Quelques éléments de Combinatoire des Mots

Cours 2014-2015

(G. Richomme)
gwenael.richomme@univ-montp3.fr

version du September 6, 2014

Toute remarque est la bienvenue.

1 Présentation générale

Modalités de contrôle des connaissances : examen + devoirs maison

Qu'est-ce que la combinatoire des mots ?

Un domaine scientifique consacré à l'étude des propriétés que peuvent avoir ou ne pas avoir des mots et aux moyens pour engendrer des mots ayant des propriétés données. Ce domaine puise ces racines dans, et tisse régulièrement des liens avec, de nombreux autres domaines scientifiques. La raison en est que de très nombreux objets peuvent être modélisés. Mieux comprendre ces mots modèles permet d'obtenir des informations sur les objets modélisés.

Dans [5], "In the latest classification applications of Mathematical Reviews combinatorics on words contitutes its own section under the chapter discrete mathematics related to computer science."

Quelques références introductives (la liste n'est pas exhaustive) :

- Lothaire [24, 25, 26] : les ouvrages de référence en combinatoire des mots (voir aussi les chapitres 6, 7, 8 du tome 1 de [36]).
- Allouche, Shallit 2003 [2], Pytheas Fogg 2002 [30], et récemment Berstel, Lauve, Reutenauer et Saliola [6].
- [7, 5] : respectivement histoire et développements récents en combinatoire des mots.
- [1, 3, 4, 10] : des exemples d'applications.

2 Quelques définitions

Alphabet. Ensemble *fini* et *non vide* de symboles.

Lettre. Élément d'un alphabet.

Mot (sur un alphabet A). Suite finie d'éléments de A.

Notation : les lettres sont accolées.

Mot vide. Noté ε, est la suite vide de lettres.

Concaténation. Étant donnés deux mots u et v, on note uv (parfois $u.v$) le mot obtenu en prenant d'abord les lettres de u puis les lettres de v.

Propriétés :

- $u\varepsilon = u = \varepsilon u$ (le mot vide est élément neutre pour l'opération de concaténation) ;
- (uv)w = u(vw) (la concaténation est associative).

Monoïde. Ensemble muni d'une opération interne associative possédant un élément neutre.

L'ensemble, noté A^*, des mots sur un alphabet A possède une structure de monoïde.

Si X est un ensemble de mots, X^* dénote l'ensemble de tous les mots obtenus par concaténation de mots de X. Le mot vide appartient à X^*. Quand X n'est pas un alphabet, X^* peut ne pas être un monoïde libre.

Monoïde libre. Tout mot se décompose de manière unique sur les lettres. En d'autres termes, si $a_1, \ldots, a_n, b_1, \ldots, b_p$ sont des lettres alors l'égalité $a_1 \ldots a_n = b_1 \ldots b_p$ implique $n = p$ et $a_i = b_i$ pour tout entier i, $1 \leq i \leq n$. Le monoïde A^* est dit libre (de base A).

Longueur. Longueur de la suite = nombre de lettres apparaissant dans le mot.
Notation : $|u|$.

Propriétés :

- $|\varepsilon| = 0$;
- $|uv| = |u| + |v|$ (pour tous mots u et v).

Nombre d'occurrences. Le nombre d'occurrences d'une lettre a dans un mot u est noté $|u|_a$. Il s'agit du nombre de fois où la lettre apparaît.

Morphisme. Soient $(M_1, \cdot_1)$ et $(M_2, \cdot_2)$ deux monoïdes. Un morphisme f de M_1 dans M_2 est une application de M_1 dans M_2 qui :

- préserve le produit : pour tous x, y dans M_1,$f(x \cdot_1 y) = f(x) \cdot_2 f(y)$
- préserve l'élément neutre : l'image de l'élément neutre de M_1 est l'élément neutre de M_2.

Morphisme de monoïdes libres. Morphisme entre deux monoïdes libres.

Propriété : un morphisme de monoïdes libres est entièrement défini par l'image des lettres.

Remarque. Dans le contexte des systèmes dynamiques discrets, la restriction d'un morphisme de monoïde libre à un alphabet s'appelle une substitution, et parfois la substitution est confondue avec le morphisme qui en est son extension naturelle. Dans le contexte des langages formels, une substitution est un morphisme de A^* dans l'ensemble des parties de A^*.

Morphisme injectif. Un morphisme f est injectif si pour tous mots u et v, $f(u) = f(v)$ implique $u = v$.

Code. Un ensemble de mots X est un code si et seulement si pour tous entiers $n \geq 0$ et $p \geq 0$, pour tous mots, $x_1, \ldots, x_n, y_1, \ldots, y_p$ de X, l'égalité $x_1 \ldots x_n = y_1 \ldots y_p$ implique $n = p$ et pour tout i entre 1 et p, $x_i = y_i$.

Propriété : un morphisme f défini sur un alphabet A est injectif si et seulement si l'ensemble $f(A)$ a même cardinalité que A et est un code.

Facteur. Un mot v est un *facteur* d'un mot u s'il existe des mots p et s tels que $u = pvs$.

Préfixe. (facteur gauche) Un mot v est un *préfixe* d'un mot u s'il existe un mot s tel que $u = vs$

La relation "est préfixe de", souvent notée $v \preceq u$ ou $v \leq u$ ($v \prec u$ ou $v < u$ si $u \neq v$), est une relation d'ordre appelée *ordre préfixiel.*

Rappel. Une relation d'ordre est une relation transitive ($x \leq y$ et $y \leq z$ implique $x \leq z$), réflexive ($\forall x$, $x \leq x$), antisymétrique ($x \leq y$ et $y \leq x$ implique $x = y$).

Suffixe. (facteur droit) Un mot v est un *suffixe* d'un mot u s'il existe un mot p tels que $u = pv$

Facteur, préfixe, suffixe *propre.* Un mot u est un facteur (resp. préfixe, suffixe) *propre* d'un mot v si $u \neq v$ et u est un facteur (resp. préfixe, suffixe) de v.

Bord. Un mot u est un *bord* d'un mot v si u est un préfixe **propre** et un suffixe de v.

Mot sans bord. Un mot est *sans bord* s'il ne possède que le mot vide comme bord.

Préfixes disjoints. Soient u et v deux préfixes différents d'un mot $x = x_1 \ldots x_{|x|}$ (x_i lettre pour tout $1 \leq i \leq |x|$) : u et v sont des *préfixes disjoints* (dans x) si $u = x$ ou $v = x$ ou $x_{|u|+1} \neq x_{|v|+1}$.

