Theoretical Computer Science 71 (1990) 381-400
North-Holland

SUR LES FACTEURS DES SUITES DE STURM

S. DULUCQ et D. GOUYOU-BEAUCHAMPS

Université de Bordeaux I, UER Mathématiques et Informatique, 351 Cours de la Libération, 33405 Talence Cedex, France

Communicated by M. Nivat
Received February 1988

Résumé. Cet article a pour objet l'étude d'une construction associant à toute droite de pente p/q (p et q premiers entre eux et $q \leqslant n$) un mot de longueur n sur l'alphabet $\{0, 1\}$. Nous montrons que nous obtenons par cette construction le langage constitué de tous les facteurs des suites de Sturm. Nous formulons, après avoir obtenu une équation fonctionnelle dont la solution est la série génératrice de ce langage, une conjecture reliant cette série génératrice à la fonction d'Euler.

Abstract. In this paper, we study a construction which connects to each line with slope p/q (such that $gcd(p, q) = 1$ and $q \leqslant n$) a word of length n over the alphabet $\{0, 1\}$. We show that this construction yields the language of all the factors of the sturmian sequences. We first obtain a functional equation whose solution is the generating function of this language, and then we give a conjecture relating this generating function to the Euler function.

Etant donné un entier $n \geqslant 2$, nous considérons l'ensemble des entiers p, q et r vérifiant les trois conditions

(a) $0 < p < q \leqslant n$,
(b) $0 \leqslant r < q \leqslant n$,
(c) $(p, q) = 1$ c'est à dire les entiers p et q sont premiers entre eux.

Nous étudions alors le mot $f = f_1 f_2 \ldots f_n$ écrit sur l'alphabet $\{0, 1\}$ et défini par

$$f_i = \left[\frac{pi + r}{q}\right] - \left[\frac{p(i-1) + r}{q}\right]$$

pour tout $1 \leqslant i \leqslant n$, $[x]$ désignant la partie entière du rationnel x.

Le mot f ainsi obtenu est le mot associé à la droite d'équation

$$y = \frac{p}{q}x + \frac{r}{q}.$$

Notre but est d'étudier le langage formé par l'ensemble des mots de longueur n associés à ces droites dans lesquelles p, q et r vérifient les conditions a, b et c. Les motivations de cette étude sont de nature arithmétique et Cohen [5] nous a signalé qu'il serait intéressant de dénombrer ces mots.

Cette construction associant à une droite un mot est classique et apparait régulièrement dans la littérature sous diverses formes.

Elle a été considérée pour la première fois, semble-t-il, par l'astronome Johan Bernoulli (1772) qui voulait calculer la suite de nombres $[m\xi+\frac{1}{2}]$ pour toutes les valeurs entières de m connaissant le développement en fraction continue du nombre irrationnel ξ, et en utilisant seulement l'opération d'addition (voir [20]).

Le plus souvent, les auteurs de ces travaux se placent dans le cas de droites à pente irrationnelle qu'ils codent par des mots infinis. On pourra se référer pour cela aux travaux dans [14, 4, 18] pour le 19$^{\text{ième}}$ siècle et aux articles dans [6, 9, 10, 11, 15, 16, 19, 12, 13] pour le 20$^{\text{ième}}$.

Les mots infinis ainsi obtenus portent des noms très variés tels que caractéristique du nombre irrationnel ξ, suite de Beatty, suite de Sturm, suite de Bernoulli, suite "two-distance", suite linéaire et, dans un cas particulier, suite musicale [8, 17].

Ils interviennent aussi dans de nombreux domaines comme la géométrie hyperbolique, l'approximation Diophantienne et les équations Diophantiennes (on pourra consulter l'article de Series pour plus de développements) et même le pavage du plan ou le calcul des cycles lunaires (voir [8, 17]).

Coven et Hedlund [6] ont montré que les suites de Sturm codant les droites à pente irrationnelle passant par l'origine (il s'agit dans ce cas d'un codage) sont les mots infinis non périodiques dont tout facteur de longueur finie f vérifie la condition (C) suivante.

(C) $\left\{\begin{array}{l}\text{pour toute factorisation } f=f_1uf_2vf_3 \text{ où } u \text{ et } v \text{ sont de même}\\ \text{longueur, les nombres de lettres 1 dans } u \text{ et dans } v \text{ diffèrent au}\\ \text{plus d'une unité.}\end{array}\right.$

Nous montrons que l'ensemble des mots de longueur n construits (il ne s'agit plus ici d'un codage) à partir des droites "rationnelles" $y=(p/q)x+(r/q)$ où p, q et r vérifient les conditions a, b et c est exactement l'ensemble des facteurs de longueur n des suites de Sturm.

De plus, la démonstration que nous donnons est constructive puisqu'elle permet de retrouver, à partir d'un facteur de longueur n d'une suite de Sturm, une droite "rationnelle" donnant ce mot.

Remarquons que nous n'obtiendrions pas tous ces facteurs si nous avions considéré une autre famille de droites, comme par exemple les droites à pente rationnelle passant par l'origine ($r=0$). Par contre, on pourrait montrer que dans ce cas cette construction est un codage.

Ensuite, nous évoquons les liens entre les mots obtenus à partir de ces droites, les développements en fraction continue et les morphismes itérés.

Nous nous intéressons également aux propriétés algébriques de ce langage D constitué par l'ensemble des facteurs finis des suites de Sturm. Nous établissons la non algébricité de ce langage et nous montrons que son complémentaire est algébrique, résultat analogue à un théorème de Berstel sur les morphismes itérés [3]. Par contre, la question de l'ambiguïté du langage complémentaire reste ouverte. Nous montrons que ce problème est très proche d'une conjecture dans [1] et de problèmes d'ambiguïté de langages algébriques résolus dans [7].

Enfin, nous obtenons une équation fonctionnelle dont est solution la série génératrice des mots du langage D. Cette équation fonctionnelle ne permet hélas pas d'énumérer les mots de longueur n de D. Nous conjecturons que ce nombre de mots est donné par

$$-1+\sum_{i=1}^{n}(n-i+1)\varphi(i),$$

dans laquelle φ est la fonction d'Euler $\varphi(p)=\{m \mid m<p \text{ et } (m,p)=1\}$.

L'exactitude de cette conjecture permettrait alors de confirmer l'ambiguïté du langage complémentaire du langage D.

1. Un langage défini par "codage" de droites

Considérons, pour tout entier $n\geqslant 2$, *l'ensemble* $\Delta(n)$ constitué par les droites d'équation $y=\delta(x)$ où

$$\left.\begin{array}{l} \delta(x)=\dfrac{p}{q}x+\dfrac{r}{q}, \quad \text{les entiers } p, q \text{ et } r \text{ vérifiant} \\ (\mathrm{D}_1)\quad 0<p<q\leqslant n, \\ (\mathrm{D}_2)\quad 0\leqslant r<q\leqslant n, \\ (\mathrm{D}_3)\quad (p,q)=1 \quad (p \text{ et } q \text{ sont premiers entre eux}). \end{array}\right\} \qquad (1)$$

Par une construction analogue à celle considérée dans [16] pour les droites de pente irrationnelle, nous allons associer à chaque droite $y=\delta(x)$ un mot sur l'alphabet $X=\{0,1\}$.

