Quelques propriétés des mots substitutifs

Julien Cassaigne François Nicolas

Résumé

Les mots infinis engendrés par des morphismes itérés (ou mots purement substitutifs) jouent un rôle important dans la combinatoire des mots. Dans cet article, nous considérons la classe plus large des mots substitutifs, c'est à dire l'ensemble des mots infinis obtenus comme images morphiques de mots purement substitutifs. Nous montrons que cette nouvelle classe est strictement plus large que la précédente puis que les mots substitutifs s'écrivent comme les images par des morphismes remarquables de mots infinis engendrés par des endomorphismes remarquables. Nous montrons enfin que ces morphismes remarquables permettent de contrôler la complexité en facteurs des mots infinis auxquels ils sont appliqués.

Abstract

Infinite words generated by iterated morphisms (or purely substitutive words) play an important part in combinatorics on words. In this article, we consider the larger class of substitutive words, that is infinite words obtained as morphic images of purely substitutive words. We show that this new class is strictly larger than that of purely substitutive words, and then that substitutive words can be written as images under a certain type of morphisms of infinite words generated by iterating a certain type of morphisms. Finally we show that these remarkable morphisms allow to control the subword complexity of the infinite words they are applied to.

1991 *Mathematics Subject Classification :* 68R15.

Key words and phrases : combinatorics on words, infinite word, substitutive word, subword complexity.

Bull. Belg. Math. Soc. 10 (2003), 661–677

1 Introduction

Les mots infinis engendrés par des morphismes itérés (que nous appellerons mots *purement substitutifs*) jouent un rôle important dans la combinatoire des mots. On peut citer par exemple les mots de Thue-Morse [14, 9] et de Fibonacci [8].

Pour profiter de bonnes propriétés, il est intéressant d'élargir la classe précédente en prenant à la place l'ensemble des mots infinis qui sont des images morphiques de mots purement substitutifs. Nous appellerons ces mots les *mots substitutifs*. Ils possèdent effectivement plusieurs propriétés intéressantes de stabilité : par exemple, les suffixes de mots substitutifs, les mots infinis dont un suffixe est substitutif et les images morphiques de mots substitutifs sont des mots substitutifs (on trouvera dans [1] un récapitulatif de toutes ces propriétés).

Dans la section suivante nous fixons les notations et la terminologie et nous rappelons quelques résultats que nous utiliserons par la suite.

Dans la section 3 nous montrons que la classe des mots substitutifs est strictement plus grande que la classe des mots purement substitutifs à l'aide de constructions explicites.

Dans les sections 4 et 5, nous montrons qu'il est toujours possible de se restreindre à des morphismes ayant certaines propriétés. Ainsi, nous montrons que tout mot substitutif est l'image par un morphisme lettre à lettre d'un mot infini engendré par un endomorphisme non effaçant, puis que tout mot purement substitutif est l'image par un morphisme bifixe d'un mot infini engendré par un endomorphisme ω-injectif.

Avant de conclure à la section 7 nous montrons que les morphismes "remarquables" des sections 4 et 5 ont effectivement de bonnes propriétés vis à vis de la complexité en facteurs des mots infinis. Pour ce faire, nous établissons trois inégalités.

2 Préliminaires

On note $\mathbb{N}$ l'ensemble des entiers positifs ou nuls. Pour tous a, $b \in \mathbb{N}$, on note $[a, b]$ l'ensemble des $n \in \mathbb{N}$ vérifiant $a \leq n \leq b$. La terminologie et les notations concernant la combinatoire des mots sont essentiellement celles de Lothaire [7, 8].

2.1 Mots

Un *alphabet* est un ensemble fini de symboles. Dans toute la suite les lettres Σ, Λ et Π désigent des alphabets. Un *mot* (sur l'alphabet Σ) est une suite de symboles (pris dans Σ) de longueur finie éventuellement nulle. La suite de longueur 0 est appelée *mot vide* et notée ε. On note $|x|$ la *longueur* du mot x. On appelle *langage* (sur Σ) tout ensemble de mots (sur Σ). Un *mot infini* (sur Σ) est un suite infinie de symboles (pris dans Σ).

Soient un mot (fini) x et un mot y éventuellement infini sur un alphabet Σ. On note xy la *concaténation* des mots x et y. La concaténation de k copies de x est notée x^k et la concaténation d'une infinité de copies de x est noté x^ω (remarquons que x^ω est vide ou infini). Soient X un langage et un ensemble Y de mots dont certains peuvent être infinis. On appelle concaténation des ensembles X et Y (notée

XY) l'ensemble des concaténations de la forme xy avec $x \in X$ et $y \in Y$. Pour alléger les notations, on notera xY à la place de $\{x\}Y$. On note X^+ l'ensemble des mots obtenus en concaténant une suite finie (non vide) de mots appartenant à X, $X^\star := X^+ \cup \{\varepsilon\}$, X^ω l'ensemble des mots obtenus par concaténation d'une suite infinie de mots appartenant à X et $X^\infty := X^\star \cup X^\omega$.

On dit que x est *facteur* de y lorsque $y \in \Sigma^\star x \Sigma^\infty$. On dit que x apparaît dans y à la *position* n lorsque $y \in \Sigma^n x \Sigma^\infty$. On note $\mathsf{F}(y)$ (resp. $\mathsf{F}_n(y)$) l'ensemble des facteurs (resp. l'ensemble des facteurs de longueur n) de y. On dit que x est *préfixe* (*propre*) de y lorsque $y \in x\Sigma^\infty$ (et $x \neq y$). On définit de la même manière les notions de *suffixe* et de *suffixe propre*.

On appelle (fonction de) *complexité* (en facteurs) de y l'application $\mathsf{p}_y : \mathbb{N} \to \mathbb{N}$ définie en posant $\mathsf{p}_y(n) := \#\,\mathsf{F}_n(y)$ pour tout $n \in \mathbb{N}$.

On dit que y est *ultimement périodique* lorsqu'il existe $y_0 \in \Sigma^\star$ et $y_1 \in \Sigma^+$ tels que $y = y_0 y_1^\omega$.

Théorème 1 (Morse-Hedlund [10]**).** *Un mot infini u est ultimement périodique si et seulement si sa fonction de complexité en facteurs p_u est bornée.*

2.2 Codes

Soit $X \subseteq \Sigma^\star$.

On dit que X est un *code* (resp. un *ω-code*) lorsque tout mot appartenant à $X^\star$ (resp. X^ω) s'écrit de manière unique comme concaténation d'éléments de X. On dit que X est un *code préfixe* (resp. *code suffixe*) lorsqu'aucun mot de X n'est préfixe (resp. suffixe) propre d'un autre mot de X. Un *code bifixe* est un code à la fois préfixe et suffixe.

On vérifie facilement que tout code préfixe est un ω-code, que tout ω-code est un code et que tout code suffixe est un code. Mais, posant $\Sigma := \{\mathtt{0}, \mathtt{1}\}$, $E_0 := \{\mathtt{0}, \mathtt{01}, \mathtt{10}\}$, $E_1 := \{\mathtt{0}, \mathtt{01}, \mathtt{11}\}$, $E_2 := \{\mathtt{0}, \mathtt{01}\}$, $E_3 := \{\mathtt{0}, \mathtt{010}\}$, $E_4 := \{\mathtt{01}, \mathtt{011}, \mathtt{11}\}$ et $E_5 := \{\mathtt{0}, \mathtt{10}\}$ on peut montrer que le langage E_0 n'est pas un code, le langage E_1 est un code suffixe mais pas un ω-code, le langage E_2 est un code suffixe et un ω-code mais pas un code préfixe, le langage E_3 est un ω-code mais pas un code préfixe ni un code suffixe, le langage E_4 est un code mais pas un code suffixe ni un ω-code, le langage E_5 est un code préfixe mais pas un code suffixe, le langage Σ est un code bifixe.