Bord disjoint. Soient x un mot, u et v deux préfixes de x avec $u \neq v$: u est un *bord disjoint* de v (dans x) si u est un bord de v, et, u et v sont des préfixes disjoints de x.

Périodes. Un entier p, $1 \leq p \leq |w|$, est une période de $w = a_1 \ldots a_n$ (a_i lettre, $1 \leq i \leq n$) si et seulement si $a_i = a_{i+p}$ pour tout entier i entre 1 et $|w| - p$.

La *période* d'un mot w est la plus petite période de w. Elle sera notée ci-après $p(v)$.

Ordre. Le rationnel $|w|/p(w)$ est appelé l'*ordre* ou l'*exposant* de w. Il sera noté $\mathrm{ord}(w)$.

Puissance. Soit k un entier, u un mot, la *puissance $k$ème* de u est le mot, noté u^k, défini par $u^0 = \varepsilon$, et par $u^k = u^{k-1}u$ si $k \geq 1$.

Propriétés :

- Pour tout entier $k \geq 1$, $u^k = uu^{k-1}$.
- Pour tout entier $k \geq 0$, $|u^k| = k|u|$.

Puissance k. Quand k est un entier supérieur ou égal à 2, une *puissance k* est un mot u^k avec u mot non vide.

Carré = puissance 2.

Cube = puissance 3.

Mot primitif. Mot qui n'est puissance que de lui-même.

Propriété : le préfixe x de longueur $p(w)$ de w est un mot primitif. De plus, $w = x^k y$ avec y préfixe propre de x et $k \geq 1$ un entier.

Racine primitive. Pour un mot non vide w, la racine primitive est l'unique mot primitif dont w est puissance (voir lemme 3.6).

Puissance fractionnaire. Soient x un mot, y un préfixe propre et $k \geq 1$. Soit $\rho = \frac{k|x|+|y|}{|x|}$. Le mot $x^k y$ est parfois noté x^ρ et est appelé puissance fractionnaire de x (Remarque : ρ est l'exposant de $x^k y$).

Conjugaison. Étant donné un monoïde M, deux éléments e_1 et e_2 sont dits *conjugués* s'il existe un mot λ tel que $e_1 \lambda = \lambda e_2$.

Transposition. Étant donné un monoïde M, deux éléments e_1 et e_2 sont dits *transposés* s'il existe des éléments x et y tels que $e_1 = xy$ et $e_2 = yx$.

Image miroir. L'image miroir d'un mot u est le mot obtenu en lisant u de droite à gauche. Si $u = a_1 \ldots a_n$ est un mot avec a_i lettre pour tout i, alors l'*image miroir* de u, notée $\tilde{u}$, est le mot $a_n \ldots a_1$. *Remarque* : $\tilde{\varepsilon} = \varepsilon$

Palindrome. Mot égal à son image miroir. En d'autres termes, un palindrome est un mot qui peut être lu indifféremment de gauche à droite ou de droite à gauche (e.g.: LAVAL, ÉTÉ, ESSAYASSE, MALAYALAM, RESSASSER). Formellement, un mot u sur l'alphabet A est un palindrome si $u = \varepsilon$ ou $u = u_1 ... u_n = u_n ... u_1$ avec $n = |u|$ et $u_i \in A, 1 \leq i \leq n$.

3 Quelques résultats de base

3.1 Le lemme de Levi et quelques conséquences

Lemme 3.1 (Lemme de Levi – propriété d'équidivisibilité)
Soient t, u, v, w quatre mots sur l'alphabet A.
Si $tu = vw$ et $|t| \leq |v|$
alors il existe un unique mot z sur A tel que $v = tz$ et $u = zw$.

Corollaire 3.2 *Soient $u, v, w, t \in A^*$.*

1) Si $u\,v = w\,t$ et $|u| = |w|$ alors $u = w$ et $v = t$.

2) Pour tout $i \in \mathbb{N} \setminus \{0\}$, ($u^i = v^i \Rightarrow u = v$.)

Lemme 3.3 *Pour A alphabet, le monoïde libre A^* est* simplifiable.
C'est à dire pour tous mots $u, v, w, t \in A^*$:

a) $u\,v = u\,w \Rightarrow v = w$;

b) $u\,v = w\,v \Rightarrow u = w$;

c) $u\,v\,w = u\,t\,w \Rightarrow v = t$.

Lemme 3.4 *Soient A un alphabet et u_1, u_2, v trois mots de A^*.*
Si $u_1 \leq v$ et $u_2 \leq v$ alors $u_1 \leq u_2$ ou $u_2 \leq u_1$.

3.2 Equations courantes

Lemme 3.5 *Soient A un alphabet et x, y des mots sur A. Sont équivalents :*

1. *$xy = yx$,*

2. *il existe deux entiers n et m non tous deux nuls tels que $x^n = y^m$.*

3. *il existe un mot z et deux entiers p et q tels que $x = z^p$ et $y = z^q$.*

Lemme 3.6 *Pour tout mot w non vide, il existe un unique mot primitif z tel que $w = z^n$ pour un entier $n \geq 1$.*

Lemme 3.7 *Les assertions suivantes sont équivalentes pour des* **mots** *u, $x \neq \varepsilon$ et $y \neq \varepsilon$:*

1. *$ux = yu$*

2. *il existe des mots r et s et un entier $k \geq 0$ tels que $u = (rs)^k r$, $x = sr$, $y = rs$*

Conséquences de la proposition :

- x et y sont donc conjugués si et seulement si y et x sont conjugués.
- La relation de conjugaison est une relation d'équivalence (non dit en 2004-2005)
- Un mot u est un conjugué d'un mot v si et seulement s'il existe des mots r, s et un entier k tel que $u = (rs)^k r$ et $v = (rs)^{k+1} r$. (le cas $u = \varepsilon$ correspond au cas $k = 0$ et $r = \varepsilon$.

Lemme 3.8 *Soient u et v deux mots conjugués.*
Le mot u est primitif si et seulement si v est primitif.

3.3 Un peu plus au sujet des équations

Proposition 3.9 (une version "light" du théorème du défaut)
Soit X un ensemble fini de mots.
S'il existe une équation non triviale entre les mots de X,
alors il existe un ensemble Y de cardinal plus petit que celui de X, tel que pour tout x dans X, $x \in Y^*$.

