Etude de la complexité du mot de Fibonacci généralisé

Cassaigne Julien [1] et Kaboré Idrissa [2]

(1) Inst. de Math. de Luminy, Université d'Aix-Marseille, Marseille, France
(2) Inst. des Sci. Exactes et Appliquées, Univ. polytech. de Bobo-Dioulasso, Burkina Faso
Julien.Cassaigne@iml.univ-mrs.fr, ikaborei@yahoo.fr

RÉSUMÉ. Dans ce papier, nous menons une étude générale de la complexité du mot de Fibonacci généralisé, $F_{l,\,m}$, engendré par le morphisme $\sigma_{l,\,m}$ donné par $\sigma_{l,\,m}(a) = a^l b^m$ et $\sigma_{l,\,m}(b) = a$.

ABSTRACT. In this paper we undertake a general study of the complexity function of the generalized Fibonacci word which is generated by the morphism defined by $\sigma_{l,\,m}(a) = a^l b^m$ and $\sigma_{l,\,m}(b) = a$.

MOTS-CLÉS : Mots infinis, morphismes, complexité, facteurs spéciaux

KEYWORDS : Infinite words, morphisms, complexity, special factors

1. Introduction

La fonction de complexité $\mathbf{p}$, qui compte le nombre de facteurs de longueur donnée dans un mot infini, est une notion centrale en combinatoire de mots. Elle a été introduite en 1975 par Ehrenfeucht et al. [7]. Elle permet de mesurer la diversité des motifs dans un mot infini. Elle est souvent utilisée pour caractériser certains mots ou familles de mots ; par exemple les mots ultimement périodiques sont les seuls mots dont la fonction de complexité est bornée. Nous référons le lecteur à [5, 3] pour plus de détails sur cette notion.

Soit σ le morphisme du monoïde libre $\{a,\, b\}^*$ défini par $\sigma(a) = ab$ et $\sigma(b) = a$. En itérant indéfiniment le morphisme σ à partir de a on obtient un mot infini appelé le mot de Fibonacci $F = abaababaabaababaab \cdots$. Le mot de Fibonacci a été largement étudié [2, 6, 9, 10] et est devenu célèbre par ses nombreuses propriétés remarquables. Nous référons le lecteur à [4] pour plus de détails. Sa fonction de complexité est connue : il possède, pour tout entier n, exactement $n + 1$ facteurs de longueur n. Le morphisme de Fibonacci généralisé du monoïde libre $\{a,\, b\}^*$ est le morphisme $\sigma_{l,\, m}$ défini par $\sigma_{l,\, m}(a) = a^l b^m$ et $\sigma_{l,\, m}(b) = a$, pour $l \geq 1$ et $m \geq 2$. En itérant le morphisme $\sigma_{l,\, m}$ on obtient un mot infini appelé mot de Fibonacci généralisé $F_{l,\, m}$ (voir [1], page 336). Nous nous proposons dans ce travail de mener une étude de la complexité de ces mots.

Le papier est organisé de la façon suivante.

Nous rappelons, à la section 2, les définitions et notations de base. Nous décrivons ensuite, à la section 3, les facteurs bispéciaux non ordinaires de $F_{l,\, m}$. Enfin, à la section 4, nous étudions sa complexité.

2. Préliminaires

La plupart des notations utilisées ici peuvent être retrouvées dans le livre de M. Lothaire [8].

Soit $\mathcal{A} = \{a,\, b\}$ un alphabet fixé. $\mathcal{A}^*$, l'ensemble des mots finis sur $\mathcal{A}$, est le monoïde libre engendré par $\mathcal{A}$; ε le mot vide étant l'élément neutre. Pour tout $u \in \mathcal{A}^*$, $|u|$ désigne le nombre de lettres du mot u ($|\varepsilon| = 0$) et pour toute lettre x de $\mathcal{A}$, $|u|_x$ est le nombre d'occurrences de x dans u. Un mot u de longueur n formé d'une seule lettre x est simplement noté $u = x^n$; par extension $x^0 = \varepsilon$.