Nous utiliserons la version suivante du célèbre *théorème de défaut* (chapitre 6 de [8]) :

Théorème 2 (Théorème de défaut). *Soit X un langage fini sur Σ qui n'est pas un ω-code. Alors, il existe un langage fini Y sur Σ tel que $X^\star \subseteq Y^\star$ et $\#\,Y < \#\,X$.*

On dit que X est à *délai de déchiffrement borné* lorsqu'il existe $d \in \mathbb{N}$ vérifiant : pour tous x, $x' \in X$, tout $u \in X^d$ et tout $u' \in X^\infty$ tels que xu est préfixe de $x'u'$ on a $x = x'$. On appelle *délai de déchiffrement* de X le plus petit entier d (s'il en existe) vérifiant la condition précédente.

Par exemple, les codes préfixes sont les langages admettant 0 pour délai de déchiffrement et pour tout $d \in \mathbb{N}$, le langage $\left\{\mathtt{0}, \mathtt{0}^d\mathtt{1}\right\}$ admet d pour délai de déchiffrement.

Tout langage à délai de déchiffrement borné est un ω-code. Réciproquement, on peut montrer que les ω-codes *finis* ont un délai de déchiffrement borné. Mais, le langage infini $\{0\} \cup \{0^n 1 0^n 1\}_{n \in \mathbb{N}}$ est un ω-code qui n'est pas à délai de déchiffrement borné.

2.3 Morphismes

Nous appellerons ici *morphismes* les morphismes de monoïdes. Etant donnés deux monoïdes M et N, on note $\hom(M, N)$ l'ensemble des *morphismes* de M vers N et $\hom(M)$ l'ensemble des *endomorphismes* de M. Un morphisme dont l'ensemble de départ est le monoïde libre $\Lambda^\star$ est entièrement défini par sa restriction à Λ.

Soit $f \in \hom(\Lambda^\star, \Sigma^\star)$. On appelle *substitution* induite par f, et l'on note également f, le prolongement naturel de f en une application $\Lambda^\infty \to \Sigma^\infty$: l'image par f de $u \in \Sigma^\omega$ est obtenue en remplaçant chaque lettre de u par son image par f.

Définition 1 (Morphismes remarquables). *Soit $f \in \hom(\Lambda^\star, \Sigma^\star)$.*
On dit que f est :
- lettre à lettre *lorsque $f(\Lambda) \subseteq \Sigma$,*
- uniforme *lorsque pour tous a, $b \in \Lambda$, on a $|f(a)| = |f(b)|$,*
- non effaçant *lorsque pour tout $a \in \Lambda$ on a $f(a) \neq \varepsilon$,*
- ω-injectif *lorsque la substitution $\Lambda^\infty \to \Sigma^\infty$ qu'il induit est injective,*
- préfixe *(resp.* suffixe*) lorsque pour tous x, $y \in \Lambda^\star$ tels que $f(x)$ soit préfixe (resp. suffixe) de $f(y)$, on a x préfixe (resp. suffixe) de y,*
- bifixe *lorsque f est à la fois préfixe et suffixe.*

On vérifie facilement que la classe des morphismes non effaçants (resp. injectifs, resp. ω-injectifs, resp. préfixes, resp. suffixes, resp. bifixes) est stable par composition.

Proposition 1. *Soit $f \in \hom(\Lambda^\star, \Sigma^\star)$.*
Le morphisme f est injectif (resp. ω-injectif, resp. préfixe, resp. suffixe, resp. bifixe) si et seulement si la restriction de f à Λ est injective et $f(\Lambda)$ est un code (resp. un ω-code, resp. un code préfixe, resp. un code suffixe, resp. un code bifixe) sur Σ.

En particulier, tout morphisme préfixe est ω-injectif et tout morphisme suffixe est injectif. Par ailleurs, remarquons qu'un morphisme uniforme est bifixe si et seulement s'il est injectif.

2.4 Mots substitutifs

Soit $g \in \hom(\Sigma^\omega)$ et $a \in \Sigma$.

Supposons qu'il existe $x \in \Sigma^\star$ tel que $g(a) = ax$. On note alors $g^\omega(a) := axg(x)g^2(x)g^3(x)\ldots$ et si, de plus, le mot $g^\omega(a)$ est infini, alors on dit que g est *prolongeable* sur a.

On dit que $u \in \Sigma^\omega$ est *engendré* par g lorsqu'on peut écrire $u = g^\omega(u_0)$, u_0 désignant la première lettre de u : u est alors l'unique mot infini admettant $g^n(u_0)$ pour préfixe quel que soit $n \in \mathbb{N}$. De plus, u est *point fixe* de g, i.e., $g(u) = u$. Réciproquement, si g est non effaçant, $|g(u_0)| \geq 2$ et $g(u) = u$, alors g engendre u.

Un *mot purement substitutif* est un mot infini engendré par un endomorphisme. Un *mot substitutif* est l'image par un morphisme d'un mot purement substitutif. Il a été montré par PANSIOT [12] que les fonctions de complexité des mots purement substitutifs avaient un comportement asymptotique très contraint :

Théorème 3 (PANSIOT). *Pour tout mot purement substitutif u, on a $\mathtt{p}_u(n) = \Theta(\varphi(n))$ où $\varphi(n)$ est l'un des 5 fonctions suivantes : 1, n, $n \ln\ln n$, $n \ln n$ ou n^2.*

Au cours de la démonstration du théorème de PANSIOT [12, 5], on établit que tout mot infini u engendré par un endomorphisme uniforme vérifie $\mathtt{p}_u(n) = \mathrm{O}(n)$ résultat que l'on réutilisera plus loin.

Nous utiliserons également deux fois le mot de THUE-MORSE [14, 9] noté $t = \mathtt{01101001}\ldots$: $t = \mu^\omega(\mathtt{0})$ où μ désigne l'unique endomorphisme de $\{\mathtt{0},\mathtt{1}\}^\star$ tel que $\mu(\mathtt{0}) = \mathtt{01}$ et $\mu(\mathtt{1}) = \mathtt{10}$. Le mot THUE-MORSE a pour propriété de n'admettre aucun facteur cubique non vide, i.e., aucun facteur de la forme w^3 avec $w \in \{\mathtt{0},\mathtt{1}\}^+$.

3 Mots substitutifs contre mots purement substitutifs

Nous allons, dans cette section, illustrer le fait que la classe des mots substitutifs contient strictement la classe des mots purement substitutifs en construisant des mots substitutifs non purement substitutifs. Commençons par un résultat facile :

Lemme 1. *Soient $u \in \Sigma^\omega$ et une lettre $a \notin \Sigma$. Alors, le mot infini aau n'est pas purement substitutif.*

Preuve : Supposons (absurde) qu'il existe $f \in \hom((\Sigma \cup \{a\})^\star)$ engendrant aau. Alors, il existe $s \in (\Sigma \cup \{a\})^+$ tel que $f(a) = as$. Comme $aau = f(aau) = asasf(u)$, on a $au = sasf(u)$ donc la lettre a apparaît dans u à la position $|s| - 1$, ce qui ne se peut pas. ∎

Il s'en suit le résultat facile suivant :

Exemple 1. *Le mot infini $\mathtt{001}^\omega$ est substitutif mais pas purement substitutif.*

Preuve : En appliquant le lemme 1 avec $u = \mathtt{1}^\omega$ et $a = \mathtt{0}$, on obtient que le mot infini $\mathtt{001}^\omega$ est n'est pas purement substitutif.