Corollaire 3.10 *S'il existe une équation non triviale entre deux mots x et y alors x et y sont puissances d'un même mot.*

Lemme 3.11 (Lyndon and Schützenberger 1962) *Si des mots x, y et z vérifient l'équation $x^m y^n = z^q$ pour des entiers $m, n, q \geq 2$, alors ils sont des puissances d'un même mot.*

Lemme 3.12 *Soient u, v, w des mots tels que $uv \neq \varepsilon$ et $vw \neq \varepsilon$. Montrer que $uvw = wvu$ si et seulement s'il existe des mots x, y et des entiers n, p, q tels que $u = (xy)^n x$, $v = (yx)^p y$ et $w = (xy)^q x$.*

3.4 Théorème de Fine et Wilf

Théoreme 3.13 *Soient x et y deux mots. Si deux puissances x^p et y^q de x et y ont un préfixe commun de longueur au moins égal à $|x|+|y|$-pgcd($|x|$,$|y|$), alors x et y sont des puissances d'un même troisième mot.*

Le corollaire suivant du théorème de Fine et Wilf est souvent appelé *lemme de périodicité* :

Corollaire 3.14 *Si p et q sont deux périodes d'un mot non vide x telles que $p + q - \mathrm{pgcd}(p, q) \leq |x|$ alors $\mathrm{pgcd}(p, q)$ est aussi une période de x.*

Remarque :

- M.G. Castelli, F. Mignosi et A. Restivo [12] puis J. Justin [20] ont généralisé le théorème de Fine et Wilf à un nombre quelconque de mots. (voir aussi [25, chapitre 8])

4 Quelques exercices

Exercice 4.1

1. Quel est le nombre de préfixes d'un mot u en fonction de la longueur de ce mot ? Idem pour suffixe ?
2. Quel est le nombre maximal possible de facteurs d'un mot u en fonction de sa longueur ?
3. Quel est le nombre minimal possible de facteurs d'un mot u en fonction de sa longueur ?

Exercice 4.2 Soient A un alphabet et $a, b \in A$, $u \in A^*$.
Montrer que si $ua = bu$ alors $a = b$ et $u \in \{a\}^*$.

Exercice 4.3 Soient $x, y, z \in A^*$. Montrer que $x^2 = y^2 z^2$ si et seulement s'il existe un mot t tel que $x, y, z \in t^*$ et $x = yz$.
(On pourra réfléchir à deux preuves différentes : une utilisant le lemme 3.11, l'autre s'en passant)

Exercice 4.4 Soient $f_1, f_2, f_3, f_4, g_1, g_2, g_3, g_4 \in A^*$. Prouver que si $f_1 f_4 = g_1 g_4$, $f_1 f_2 f_4 = g_1 g_2 g_4$ et $f_1 f_3 f_4 = g_1 g_3 g_4$ alors $f_1 f_2 f_3 f_4 = g_1 g_2 g_3 g_4$. Est-ce que la condition $f_1 f_4 = g_1 g_4$ est nécessaire ?

Exercice 4.5 Pour deux mots u et v, montrer que $\tilde{\tilde{u}} = u$ et $\widetilde{uv} = \tilde{v}\tilde{u}$.

Exercice 4.6 Soit $x \in A^*$ ayant un bord non vide. Soit u le plus court bord non vide de x.

1. Montrer que u est sans bord et qu'il existe un mot $v \in A^*$ tel que $x = uvu$.
2. La propriété précédente est-elle vraie pour n'importe quel bord u de x ?

Exercice 4.7 Soient u, v, w, x quatre mots tels que u est un bord disjoint de v dans x et v est un bord disjoint de w dans x. A-t-on u bord disjoint de w ?

Exercice 4.8 Montrer qu'un mot w est primitif si et seulement s'il a exactement $|w|$ conjugués.

Exercice 4.9 Montrer que le nombre de conjugués d'un mot u est la longueur de sa racine primitive.

Exercice 4.10 Soient u et v deux mots de même longueur.
Soient z_1 et z_2 la racine primitive de u et v.
Montrer que u et v sont conjugués si et seulement si z_1 et z_2 sont conjugués.
Montrer de plus que $|z_1| = |z_2|$ et il existe un entier k tel que $u = z_1^k$ et $v = z_2^k$.

Exercice 4.11 Soient u et v deux mots primitifs. Montrez que la période de uv est plus grande que ou égale à la plus grande des périodes de u et v.

Exercice 4.12 Montrer que si u et v sont deux mots de période p alors uv n'est pas nécessairement de période p.

Exercice 4.13 Proposer une construction d'une suite de mots $(u_n)_{n \geq 0}$ tous différents pour lesquels $p(u_n) = 1$.

Exercice 4.14 Proposer une construction d'une suite de mots $(u_n)_{n \geq 0}$ tous différents pour lesquels $p(u_n) = |u_n|$.

Exercice 4.15 Montrer que si une puissance d'un mot x et une puissance d'un mot y ont un facteur commun de longueur au moins $|x| + |y| - pgcd(|x|, |y|)$ alors il existe deux mots conjugués u et v tels que x est une puissance de u et y est une puissance de v.

Exercice 4.16 Montrer que la propriété "*pour tous mots u, v,* $f(uv) = f(u)f(v)$" implique $f(\varepsilon) = \varepsilon$

5 Mots de Fibonacci

5.1 Définition de la suite de Fibonacci

Les *mots de Fibonacci*, que nous noterons $(f_n)_{n\geq 0}$, sont les mots définis par :

$$f_0 = a,$$

$$f_1 = ab,$$

$$f_{n+2} = f_{n+1}f_n, \text{ pour } n \geq 0.$$

Remarque : certains auteurs commencent la suite de Fibonacci à l'indice -1 en posant $f_{-1} = b$. Soyez vigilants et signalez tout mélange que vous rencontreriez dans ce cours.

Nous énonçons tout au long de ce chapitre des propriétés qui sont autant d'exercices.

5.2 Pourquoi Fibonacci ?

La suite des entiers de Fibonacci est définie[1] par :

$$\begin{cases} F_0 = 1, \\ F_1 = 2, \\ F_n = F_{n-1} + F_{n-2} \text{ pour } n \geq 2. \end{cases}$$

Nous poserons quand nécessaire : $F_{-1} = 1$, $F_{-2} = 0$.