Soient donc un entier $n\geqslant 2$ et $y=\delta(x)$ une droite appartenant à $\Delta(n)$. Considérons *l'application* Ψ associant à cette droite le mot $\Psi(\delta)=f=f_1f_2\ldots f_n$ de longueur n ($f_i\in X$) défini visuellement de la manière suivante:

$$\left.\begin{array}{l} \text{pour tout } k,\ 1\leqslant k\leqslant n,\ I_k \text{ désignant l'intervalle }]k-1,k], \text{ si } \delta \\ \text{coupe une droite d'équation } y=c\ (c\in\mathbb{N}) \text{ sur la bande verticale} \\ \text{ayant pour projection } I_k, \\ \quad \text{alors } f_k=1, \qquad \text{sinon } f_k=0. \end{array}\right\} \qquad (2)$$

Plus formellement, les lettres f_k du mot $\Psi(\delta)=f_1f_2\ldots f_n$ vérifient

$$f_k=[\delta(k)]-[\delta(k-1)]$$

où $[x]$ désigne la partie entière du nombre x.

Exemple 1.1. La Figure 1 donne un exemple de "codage" pour $n=10$ et la droite appartenant à $\Delta(10)$ d'équation $y=\delta(x)$ où

$$\delta(x)=\tfrac{2}{7}x+\tfrac{5}{7}.$$

On obtient pour cette droite $y=\delta(x)$ le mot $\Psi(\delta)=1000100100$.

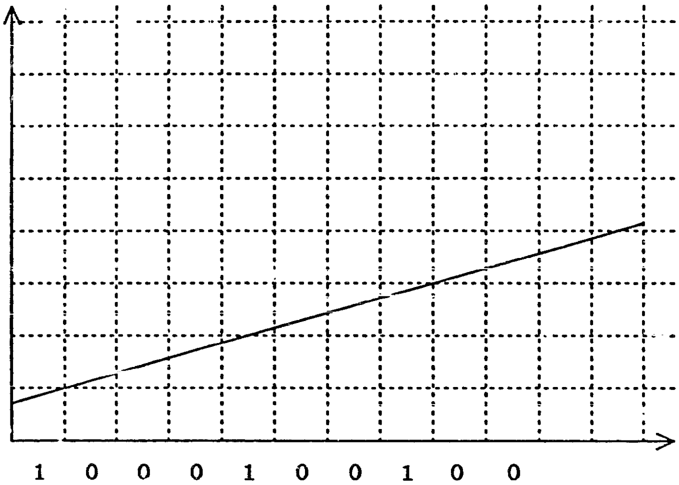

Fig. 1. Le mot associé à $y = \frac{2}{7}x + \frac{5}{7}$.

Remarque 1.2. L'application Ψ ne définit pas un codage de ces droites. En effet Ψ n'est pas injective. Par exemple, pour $n = 3$, considérons les droites d'équations $y = \delta(x)$ et $y = \lambda(x)$ où

$$\delta(x) = \tfrac{1}{2}x \quad \text{et} \quad \lambda(x) = \tfrac{1}{3}x + \tfrac{1}{3}.$$

Ces deux droites appartiennent à $\Delta(3)$ et vérifient $\Psi(\delta) = \Psi(\lambda) = 010$.

Dans la suite de cet exposé, nous allons nous intéresser au langage constitué de tous les mots $\Psi(\delta)$ obtenus à partir des droites $y = \delta(x)$ appartenant aux ensembles $\Delta(n)$, langage que nous caractériserons dans les paragraphes suivants.

Considérons donc *le langage D* défini par

$$D = \{f \in X^* \mid \text{il existe une droite } \delta \text{ appartenant à } \Delta(n) \text{ où } n = |f| \text{ telle que } \Psi(\delta) = f\}. \tag{3}$$

Dans cette définition, $|f|$ désigne la longueur du mot f (c'est à dire son nombre de lettres).

2. Une première propriété du langage D

Considérons le morphisme Φ de X^* sur lui même défini par

$$\Phi(0) = 1, \qquad \Phi(1) = 0.$$

Nous allons montrer que le langage D est fermé par ce morphisme, et ceci en donnant une interprétation de Φ sur les droites.

Proposition 2.1. *Si $f \in D$ alors $\Phi(f) \in D$.*

Preuve. Soient $f \in D$ et $n = |f|$ sa longueur. Alors, il existe une droite appartenant à $\Delta(n)$ d'équation $y = \delta(x)$ où $\delta(x) = (p/q)x + (r/q)$, et telle que $\Psi(\delta) = f$. Considérons alors la droite d'équation $y = \lambda(x)$ où

$$\lambda(x) = \frac{q-p}{q}x + \frac{q-r-1}{q}.$$

Clairement, cette droite appartient à $\Delta(n)$ car, comme p et q sont premiers entre eux, il en est de même pour $q-p$ et q. Soit $g = \Psi(\lambda)$. Nous allons montrer que $g = \Phi(f)$. Or

$$\lambda(x) = \frac{q-p}{q}x + \frac{q-r-1}{q} = x - \left(\frac{p}{q}x + \frac{r}{q}\right) + \frac{q-1}{q}. \tag{4}$$

Soit maintenant $k \in \mathbb{N}$, $0 \leq k < n$ et posons $\delta(k) = s + (t/q)$ où $s, t \in \mathbb{N}$ et $0 \leq t < q$. Ainsi $s = [\delta(k)]$. Nous avons alors $\delta(k+1) = s + ((t+p)/q)$ et, d'après (4),

$$\lambda(k) = k - s + \frac{q-t-1}{q} \quad \text{et} \quad \lambda(k+1) = k - s + \frac{2q-t-p-1}{q}.$$

Examinons alors les deux cas possibles pouvant se présenter pour la $(k+1)$ième lettre f_{k+1} du mot $f = \Psi(\delta)$.

Cas 1: $f_{k+1} = 0$ c'est à dire, sur la bande verticale ayant pour projection $]k, k+1]$, la droite $y = \delta(x)$ ne coupe aucune droite d'équation $y = c$ ($c \in \mathbb{N}$). Ainsi

$$f_{k+1} = 0 \Leftrightarrow 0 \leq t < t+p < q.$$

Alors

$$t+p < q \Rightarrow 0 \leq q-t-p-1 \Rightarrow q \leq 2q-t-p-1,$$

et

$$t < q \Rightarrow 0 \leq q-t-1 < q.$$

Nous déduisons alors de ces deux inégalités que la droite $y = \lambda(x)$ coupe la droite d'équation $y = k-s+1$ pour x appartenant à l'intervalle $]k, k+1]$, et donc $g_{k+1} = 1 = \phi(f_{k+1})$.

Cas 2: $f_{k+1} = 1$ c'est à dire, pour x appartenant à l'intervalle $]k, k+1]$, la droite d'équation $y = \delta(x)$ coupe la droite $y = s+1$. Ainsi

$$f_{k+1} = 1 \Leftrightarrow 0 \leq t < q \leq t+p.$$

Alors

$$q \leq t+p \Rightarrow q-t-p \leq 0 \Rightarrow 2q-t-p-1 < q,$$

et

$$t < q \Rightarrow 0 \leq q-t-1 < q.$$

De ces deux dernières inégalités, nous déduisons que la droite d'équation $y=\lambda(x)$ ne coupe aucune droite $y=c$ ($c\in\mathbb{N}$) pour tout x appartenant à l'intervalle $]k, k+1]$, et donc $g_{k+1}=0=\Phi(f_{k+1})$.