D'autre part, on définit f, $g \in \hom(\{\mathtt{0},\mathtt{1}\}^\star)$ en posant $f(\mathtt{0}) := \mathtt{01}$, $f(\mathtt{1}) := \mathtt{1}$, $g(\mathtt{0}) := \mathtt{00}$ et $g(\mathtt{1}) = \mathtt{1}$: par construction, on a $f^\omega(\mathtt{0}) = \mathtt{01}^\omega$ et $\mathtt{001}^\omega = g(f^\omega(\mathtt{0}))$ donc $\mathtt{001}^\omega$ est substitutif. ∎

Comme on sait (voir [1]) que tout mot admettant pour suffixe un mot substitutif est également substitutif on peut généraliser l'exemple précédent :

Proposition 2. *Soient u un mot substitutif et une lettre a n'apparaissant pas dans u. Alors, le mot infini aau est substitutif mais pas purement substitutif.*

Mais cette généralisation n'est toujours pas satisfaisante car elle ne permet de construire qu'un seul mot binaire substitutif mais non purement substitutif qui est $\mathtt{001}^\omega$. Nous allons maintenant construire un mot substitutif binaire non ultimement périodique et non purement substitutif. Ce mot va être l'image du mot de THUE-MORSE par le morphisme qui remplace chaque lettre par son carré.

Exemple 2. *Soit δ l'unique endomorphisme de $\{0,1\}^\star$ tel que $\delta(0) = 00$ et $\delta(1) = 11$. Alors, le mot substitutif $\delta(t)$ n'est pas ultimement périodique et n'est pas purement substitutif.*

Preuve : Comme t est engendré par le morphisme μ, $\delta(t)$ est substitutif.

Par ailleurs, supposons (absurde) qu'il existe un suffixe de $\delta(t)$ admettant une période $p \geq 1$. Quitte à effacer sa première lettre, on peut supposer que ce suffixe est de la forme $\delta(t')$ où t' est suffixe de t et, quitte à remplacer p par $2p$, on peut supposer que p est pair. Comme on retrouve t' à partir de $\delta(t')$ en effaçant une lettre sur deux, t' admet pour période $p/2$ donc t est ultimement périodique : contradiction (t ne contient même pas de cube non vide !). On en déduit que $\delta(t)$ n'est pas ultimement périodique[1].

Il ne reste plus qu'à montrer que $\delta(t)$ n'est pas purement substitutif. Pour cela, on aura besoin de la propriété de synchronisation suivante facile à vérifier :

$$\forall x \in \{0,1\}^\star \ \forall y \in \{0,1\}^\infty \quad x01y \in \{00,11\}^\infty \iff \{x0, 1y\} \subseteq \{00,11\}^\infty \qquad (1)$$

Supposons (absurde) qu'il existe $f \in \hom(\{0,1\}^\star)$ engendrant $\delta(t)$.

On va tout d'abord vérifier que 001 est préfixe de $f(0)$. Pour cela, il suffit de voir que $f(0)$ et $0011 = \delta(01)$ sont deux préfixes de $\delta(t)$ et qu'il est impossible que $f(0)$ soit égal à ε (resp. à 0, resp. à 00) car sinon, $\delta(t) = f^\omega(0)$ serait égal à ε (resp. à 0, resp. à 0^ω) ce qui ne se peut pas. Ainsi, il existe $s \in \{0,1\}^\star$ tel que $f(0) = 001s$.

Posons $w := 01101$ et soit $v \in \{0,1\}^\omega$ tel que $t = w00v$. On a $\delta(t) = \delta(w)0000\delta(v)$ d'où :

$$\begin{aligned} \delta(t) &= f(\delta(t)) = f(\delta(w))f(0)f(0)f(0)f(0)f(\delta(v)) \\ &= f(\delta(w))00 \ \cdot 1s00 \cdot 1s00 \cdot 1s00 \cdot 1sf(\delta(v)) \end{aligned}$$

En utilisant plusieurs fois la propriété (1), on obtient que les mots $f(\delta(w))00$, $1s00$ et $1sf(\delta(v))$ isolés ci-dessus sont dans $\{00,11\}^\infty$.

Notant (abusivement) $\delta^{-1} : \{00,11\}^\infty \to \{0,1\}^\infty$ la réciproque de l'application bijective $\{0,1\}^\infty \to \{00,11\}^\infty$ induite par δ, on peut écrire :

$$\begin{aligned} t &= \delta^{-1}(\delta(t)) \\ &= \delta^{-1}(f(\delta(w))00)\ \delta^{-1}(1s00)\ \delta^{-1}(1s00)\ \delta^{-1}(1s00)\ \delta^{-1}(1sf(\delta(v))) \end{aligned}$$

On vient ainsi d'exhiber un facteur cubique non vide de t : contradiction. ∎

Note. Un autre exemple de mot substitutif mais non purement substitutif est le mot d'Aršon [2].

4 Autour d'une réduction de Cobham pour les mots substitutifs

En 1968, Cobham a obtenu le résultat suivant ([3], voir aussi [1]), retrouvé par Pansiot [11] :

[1] On pourrait aussi montrer que $\delta(t)$ n'est pas ultimement périodique en appliquant le corollaire 1 ci-dessous.

Théorème 4. *Soit s un mot substitutif sur Σ. Alors, s est l'image par un morphisme lettre à lettre d'un mot infini engendré par un endomorphisme non effaçant.*

Les preuves de Cobham et Pansiot utilisent des méthodes compliquées. Nous allons en donner une preuve *directe* inspirée par [4]. Pour cela, nous avons besoin de quelques lemmes techniques :

Lemme 2. *De toute suite d'entiers naturels, on peut extraire une sous-suite constante ou une sous-suite strictement croissante.*

Preuve : Soit $\nu \in \mathbb{N}^{\mathbb{N}}$.

Supposons tout d'abord que ν soit majorée. Alors, $\nu(n)$ ne prend qu'un nombre fini de valeurs distinctes lorsque n décrit $\mathbb{N}$. Par suite, il existe $m \in \mathbb{N}$ tel que $\nu^{-1}(\{m\})$ soit infini. On peut alors extraire de ν une sous-suite constante égale à m.

Supposons maintenant que ν ne soit pas majorée. Alors, pour tout $p \in \mathbb{N}$, $\max\{\nu(0), \nu(1), \ldots, \nu(p)\}$ ne peut pas majorer ν donc il existe un entier $q > p$ tel que $\nu(q) > \nu(p)$. Ceci permet de construire par récurrence une sous-suite strictement croissante de ν. ∎

Généralisons le lemme précédent :

Lemme 3. *Soient un entier $d \geq 1$ et $\nu_1, \nu_2, \ldots, \nu_d \in \mathbb{N}^{\mathbb{N}}$. Alors, il existe $\varphi \in \mathbb{N}^{\mathbb{N}}$ strictement croissante telle que pour tout $i \in [1, d]$, $\nu_i \circ \varphi$ soit strictement croissante ou constante.*

Preuve : On procéde par récurrence sur d.

Si $d = 1$ alors le lemme 2 fournit le résultat donc la récurrence s'initialise bien et on peut supposer désormais $d \geq 2$.

En appliquant l'hypothèse de récurrence aux suites $\nu_1, \nu_2, \ldots, \nu_{d-1}$ on récolte $\psi \in \mathbb{N}^{\mathbb{N}}$ strictement croissante telle que pour tout $i \in [1, d-1]$, $\nu_i \circ \psi$ soit strictement croissante ou constante.