Cette suite d'entier est la suite d'entier A000045 dans *the On-Line Encyclopedia of Integer Sequences* (maintenue par N. J. A. Sloane),
voir http://www.research.att.com/∼njas/sequences/.

La propriété suivante montre que la suite des longueurs des mots de Fibonacci est la suite des entiers de Fibonacci, expliquant ainsi le choix de terminologie.

Propriété 5.1 *Pour tout entier $n \geq 0$, $|f_n| = F_n$.*

Une manière plus ou moins équivalente d'expliquer la terminologie vient du remplacement de l'opération d'addition par l'opération de concaténation (la fonction longueur étant alors un morphisme de monoïdes du monoïde libre dans l'ensemble des entiers munis de l'addition).

Deux autres propriétés importantes :

Propriété 5.2 *Pour tout entier $n \geq 0$, $|f_n|_a = F_{n-1}$.*

Propriété 5.3 *Pour tout entier $n \geq 0$, $|f_n|_b = F_{n-2}$.*

[1]Selon les auteurs et les contextes, les valeurs des premiers termes peuvent varier. Par exemple dans [19] (et le plus souvent), les premiers termes sont 0 et 1 (puis 1 et 2). Notre choix de premières valeurs pour cette suite d'entiers permet de rendre la propriété 5.1 plus *belle*.

5.3 Annexe : quelques informations sur les entiers de Fibonacci

Ces entiers (voir [19] pour plus d'informations) sont très liés au nombre d'or $\boxed{\Phi = \frac{1+\sqrt{5}}{2}}$ et le nombre (appelé le conjugué du nombre d'or) $\boxed{\overline{\Phi} = \frac{1-\sqrt{5}}{2}}$.

Rappelons que Φ et $\overline{\Phi}$ sont les deux solutions de l'équation $x^2 = x + 1$.

Valeurs des termes de la suite :

$$\boxed{\text{Pour } n \geq 0, \quad F_n = \frac{1}{\sqrt{5}}\left[\Phi^{n+2} - \overline{\Phi}^{n+2}\right]}$$

Somme des premières valeurs de la suite :

$$\boxed{\text{Pour } n \geq 0, \quad \sum_{i=0}^{n} F_i = F_{n+2} - 2}$$

Ainsi : $$\boxed{\lim_{n\to\infty} \frac{\sum_{i=0}^{n} F_i}{F_n} = \Phi^2 = \frac{3+\sqrt{5}}{2}}$$

Exercice 5.4

1. Vérifier les formules $\Phi + \bar{\Phi} = 1$, $\Phi\bar{\Phi} = -1$ (Remarque : elles sont liées à l'équation vérifiée par Φ et $\bar{\Phi}$).
2. Vérifier $\frac{(\Phi - \bar{\Phi})}{\sqrt{5}} = 1$.
3. Vérifier par récurrence la formule close vérifiée par F_n.
4. Vérifier par récurrence la formule close vérifiée par $\sum_{i=0}^{n} F_i$.

5.4 Une définition morphique

Soit φ le morphisme, appelé *morphisme de Fibonacci*, défini par :

$$\varphi \begin{cases} a \mapsto ab \\ b \mapsto a \end{cases}$$

La propriété suivante peut être montrée par une simple récurrence :

Propriété 5.5 *Pour tout entier $n \geq 0$, $f_n = \varphi^n(a)$.*

Exercice 5.6 Montrer que le morphisme φ est injectif.

5.5 Primitivité

Un examen des premiers mots de Fibonacci laisse supposer que ces mots ne sont pas puissances d'un autre, i.e. :

Propriété 5.7 *Pour tout $n \geq 0$, f_n est un mot primitif.*

Un moyen simple de démontrer la propriété précédente consiste à s'appuyer sur la propriété suivante du morphisme de Fibonacci.

Propriété 5.8 *Pour tout mot w, w est primitif si et seulement si $\varphi(w)$ est primitif.*

Exercice 5.9 Montrer que pour tout entier $n \geq 1$, $\text{pgcd}(|f_n|, |f_{n+1}|) = 1$

Exercice 5.10 Est-ce que $|f_n|$ est un entier premier pour tout entier $n \geq 2$?

5.6 Apériodicité

Plus que d'être primitif, nous pouvons observer que les mots de Fibonacci sont apériodiques. Plus précisément :

Propriété 5.11 *Soit $n \geq 2$.*
La plus petite période de f_n est $|F_{n-1}|$.
De manière équivalente, le plus grand bord de f_n est f_{n-2}.
En conséquence, $\exp f_n = \frac{F_n}{F_{n-1}} < 2$.

Il est possible de montrer un peu plus.

5.7 Périodes et bords

Pour les mots de Fibonacci, eux-mêmes :

Propriété 5.12 *Pour tout entier $n \geq 0$, les bords de f_n sont ε et les mots f_{n-2i} pour $1 \leq i < \lfloor \frac{n}{2} \rfloor$.*

5.8 Un grand préfixe périodique

Définition. Pour $n \geq 1$, notons g_n le préfixe de longueur $F_n - 2$ de f_n.

Propriété 5.13 *Pour tout $n \geq 3$, f_n^2 et f_{n-1}^3 ont g_{n+1} comme préfixe commun. En d'autres termes, g_{n+1} possède f_n et f_{n-1} comme préfixes et a les entiers $|f_n|$ et $|f_{n-1}|$ comme périodes.*

Cette propriété est une conséquence de :

- pour $n \geq 1$, $g_{n+1} = f_{n-1}g_n$ et

- pour $n \geq 3$, $g_{n+1} = f_{n-1}^2 g_{n-2}$.

Remarque 5.14 Les mots g_n montrent l'optimalité de la borne apparaissant dans le théorème de Fine et Wilf (Théorème 3.13). C'est une conséquence de la propriété 5.13, du fait que les mots de Fibonacci sont primitifs, et de la relation $|g_{n+1}| = |f_n| + |f_{n-1}| - 2$.

Ces mots permettent de calculer le délai maximal pouvant se produire dans l'algorithme de recherche de motifs de Knuth, Morris et Pratt.

5.9 Bords et bords disjoints

Commençons par les bords des mots $(g_n)_{n\geq 0}$:

Propriété 5.15 *Pour tout entier $n \geq 2$,*

1. $g_{n-1} = Bord(g_n)$
2. *Les bords de g_n sont $g_{n-1}, \ldots, g_2, g_1 = \varepsilon$.*

Pour montrer que g_n n'a pas de bord, il peut être utile de voir que :

Exercice 5.16 Si u est un bord d'un palindrome p, alors u est un palindrome.