De ces deux cas, nous déduisons la propriété annoncée, à savoir la droite d'équation

$$y=\frac{q-p}{q}x+\frac{q-r-1}{q}$$

donne par Ψ le "mot opposé" à celui obtenu pour la droite d'équation $y=(p/q)x+(r/q)$. □

Exemple 2.2. Soient $n=10$ et la droite appartenant à $\Delta(10)$ d'équation $y=\delta(x)$ où $\delta(x)=\frac{5}{7}x+\frac{1}{7}$. Alors $\Psi(\delta)=0111011011$. La droite d'équation $y=\frac{2}{7}x+\frac{5}{7}$ admet alors pour codage le mot $f=1000100100=\Phi(\Psi(\delta))$.

3. Caractérisation du langage D

Dans ce paragraphe, nous donnons une caractérisation des mots du langage D en montrant que ce sont les mots de X^* dont tout couple de facteurs de même longueur vérifient une condition simple sur le nombre de lettres 1 qu'ils contiennent.

Soit f un mot de X^*. Comme on désigne far $|f|$ la longueur du mot f, on note $|f|_x$ le nombre d'occurrences de la lettre x dans le mot f.

Exprimons maintenant cette condition. Soit $f\in X^*$. Nous dirons que *le mot f vérifie la condition* (C) si et seulement si

$$\text{(C)}\qquad \begin{cases}\text{pour toute factorisation } f=f_1uf_2vf_3 \text{ où } |u|=|v|,\\ \text{on a } -1\leqslant |u|_1-|v|_1\leqslant 1.\end{cases}$$

Considérons maintenant le langage L constitué de tous les mots autres que 0^m et 1^m ($m\geqslant 0$) et satisfaisant cette condition (C), c'est à dire

$$L=\{f\in X^*\mid f \text{ vérifie (C)}\}\setminus\{0^m, 1^m\mid m\geqslant 0\}. \tag{5}$$

Théorème 3.1. *Les langages D et L sont égaux.*

Nous démontrerons ce théorème en montrant la double inclusion de ces langages. Pour montrer que $D\subseteq L$, nous utiliserons le lemme suivant.

Lemme 3.2. *Soit f un mot de D tel que $f=ugv$ avec $|g|=n$ et $|g|_1=m$. Alors, toute droite $y=\delta(x)$ appartenant à $\Delta(|f|)$, telle que $\Psi(\delta)=f$, a une pente α vérifiant*

$$\frac{m-1}{n}<\alpha<\frac{m+1}{n}. \tag{6}$$

Preuve. Soit $f \in D$ où $f = ugv$ avec $|g| = n$ et $|g|_1 = m \leq n$. Alors, d'après la définition de Ψ, toute droite $y = \delta(x)$ appartenant à $\Delta(|f|)$, telle que $\Psi(\delta) = f$, vérifie

$$(\delta|u| + n) - \delta(|u|) < m + 1.$$

En effet, si cela n'était pas, nous aurions $|g|_1 \geq m + 1$. Ainsi, la pente α de la droite $y = \delta(x)$ vérifie $\alpha < (m+1)/n$. Pour des raisons analogues, nous avons

$$\delta(|u| + n) - \delta(|u|) > m - 1.$$

Et la pente α de la droite $y = \delta(x)$ vérifie alors $\alpha > (m-1)/n$. □

Lemme 3.3. *Le langage D est inclus dans le langage L.*

Preuve. Soit f un mot de D tel que $|f| = n$ et soit une factorisation de f en $f = f_1 u f_2 v f_3$ où

$$q = |u| = |v|, \qquad |u|_1 = p_1, \qquad |v|_1 = p_2.$$

Soit $y = \delta(x)$ la droite appartenant à $\Delta(n)$, de pente α, telle que $\Psi(\delta) = f$. D'après le Lemme 3.2, nous avons

$$\frac{p_1 - 1}{q} < \alpha < \frac{p_1 + 1}{q} \quad \text{et} \quad \frac{p_2 - 1}{q} < \alpha < \frac{p_2 + 1}{q}.$$

Nous en déduisons $-1 \leq p_1 - p_2 \leq +1$.

Ceci étant vrai pour toute factorisation de f où les facteurs u et v ont même longueur, le mot f vérifie alors la condition (C) et donc $f \in L$. Ainsi $D \subseteq L$. □

Remarque 3.4. Il est immédiat de constater, d'après la définition du langage L, qu'un mot f appartient à L si et seulement si le mot $\Phi(f)$ appartient à L où Φ est le morphisme échangeant les lettres 0 et 1. De plus, f étant un mot de L, ce mot f ne peut avoir à la fois deux lettres 0 et deux lettres 1 consécutives.

Compte-tenu de cette remarque, considérons un mot f appartenant à L et débutant par la lettre 0. Alors, il existe un entier $m \geq 0$ tel que le mot f s'écrive

(E) $f = 0u_1 0u_2 0 \ldots 0u_n 0u_{n+1}$ avec

(E$_1$) pour tout i, $1 \leq i \leq n$, $u_i = 1^m$ ou $u_i = 1^{m+1}$,

(E$_2$) $u_{n+1} = 1^s$ où $0 \leq s \leq m+1$.

Considérons alors *le mot réduit de f, noté $r_m(f)$*, obtenu en effectuant les substitutions suivantes:

(R$_1$) à chaque facteur $0u_i$ $(1 \leq i \leq n)$ est substituée la lettre 0 si $|u_i| = m$, 1 si $|u_i| = m+1$,

(R$_2$) au facteur $0u_{n+1}$ est substitué soit la lettre 1 si $|u_{n+1}| = m+1$, soit le mot vide ε si $|u_{n+1}| \leq m$.

Nous obtenons alors la propriété suivante qu'il est immédiat de vérifier compte-tenu de la condition (C) que vérifient les mots de L.

Proposition 3.5. *Si f est un mot débutant par la lettre* 0 *et autre que* $(01^m)^n01^s$ $(s \leqslant m)$, *alors f appartient à L si et seulement si le réduit $r_p(f)$ de f appartient à L (la réduction r_p dépendant de l'écriture* (E), (E_1) *et* (E_2) *de f).*

Remarque 3.6. La Proposition 3.5 serait fausse si, dans la définition de la réduction r_m, au dernier facteur $0u_{n+1}$ nous substituions la lettre 0 au lieu du mot vide dans le cas où $0u_{n+1} = 01^m$. En effect, considérons par exemple le mot

$$f = 01110111011011.$$

On vérifierait aisément que ce mot est bien un mot du langage L. Le mot $r_2(f)$, d'après notre définition, est le mot 110 qui appartient bien à L. Par contre, si l'on substitue au dernier facteur 011 la lettre 0, nous obtenons le mot 1100 qui, lui, n'appartient pas à L.

Lemme 3.7. *Le langage L est inclus dans le langage D.*

Nous allons donner une interprétation de cette inclusion, c'est à dire une démonstration qui indiquera comment, étant donné un mot de L, retrouver l'équation d'une droite dont l'image par Ψ est ce mot.

Preuve. Par récurrence sur la longueur des mots de L. Soit f un mot de L. Si $|f| = 2$ alors les seuls mots possibles sont $f = 01$ et $f = 10$. Alors la droite d'équation $y = \frac{1}{2}x$ donne par Ψ le mot 01, et la droite d'équation $y = \frac{1}{2}x + \frac{1}{2}$ donne par Ψ le mot 10. Ces deux mots 01 et 10 appartiennent donc bien à D.

