline{S}'_2 = \underline{R}\,(\underline{S}_{>j} + \underline{U})\,\underline{S}'_2 \\ &= \underline{R}_{\geqq j}\,\underline{R}_{<j}\,\underline{S}_{>j}\,\underline{S}'_2 + \underline{R}\,\underline{U}'\,\underline{R}_{<j}\,\underline{S}'_2 \\ &= (\underline{R}_{\geqq j}\,\underline{S}_{>j} + \underline{R}\,\underline{U}')\,\underline{R}_{<j}\,\underline{S}'_2 \\ &= \underline{P}\,\underline{Q}\end{aligned}$$

où $P = R_{\geqq j}S_{>j} + R\,U'$. P est différent de 1 car $j \leqq k-2$ et donc $p_{n_k, n_{k-1}}(t)$ divise $\underline{S}_{>j}$.

Il reste à prouver que Q est préfixiel. Soit $q \in Q : q = rs$ où $r \in \mathrm{R}_{<j}$ et $s \in S'_2$.

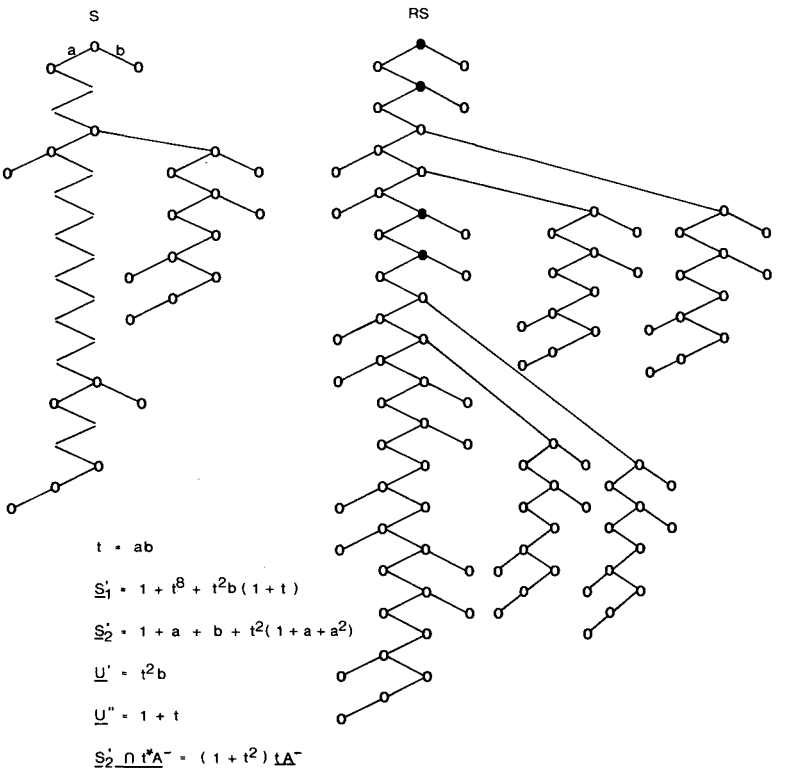
On factorise s en uav où $u \in t^* A^-$ est de longueur maximale, $a \in A$ et $v \in A^*$. Soit q' un facteur gauche de q. Si $ru < q'$, alors $q' \in Q$ par l'hypothèse (4). Supposons q' facteur gauche de ru. Le mot u appartient à $S_{\leqq j}(A^*)^{-1}$. En effet, soit $s_{\max}$ le plus long mot de $S_{>j}$, alors

$$s_{\max}\,uav \in S_{>j}S'_2 \subset S'_1\,S'_2 = S.$$

Par le théorème 11, $s_{\max}u \in S_t(A^*)^{-1}$ et donc $u \in S_{\leqq j}(A^*)^{-1}$. Comme $R_{<j}S_{\leqq j}$ est préfixiel, ru et par conséquent q' appartiennent à $R_{<j}S_{\leqq j} \subset R_{<j}S'_2$. ■

L'exemple suivant illustre la proposition 13. On voit nettement apparaître le découpage de RS en huit copies Q identiques, ce qui indique la façon de décomposer le code $RSA - RS$.

Exemple 14 :



Notons que les deux familles décrites ne regroupent pas tous les cas d'ensembles préfixiels finis qui se factorisent de façon non ambiguë. L'exemple $RS \subset \{a, b\}^*$ tel que $R = \{1, ab\}$ et $S = \{1, a, abab, ababa, b, abbb, ababb\}$ ne peut être classé ni dans la première famille (le code préfixe associé à RS est indécomposable), ni dans la seconde (les mots $\{ab, aba, abb\}$ qui manquent à S pour être préfixiel ne sont pas tous facteur gauche d'une puissance de $t = ab$).

Nous ne connaissons pas la forme générale des produits non ambigus RS préfixiels et finis.

REMERCIEMENTS

Je remercie vivement Dominique Perrin pour ses nombreuses suggestions et discussions encourageantes.

BIBLIOGRAPHIE

1. J. BERSTEL et D. PERRIN, *Theory of Codes*, Academic Press, 1985.
2. N. J. FINE et H. S. WILF, *Uniqueness Theorems for Periodic Functions*, Proc. Amer. Math. Soc., vol. 16, 1965, p. 109-114.
3. M. KRASNER et B. RANULAC, *Sur une propriété des polynômes de la division du cercle*, C. R. Acad. Sci. Paris, 240, 1937, p. 397-399.
4. M. LOTHAIRE, *Combinatorics on Words*, Reading, Massachusetts, Addison-Wesley, 1983.
5. D. PERRIN, *Codes asynchrones*, Bull. Soc. Math. Fr., tome 105, 1977, p. 385-404.