Regardons à présent les bords disjoints des mots $(g_n)_{n\geq 0}$. Observons :

Propriété 5.17 *Pour tout entier $p \geq 1$,*

1. *$g_{2p-1}a$ est un préfixe de g_{2p},*
2. *$g_{2p}b$ est un préfixe de g_{2p+1}.*

Par suite[2] :

Propriété 5.18 *Pour tout entier $n \geq 2$, dans g_{n+1} :*

1. *g_{n-1} est le plus grand bord disjoint de g_n ;*
2. *les bords disjoints de g_n sont les mots g_{n-2i-1} avec $0 \leq i < \lfloor \frac{n}{2} \rfloor$.*

Nous avons également :

Propriété 5.19 *Pour $n \geq 2$, la suite $g_n, g_{n-1}, g_{n-2}, \ldots g_1$ est telle que g_{i+1} est un bord disjoint de g_i. En conséquence le délai dans l'algorithme de Knuth, Morris et Pratt est au moins de $log_\Phi(w)$ où w est le motif cherché.*

Pour les mots de Fibonacci, eux-mêmes :
(Indication : si $n \geq 1$, un bord suffisamment long de f_n est un mot de la forme uab ou uba avec u bord de g_n)

[2]Dans la propriété suivante, $\lfloor x \rfloor$ désigne la partie entière inférieure de x

5.10 Palindromes

Une autre propriété remarquable des mots g_n est :

Propriété 5.20 *Pour tout $n \geq 2$, le mot g_n est un palindrome.*

En d'autres mots :

Propriété 5.21 *Pour tout $i \geq 1$, il existe un palindrome g_i tel que*
$f_i = \begin{cases} g_i ab \text{ si } i \text{ est pair} \\ g_i ba \text{ si } i \text{ est impair} \end{cases}$
où les mots $(g_n)_{n\geq 1}$ vérifient $g_1 = \varepsilon$ et $g_{n+1} = \varphi(g_n)a$ pour tout entier $n \geq 1$.

La propriété 5.20 peut être démontrée en utilisant le résultat suivant :

Propriété 5.22 *Soit w un palindrome.*

1. *w est un palindrome si et seulement si $\varphi(w)a$ est un palindrome.*
2. *De plus, si $w \neq \varepsilon$,*
   w est un palindrome si et seulement si $a^{-1}\varphi(w)$ est un palindrome,
   où $a^{-1}u$ dénote le mot u sans son premier a (définition valide uniquement si u est non vide et commence par a).

5.11 Plus sur les palindromes

Exercice 5.23 Montrer que pour tout entier $n \geq 2$, $f_n = u_n v_n$ avec u_n et v_n deux palindromes.

Montrer qu'une telle décomposition est unique.
Étudier les liens entre u_n et g_n ?

Exercice 5.24 Soit u un mot, appelons *clôture palindromique* et notons $u^{(+)}$ le plus petit palindrome dont u est préfixe.

Pour tout entier $p \geq 1$, $g_{2p} = (g_{2p-1}a)^{(+)}$ et $g_{2p+1} = (g_{2p}a)^{(+)}$

6 Mots infinis

6.1 Quelques définitions

Mot infini. Un *mot infini* sur un alphabet A est une suite infinie de lettres indicée par les entiers naturels (resp. entiers naturels non nul selon auteurs et/ou contexte). Il peut être vu comme une fonction de $\mathbb{N}$ dans A (resp. de $\mathbb{N}^+$ dans A). La $(i+1)^{\text{ème}}$ lettre (resp. la $i^{\text{ème}}$ lettre) d'un mot infini $\mathbf{w}$ est notée $\mathbf{w}_i$ ou $\mathbf{w}(i)$.

A^ω L'ensemble des mots infinis sur A sera noté A^ω (autre notation fréquente $A^{\mathbb{N}}$).

Périodicité. Un mot infini $\mathbf{w}$ est *périodique* s'il existe un entier $p > 0$ tel que, pour tout entier i, $\mathbf{w}_i = \mathbf{w}_{i+p}$. L'entier p appelé une *période* de $\mathbf{w}$.

u^ω. Si $\mathbf{w}$ est périodique de période p et si u est le préfixe de longueur p de $\mathbf{w}$, alors $\mathbf{w}$ sera noté u^ω (le mot u est aussi parfois appelé *période* de $\mathbf{w}$.

Concaténation. Pour u un mot fini et $\mathbf{v}$ un mot infini, le mot uv est le mot tel que pour $0 \leq i < |u|$, $(uv)(i) = u_i$ et pour $i \geq |u|$, $(uv)(i) = v(i - |u|)$.

Propriété : pour u, v mots finis et $\mathbf{w}$ infini, $(uv)w = u(vw)$.

Ultime périodicité. Un mot infini $\mathbf{w}$ est *ultimement périodique* s'il est de la forme uv^ω pour des mots u et v avec $v \neq \varepsilon$.

Apériodicité. Un mot infini $\mathbf{w}$ est apériodique s'il n'est pas ultimement périodique.

Facteur, préfixe. Un mot fini u est *facteur* d'un mot infini $\mathbf{w}$, s'il existe un mot fini p et un mot infini $\mathbf{s}$ tel que $\mathbf{w} = pu\mathbf{s}$. Si $p = \varepsilon$, u est un *préfixe* de $\mathbf{w}$.

L'ensemble des facteurs d'un mot $\mathbf{w}$ sera noté $\mathcal{F}(\mathbf{w})$. L'ensemble des facteurs de longueur n sera noté $\mathcal{F}_n(\mathbf{w})$.

Suffixe. Un mot infini $\mathbf{s}$ est un *suffixe* d'un mot infini $\mathbf{w}$, s'il existe un mot fini p tel que $\mathbf{w} = p\mathbf{s}$. Le mot $\mathbf{s}$ est un *suffixe propre* de $\mathbf{w}$ si $p \neq \varepsilon$.

Exercice 6.1 Montrer qu'un mot $\mathbf{w}$ est périodique si et seulement si $\mathbf{w}$ est un suffixe propre de lui-même.