Un mot infini est une suite de lettres de $\mathcal{A}$ indexée par $\mathbb{N}$. On désigne par $\mathcal{A}^\omega$ l'ensemble des mots infinis sur $\mathcal{A}$. Un mot fini v est facteur d'un mot u s'il existe deux mots u_1 et u_2 sur l'alphabet $\mathcal{A}$ tels que $u = u_1 v u_2$; on dit aussi que u contient v. Le facteur v est dit préfixe (resp. suffixe) si u_1 (resp. u_2) est le mot vide.

Soient $u \in \mathcal{A}^\omega$, w un facteur de u et x une lettre de $\mathcal{A}$. Le langage des facteurs de longueur n de u est noté $\mathcal{L}_n(u)$ et l'ensemble de tous les facteurs de u est noté $\mathcal{L}(u)$. La lettre x est un prolongement à gauche (resp. à droite) de w si xw (resp. wx) appartient à $\mathcal{L}(u)$. Un facteur v de u est dit spécial à gauche (resp. à droite) si av et bv (resp. va et vb) sont dans u. Un facteur à la fois spécial à gauche et à droite est dit bispécial.

La fonction de complexité de u est l'application de $\mathbb{N}$ dans $\mathbb{N}^*$ définie par $\mathbf{p}_u(n) = \#\mathcal{L}_n(u)$, où $\#\mathcal{L}_n(u)$ désigne le cardinal de $\mathcal{L}_n(u)$. Dans tout ce qui suit, la fonction de complexité, $\mathbf{p}_u$, d'un mot u sera simplement notée $\mathbf{p}$.

On appelle différence première de la fonction de complexité de u, la fonction notée $\mathbf{s}$ et définie par : $\mathbf{s}(n) = \mathbf{p}(n+1) - \mathbf{p}(n)$. On en déduit la formule $\mathbf{p}(n) = \mathbf{p}(k_0) +$

$\sum_{k=k_0}^{n-1} \mathbf{s}\,(k)$. Sur un alphabet à deux lettres la fonction $\mathbf{s}$ compte le nombre de facteurs spéciaux à droite d'une longueur donnée dans u. Il se trouve que le dénombrement de certains facteurs bispéciaux intervient dans la détermination de la fonction $\mathbf{s}$ (voir section 3 et 4).

Un morphisme f est une application de $\mathcal{A}^*$ dans lui-même telle que $f(uv) = f(u)f(v)$ pour tous $u,\, v \in \mathcal{A}^*$.

On dit qu'un mot infini u est engendré par un morphisme f s'il existe une lettre a telle que les mots a, $f(a)$, $f^2(a)$, $\cdots$, $f^n(a)$, $\cdots$ sont des préfixes de plus en plus longs de u. On note alors $u = f^\omega(a)$.

Soit u un mot infini sur $\mathcal{A}$ et v un facteur de u. Le vecteur de Parikh de v est $\chi(v) = (|v|_a\,,\,|v|_b)$. On appelle matrice d'incidence d'un morphisme φ la matrice

$$M_\varphi = \begin{pmatrix} |\varphi\,(a)|_a & |\varphi\,(b)|_a \\ |\varphi\,(a)|_b & |\varphi\,(b)|_b \end{pmatrix}.$$

3. Facteurs bispéciaux non ordinaires de $F_{l,\,m}$

Définition 3.1. *Soit u un mot infini sur $\mathcal{A}$ et v un facteur bispécial de u.*

– v est dit bispécial fort si $ava,\ avb,\ bva,\ bvb$ sont des facteurs de u.

– v est dit bispécial faible si uniquement ava et bvb, ou avb et bva, sont des facteurs de u.

– v est dit bispécial bispécial ordinaire si v n'est ni fort ni faible.

Définition 3.2. *Un facteur bispécial de $F_{l,\,m}$ est dit court s'il ne contient aucun des trois mots a^{l+1}, b^m et ba. Un facteur bispécial de $F_{l,\,m}$ qui n'est pas court sera dit bispécial long.*

Lemme 3.1.

1) $F_{l,\,m}$ possède exactement un facteur bispécial court et faible : b^{m-1}.

2) $F_{l,\,m}$ possède exactement deux facteurs bispéciaux courts et forts qui sont : $\epsilon,\ a^l$.

Lemme 3.2. *Soit w un facteur de $F_{l,\,m}$. Les assertions suivantes sont équivalentes :*

1) w est bispécial long de $F_{l,\,m}$.