Soit maintenant f un mot de L de longueur $n > 2$. Supposons que le mot f débute par la lettre 0. Si tel n'était pas le cas, on raisonnerait sur le mot $\Phi(f)$ en utilisant la Proposition 2.1. Si f est un mot de la forme $(01^m)^k01^s$ où $s \leqslant m$, il est immédiat de constater que la droite d'équation $y = \delta(x)$, où $\delta(x) = (m/m+1)x$, appartient à $\Delta(|f|)$ et vérifie $\Psi(\delta) = f$. Ainsi $f \in D$. Sinon, soit $g = r_m(f)$ le mot f réduit $(m \geqslant 0)$ où la réduction r_m est définie par les substitutions:

$$01^m \text{ se réécrit en } 0, \qquad 01^{m+1} \text{ se réécrit en } 1.$$

D'après la remarque et la proposition précédentes, $g \in L$ et $|g| < |f|$. Ainsi l'hypothèse de récurrence nous assure qu'il existe une droite appartenant à $\Delta(|g|)$ d'équation $y = \delta(x)$ où $\delta(x) = (p/q)x + (r/q)$, et vérifiant $\Psi(\delta) = g$. Considérons alors la droite d'équation $y = \lambda(x)$ où

$$\lambda(x) = \frac{p + qm}{p + q(m+1)}x + \frac{r}{p + q(m+1)}. \tag{7}$$

Nous allons montrer que cette droite $y = \lambda(x)$ vérifie $\Psi(\lambda) = f$ et donc que le mot f appartient au langage D.

Montrons tout d'abord que $y = \lambda(x)$ appartient à $\Delta(|f|)$. Nous obtenons:

(a) $(p + qm, p + q(m+1)) = 1$ car $(p, q) = 1$.

(b) Sachant que $g = r_m(f) = \Psi(\delta)$, nous en déduisons $q < |g|$, et donc $|g|_1 \geqslant p$ car $\delta(q) = p + (r/q)$.

De plus, comme $g = r_m(f)$ et d'après la forme (R_1) de cette réduction, le mot f de longueur minimale est obtenu lorsque $|g|_1 = p$. Ainsi

$$\begin{aligned} n = |f| &\geqslant p(m+2) + (|g|-1)(m+1) \\ &\geqslant |g|(m+1) + p \\ &\geqslant q(m+1) + p \quad \text{car } |g| \geqslant q. \end{aligned}$$

Ces deux points (a) et (b) nous assurent que la droite $y = \lambda(x)$ appartient à $\Delta(n)$.

Montrons maintenant que $\Psi(\lambda) = f$. Soient $i \in \mathbb{N}$, $0 \leqslant i \leqslant |g|$ et k la partie entière de $\delta(i)$, c'est à dire

$$\delta(i) = k + \frac{\alpha}{q} \quad \text{où } k \in \mathbb{N},\ 0 \leqslant \alpha < q.$$

Ainsi, les i premières lettres de g sont k lettres 1, et $i-k$ lettres 0. Montrons alors que les premières lettres de $\Psi(\lambda)$ sont

$$\begin{aligned} &k(m+1) + (i-k)m = im + k \text{ lettres } 1, \\ &k + (i-k) = i \text{ lettres } 0, \end{aligned} \tag{8}$$

ce qui assurera la cohérence avec les i premières lettres de g compte-tenu de la réduction r_m effectuée sur f (voir (R_1)). Pour montrer cela, calculons la valeur de λ au point $(m+1)i+k$. Or

$$\lambda(x) = \frac{p+qm}{p+q(m+1)}x + \frac{r}{p+q(m+1)} = x + \frac{r-qx}{p+q(m+1)}.$$

Nous obtenons alors

$$\begin{aligned} \lambda(i(m+1)+k) &= (m+1)i + k + \frac{r - qk - q(m+1)i}{p+q(m+1)} \\ &= mi + k + \frac{r + pi - qk}{p+q(m+1)}. \end{aligned}$$

Or

$$\delta(i) = \frac{p}{q}i + \frac{r}{q} = k + \frac{\alpha}{q},$$

ce qui implique $pi + r = qk + \alpha$. Et ainsi

$$\lambda(i(m+1)+k) = mi + k + \frac{\alpha}{p+q(m+1)}. \tag{9}$$

Comme $0 \leqslant \alpha < q$, nous avons évidemment $0 \leqslant \alpha < p + q(m+1)$.

Ainsi, parmi les $(m+1)i+k$ premières lettres de $\Psi(\lambda)$, exactement $mi+k$ sont des lettres 1 et nous obtenons donc bien (8). Ce raisonnement étant vrai pour tout entier i vérifiant $1 \leqslant i \leqslant |g|$, nous déduisons de (8) la propriété suivante.

Si la $(i+1)$ième lettre de $g = \Psi(\delta)$ est

(i) la lettre 1, alors il lui correspond dans $\Psi(\lambda)$ un facteur constitué de

$m(i+1)+k+1-(mi+k) = m+1$ lettres 1,

$i+1-i=1$ lettre 0,

(ii) la lettre 0, alors il lui correspond dans $\Psi(\lambda)$ un facteur constitué de

$m(i+1)+k-(mi+k) = m$ lettres 1,

$i+1-i=1$ lettre 0.

Il nous reste maintenant à montrer que tous ces facteurs de $\Psi(\lambda)$, chacun d'eux correspondant à une lettre de g, commencent tous par la lettre 0. De manière équivalente, montrons que $\Psi(\lambda) = \Theta_m(g)$ où Θ_m est le morphisme défini par $\Theta_m(0) = 01^m$, $\Theta_m(1) = 01^{m+1}$. Pour cela, montrons que, pour tout entier i vérifiant $0 \leqslant i \leqslant |g|$, $\lambda(i(m+1)+k)$ et $\lambda(i(m+1)+k+1)$ ont même partie entière. Or, d'après (9), nous avons

$$\lambda(i(m+1)+k) = mi+k+\frac{\alpha}{p+q(m+1)}, \tag{10}$$

et

$$\lambda(i(m+1)+k) = mi+k+\frac{\alpha}{p+q(m+1)}+\frac{p+qm}{p+q(m+1)}. \tag{11}$$

Or $\alpha+p+qm = p+q(m+1)+\alpha-q$ et de plus $0 \leqslant \alpha < q$. Ainsi, (10) et (11) ont même partie entière.

Comme nous venons de montrer que $\Psi(\lambda) = \Theta_m(g)$ où Θ_m est le morphisme inverse de la réduction r_m vérifiant $g = r_m(f)$, nous obtenons, d'après la définition (R_2) de r_m, $\Psi(\lambda) = u$ où u est le facteur gauche de f tel que $f = uv$ et

$$v = \begin{cases} \text{le mot vide} & \text{si } f \text{ se termine par } 01^{m+1}, \\ 01^s & \text{sinon } (s \text{ vérifie alors } 0 \leqslant s \leqslant m). \end{cases} \tag{12}$$

Il est alors immédiat de constater que la droite $y = \delta(x)$, donnant par Ψ le mot u sur l'intervalle $]0, |u|]$, donne par prolongement le mot $f = uv$ par Ψ sur l'intervalle $]0, |f|]$.

Ceci termine la preuve du Lemme 3.7 et, compte-tenu du Lemme 3.3, nous obtenons le Théorème 3.1. □

De cette preuve du Lemme 3.7, nous déduisons immédiatement le résultat suivant.