Mot limite. Un mot infini $\mathbf{w}$ est la limite d'une suite de mots $(u_n)_{n\geq 0}$ si pour tout préfixe p de $\mathbf{w}$, il existe un entier n_p tel que pour tout entier $n \geq n_p$, p est un préfixe de u_n. Nous noterons

$$\mathbf{w} = \lim_{n\to\infty} u_n$$

En d'autres termes (définition de [25]) une suite $(u_n)_{n\geq 0}$ de mots finis converge vers un mot infini $\mathbf{w}$ si tout préfixe de $\mathbf{w}$ est le préfixe de tous les u_n à l'exception d'un nombre fini.

Récurrent. Un mot fini u est *récurrent* dans un mot infini $\mathbf{w}$ s'il apparaît infiniment souvent dans $\mathbf{w}$.

Récurrent. Un mot infini est *récurrent* si tous ses facteurs sont récurrents.

Uniformément récurrent. Un mot infini $\mathbf{w}$ est *uniformément récurrent* si tous ses facteurs apparaissent infiniment souvent à "distance bornée". En d'autres termes, pour tout entier n, il existe un entier r tel que tout facteur de $\mathbf{w}$ de longueur r contient tous les facteurs de longueur n.

Remarque. La terminologie *quasi-périodique* (almost periodic), issue de la théorie des systèmes dynamiques, remplace parfois celle de *uniformément récurrent.* Le terme *quasi-périodique* est aussi utilisé dans un contexte issu de l'algorithmique du texte (voir partie suivante).

Exercice 6.2 Est-ce que tout mot engendré par un morphisme est uniformément récurrent ?

Quasipériodicité. Un mot infini $\mathbf{w}$ est *quasi-périodique*, si et seulement si, il peut être couvert par des occurrences d'un mot q alors appelé *quasi-période.*

Facteur spécial. Un mot fini u est un *facteur spécial à droite* (resp. *facteur spécial à gauche*) d'un mot infini $\mathbf{w}$ s'il existe au moins deux lettres différentes α et β telles que les mots $u\alpha$ et $u\beta$ sont des facteurs de $\mathbf{w}$ (resp. telles que les mots αu et βu sont des facteurs de $\mathbf{w}$).

Facteur bispécial. Un mot fini qui est à la fois un facteur spécial gauche et un facteur spécial droit d'un mot $\mathbf{w}$ est appelé *facteur bispécial* de $\mathbf{w}$.

6.2 Mots engendrés par morphismes

- Un morphisme f sur A est *effaçant* s'il existe une lettre tel $f(a) = \varepsilon$ (ou de manière équivalente, s'il existe un mot non vide u tel que $f(u) = \varepsilon$).
- Un morphisme f sur A est *prolongeable* en $a \in A$ s'il existe un mot non vide s tel que $f(a) = as$.
- Un mot fini ou infini w est un *point fixe* d'un morphisme f, si $f(w) = w$.

Proposition 6.3 *Soit f un endomorphisme sur A non-effaçant et prolongeable en $a \in A$. Soit s le mot tel que $f(a) = as$.*

Pour $n \geq 0$ entier, soit $u_n = f^n(a)$ et $v_n = f^n(s)$.

1. *Pour $n \geq 0$ entier, $u_{n+1} = u_n v_n$.*
2. *Pour $n \geq 0$ entier, $u_{n+1} = av_0v_1 \ldots v_n$.*

3. *le mot infini* $x = asf(s)f^2(s)\dots f^n(s)\dots$ *est* $\lim_{n\to\infty} u_n$. *De plus, il est l'unique mot commençant par a point fixe de* f.

Le mot x dans la proposition précédente est aussi noté $f^\omega(a)$ et est appelé *mot morphique.* On dit aussi qu'il est *engendré* par morphisme.

Exemple :

- *Mot de Thue-Morse* : mot engendré par le morphisme μ défini par $\mu(a) = ab$, $\mu(b) = ba$.
- *Mot (infini) de Fibonacci* : mot engendré par le morphisme φ défini par $\varphi(a) = ab$, $\varphi(b) = a$.

Exercice 6.4 Montrer que tout mot infini périodique est engendré par un morphisme prolongeable. Est-ce que c'est le cas pour un mot ultimement périodique ?

Exercice 6.5 Montrer : pour tout endomorphisme f de $\{a, b\}$ sur $\{a, b\}$, soit f ou f^2 est prolongeable en a, soit $\forall n \geq 0$, f^n n'est pas prolongeable en a.

Est-ce que cette propriété est généralisable à des alphabets plus grands ?

Exercice 6.6 Soit E (morphisme d'échange) et μ (morphisme de Thue-Morse), les endomorphismes sur $\{a, b\}$ définis par $E(a) = b$, $E(b) = a$, $\mu(a) = ab$, $\mu(b) = ba$.

Montrer :

a) $E^2 = Id$

b) $E \circ \mu = \mu \circ E$

c) $E \circ \mu^n = \mu^n \circ E$

On se donne les suites de mots $(u_n)_{n\geq 0}$ et $(v_n)_{n\geq 0}$ définies par $u_0 = a$, $v_0 = b$ et pour $n \geq 1$ entier, $u_n = u_{n-1}v_{n-1}$ et $v_n = v_{n-1}u_{n-1}$.

d) Pour tout entier $n \geq 0$, $v_n = E(u_n)$

e) Pour tout entier $n \geq 0$, $u_n = \mu^n(a)$.

f) Montrer que $\mu^n(a)$ est un palindrome si et seulement si n est pair (définir image miroir, palindrome.

g) Pour $n \geq 0$, quels sont les bords de u_n ?

6.3 D'autres moyens de définir des mots infinis

Mot morphique. Un mot infini $\mathbf{w}$ est *morphique* s'il existe deux morphismes f et g tel que $\mathbf{w} = g(f^\omega(a))$.

L-systèmes. Les *systèmes de Lindenmayer* sont des systèmes engendrant des langages. Contrairement aux grammaires algébriques qui agissent de manière séquentielle, ces systèmes réécrivent toutes les lettres en parallèles de manière morphique.

Un *D0L* (D-zéro-L) système est la donnée d'un triplet (A, h, w) où h est un alphabet, h un morphisme de A^* dans A^* et w un mot sur A. Le langage engendré par un tel L-système est l'ensemble

$$\{h^n(w) \mid n \geq 1\}$$

Un mot infini engendré par un morphisme peut être vu comme comme la limite des mots du langage d'un D0L système.