D'autre part, en appliquant le lemme 2, on peut extraire de $\nu_d \circ \psi$ une sous-suite strictement croissante ou constante : il existe $\psi' \in \mathbb{N}^{\mathbb{N}}$ strictement croissante telle que $\nu_d \circ \psi \circ \psi'$ soit strictement croissante ou constante.

Posant $\varphi := \psi \circ \psi'$ on obtient ce qu'on voulait. ∎

Il faut remarquer que dans le lemme précédent la "fonction d'extraction" φ est *la même* pour tous les ν_i.

Lemme 4. *Soit s un mot substitutif sur Σ. Alors, il existe un alphabet Λ, $u_0 \in \Lambda$, $f \in \hom(\Lambda^\star)$ prolongeable sur u_0 et $f' \in \hom(\Lambda^\star, \Sigma^\star)$ tels que l'on ait $s = f'(f^\omega(u_0))$, $|f'(f(u_0))| > |f'(u_0)| > 0$ et $|f'(f(a))| \geq |f'(a)|$ pour tout $a \in \Lambda$.*

Preuve : Comme s est substitutif, il existe un alphabet Λ, $u_0 \in \Lambda$, $g' \in \hom(\Lambda^\star)$ prolongeable sur u_0 et $g \in \hom(\Lambda^\star, \Sigma^\star)$ tels que $s = g'(g^\omega(u_0))$,

Pour tout $a \in \Lambda$, on définit :

$$\begin{array}{rccc} \nu_a : & \mathbb{N} & \longrightarrow & \mathbb{N} \\ & n & \longmapsto & |g'(g^n(a))| \end{array}$$

et on applique le lemme 3 à la famille (finie) $(\nu_a)_{a\in\Lambda}$. On obtient alors l'existence de $\varphi \in \mathbb{N}^{\mathbb{N}}$ strictement croissante telle que pour tout $a \in \Lambda$, $\nu_a \circ \varphi$ soit strictement croissante ou constante.

Posons $p := \varphi(1)$ et $q := \varphi(2) - \varphi(1)$, $f := g^q$ ($\in \hom(\Lambda^\star)$) et $f' := g' \circ g^p$ ($\in \hom(\Lambda^\star, \Sigma^\star)$). Par construction, f est prolongeable sur u_0 et engendre $g^\omega(u_0)$ (remarquer que $\varphi(2) > \varphi(1)$ force $q \geq 1$) et par suite $s = f'(f^\omega(u_0))$.

De plus, pour tout $a \in \Sigma$, $\nu_a \circ \varphi$ est croissante donc on a :

$$|f'(f(a))| = \nu_a(q+p) = \nu_a(\varphi(2)) \geq \nu_a(\varphi(1)) = |f'(a)|$$

Enfin, comme le mot $g'(g^\omega(u_0)) = s$ est infini, on a $\lim_{n\to\infty} \nu_{u_0}(n) = \infty$ donc également $\lim_{n\to\infty}(\nu_{u_0}\circ\varphi)(n) = \infty$. On en déduit que la suite $\nu_{u_0}\circ\varphi$ ne peut pas être constante donc elle est strictement croissante. Par suite, on a les inégalités strictes $\nu_{u_0}(\varphi(2)) > \nu_{u_0}(\varphi(1)) > \nu_{u_0}(\varphi(0))$ ou encore $|f'(f(u_0))| > |f'(u_0)| > \nu_{u_0}(\varphi(0)) \geq 0$. ∎

Nous sommes maintenant en mesure de démontrer le théorème 4.

Preuve (du théorème 4) : Comme s est substitutif, il existe un alphabet Λ, $u_0 \in \Lambda$, $f \in \hom(\Lambda^\star)$ prolongeable sur u_0 et $f' \in \hom(\Lambda^\star, \Sigma^\star)$ tels que $s = f'(f^\omega(u_0))$. Par le lemme 4, on peut supposer que, de plus, on a :

$$|f'(f(u_0))| > |f'(u_0)| > 0 \tag{2}$$

$$\forall a \in \Lambda \qquad |f'(f(a))| \geq |f'(a)| \tag{3}$$

On pose alors $u := f^\omega(u_0)$ et $\Pi := \{(a,j) \in \Lambda \times \mathbb{N} : j < |f'(a)|\}$ (remarquer que Π est fini). On définit ensuite $\alpha \in \hom(\Lambda^\star, \Pi^\star)$ et $g' \in \hom(\Pi^\star, \Sigma^\star)$ en posant :

- pour tout $a \in \Lambda$, $\alpha(a) := (a,0)(a,1)\dots(a,|f'(a)|-1)$ avec la convention que $\alpha(a) = \varepsilon$ si $|f'(a)| = 0$,
- pour tout $(a,j) \in \Pi$, $g'(a,j) :=$ la lettre apparaissant à la position j dans $f'(a)$.

Par construction g' est lettre à lettre et on a $g' \circ \alpha = f'$. Ainsi, posant $v := \alpha(u)$ on a $g'(v) = f'(u) = s$ donc il ne nous reste plus qu'à construire $g \in \hom(\Pi^\star)$ non effaçant et engendrant v.

On a $|\alpha(u_0)| = |f'(u_0)| > 0$ d'après (2) et par suite la première lettre de v est la première lettre de son préfixe $\alpha(u_0)$ c'est à dire $v_0 := (u_0, 0)$.

D'autre part, on a aussi $|\alpha(w)| = |g'(\alpha(w))| = |f'(w)|$ pour tout $w \in \Lambda^\star$ donc pour tout $a \in \Lambda$, on a $|\alpha(f(a))| = |f'(f(a))| \geq |f'(a)|$ à cause de (3) et $|\alpha(f(u_0))| > |f'(u_0)|$ à cause de (2). Par suite, pour tout $a \in \Lambda$ tel que $f(a) \neq \varepsilon$, on peut factoriser $\alpha(f(a))$ sous la forme :

$$\alpha(f(a)) = w_{a,0}w_{a,1}\dots w_{a,|f'(a)|-1}$$

avec $w_{a,0}, w_{a,1}, \dots, w_{a,|f'(a)|-1} \in \Pi^+$ et $|w_{u_0,0}| \geq 2$. On peut alors définir $g \in \hom(\Pi^\star)$ en posant $g(a,j) := w_{a,j}$ pour tout $(a,j) \in \Pi$. Ainsi, g est non effaçant, $g(v_0) = w_{u_0,0}$ est de longueur au moins 2 et :

$$\forall a \in \Lambda \quad \alpha(f(a)) = g(a,0)g(a,1)\dots g(a,|f'(a)|-1) = g(\alpha(a))$$

donc $\alpha \circ f = g \circ \alpha$. Comme $f(u) = u$, il vient $g(v) = g(\alpha(u)) = \alpha(f(u)) = \alpha(u) = v$: v est bien engendré par g et $s = g'(v) = g'(g^\omega(v_0))$. ∎

Le thérorème précédent dit que : "toute image par un morphisme d'un mot engendré par un endomorphisme est l'image par un morphisme lettre à lettre d'un mot engendré par un endomorphisme non effaçant". On est alors tenté de se débarrasser des deux occurences du facteur "image par un morphisme" dans l'énoncé précédent et d'écrire : tout mot engendré par un endomorphisme est engendré par un endomorphisme non effaçant. Or, ce second énoncé est faux ! En voici un contre-exemple :

Exemple 3. *Soit f l'unique endomorphisme de $\{0,1,2\}^\star$ tel que $f(0) = 01222$, $f(1) = 10222$ et $f(2) = \varepsilon$. Alors, f est prolongeable sur 0 et le mot infini $u := f^\omega(0)$ n'est engendré par aucun endomorphisme non effaçant de $\{0,1,2\}^\star$.*

Preuve : Pour pouvoir démontrer (par l'absurde) notre résultat nous avons besoin de quelques remarques préliminaires.