2) Il existe un facteur bispécial v de $F_{l,\,m}$ tel que $w = \hat{\sigma}_{l,\,m}(v)$ où $\hat{\sigma}_{l,\,m}(v) = \sigma_{l,\,m}(v)a^l$. De plus, v et w sont de même type et $|v| < |w|$.

Les preuves des lemmes 3.1 et 3.2 sont longues et techniques mais elles utilisent une méthode générale développée dans [5] (voir aussi [3]). De ce fait, nous ommettons ces preuves.

En conséquence, nous avons :

1) Les facteurs bispéciaux faibles de $F_{l,\,m}$ sont donnés par la suite (y_n) définie par $y_1 = b^{m-1}$ et $y_{n+1} = \hat{\sigma}_{l,\,m}\,(y_n)$ pour $n \geq 1$.

2) Les facteurs bispéciaux forts de $F_{l,\,m}$ sont donnés par la suite (x_n) définie par $x_0 = \varepsilon$ et $x_{n+1} = \hat{\sigma}_{l,\,m}\,(x_n)$ pour $n \geq 0$.

4. Complexité de $F_{l,\,m}$

Définition 4.1. *Soit* $v,\, w \in \mathcal{A}^*$ *et* $\chi\left(v\right),\, \chi\left(w\right)$ *leurs vecteurs de Parikh. On dira que* $\chi\left(v\right)$ *est inférieur à* $\chi\left(w\right)$ *et on écrira* $\chi\left(v\right) \leq \chi\left(w\right)$ *lorsque* $|v|_x \leq |w|_x$ *pour tout* $x \in \mathcal{A}$. *De plus, si* $\chi(v) \neq \chi(w)$, *on note* $\chi(v) < \chi(w)$.

Proposition 4.1. *Soient* $v,\, w,\, v',\, w'$ *quatre mots finis tels que* $v' = \hat{\sigma}_{l,\,m}(v)$ *et* $w' = \hat{\sigma}_{l,\,m}(w)$. *Alors,*

$$\chi\left(v\right) < \chi\left(w\right) \Longrightarrow \chi\left(v'\right) < \chi\left(w'\right)$$

Démonstration. D'une part, nous avons $|v'|_a = l\,|v|_a + |v|_b + l$ et $|w'|_a = l\,|w|_a + |w|_b + l$; donc $|v'|_a < |w'|_a$.
D'autre part, nous avons $|v'|_b = m\,|v|_b$ et $|w'|_a = m\,|w|_a$; donc $|v'|_b \leq |w'|_b$. □

Proposition 4.2. *Pour tout* $m \geq 2$ *on a :*

$$\forall n \geq 0,\ \chi\left(x_n\right) < \chi\left(y_{n+1}\right) < \chi\left(x_{n+2}\right)$$

Démonstration. Nous avons $\chi\left(x_0\right) < \chi\left(y_1\right) < \chi\left(x_2\right)$ puisque $x_0 = \varepsilon$, $y_1 = b^{m-1}$ et $x_2 = \left(a^l b^m\right)^l a^l$. Supposons, les inégalités vraies jusqu'à l'ordre n, *i.e.*,

$$\chi\left(x_n\right) < \chi\left(y_{n+1}\right) < \chi\left(x_{n+2}\right).$$

D'après la proposition 4.1, nous obtenons $\chi\left(x_{n+1}\right) < \chi\left(y_{n+2}\right) < \chi\left(x_{n+3}\right)$. □

Le lemme suivant décrit la fonction $\mathbf{s}$.

Lemme 4.1. *Soit* $n \in \mathbb{N}$ *on a :*

– $\mathbf{s}(n) = 1$ *pour* $n = 0$.

– *Si* $n \in \mathbb{N}^*$, *on note* k *le plus grand entier tel que* $n > |x_k|$.

1) Si $k = 0$, *on a :* $\mathbf{s}(n) = \begin{cases} 2 & \text{si } 1 \leq n \leq \min(l,\, m-1) \\ 1 & \text{si } m \leq n \leq l \end{cases}$.