Un *CD0L système* (aussi appelé *tag-système*) est la donnée d'un quintuplet (A, B, h, g, w) avec (A, h, w) D0L système, B alphabet et g morphisme de A^* dans B^*. Le langage engendré est alors

$$\{g(h^n(w)) \mid n \geq 1\}$$

et le mot $\lim_{n\to\infty} g(h^n(w))$ (quand il existe) est étudié. Il existe des mots infinis engendrés par CD0L système qui ne le sont pas par D0L système.

Mots automatiques. Le cas des tag systèmes dans lequel le morphisme h est uniforme (toutes les lettres ont des images de même longueur) est un cas important car équivalent à la notion de suite automatique (voir par exemple [2]).

Automates avec sorties. Un *Automate avec sorties* (ou *transducteur séquentiel*) ou *gsm (generalized sequential machine)* est un automate fini $< A, Q, D, F, \delta >$ auquel on ajoute un nouvel alphabet B et une fonction de sortie λ de $Q \times A$ vers B^* : un tel automate permet donc, non seulement la reconnaissance d'un mot, mais aussi la production de mots.

...

Moyens ad hoc.

7 Le mot infini de Fibonacci

Definition

$$\mathbf{F} = \lim_{n\to\infty} f_n = \lim_{n\to\infty} \varphi^n(a)$$

Le mot infini de Fibonacci est le point fixe du morphisme φ.

Quelques propriétés

Lemme 7.1 *Le mot de Fibonacci n'est pas périodique.*

Lemme 7.2 *Le mot de Fibonacci n'est pas ultimement périodique.*

Lemme 7.3 *Le mot de Fibonacci est récurrent.*

Lemme 7.4 *Le mot de Fibonacci est uniformément récurrent.*

Proposition 7.5 *Le mot de Fibonacci est quasipériodique (voir* [23] *pour une caractérisation des quasi-périodes).*

Lemme 7.6 *Pour tout entier $n \geq 0$, $\mathbf{F}$ a exactement $n+1$ facteurs différents de longueur n ($p_{\mathbf{F}}(n) = n+1$).*

Lemme 7.7 *Les facteurs bispéciaux de $\mathbf{F}$ sont les mots g_n (pour $n \geq 1$).*

Corollaire 7.8 *Un mot est un facteur bispécial de $\mathbf{F}$, si et seulement si, il est un palindrome préfixe de $\mathbf{F}$.*

Corollaire 7.9 *Un mot est un facteur spécial gauche de $\mathbf{F}$, si et seulement si, il est un préfixe de $\mathbf{F}$.*

Exercice 7.10 Montrer que l'ensemble des facteurs de $\mathbf{F}$ est clos par image miroir (pour tout facteur de $\mathbf{F}$, son image miroir est aussi un facteur de $\mathbf{F}$).

Proposition 7.11 *[14] Pour tout entier $n \geq 0$, $\mathbf{F}$ a exactement 1 palindrome de longueur n si n est pair, 2 palindromes de longueur n si n est impair.*

Plus précisément, si n est impair, il existe un palindrome facteur de $\mathbf{F}$ de la forme $\tilde{u}au$ et un autre de la forme $\tilde{v}bv$.

7.1 Facteurs singuliers et décomposition en palindromes

Lemme 7.12 *Soit $n \geq 1$ un entier et α une lettre telle que $g_n\alpha$ est préfixe de $\mathbf{F}$. Le mot $\alpha g_n \alpha$ est un palindrome facteur de $\mathbf{F}$.*

Les mots a, b et les palindromes du lemme précédent ont été appelés *facteurs singuliers* par Z.X. Wen et Z.Y. Wen, qui ont également démontré :

Proposition 7.13 *[38] Le mot de Fibonacci admet la décomposition en facteurs singuliers suivante :*

$$\mathbf{F} = a.b.\prod_{n\geq 0} ag_{2n+1}a.bg_{2n+2}b$$

7.2 Runs

Mot périodique. Un mot fini u de période p est périodique si $\frac{|u|}{p} \geq 2$.

Run. Soient u, p, s, w 4 mots tels que $w = pus$ (w et s peuvent être infinis). Notons α la dernière lettre de p si $p \neq \varepsilon$ et β la première lettre de s si $s \neq \varepsilon$. Le mot u est un "*run*" ou *facteur répétif maximal* de w si c'est un mot périodique, et si p est sa (plus petite) période alors αu (si $p \neq \varepsilon$) et $u\beta$ (si $s \neq \varepsilon$) ne sont pas périodiques de période p.

Proposition 7.14 *Un mot u est un run de Fibonacci si et seulement si u vérifie un des quatre cas suivants :*

1. $u = aa$,
2. $u = ababa$,
3. $u = f_{n+1}^3 g_n$ *pour un entier* $n \geq 1$,
4. $u = f_{n+1}^2 g_n$ *pour un entier* $n \geq 1$.

En fait, le résultat peut être précisé grâce aux deux résultats suivants :

Lemme 7.15 *Un mot u est un run préfixe du mot de Fibonacci si et seulement si $u = f_{n+1}^2 g_n$ pour un entier $n \geq 1$.*

Lemme 7.16 *Un mot u est un run facteur interne du mot de Fibonacci si et seulement si u vérifie un des quatre cas suivants :*

1. $u = aa$,
2. $u = ababa$,
3. $u = f_{n+1}^3 g_n$ *pour un entier* $n \geq 1$,
4. $u = f_{n+1}^2 g_n$ *pour un entier* $n \geq 1$.

Pour démontrer le lemme précédent, le suivant peut être utile :

Lemme 7.17 *Soient α une lettre et soit $n \geq 1$ un entier. Si $g_n \alpha g_n$ est facteur de $\mathbf{F}$ alors $n = 1$, ou, $n = 2$, $\alpha = b$.*

Puissance fractionnaire. Un mot de la forme $(uv)^k u$ (avec $uv \neq \varepsilon$ et $k \geq 1$) est appelé une *puissance fractionnaire*. Son exposant est $k + \frac{|u|}{|uv|}$. En d'autres mots, une puissance fractionnaire d'exposant $\frac{p}{q}$ est un mot dont la longueur n et une période π vérifient $\frac{n}{\pi} = \frac{p}{q}$. Un tel mot est noté $u^{\frac{n}{\pi}}$ où u est le préfixe de longueur π.

Mot sans puissance k. (définition plus générale que la définition de la partie 1) Si k est un rationnel strictement supérieur à 1, on dit qu'un mot fini ou infini u est *sans puissance* k s'il ne contient aucun facteur de la forme u^k.