Dans un premier temps, on calcule que :

$$f^3(0) = \underline{01}\ \underline{222}\ \underline{10}\ 222\ 10\ 2220\ \underline{1222}$$

donc les mots soulignés ci-dessus sont facteurs de u.

Ensuite, soit $\chi \in \hom(\{0,1,2\}^\star, \{0,1\}^\star)$ donné par $\chi(0) := 0$, $\chi(1) := 1$ et $\chi(2) := \varepsilon$: χ est le morphisme qui efface les 2. On vérifie facilement que $\mu \circ \chi = \chi \circ f$ (il suffit de vérifier que ces deux morphismes coïncident sur $\{0,1,2\}$) d'où $\mu(\chi(u)) = \chi(f(u)) = \chi(u)$. Or, $\chi(u)$ commence par 0 et le mot de Thue-Morse t est le seul point fixe de μ commençant par 0. On en déduit :

$$t = \chi(u)$$

D'autre part, comme $u = f(u)$, u s'écrit comme une concaténation (infinie) de 01222 et de 10222 donc :

$$\mathsf{F}(u) \cap \{0,1\}^\star = \{\varepsilon, 0, 1, 01, 10\} \quad (4)$$

$$\mathsf{F}(u) \cap \{2\}^\star = \{\varepsilon, 2, 22, 222\} \quad (5)$$

Supposons maintenant (absurde) qu'il existe $g \in \hom(\{0,1,2\}^\star)$ non effaçant engendrant u.

L'image par g de tout facteur (resp. préfixe) u est alors facteur (resp. préfixe) de u.

En particulier $g(222) = g(2)^3$ est facteur de u donc $\chi(g(2))^3 = \chi\left(g(2)^3\right)$ est facteur de $\chi(u) = t$ qui est sans cube non vide. Par suite, on a nécessairement $\chi(g(2)) = \varepsilon$ ou encore $g(2) \in \{2\}^\star$. Ainsi, (5) garantit que $g(2)^3$ est dans $\{\varepsilon, 2, 22, 222\}$. Comme ε et 222 sont les seuls cubes de ce dernier langage [2] on obtient que $g(2)^3 = 222$ (car $g(2) \neq \varepsilon$ par hypothèse) ou encore :

$$g(2) = 2$$

Comme $g(0)$ est un préfixe de u de longueur au moins deux et comme 01 est le préfixe de u de longueur deux, il existe $x \in \{0,1\}^\star$ tel que :

$$g(0) = 01x$$

[2] Ce sont même les seuls cubes facteurs de u tout court...

Comme $g(\mathtt{1222}) = g(\mathtt{1})\mathtt{222}$ est facteur de u mais pas $\mathtt{2222}$ (à cause de (5)), la dernière lettre de $g(\mathtt{1})$ (qui existe car $g(\mathtt{1}) \neq \varepsilon$) n'est pas un $\mathtt{2}$. On en déduit qu'il existe $a \in \{\mathtt{0},\mathtt{1}\}$ et $y \in \{\mathtt{0},\mathtt{1},\mathtt{2}\}^\star$ tels que :

$$g(\mathtt{1}) = ya$$

Enfin, $g(\mathtt{10}) = ya\mathtt{01}x$ est facteur de u donc il en est de même pour $a\mathtt{01}$: contradiction avec (4). ∎

Note. Un autre contre-exemple, également basé sur le mot de Thue-Morse, est donné par Pansiot [11, Propriété 2.12].

5 Une réduction pour les mots purement substitutifs

Nous allons dans cette section renforcer le théorème 4 dans le cas particulier où le mot substitutif s de l'énoncé est purement substitutif :

Théorème 5. *Tout mot purement substitutif est l'image par un morphisme bifixe d'un mot infini engendré par un endomorphisme ω-injectif*[3]*.*

Pour cela, nous allons avoir besoin de deux lemmes techniques permettant de renforcer légèrement le théorème de défaut (théorème 2).

Lemme 5. *Soit Y un langage fini sur Σ. Alors, il existe un code bifixe Z sur Σ tel que $Y^\star \subseteq Z^\star$ et $\#Z \leq \#Y$.*

Preuve : Soit $\mathcal{L}$ l'ensemble des langages finis $L \subseteq \Sigma^+$ vérifiant $X^\star \subseteq L^\star$ et $\#L \leq \#Y$.

Par construction $\mathcal{L}$ est non vide car $Y \setminus \{\varepsilon\} \in \mathcal{L}$ donc il existe $Z \in \mathcal{L}$ tel que :

$$\sum_{z\in Z} |z| = \min_{L\in\mathcal{L}} \sum_{w\in L} |w|$$

Supposons (absurde) que Z ne soit pas un code préfixe (resp. pas suffixe). Alors, il existe z_1, $z_2 \in Z$ et $z' \in \Sigma^+$ tels que $z_2 = z_1 z'$ (resp. $z_2 = z' z_1$). Posant $Z' := (Z \setminus \{z_2\}) \cup \{z'\}$, on a, par construction, $Z' \in \mathcal{L}$ et $|z_2| > |z'|$ d'où :

$$\sum_{z'\in Z'} |z'| \overset{(*)}{\leq} \sum_{z\in Z} |z| - |z_2| + |z'| < \sum_{z\in Z} |z|$$

ce qui ne se peut pas (remarquer que l'on ne peut pas remplacer l'inégalité $(*)$ par une égalité car rien ne garantit que la réunion $(Z \setminus \{z_2\}) \cup \{z'\}$ soit disjointe). ∎

On déduit du lemme précédent et du théorème de défaut (théorème 2) :

Proposition 3 (Théorème de défaut renforcé). *Soit X un langage fini sur Σ qui n'est pas un ω-code. Alors, il existe un code bifixe Z sur Σ tel que $X^\star \subseteq Z^\star$ et $\#Z < \#X$.*

[3] En fait, en adaptant légèrement la démonstration donnée ci-dessous, on pourrait même supposer que l'endomorphisme itéré est *élémentaire* ce qui est plus fort que ω-injectif (voir [13]).

Lemme 6. *Soit $f \in \hom(\Lambda^\star, \Sigma^\star)$ un morphisme qui n'est* pas *ω-injectif. Alors, il existe un code bifixe Z sur Σ tel que $\#Z < \#\Lambda$ et $f(\Lambda) \subseteq Z^\star$.*

Preuve : Par la proposition 1, il se présente deux cas : soit la restiction de f à Λ n'est pas injective, soit $f(\Lambda)$ n'est pas un ω-code.

Supposons tout d'abord que la restriction de f à Λ ne soit pas injective. Alors, on a $\#f(\Lambda) < \#\Lambda$ et on peut conclure en appliquant le lemme 5 où on a remplacé Y par $f(\Lambda)$.

Supposons maintenant que $f(\Lambda)$ ne soit pas un ω-code. La proposition 3 appliquée en remplaçant X par $f(\Lambda)$ garantit l'existence d'un code bifixe Z sur Σ tel que $\#Z < \#f(\Lambda)$ et $f(\Lambda) \subseteq Z^\star$. Comme, trivialement, on a $\#f(\Lambda) \leq \#\Lambda$, tout va bien. ∎

Nous sommes maintenant en mesure de démontrer le théorème 5.

Preuve (du théorème 5) : On va montrer par récurrence sur $\#\Sigma$ que tout mot purement substitutif sur l'alphabet Σ, est l'image par un morphisme bifixe d'un mot engendré par un endomorphisme ω-injectif.