Corollaire 3.8. *Soit g un mot du langage D. Soit Θ_m la substitution (le morphisme) définie par $\Theta_m(0)=01^m$ et $\Theta_m(1)=01^{m+1}$.*

Si la droite d'équation $y=\delta(x)$ où $\delta(x)=(p/q)x+(r/q)$ vérifie $\Psi(\delta)=g$, alors la droite d'équation $y=\lambda(x)$ où

$$\lambda(x)=\frac{p+qm}{p+q(m+1)}x+\frac{r}{p+q(m+1)}$$

appartient à $\Delta(|f|)$ où f est le mot $f=\Psi(\lambda)=\Theta_m(g)$.

Exemple 3.9. Soit $g=01$ obtenu à partir de la droite $\delta_0(x)=(\frac{1}{2})x$. Considérons le morphisme $\Theta=\Theta_1$ ($\Theta(0)=01$ et $\Theta(1)=011$). Alors $\Theta(g)=01011$ est le mot obtenu à partir de la droite d'équation $y=\delta_1(x)$ où $\delta_1(x)=(\frac{3}{5})x$.

De même $\Theta^2(g)=0101101011011$ est obtenu à partir de la droite d'équation $y=\delta_2(x)$ où $\delta_2(x)=(\frac{8}{13})x$. Nous en déduisons que, pour tout $n\geq 0$, $\Theta^n(g)$ est le mot

$$0101101011011010110101101101011011\ldots$$

de longueur F_{2n+2} obtenu à partir de la droite d'équation $y=\delta_n(x)$ où

$$\delta_n(x)=\frac{F_{2n+1}}{F_{2n+2}}x,$$

la suite F_n étant la suite de Fibonacci $F_0=1$, $F_1=1$, $F_{n+1}=F_n+F_{n-1}$.

Le mot $\Theta^n(g)$ converge alors vers le mot infini de Fibonacci obtenu, en utilisant la même construction Ψ, à partir de la droite d'équation $y=\alpha x$ dans laquelle $\alpha=\frac{1}{2}(-1+\sqrt{5})$. Remarquons que α admet pour développement en fraction continue $\alpha=(0;1,1,1,1,\ldots)$.

4. Une propriété arithmétique des mots du langage *D*

Nous déduisons des résultats du paragraphe précédent le théorème suivant.

Théorème 4.1. *Soient f_0 un mot de D et Θ_m le morphisme défini par $\Theta_m(0)=01^m$ et $\Theta_m(1)=01^{m+1}$ où $m\geq 0$. Alors, la suite $(f_n)_{n\geq 0}$ de mots de D définie par $f_n=\Theta_m^n(f_0)$ converge vers le mot infini f obtenu, par la même construction Ψ, à partir de la droite d'équation $y=\alpha x$ où $\alpha=\frac{1}{2}(-m+(m(m+4))^{1/2})$. De plus, le développement en fraction continue de α est $\alpha=(0;1,m,1,m,1,m,\ldots)$.*

Preuve. Soient f_0 un mot de D et Θ_m le morphisme du théorème. Soit la droite $y=\delta_0(x)$, où $\delta_0(x)=(p_0/q_0)x+(r/q_0)$, vérifiant $\Psi(\delta_0)=f_0$. Posons $t_0=p_0/q_0$ représentant la pente de cette droite. Nous déduisons du Corollaire 3.8 que la droite d'équation $y=\delta_1(x)$ où

$$\delta_1(x)=\frac{p_0+q_0m}{p_0+q_0(m+1)}x+\frac{r}{p_0+q_0(m+1)}$$

vérifie $f_1=\Psi(\delta_1)=\Theta_m(f_0)$. Soit

$$t_1=\frac{p_1}{q_1}=\frac{p_0+q_0m}{p_0+q_0(m+1)}$$

la pente de la droite δ_1. Alors

$$t_1=\frac{1}{1+q_0/(p_0+q_0m)}=\frac{1}{1+1/(m+t_0)}.$$

Ce raisonnement nous assure que, pour tout $n\geqslant 1$, le mot $f_n=\Theta_m^n(f_0)$ vérifie $f_n=\Psi(\delta_n)$ où la droite d'équation $y=\delta_n(x)$ est donnée par $\delta_n(x)=(p_n/q_n)x+(r/q_n)$ avec

$$\frac{p_n}{q_n}=\frac{1}{1+1/[m+(p_{n-1}/q_{n-1})]}.$$

Il est alors clair que la suite $(f_n)_{n\geqslant 0}$ de mots de D converge vers le mot infini f obtenu à partir de la droite d'équation $y=\alpha x$ où α est le nombre ayant pour développement en fraction continue $\alpha=(0;1,m,1,m,\ldots)$. Ainsi α vérifie $\alpha=(m+\alpha)/(m+\alpha+1)$ d'où nous déduisons $\alpha=\frac{1}{2}(-m+\sqrt{m(m+4)})$. $\square$

5. Propriétés algébriques du langage D

Dans ce paragraphe, nous utiliserons la terminologie habituellement employée en théorie des langages (voir [2]). D'après la caractérisation donnée précédemment pour le langage D (ensemble des mots vérifiant la condition (C)), nous obtenons immédiatement la propriété suivante.

Proposition 5.1. *Tout facteur f d'un mot de D, autre que 0^k ou 1^k, appartient aussi au langage D.*

Remarque 5.2. Cette dernière propriété, évidente sur les mots de D, ne l'est pas si l'on raisonne sur les droites considérées ici et leur "codage" Ψ. En effet, si f est un mot du langage D, alors il existe une droite δ appartenant à $\Delta(|f|)$ telle que $f=\Psi(\delta)$. Cette droite a alors pour équation $y=\delta(x)$ où $\delta(x)=(p/q)x+(r/q)$ avec $q\leqslant|f|$.

Si g est facteur de f, c'est à dire $f=ugv$, il n'est nullement évident qu'il existe une droite d'équation $y=\lambda(x)$ où $\lambda(x)=(p'/q')x+(r'/q')$ avec $q'\leqslant|g|$, et telle que $\Psi(\lambda)=g$. Ce résultat (Proposition 5.1) n'est qu'une conséquence de la caractérisation du langage D que nous avons donnée.

Dans [6], Coven et Hedlund ont considéré et caractérisé *les mots infinis "Strumiens"* ou *suites de Sturm*, mots obtenus par "codage" Ψ des droites d'équation $y = \alpha x$ où α est un nombre irrationnel. Coven et Hedlund ont montré que *les suites de Sturm sont les mots infinis non périodiques* dont tout facteur de longueur finie vérifie la condition (C).

Compte-tenu de cette caractérisation des suites de Sturm et de la définition du langage D, nous obtenons alors le théorème suivant.

Théorème 5.3. *Le langage D est constitué de l'ensemble des facteurs (finis) de toutes les suites de Sturm, excepté les mots ne comportant que la lettre* 0 *(respectivement la lettre* 1*).*

Preuve. Si le mot f est un facteur de longueur finie d'une suite de Sturm, alors, d'après [6], f est un mot du langage D.

Réciproquement, soit f un mot du langage D. Montrons, par récurrence sur la longueur du mot f, qu'il existe une suite de Sturm ayant f pour facteur. Si le mot f est de longueur 2, f est soit le mot 01, soit le mot 10 qui sont tous deux facteurs de la suite de Fibonacci (exemple précédent), et en fait facteurs de toute suite de Sturm. Soit f un mot du langage D tel que $|f| > 2$. Considérons alors le mot g, réduit du mot f, $g = r_m(f)$. Le mot g est un mot de longueur $|g| < |f|$. Suivant les cas, g est soit un mot du langage D, soit le mot 0^k ou le mot 1^k.