2) Sinon, on a : $\mathbf{s}(n) = \begin{cases} 3 & \text{si } n \leq |y_k| \\ 2 & \text{si } |y_k| < n \leq |y_{k+1}| \\ 1 & \text{si } |y_{k+1}| < n \end{cases}$.

Démonstration. La fonction $\mathbf{s}$ est donnée par la formule suivante

$$\mathbf{s}(n) = 1 + \#\left\{w \text{ bispécial fort : } |w| < n\right\} - \#\left\{w \text{ bispécial faible : } |w| < n\right\}.$$

Nous savons que $\mathbf{s}(0) = 1$. On observe aussi que $\mathbf{s}(n) = 2$ si $1 \leq n \leq \min(l,\, m-1)$ et $\mathbf{s}(n) = 1$ si $m \leq n \leq l$.

Supposons $n \geq |x_1|$. On note k le plus grand entier tel que $n > |x_k|$. Alors, on a :

$$\mathbf{s}(n) = 1 + (k+1) - \begin{cases} k-1 & \text{si } n \leq |y_k| \\ k & \text{si } |y_k| < n \leq |y_{k+1}| \\ k+1 & \text{si } |y_{k+1}| < n \end{cases}.$$

Ce qui prouve le lemme. □

Théorème 4.3. *La complexité de* $F_{l,\,m}$ *satisfait les inégalités suivantes :*

$$n + 1 \leq \mathbf{p}(n) \leq 3n + 1.$$

Démonstration. D'après le lemme 4.1, nous avons

$$\forall n \geq 0,\ 1 \leq \mathbf{s}(n) \leq 3.$$

D'où,

$$1 + \sum_{k=0}^{n-1} 1 \leq \mathbf{p}(n) \leq 1 + \sum_{k=0}^{n-1} 3.$$

□

Lemme 4.2. *Nous avons les équivalences suivantes.*

1) $m \in \left[2,\ 2l^2+1\right] \Longleftrightarrow \exists k_0 \in \mathbb{N} : \forall k \geq k_0,\ |x_k| - |y_k| > 0.$

2) $m \in \left[2l^2+2,\ \infty\right[\Longleftrightarrow \exists k_0 \in \mathbb{N} : \forall k \geq k_0,\ |x_k| - |y_k| < 0.$

Démonstration. Considérons les suites $(v_k)_{k\geq 1}$ et $(V_k)_{k\geq 1}$ définies par $v_k = |x_k| - |y_k|$ et $V_k = \chi(x_k) - \chi(y_k)$. Nous avons :

$$V_1 = \begin{pmatrix} l \\ -m+1 \end{pmatrix} \text{ et } V_{k+1} = AV_k$$

où $A = \begin{pmatrix} l & 1 \\ m & 0 \end{pmatrix}$ désigne la matrice d'incidence de $\sigma_{l,m}$. Les valeurs propres de A, étant racines de $X^2 - lX - m$, sont

$$\lambda_1 = \frac{l + \sqrt{l^2+4m}}{2} \text{ et } \lambda_2 = \frac{l - \sqrt{l^2+4m}}{2}.$$

Observons que $\lambda_1 > l \geq 1$ et $|\lambda_2| < \lambda_1$. Par ailleurs, nous avons

$$\forall k \geq 1,\ v_k = \begin{pmatrix} 1 & 1 \end{pmatrix} V_k.$$

D'où,

$$\begin{aligned} v_k &= \begin{pmatrix} 1 & 1 \end{pmatrix} A^{k-1} \begin{pmatrix} l \\ -m+1 \end{pmatrix} \\ &= \alpha_1 \lambda_1^{k-1} + \alpha_2 \lambda_2^{k-1} \end{aligned}$$

avec α_1 et α_2 vérifiant le système

$$\begin{cases} \alpha_1 + \alpha_2 & = & l - m + 1 \\ \alpha_1\lambda_1 + \alpha_2\lambda_2 & = & l^2 + lm - m + 1 \end{cases}.$$

Nous avons

$$\alpha_1 = \frac{l^2 + lm - m + 1 - \lambda_2(l - m + 1)}{(\lambda_1 - \lambda_2)},$$

$$\alpha_2 = \frac{l^2 + lm - m + 1 - \lambda_1(l - m + 1)}{(\lambda_2 - \lambda_1)}.$$

Puisque $|\lambda_2| < \lambda_1$, alors v_k est du signe de α_1 pour k assez grand.