Si Σ était vide alors il n'existerait pas de mot infini sur Σ. Par suite, le cas $\#\Sigma = 0$ ne se présente pas et donc la récurrence s'initialise trivialement.

Soit $f \in \hom(\Sigma^\star)$ engendrant u. Si f est ω-injectif, on termine en disant que u est l'image de u par la substitution induite par l'identité sur $\Sigma^\star$ qui est bien un morphisme bifixe.

On peut donc désormais se placer dans le cas où f n'est pas ω-injectif.

Par le lemme 6, il existe un code bifixe Z sur Σ tel que $\#Z < \#\Sigma$ et $f(\Sigma) \subseteq Z^\star$. Soient alors $\tilde{\Sigma}$ un alphabet de même cardinal que Z et $\tilde{h} \in \hom(\tilde{\Sigma}^\star, Z^\star)$ induisant une bijection $\tilde{\Sigma} \to Z$. Par la proposition 1, $\tilde{h}$ *est bifixe* donc en particulier (ω-)injectif. Ainsi, le morphisme $\tilde{h}$:

– est bijectif et $\tilde{h}^{-1} \in \hom(Z^\star, \tilde{\Sigma}^\star)$,

– induit une substitution bijective $\tilde{\Sigma}^\infty \to Z^\infty$ dont l'inverse prolonge $\tilde{h}^{-1}$.

Comme $u = f(u) \in f(\Sigma)^\omega \subseteq Z^\omega$ on peut poser $\tilde{u} := \tilde{h}^{-1}(u)$ $(\in \tilde{\Sigma}^\omega)$ et comme $f(Z^\star) \subseteq f(\Sigma^\star) = f(\Sigma)^\star \subseteq Z^\star$, on peut poser $\tilde{f} := \tilde{h}^{-1} \circ f \circ \tilde{h}$ $(\in \hom(\tilde{\Sigma}^\star))$.

Montrons que $\tilde{u}$ est engendré par $\tilde{f}$.

Notant $\tilde{u}_0$ la première lettre de $\tilde{u}$, u_0 la première lettre de u et $m := \max_{z \in Z} |z|$, on a pour tout $n \in \mathbb{N}$:

$$\left|\tilde{f}^n(\tilde{u}_0)\right| \overset{(i)}{=} \left|\tilde{h}^{-1}\left(f^n\left(\tilde{h}(\tilde{u}_0)\right)\right)\right| \overset{(ii)}{\geq} \frac{1}{m}\left|f^n\left(\tilde{h}(\tilde{u}_0)\right)\right| \overset{(iii)}{\geq} \frac{1}{m}\left|f^n(u_0)\right| \tag{6}$$

En effet, on a $\tilde{f}^n = \tilde{h}^{-1} \circ f^n \circ \tilde{h}$ (récurrence triviale sur n) ce qui justifie (i) et on montre facilement que pour tout $w \in Z^\star$, on a $\left|\tilde{h}^{-1}(w)\right| \geq \frac{1}{m}|w|$ ce qui justifie (ii). De plus, $\tilde{h}(\tilde{u}_0)$ est un préfixe de $\tilde{h}(\tilde{u}) = u$ et $\tilde{h}(\tilde{u}_0)$ est non vide puisque appartenant au code Z. Par suite, la première lettre de $\tilde{h}(\tilde{u}_0)$ est la même que la première lettre de u à savoir u_0. On en déduit que $f^n(u_0)$ est préfixe de $f^n(h(\tilde{u}_0))$ d'où (iii).

Lorsque $n \to \infty$, on a $|f^n(u_0)| \to \infty$ donc, par (6), on a également $\left|\tilde{f}^n(\tilde{u}_0)\right| \to \infty$. Comme de plus :

$$\tilde{f}(\tilde{u}) = \tilde{h}^{-1}(f(\tilde{h}(\tilde{u}))) = \tilde{h}^{-1}(f(u)) = \tilde{h}^{-1}(u) = \tilde{u}$$

on a bien que $\tilde{u}$ est engendré par $\tilde{f}$.

Comme $\tilde{u} \in \tilde{\Sigma}^\omega$ est purement substitutif et comme $\#\tilde{\Sigma} = \#Z < \#\Sigma$, on peut appliquer l'hypothèse de récurrence à $\tilde{u}$: il existe un alphabet Λ, un mot infini $v \in \Lambda^\omega$ engendré par un endomorphisme ω-injectif de $\Lambda^\star$ et un morphisme bifixe $h' \in \hom(\Lambda^\star, \tilde{\Sigma}^\star)$ tels que $h'(v) = \tilde{u}$. Posant $h := \tilde{h} \circ h'$, on a $h(v) = \tilde{h}(\tilde{u}) = u$ et h est bifixe comme composée de morphismes bifixes. ∎

6 Complexité des images morphiques de mots infinis

Étant donnés un mot infini u sur Λ et un morphisme $f \in \hom(\Lambda^\star, \Sigma^\star)$ tel que $f(u)$ soit infini, nous cherchons, dans cette section, à comparer p_u et $\mathsf{p}_{f(u)}$. Suivant que f est non effaçant, injectif ou ω-injectif nous pouvons établir trois inégalités. Posons $M := \max_{a\in\Lambda} |f(a)|$.

6.1 Cas où f est non effaçant

L'inégalité suivante permet de majorer $\mathsf{p}_{f(u)}$ en fontion de p_u :

Proposition 4. *Soient $u \in \Lambda^\omega$, $f \in \hom(\Lambda^\star, \Sigma^\star)$ et $M := \max_{a\in\Lambda} |f(a)|$.*
Si f est non effaçant, alors, on a :

$$\forall n \in \mathbb{N} \qquad \mathsf{p}_{f(u)}(n) \leq M\, \mathsf{p}_u(n)$$

Preuve : Soit $w' \in \mathsf{F}_n(f(u))$.

Soit $q \in \mathbb{N}$ tel qu'une occurence de w' apparaisse dans $f(u)$ à la position q. Soit x le plus long préfixe de de u tel que $|f(x)| \leq q$. Soit $w \in \mathsf{F}_n(u)$ tel que xw soit préfixe de u. Soient $a \in \Lambda$ et $s \in \Lambda^{n-1}$ tels que $w = as$.

Comme $xas = xw$ est préfixe de u, il en est de même pour xa donc par maximalité de $|x|$, on a $|f(xa)| \geq q + 1$. Comme f est non effaçant, on a aussi $|f(s)| \geq n - 1$ et par suite, il vient :

$$|f(xw)| = |f(xa)| + |f(s)| \geq (q + 1) + (n - 1) = q + n$$

Comme $f(x)f(w) = f(xw)$ est préfixe de $f(u)$, $f(w)$ est le facteur de $f(u)$ débutant à la position $|f(w)|$ et se terminant à la position $|f(xw)| - 1$. Comme on a les inégalités :

$$|f(x)| \leq q \leq q + n - 1 \leq |f(xw)| - 1$$

il en résulte que $f(w)$ contient le facteur de $f(u)$ débutant à la position q et se terminant à la position $q + n - 1$ c'est à dire $w' : w' \in \mathsf{F}_n(f(w))$. Plus précisément, l'encadrement $|f(x)| \leq q < |f(xa)| = |f(x)| + |f(a)|$ garantit qu'une occurence de w' dans $f(w)$ débute à une position strictement inférieure à $|f(a)|$ donc strictement inférieure à M.