Si g est un mot du langage D, g est facteur d'une suite de Sturm correspondant à une droite d'équation $y = \alpha x$ où α est irrationnel. Si g est le mot 1^k (respectivement 0^k), alors g est facteur de la suite de Sturm obtenue à partir de la droite d'équation $y = \alpha x$ où α est le nombre irrationnel dont le développement en fraction continue est $\alpha = (0; 1, k, 1, k, 1, k, \ldots)$ (resp. $\alpha = (0; k, 1, k, 1, k, 1, \ldots)$). Ainsi, dans tous les cas, le réduit $g = r_m(f)$ du mot f est facteur d'une suite de Sturm associée à la droite d'équation $y = \alpha x$. Considérons maintenant le nombre irrationnel β défini par

$$\beta = \frac{1}{1 + 1/(m+\alpha)},$$

c'est à dire le nombre irrationnel dont le développement en fraction continue est $\beta = (0; 1, m, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \ldots)$ où $\alpha = (0; \alpha_0, \alpha_1, \alpha_2, \ldots)$.

D'après le théorème 4.1 (cf. preuve), à la droite d'équation $y = \beta x$ correspond le mot infini obtenu en appliquant le morphisme Θ_m au mot infini obtenu à partir de la droite d'équation $y = \alpha x$. Ainsi, le mot f est facteur de la suite de Sturm correspondant à la droite d'équation $y = \beta x$. $\square$

Remarque 5.4. Si l'on restreint les droites considérées ici (ensemble $\Delta(n)$) aux droites dont la pente est rationnelle et qui passent par l'origine ($r = 0$ dans (1)) alors le langage obtenu par "codage" de ces droites n'est pas le langage L (relation (5)). Nous n'obtiendrions pas avec cet ensemble de droites tous les facteurs des suites de Sturm.

Intéressons nous maintenant aux propriétés algébriques du langage D.

Nous obtenons un résultat analogue à celui donné par Berstel dans [3]. Berstel montre en effet que l'ensemble des facteurs gauches d'un mot infini obtenu par morphisme itéré à partir d'une lettre est un langage dont le complémentaire est un langage algébrique. Or, compte-tenu de la caractérisation donnée du langage D, nous constatons immédiatement le fait suivant.

Le langage D restreint aux mots débutant par la lettre 0 est constitué de l'ensemble des facteurs gauches de l'ensemble des mots infinis obtenus par itérations successives des morphismes Θ_m ($\Theta_m(0)=01^m$, $\Theta_m(1)=01^{m+1}$ où $m\geqslant 0$).

Nous obtenons alors le résultat suivant, analogue à celui de [3].

Théorème 5.5. *Le langage D est un langage non algébrique dont le complémentaire est un langage algébrique.*

Preuve. (1) Clairement, D n'est pas un langage algébrique. En effet, considérons le langage $M=D\cap 01^+01^+01^+$ où 1^+ désigne le langage rationnel $\{1^k \mid k\geqslant 1\}$. Nous avons

$$M=\{01^i01^j01^k \mid \text{il existe } m\geqslant 0 \text{ tel que } i,j,k\in\{m,m+1\}\}.$$

Or, il est bien connu que ce langage ne peut être engendré par une grammaire algébrique. En conséquence, le langage D n'est pas un langage algébrique car autrement le langage M le serait également.

(2) Soit maintenant $C=X^*\backslash D$. C est alors le langage $C=0^*\cup 1^*\cup A$ où A est le langage ($\|x\|$ désignant la valeur absolue de x)

$$A=\{f\in X^* \mid \text{il existe une factorisation } f=f_1uf_2vf_3$$
$$\text{avec } |u|=|v| \text{ et } \| |u|_1-|v|_1 \|\geqslant 2\}.$$

Considérons alors la grammaire algébrique dont les règles de dérivation sont les suivantes ($\mathbb{1}$ désignant le mot vide):

$$\xi\to\tau\chi\tau$$
$$\tau\to 0\tau \mid 1\tau \mid \mathbb{1}$$
$$\chi\to 0\chi 0 \mid 1\chi 1 \mid 1\rho 0 \mid 0\sigma 1$$
$$\rho\to 0\rho 0 \mid 1\rho 1 \mid 1\tau 0 \mid 0\chi 1$$
$$\sigma\to 0\sigma 0 \mid 1\sigma 1 \mid 0\tau 1 \mid 1\chi 0.$$

Ainsi, tout mot appartenant à A peut être engendré par cette grammaire à partir de l'axiome ξ, et réciproquement, cette grammaire n'engendre que des mots de A. Les symboles non terminaux ρ et σ permettent d'engendrer deux facteurs u et v de même longueur dont les nombres de lettres 1 qu'ils comportent diffèrent d'au moins deux unités. Le langage A étant ainsi un langage algébrique, il en est de même pour le langage C, complémentaire du langage D, car il est l'union des deux langages rationnels 0^* et 1^* avec le langage A. Ceci termine la preuve du Théorème 5.5. □

Intéressons nous au problème de l'ambiguïté du langage C complementaire du langage D. Considérons le langage $E = C \cap 01^+01^+01^+0$. Nous constatons, d'après la condition (C) vérifiée par les mots du langage D, que

$$E = \{01^m01^n01^p0 \mid m, n, p > 0 \text{ et } (|m-n| \geq 2 \text{ ou } |n-p| \geq 2 \text{ ou } |p-m| \geq 2)\}.$$

Ce langage E est analogue au langage $\{a^m b^n c^p \mid |m-n| \geq 1 \text{ ou } |n-p| \geq 1\}$ dans [1], que les auteurs ont conjecturé être un langage ambigu. De même, le langage C est voisin de certains langages algébriques considérés par Flajolet [7] qui prouve leur ambiguïté en montrant que les séries génératrices de ces langages admettent une infinité de pôles. Nous formulons donc la conjecture suivante.

Conjecture 5.6. *Le langage algébrique complémentaire du langage D est un langage ambigu.*

6. Equation fonctionnelle donnant la série génératrice des mots du langage D

Dans ce paragraphe, nous donnons une équation fonctionnelle dont la série génératrice des mots du langage D est solution. Considérons la série génératrice $f(x, y)$ des mots du langage D se terminant par la lettre 1. Elle est définie par

$$f(x, y) = \sum_{n \geq 1, p \geq 1} a_{n,p} x^n y^p, \tag{13}$$

où $a_{n,p}$ est le nombre de mots de D ayant exactement n lettres 0, p lettres 1, et se terminant par la lettre 1.

La série génératrice $d(t)$ des mots de D définie par

$$d(t) = \sum_{n \geq 2} d_n t^n \tag{14}$$

où d_n est le nombre de mots de D de longueur n, se déduit alors simplement de la série génératrice $f(x, y)$.

Proposition 6.1. *La série génératrice $d(t)$ est donnée par $d(t) = 2f(t, t)$.*

Preuve. La série d étant la série génératrice des mots de D, nous avons $d_m = \sum_{n+p=m} b_{n,p}$ où $b_{n,p}$ est le nombre de mots de D ayant exactement n lettres 0 et p lettres 1. Ainsi, d'après la Proposition 2.1, nous avons $b_{n,p} = a_{n,p} + a_{p,n}$, et nous en déduisons, d'après la définition de la série $f(x, y)$, le résultat annoncé, à savoir $d(t) = 2f(t, t)$. □

En considérant l'ensemble des substitutions permettant d'obtenir les mots du langage D, nous obtenons l'équation fonctionnelle suivante dont est solution la série génératrice $f(x, y)$.