Cas 1. $l - m + 1 \geq 0$. Alors, nous avons $-\lambda_2(l - m + 1) \geq 0$ car $\lambda_2 < 0$. Donc, $\alpha_1 > 0$ puisque $\lambda_1 - \lambda_2 > 0$ et $l^2 + lm - m + 1 = l^2 + l(m-1) + 1 > 0$.

Cas 2. $l-m+1<0$. Nous avons :

$$\begin{aligned}\alpha_1>0 &\Longleftrightarrow \lambda_2>\frac{l^2+lm-m+1}{l-m+1}\\ &\Longleftrightarrow \frac{l-\sqrt{l^2+4m}}{2}>\frac{l^2+lm-m+1}{l-m+1}\\ &\Longleftrightarrow l^2+4m<\frac{\left(-l^2-3lm+l+2m-2\right)^2}{\left(l-m+1\right)^2}\end{aligned}$$

car $-l^2-3lm+l+2m-2=m(2-3l)-l(l-1)-2<0$ et $l-m+1<0$. La dernière inégalité peut se ramener à l'inégalité suivante :

$$P_l\left(m\right)=m^3+m^2\left(-2l^2+l-3\right)+m\left(-2l^3+3l^2-2l+3\right)+l^3-l^2+l-1<0$$

Si $l=1$ alors $P_l(m)=m(m^2-4m+2)$. Donc, $P_l(m)<0 \Longleftrightarrow m\in\{2,\,3\}$. Supposons désormais $l\geq 2$. La dérivée

$$P_l'\left(m\right)=3m^2+2m\left(-2l^2+l-3\right)-2l^3+3l^2-2l+3$$

admet deux racines de signes contraires

$$\beta_1=\frac{2l^2-l+3+\sqrt{4l^4+2l^3+4l^2}}{3}>0,$$

$$\beta_2=\frac{2l^2-l+3-\sqrt{4l^4+2l^3+4l^2}}{3}<0$$

et est négative entre ces deux racines. Donc, P_l est décroissante sur $[\beta_1,\,\beta_2]$ qui contient 0 et croît sur $]-\infty,\,\beta_1]\cup[\beta_2,\,+\infty[$. De plus, on vérifie que $P_l(0)>0$, $P_l(1)<0$, $P_l(2l^2+1)<0$ et $P_l(2l^2+2)>0$. Ce qui implique que, P_l admet deux racines positives m_1 et m_2 et une troisième racine négative m_3 avec $0<m_1<1$, $2l^2+1<m_2<2l^2+2$ et $m_1\leq\beta_2\leq m_2$. Donc, P_l est strictement négative sur $]m_1,\,m_2[$ et strictement positive sur $]m_2,\,+\infty[$. Ainsi, $\alpha_1>0$ si $m\in[2,\,2l^2+1]$ et $\alpha_1<0$ si $m\geq 2l^2+2$.

En définitive, nous avons :

1) $m\in\left[2,\;2l^2+1\right]\Longleftrightarrow \exists k_0\in\mathbb{N}:\forall k\geq k_0,\;|x_k|-|y_k|>0.$

2) $m\in\left[2l^2+2,\;\infty\right[\Longleftrightarrow \exists k_0\in\mathbb{N}:\forall k\geq k_0,\;|x_k|-|y_k|<0.$

□

Théorème 4.4. *1) Si $m\in\left[2,\;2l^2+1\right]$ alors, il existe une constante c et un entier naturel n_0 tels que pour tout $n>n_0$*

$$n+1\leq\mathbf{p}(n)\leq 2n+c.$$

2) Si $m\in\left[2l^2+2,\;+\infty\right[$ alors, il existe une constante c et un entier naturel n_1 tels que pour tout $n>n_1$

$$2n+c\leq\mathbf{p}(n)\leq 3n+1.$$

Démonstration. Supposons $m \in [2,\, 2l^2+1]$. Alors, il existe k_0 tel que pour tout $k \geq k_0$, $v_k > 0$. Dans ce cas, pour $n \in]|x_k|\,,\, |x_{k+1}|]$ nous avons

$$\mathbf{s}(n) = 1 + (k+1) - \left\{ \begin{array}{ll} k & \text{si } n \leq |y_{k+1}| \\ k+1 & \text{si } |y_{k+1}| < n \end{array} \right. .$$