Posant $D := \{(k, w) \in [0, M - 1] \times \mathsf{F}_n(u) : |f(w)| \geq n + k\}$ on a montré que l'image de l'application :

$$\begin{array}{ccl} D & \longrightarrow & \Sigma^n \\ (k, w) & \longmapsto & \text{le facteur de } f(w) \text{ de longueur } n \text{ débutant à la position } k \end{array}$$

contenait $\mathsf{F}_n(f(u))$ et comme D est trivialement de cardinal inférieur ou égal à $M\, \mathsf{p}_u(n)$ notre proposition est démontrée. ∎

Dans la proposition 4, f doit être supposé non effaçant. Nous allons en effet montrer que dans le cas contraire, il se peut que la complexité de $f(u)$ soit d'un ordre de grandeur supérieur à la complexité de u. Pour cela, nous avons besoin de la "réduction" suivante [6] :

Proposition 5. *Tout mot purement substitutif est l'image par un morphisme effaçant d'un mot engendré par un morphisme uniforme.*

Preuve : Soient $v \in \Sigma^\omega$ un mot purement substitutif, $g \in \hom(\Sigma^\omega)$ engendrant v, v_0 la première lettre de v, b une lettre n'appartenant pas à Σ, $\Lambda := \Sigma \cup \{b\}$ et $L := \max_{a\in\Sigma} |g(a)|$.

On définit $h \in \hom(\Lambda^\star)$ et $f \in \hom(\Lambda^\star, \Sigma^\star)$ en posant : $h(a) := g(a)b^{L-|g(a)|}$ et $f(a) := a$ pour tout $a \in \Sigma$ et $g(b) := b^L$ et $f(b) := \varepsilon$. Par construction, f est effaçant, h est uniforme et $h(v_0)$ admet $g(v_0)$ pour préfixe donc $h(v_0)$ admet v_0 comme première lettre. Ceci permet de poser $u := h^\omega(v_0)$.

Or, on vérifie facilement que $f\circ h = g\circ f$ donc, pour tout $n \in \mathbb{N}$, on a $f\circ h^n = g^n\circ f$ puis $f(h^n(v_0)) = g^n(v_0)$ et enfin $f(u) = v$. ∎

Utilisant le fait que les mots engendrés par des endomorphismes uniformes ont des complexités en $\mathrm{O}(n)$ (voir [5]) et le fait qu'il existe des mots purement substitutifs dont les complexités sont en $\Omega(n \ln\ln n)$ (on trouve dans [12] des critères assez pratiques permettant de construire de tels mots) on déduit de la réduction précédente :

Proposition 6. *Il existe un mot purement substitutif u et un morphisme effaçant f tels que $\mathsf{p}_u(n) = \mathrm{O}(n)$ et $\mathsf{p}_{f(u)}(n) = \Omega(n \ln\ln n)$.*

6.2 Cas où f est injectif et ω-injectif

Nous donnons dans cette section deux inégalités permettant de majorer p_u en fonction de $\mathsf{p}_{f(u)}$.

Proposition 7. *Soient $u \in \Lambda^\omega$, $f \in \hom(\Lambda^\star, \Sigma^\star)$ et $M := \max_{a\in\Lambda} |f(a)|$.*

Si f est injectif alors on a :

$$\forall n \in \mathbb{N} \qquad \mathsf{p}_u(n) \le M\, \mathsf{p}_{f(u)}(Mn+1)$$

Preuve : Soit $w \in \mathsf{F}_n(u)$, alors $f(w) \in \mathsf{F}(f(u))$ et $|f(w)| \le Mn$. Comme u est un mot infini à droite, w est préfixe de facteurs de u arbitrairement longs dont les images par f sont arbitrairement longues et, par suite, il existe $e(w) \in \mathsf{F}(u)$ que l'on va supposer de longueur minimale vérifiant : w est un préfixe *propre* de $e(w)$ et $|f(e(w))| \ge Mn+1$. Soient $\lambda(w) \in \Lambda$ et $p(w) \in \Lambda^\star$ tels que $e(w) = p(w)\lambda(w)$. Par minimalité de $|e(w)|$, on a $|f(p(w))| < Mn+1 \le |f(e(w))| = |f(p(w))f(\lambda(w))|$ donc il existe un préfixe non vide $l(w)$ de $f(\lambda(w))$ tel que :

$$|f(p(w))l(w)| = Mn+1$$

Ainsi, on a :

$$1 \le |l(w)| \le |f(\lambda(w))| \le M$$

On peut alors définir de manière cohérente la fonction :

$$\begin{array}{rccc} \phi: & \mathsf{F}_n(u) & \longrightarrow & [1,M]\times \mathsf{F}_{Mn+1}(f(u)) \\ & w & \longmapsto & (|l(w)|\,, f(p(w))l(w)) \end{array}$$

et pour démontrer l'inégalité souhaitée, il suffit de montrer que ϕ est injective.

Soient x, $y \in \mathsf{F}_n(u)$ tels que $\phi(x) = \phi(y)$.

Alors, $l(x)$ et $l(y)$ sont de même longueur et suffixes respectivement de $f(p(x))l(x)$ et $f(p(y))l(y)$ qui sont égaux. Par suite, on a $l(x) = l(y)$ puis $f(p(x)) = f(p(y))$ et, comme f est injectif, il vient $p(x) = p(y)$. Or, x et y sont de même longueur et sont préfixes propres respectivement de $e(x)$ et $e(y)$ donc préfixes respectivement de $p(x)$ et $p(y)$. On en déduit que $x = y$, ce qu'on voulait. ∎

Supposant f injectif, on a, pour tout entier $n \geq 1$:

$$\mathsf{p}_{f(u)}(n) = \mathsf{p}_{f(u)}\left(M\frac{n-1}{M}+1\right) \geq \mathsf{p}_{f(u)}\left(M\left\lfloor\frac{n-1}{M}\right\rfloor+1\right) \geq \frac{1}{M}\,\mathsf{p}_u\left(\left\lfloor\frac{n-1}{M}\right\rfloor\right)$$

donc on tire de la proposition 7 un minoration de $\mathsf{p}_{f(u)}$ en fonction de p_u.

En appliquant le théorème de Pansiot (théorème 3), l'inégalité ci-dessus et la proposition 4, on démontre facilement :

Proposition 8. *Soient $u \in \Lambda^\omega$ et $f \in \hom(\Lambda^\star, \Sigma^\star)$.*

Si u est purement substitutif et si f est injectif alors on a $\mathsf{p}_{f(u)} = \Theta(\mathsf{p}_u)$.

On déduit de la proposition 7 et du théorème 1 (Morse-Hedlund) une caractérisation des morphismes injectifs.

Corollaire 1. *Soit $f \in \hom(\Lambda^\star, \Sigma^\star)$ tel que $f(\Lambda^\star) \neq \{\varepsilon\}$.*

Alors, f est injectif si et seulement si pour tout $u \in \Lambda^\omega$ non ultimement périodique et tel que $f(u) \in \Sigma^\omega$, $f(u)$ est non ultimement périodique.

Preuve : ($\Rightarrow$) Supposons que f soit injectif.

Soit $u \in \Lambda^\omega$ non ultimement périodique (f est en particulier non effaçant donc, nécessairement, $f(u) \in \Sigma^\omega$). Le théorème 1 garantit que la fonction p_u est non bornée et la proposition 7 permet d'en déduire que $\mathsf{p}_{f(u)}$ n'est pas bornée non plus. En appliquant une nouvelle fois le théorème 1, on obtient que $f(u)$ n'est pas ultimement périodique : ce qu'on voulait.

($\Leftarrow$) Réciproquement, supposons que f soit non injectif et construisons un mot infini $u \in \Lambda^\omega$ non ultimement périodique tel que que $f(u)$ soit infini et (ultimement) périodique.