Théorème 6.2. *La série génératrice $f(x, y)$ est donnée par*

$$f(x,y)=g(x,y)+\sum_{m\geq 1} xy\frac{1-y^m}{1-y}(f(xy^m, xy^{m+1})+f(xy^{m+1}, xy^m))$$
$$+\sum_{m\geq 0} y(f(yx^m, yx^{m+1})+f(yx^{m+1}, yx^m))$$
$$+\sum_{m\geq 0} f(xy^m, xy^{m+1}),$$

où $g(x, y)$ est la série

$$g(x,y)=y\frac{xy}{1-xy}+\sum_{m\geq 1}\left(xy^m\frac{xy^{m+1}}{1-xy^{m+1}}+\frac{xy^m}{1-xy^m}\right)$$
$$+\sum_{m\geq 1}\left(y\frac{yx^{m+1}}{1-yx^{m+1}}+xy\frac{xy^{m+2}(1-y^m)}{(1-y)(1-xy^{m+2})}\right).$$

Preuve. Soit D_1 l'ensemble des mots de D se terminant par la lettre 1, langage dont $f(x, y)$ est la série génératrice. Ce langage D_1 se décompose alors en la réunion disjointe de deux langages A et D' définis par $A = A_1 \cup A_2 \cup A_3$ et $D' = D_1 \backslash A$, où les langages A_1, A_2 et A_3 sont les suivants:

$$A_1=\{(01^m)^r \mid r\geq 1 \text{ et } m\geq 1\},$$
$$A_2=\{(10^{m+1})^r 1 \mid r\geq 1 \text{ et } m\geq 1\},$$
$$A_3=\{(01^{m+2})^r 01^s \mid r\geq 1 \text{ et } 1\leq s\leq m\}.$$

La série génératrice $a(x, y)$ des mots du langage A est donnée par

$$a(x,y)=\sum_{m\geq 1}\left(\frac{xy^m}{1-xy^m}+y\frac{yx^{m+1}}{1-yx^{m+1}}+xy\frac{xy^{m+2}(1-y^m)}{(1-y)(1-xy^{m+2})}\right). \tag{15}$$

De plus, tout mot de D' s'obtient à partir d'un mot de D_1 de longueur strictement inférieure en effectuant l'une des deux substitutions Θ_m ou Γ_m sur ce mot. Ces substitutions sont définies dans les quatre cas (suivant que $m \geq 1$ ou $m = 0$) ci-dessous. Plus précisément, les substitutions Θ_m (respectivement Γ_m) permettent d'obtenir tous les mots de D' commençant par la lettre 0 (respectivement la lettre 1).

Cas 1: Pour tout $m \geq 1$ $\Theta_m(0) = 01^m$, $\Theta_m(1) = 01^{m+1}$. Cette substitution est effectuée pour tous les mots de D_1 ayant n lettres 0 et p lettres 1 en concaténant à droite des mots obtenus

- soit le mot vide,
- soit tous les mots de la form 01^q pour $1 \leq q \leq m$.

Il faut alors retrancher tous les mots de D_1 ayant $(n+1)$ lettres 0, p lettres 1 ($n \geq 1$ et $p \geq 1$) et se terminant par la lettre 0 auxquels on applique la même substitution. Nous obtenons alors le terme

$$S_1=\sum (a_{n,p}+a_{p,n})x^{n+p}y^{(n+p)m+p}\left(1+xy\frac{1-y^m}{1-y}\right)$$
$$-\sum a_{p,n+1}x^{n+p}y^{(n+p)m+p}xy^m, \tag{16}$$

où les deux sommations sont effectuées pour toutes les valeurs de n, p et m vérifiant $n \geq 1$, $p \geq 1$ et $m \geq 1$.

Ainsi, après simplification, et comme $a_{p,1} = 1$ pour toute valeur de p (d'après la définition de la suite $a_{n,p}$), nous avons

$$S_1 = \sum (a_{n,p} + a_{p,n}) x^{n+p} y^{(n+p)m+p} xy \frac{1-y^m}{1-y}$$
$$+ \sum a_{n,p} x^{n+p} y^{(n+p)m+p}$$
$$+ \sum xy^m \frac{xy^{m+1}}{1-xy^{m+1}}, \tag{17}$$

les sommations se faisant toujours pour $n \geq 1$, $p \geq 1$ et $m \geq 1$.

Cas 2: Pour tout $m \geq 1$ $\Gamma_m(0) = 10^m$, $\Gamma_m(1) = 10^{m+1}$. Cette substitution est effectuée pour tous les mots de D_1 ayant n lettres 0 et p lettres 1 en concaténant à droite des mots obtenus la lettre 1. Nous obtenons alors le terme suivant:

$$S_2 = \sum_{n \geq 1, p \geq 1, m \geq 1} (a_{n,p} + a_{p,n}) x^{(n+p)m+p} y^{n+p+1}. \tag{18}$$

Cas 3: $\Theta_0(0) = 0$, $\Theta_0(1) = 01$. Cette substitution est effectuée pour tous les mots de D_1 ayant n lettres 0, p lettres 1 ($n \geq 1$ et $p \geq 1$) et se terminant par la lettre 1. Nous obtenons alors le terme

$$S_3 = \sum_{n \geq 1, p \geq 1} a_{n,p} x^{n+p} y^p. \tag{19}$$

Cas 4: $\Gamma_0(0) = 1$, $\Gamma_1(1) = 10$. Cette substitution doit être effectuée pour tous les mots de D_1

(a) ayant n lettres 0, p lettres 1 et se terminant par la lettre 0, ce qui nous donne le terme S_{41}.

(b) ayant n lettres 0 et p lettres 1 et en concaténant aux mots obtenus la lettre 1, ce qui nous donne le terme S_{42}.

Il faut alors retrancher tous les mots de D_1 ayant $(n+1)$ lettres 0, p lettres 1 ($n \geq 1$ et $p \geq 1$) et se terminant par la lettre 0 auxquels on applique la même substitution, ce qui nous donne le terme S_{43}.

Ces trois termes et le terme résultant sont alors les suivants

$$S_{41} = \sum_{n \geq 1, p \geq 1} a_{p,n} x^p y^{n+p}, \tag{20}$$

$$S_{42} = \sum_{n \geq 1, p \geq 1} (a_{n,p} + a_{p,n}) x^p y^{n+p+1}, \tag{21}$$

$$S_{43} = \sum_{n \geq 1, p \geq 1} a_{p,n+1} x^p y^{n+p+1}, \tag{22}$$

$$S_4 = S_{41} + S_{42} - S_{43}. \tag{23}$$

Ainsi, après simplification, et compte-tenu du fait que $a_{p,1} = 1$ pour toute valeur de p (d'après la définition de la suite $a_{n,p}$), nous obtenons

$$S_4 = y \frac{xy}{1-xy} + \sum_{n \geq 1, p \geq 1} (a_{n,p} + a_{p,n}) x^p y^{n+p+1}. \tag{24}$$

Ainsi, à partir des mots du langage A et en itérant ces quatre substitutions correspondant à ces quatre cas, nous obtenons tous les mots du langage D_1.