D'où,

$$\mathbf{s}(n) = \left\{ \begin{array}{ll} 2 & \text{si } |y_k| < n \leq |y_{k+1}| \\ 1 & \text{si } |y_{k+1}| < n \end{array} \right. .$$

Donc,

$$\forall n \geq |x_{k_0}|,\ 1 \leq \mathbf{s}\,(n) \leq 2.$$

En prenant la somme, on obtient les inégalités

$$\sum_{k=|x_{k_0}|}^{n-1} 1 \leq \sum_{k=|x_{k_0}|}^{n-1} \mathbf{s}\,(k) \leq \sum_{k=|x_{k_0}|}^{n-1} 2.$$

Soit encore

$$\mathbf{p}\,(|x_{k_0}|) + n - |x_{k_0}| \leq \mathbf{p}(n) \leq \mathbf{p}\,(|x_{k_0}|) + 2\,(n - |x_{k_0}|)\,.$$

D'où,

$$n + 1 \leq \mathbf{p}(n) \leq 2n + c,\ \text{avec}\ \ c \in \mathbb{Z}.$$

Supposons $m \geq 2l^2+2$. Alors, d'après le lemme 4.2, il existe k_0 tel que pour tout $k \geq k_0$, $v_k < 0$.

Dans ce cas, pour $n \in]|x_k|\,,\, |x_{k+1}|]$ nous avons

$$\mathbf{s}(n) = 1 + (k+1) - \left\{ \begin{array}{ll} k-1 & \text{si } n \leq |y_k| \\ k & \text{si } |y_k| < n \end{array} \right. .$$

D'où,

$$\mathbf{s}(n) = \left\{ \begin{array}{ll} 3 & \text{si } n \leq |y_k| \\ 2 & \text{si } |y_k| < n \leq |x_{k+1}| \end{array} \right. .$$

Donc,

$$\forall n \geq |x_{k_0}|,\ 2 \leq \mathbf{s}\,(n) \leq 3.$$

Ainsi, en procédant de façon similaire que précédemment, on obtient

$$2n + c \leq \mathbf{p}(n) \leq 3n + 1,\ \text{avec}\ \ c \in \mathbb{Z}.$$

□

5. Bibliographie

[1] J.-P. ALLOUCHE, J. SHALLIT, « Automatic Sequences : Theory, Applications, Generalizations », *Cambridge University Press, UK*, 2003.

[2] J. BERSTEL, « Mots de Fibonacci », *L.I.T.P. Séminaire d'Informatique Théorique, Paris*, (1980-1981).

[3] J. CASSAIGNE, F. NICOLAS, « Factor complexity », in *Combinatorics, Automata and Number Theory, V. Berthé, M. Rigo (Eds), Encyclopedia of Mathematics and its Applications* n° 135, Cambridge University Press, 2010.

[4] J. CASSAIGNE, « On extremal properties of the Fibonacci word », *Theoret. Inform. Appl.*, vol. 42, n° 4, 2008.

[5] J. CASSAIGNE, « Complexité et facteurs spéciaux », *Bull. Belg. Math. Soc.*, vol. 4, 1997.

[6] A. DE LUCA, « A combinatorial property of the Fibonacci words », *Inform.Process. Lett.*, vol. 12, n° 4, 1981.

[7] A. EHRENFEUCHT, K.P. LEE, G. ROZENBERG, « Subword complexities of various classes of deterministic developmental languages without interaction », *Theoret. comput. Sci.* vol. 1, 1975.

[8] M. LOTHAIRE, « Algebraic combinatorics on words », *Cambridge University Press*, 2002.

[9] F. MIGNOSI, G. PIRILLO, « Repetitions in the Fibonacci infinite word », *RAIRO Inform. Théor. Appl.*, vol. 26, n° 3, 1992.

[10] G. PIRILLO, « From the Fibonacci word to Sturmian words », *Publ. Math. Debrecen*, vol. 54(suppl.), 1999.