Par hypothèse, il existe x, $y \in \Lambda^\star$ vérifiant $x \neq y$ et $f(x) = f(y)$. De plus, comme on a supposé $f(\Lambda^\star) \neq \{\varepsilon\}$, il existe $a \in \Lambda$ tel que $f(a) \neq \varepsilon$. Quitte à remplacer x et y respectivement par ax et ay, on peut supposer que $f(x) = f(y)$ est non vide.

Soit alors $v \in \{\mathtt{0},\mathtt{1}\}^\omega$ un mot non ultimement périodique, $g \in \hom(\{\mathtt{0},\mathtt{1}\}^\star, \Lambda^\star)$ donné par $g(\mathtt{0}) := x$ et $g(\mathtt{1}) := y$ et $u := g(v)$.

Par construction, on a $u \in \{x,y\}^\omega$ donc $f(u) = f\,(x)^\omega$: ainsi, $f(u)$ est infini (car $f(x)$ est non vide) et périodique.

Si l'on arrive à montrer que g est injectif alors la proposition 7 garantira que u n'est pas ultimement périodique et on aura fini. Comme g est, par construction,

injectif sur $\{\mathtt{0},\mathtt{1}\}$, il ne reste plus qu'à vérifier que $g(\{\mathtt{0},\mathtt{1}\}) = \{x, y\}$ est un code (proposition 1).

Supposons (absurde) que $\{x, y\}$ ne soit pas un code. Alors, le théorème de défaut (théorème 2) garantit l'existence de $z \in \Lambda^\star$ et de $m, n \in \mathbb{N}$ tels que $x = z^m$ et $y = z^n$. On a ainsi $f(z)^m = f(z^m) = f(x) = f(y) = f(z)^n$ donc en prenant les longueurs, il vient $m\,|f(z)| = n\,|f(z)|$ et les deux membres de cette dernière égalité sont égaux à $|f(x)|$ qui est non nul. On peut par conséquent les simplifier par $|f(z)|$ pour obtenir $m = n$ ce qui garantit que $x = y$: contradiction. ∎

Si l'on renforce l'hypothèse d'injectivité faite sur f on obtient une nouvelle majoration, plus fine si $\mathsf{p}_{f(u)}$ croît lentement :

Proposition 9. *Soient $u \in \Lambda^\omega$, $f \in \hom(\Lambda^\star, \Sigma^\star)$ et $M := \max_{a\in\Lambda} |f(a)|$.*

Si f est ω-injectif alors on a :

$$\forall n \in \mathbb{N} \qquad \mathsf{p}_u(n) \leq \mathsf{p}_{f(u)}(M(n+d))$$

où d est le délai de déchiffrement de $f(\Lambda)$.

Preuve : Soit $w \in \mathsf{F}_n(u)$. Comme u est un mot infini à droite, w est préfixe de facteurs de u arbitrairement longs donc, en particulier, il existe $e(w) \in \mathsf{F}_{n+d}(u)$ tel que w soit préfixe de $e(w)$. Alors, $f(e(w))$ est facteur de $f(u)$ et on a $|f(e(w))| \leq M\,|e(w)| = M(n+d)$. Comme $f(u)$ est un mot infini à droite, $f(e(w))$ est préfixe de facteurs de $f(u)$ arbitrairement longs donc il existe $\phi(w) \in \mathsf{F}_{M(n+d)}(f(u))$ tel que $f(e(w))$ soit préfixe de $\phi(w)$.

Il ne reste plus qu'à montrer que ϕ est injective.

Soient x, $y \in \mathsf{F}_n(w)$ tels que $\phi(x) = \phi(y)$.

Alors, $f(e(x))$ et $f(e(y))$ sont tous les deux préfixes d'un même mot donc on peut supposer, par exemple, que $f(e(x))$ est préfixe de $f(e(y))$. En particulier, $f(x) \in f(\Lambda)^n$ est préfixe de $f(e(y)) \in f(\Lambda)^{n+d}$. Comme $f(\Lambda)$ est un ω-code admettant d pour délai de décodage, il vient $f(x_i) = f(y_i)$ pour tout $i \in [0, n-1]$ où x_i (resp. y_i) désigne la lettre de x (resp. y) apparaissant à la position i. Comme f est injective sur Λ, on obtient $x = y$: ce qu'on voulait. ∎

Supposons que f soit ω-injectif. Commme précédemment, on déduit de la proposition 9 que pour tout entier $n \geq Md$ on a :

$$\mathsf{p}_{f(u)}(n) \geq \mathsf{p}_u\left(\left\lfloor \frac{n}{M} \right\rfloor - d\right)$$

7 Conclusion

Nous avons dans cet article donné quelques résultats concernant les mots substitutifs et justifiant de l'intérêt de leur étude. Néanmoins une question importante reste à traiter qui est la classification des comportements asymptotiques des fonctions de complexité des mots substitutifs. Un résultat surprenant est que le théorème de Pansiot est faux si l'on essaye de l'étendre tel quel aux mots substitutifs quelconques. En effet, pour tout entier $k \geq 1$, on peut construire un mot substitutif dont la complexité est $\Theta(n^{1+1/k})$ ce que nous expliciterons dans un prochain article.

Remerciements

Les auteurs remercient Eric Rivals pour avoir relu une version préliminaire de cet article et Jean-Paul Allouche pour nous avoir fourni des références utiles.

Références

[1] J.-P. Allouche et J. Shallit, *Automatic Sequences. Theory, Applications, Generalizations*, à paraître.

[2] J. Berstel, Mots sans carré et morphismes itérés, *Discrete Math.* **29** (1979), 235–244.

[3] A. Cobham, On the Hartmanis-Stearns problem for a class of tag machines, *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, 1968.

[4] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* **179** (1998), 89–101.

[5] A. Ehrenfeucht, K. P. Lee et G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interaction, *Theoret. Comput. Sci.* **1** (1975), 59–75.

[6] A. Ehrenfeucht et G. Rozenberg, On subword complexities of homomorphic images of langages, *R.A.I.R.O. Informatique théorique* **16** (1982), 108–113.

[7] M. Lothaire, *Combinatorics on Words*, *Encyclopedia of Mathematics and its Applications* vol. 17, Addison-Wesley, 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, 1997.

[8] M. Lothaire, *Algebraic Combinatorics on Words*, *Encyclopedia of Mathematics and its Applications* vol. 90, Cambridge University Press, 2002.

[9] M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.* **22** (1921), 84–100.

[10] M. Morse et G. A. Hedlund, Symbolic dynamics, *American J. Math.* **60** (1938), 815–866.

[11] J.-J. Pansiot, Hiérarchie et fermeture de certaines classes de tag-systèmes, *Acta Inform.* **20** (1983), 179–196.

[12] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in *ICALP '84*, pp. 380–389, *Lecture Notes in Computer Science* **172**, Springer-Verlag, 1984.

[13] A. Salomaa, *Jewels of Formal Language Theory*, Pitman publishing limited, 39 Parker Street, London WC2B 5PB, 1981.

[14] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Kra. Vidensk. Selsk. Skrifter, I. Mat. Nat. Kl.* **1** (1912), 1–67.

Julien Cassaigne
Institut de Mathématiques de Luminy — CNRS U.P.R. 9016
Case 907, 163, avenue de Luminy, F-13288 Marseille Cedex 9, France
cassaigne@iml.univ-mrs.fr

François Nicolas
Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier — CNRS U.M.R. 5506
161, rue Ada, F-34392 Montpellier Cedex 5, France
nicolas@lirmm.fr