On constate également que seuls les mots du langage A ne peuvent être obtenus à partir d'un mot de D_1 par l'une de ces substitutions, ceci d'après la définition de D_1 (tout mot de D_1 possède au moins une lettre 0 et une lettre 1) et d'après la définition du langage A et des substitutions considérées. La série génératrice $f(x, y)$ du langage D_1 est alors

$$f(x, y) = a(x, y) + S_1 + S_2 + S_3 + S_4. \tag{25}$$

En reportant dans cette équation les valeurs obtenues dans les formules (15), (17), (18), (19) et (24), nous déduisons de la définition de la série génératrice $f(x, y)$ l'équation fonctionnelle du Théorème 6.2. □

A l'aide du langage de calcul symbolique MACSYMA, nous avons obtenu les premières valeurs suivantes pour les nombres $a_{n,p}$ à partir de cette équation fonctionnelle (Table 1).

Le nombre d_n de mots de D de longueur n comprise entre 2 et 30 est donné dans la Table 2.

L'équation fonctionnelle dont $f(x, y)$, la série génératrice des mots du langage D se terminant par la lettre 1, est l'unique solution ne permet pas d'obtenir une expression simple de la série génératrice $d(x)$ du langage D.

Table 1
Nombres $a_{n,p}$ $(n + p \leq 21)$.

n \ p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	1	2	3	5	7	9	12	15	18	22	26	30	35	40	45	51	57	63	70	
3	1	3	3	4	7	9	10	13	16	19	22	25	29	34	37	41	47	52		
4	1	3	4	4	5	10	12	12	15	18	23	26	27	32	37	41	45			
5	1	4	5	7	5	6	13	18	17	17	19	26	32	35	35	39				
6	1	4	4	6	9	6	7	16	22	23	22	21	24	34	45					
7	1	5	6	7	11	12	7	8	20	27	28	32	31	26						
8	1	5	6	7	8	15	15	8	9	24	33	31	37							
9	1	6	6	8	9	14	18	19	9	10	28	40								
10	1	6	7	9	9	11	18	23	22	10	11									
11	1	7	8	10	13	11	18	21	30	27										
12	1	7	7	8	12	12	14	23	24											
13	1	8	9	11	12	18	13	25												
14	1	8	9	12	13	20	15													
15	1	9	9	11	13	15														
16	1	9	10	11	14															
17	1	10	11	14																
18	1	10	10																	
19	1	11																		
20	1																			

Table 2
Nombres d_n $(2 \leqslant n \leqslant 30)$.

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
d_n	2	6	12	22	34	52	74	102	134	176	222	280	344	416	496

n	17	18	19	20	21	22	23	24	25	26	27	28	29	30
d_n	592	694	814	942	1082	1232	1404	1584	1784	1996	2226	2468	2738	3016

En exploitant les tableaux de nombres ci-dessus, nous avons constaté que la série $d(x)$ pouvait s'exprimer en fonction de *l'indicatrice d'Euler*. Considérons la série génératrice du langage D

$$d(x) = \sum_{n \geqslant 2} d_n x^n,$$

et la série d'Euler définie par

$$\Phi(x) = \sum_{n \geqslant 1} \varphi(n) x^n, \tag{26}$$

dans laquelle $\varphi(n)$ est le nombre d'entiers inférieurs à n et premiers avec n.

Remarque 6.3. Pour tout $n \geqslant 2$, le nombre de droites de $\Delta(n)$ est

$$\operatorname{card}(\Delta(n)) = \sum_{i=2}^{n} i\varphi(i). \tag{27}$$

Nous formulons la conjecture suivante, reliant la série génératrice du langage D à la série d'Euler.

Conjecture 6.4. *La série génératrice $d(x)$ du langage D est donnée par*

$$d(x) = \frac{x(x-1) + \Phi(x)}{(1-x)^2}. \tag{28}$$

De façon équivalente, le nombre de mots de longueur $n \geqslant 2$ de D est

$$d_n = -1 + \sum_{i=1}^{n} (n-i+1)\varphi(i). \tag{29}$$

Remarque 6.5. La Conjecture 6.4 implique la Conjecture 5.6 du fait de la non algébricité de la série indicatrice d'Euler $\Phi(x)$ et de la relation (28). En effet, $d(x)$ n'est donc pas une série algébrique si la Conjecture 6.4 est vraie, et il en est alors de même pour la série génératrice du langage complémentaire de D. Or, le langage complémentaire de D étant algébrique, ce langage est alors ambigu.

Remarque 6.6. Sachant que

$$\sum_{i=1}^{n} \varphi(i) = \frac{3n^2}{\pi^2} + O(n \log n),$$

nous en déduisons, si la Conjecture 6.4 est exacte, que d_n est asymptotiquement équivalent à n^3/π^2 et que le rapport card$(\Delta(n))/d_n$ a pour limite 2 pour n tendant vers l'infini.

Remerciements

Nour remercions Messieurs Cohen et Laubie pour avoir attiré notre attention sur ce problème et Messieurs Allouche et Rauzy pour leur aide dans les recherches bibliographiques.

Références

[1] J.M. Autebert, J. Beququier, L. Boasson et M. Nivat, Quelques problèmes ouverts en théorie des langages algébriques, *RAIRO Theoret. Comput. Sci.* **13** (1979) 363-379.
[2] J. Berstel, *Transductions and Context-free Languages* (Teubner, Stuttgart, 1979).
[3] J. Berstel, Every iterated morphism yields a co-CFL, *Inform. Process. Lett.* **22** (1986) 7-9.
[4] E.B. Christoffel, Observatio Arithmetica, *Ann. Math. (2)* **6** (1875) 148-152.
[5] H. Cohen, Communication personnelle, 1987.
[6] E.M. Coven and G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* **7** (1983) 138-153.
[7] P. Flajolet, Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.* **49** (1987) 283-309.
[8] B. Grunbaum and G.C. Shephard, *Tilings and Patterns* (Freeman, San Francisco, 1986).
[9] G.A. Hedlund, Sturmian minimal sets, *Amer. J. Math.* **66** (1944) 605-620.
[10] G.A. Hedlund and M. Morse, Symbolic dynamics, *Amer. J. Math.* **60** (1938) 815-866.
[11] G.A. Hedlund and M. Morse, Symbolic dynamics, part II: Sturmian trajectories, *Amer. J. Math.* **62** (1940) 1-42.
[12] W.F. Lunnon and P.A.B. Pleasants, Characterization of two distance sequences, Preprint.
[13] W.F. Lunnon and P.A.B. Pleasants, Quasicrystallographie Tilings, *J. Math. Pure Appl.*, **66** (1987) 217-263.
[14] A.A. Markoff, Sur une question de Jean Bernoulli, *Math. Ann.* **19** (1882) 27-36.
[15] G. Rauzy, Suites à termes dans un alphabet fini, dans: *Séminaire de Théorie des Nombres de Bordeaux* (1982-1983) 25-01, 25-16.
[16] G. Rauzy, Mots infinis en arithmétique, dans: *Automata on Infinite Words*, Ecole de Printemps d'Informatique Théorique, Le Mont Doré 1984, Lecture Notes in Computer Science **192** (Springer, Berlin, 1985) 165-171.
[17] C. Series, The geometry of Markoff numbers, *Math. Intelligencer* **7** (1985) 20-29.
[18] H.J.S. Smith, Note on continued fractions, *Messenger Math. (2)* **6** (1876) 1-14.
[19] K.B. Stolarsky, Beatty sequences, continued fractions and certain shift operators, *Canad. Math. Bull.* **19** (1976) 473-482.
[20] B.A. Venkov, *Elementary Number Theory* (Wolters-Noordhoff, Groningen, The Netherlands, 1970) 67.