Mot sans puissance k^+. Ici k est un réel positif quelconque. Un mot *sans puissance* k^+ est un mot sans-puissance k' pour tout rationnel $k' > k$.

Mot sans puissance k^-. Ici k est un réel positif quelconque. Un mot *sans puissance* k^- est un mot sans-puissance k mais qui, pour tout rationnel $k' < k$, n'est pas sans puissance k'.

Corollaire 7.18 [22] *Le mot de Fibonacci ne contient aucune puissance* 4.

Corollaire 7.19 [27] *Le mot de Fibonacci est sans puissance* $(2 + \Phi)^-$.

Remarque 7.20 Dans le cadre des mots finis, en 1999, G. Kolpakov et R. Kucherov ont conjecturé (voir [26]) : le nombre de runs apparaissant dans un mot fini est strictement inférieur à la longueur de ce mot. Les résultats s'améliorent d'article en article Après une série d'articles sur le sujet, les meilleurs résultats obtenus ont permis de montrer que le nombre de runs dans un mot w est inférieur ou égal à $1.029|w|$ (M. Crochemore, L. Ilie, L. Tinta, *The "runs" conjecture*, actes de la conférence Words 2009, septembre 2009, Italie). Depuis le résultat a certainement été amélioré.

Remarque 7.21 Il est également possible de caractériser les plus grands runs préfixes de Fibonacci (voir par exemple [9]). Ce sont les mots de la forme $f_{n+1}^2 g_n$.

7.3 Équilibre

Équilibre. Un mot sur $\{a, b\}$ est équilibré si pour tous facteurs x et y avec $|x| = |y|$, $||x|_a - |y|_a| \leq 1$.

Exercice 7.22 Montrer qu'un mot est équilibré si et seulement si pour tous facteurs x et y avec $|x| = |y|$, $||x|_b - |y|_b| \leq 1$.

Lemme 7.23 *Le mot de Fibonacci est équilibré.*

7.4 Mots de Lyndon

Dans cette partie, nous notons $<$ la relation d'ordre lexicographique sur les mots de Lyndon (voir [35] pour une introduction ludique aux mots de Lyndon).

Mot de Lyndon. Un mot fini non vide ou infini w est un mot de Lyndon si et seulement si il est strictement plus petit dans l'ordre lexicographique que tous ses suffixes propres non vides.

Proposition 7.24 *Le mot* $a\mathbf{F}$ *est un mot de Lyndon sur l'alphabet* $\{a, b\}$ *ordonné par* $a < b$. *Le mot* $b\mathbf{F}$ *est un mot de Lyndon ordonné par* $b < a$.

Une des preuves connues peut utiliser les faits suivants [33, 32] :
- Le mot $a\mathbf{F}$ est le point fixe du morphisme f défini par $f(a) = aab$ et $f(b) = ab$.
- Le mot $b\mathbf{F}$ est le point fixe du morphisme g défini par $g(b) = ba$, $g(a) = baa$.
- Le morphisme f préserve les mots de Lyndon sur l'alphabet $\{a, b\}$ ordonné par $a < b$.
- Le morphisme g préserve les mots de Lyndon sur l'alphabet $\{a, b\}$ ordonné par $b < a$.

7.5 Remarque

Le cours est loin d'être exhaustif sur les propriétés du mot de Fibonacci (voir par exemple [11]). Nous verrons par la suite que beaucoup de ces propriétés sont partagées par les mots sturmiens (il en est un cas particulier) et souvent elles deviennent caractéristiques de ces mots.

Exercice 7.25 Est-ce que le mot de Fibonacci peut être décomposé en carrés ? Si oui, une telle décomposition peut elle être obtenue avec des carrés de longueur bornée ?

Même questions pour une décomposition en cubes.

8 Mots et morphismes sturmiens

Voir par exemple [25, chapitre 2] pour plus de détails.

8.1 Mots sturmiens

Définition [28] Un mot *binaire* infini $\mathbf{w}$ est *sturmien* s'il est apériodique et équilibré.

De nombreuses propriétés des mots sturmiens peuvent être obtenues par une analyse morphique des mots sturmiens. Tous les mots sturmiens ne sont pas des points fixes de morphismes, mais ils peuvent tous être "désubstituer" en un mot sturmien via un nombre fini de morphismes. Pour cela, deux familles de morphismes jouent un rôle important :

Morphismes L. Soit A un alphabet et α une lettre. Nous noterons L_α le morphisme défini par $L_\alpha(\alpha) = \alpha$ et $L_\alpha(\beta) = \alpha\beta$ quand β est une lettre différente de α.

Morphismes R. Soit A un alphabet et α une lettre. Nous noterons R_α le morphisme défini par $R_\alpha(\alpha) = \alpha$ et $R_\alpha(\beta) = \beta\alpha$ quand β est une lettre différente de α.

Proposition 8.1 *Soit* $\mathbf{w}$ *un mot sur* $\{a,b\}$*,* α *sa première lettre et* β *la lettre telle que* $\{a,b\} = \{\alpha,\beta\}$*. Ce mot* $\mathbf{w}$ *est sturmien, si et seulement si, il existe un mot sturmien* $\mathbf{w}'$ *tel que :*
- $\mathbf{w} = L_\alpha(\mathbf{w}')$ *si* $\mathbf{w}$ *contient* $\alpha\alpha$*,*
- $\mathbf{w} = R_\beta(\mathbf{w}')$ *sinon.*

Ce résultat (aussi valide pour n'importe quel mot équilibré) est une conséquence de :

Lemme 8.2 *Pour* $\alpha \in \{a,b\}$ *lettre, un mot* $\mathbf{w}$ *est apériodique, si et seulement si,* $L_\alpha(\mathbf{w})$ *est apériodique, si et seulement si,* $R_\alpha(\mathbf{w})$ *est apériodique.*

Exercice 8.3 Est-il vrai, pour α lettre que : un mot $\mathbf{w}$ est périodique, si et seulement si, $L_\alpha(\mathbf{w})$ est périodique, si et seulement si, $R_\alpha(\mathbf{w})$ est périodique.

Lemme 8.4 *Un mot (fini ou infini)* $\mathbf{w}$ *sur* $\{a,b\}$ *n'est pas équilibré, si et seulement si, il existe un mot* u *tel que* aua *et* bub *sont deux facteurs de* $\mathbf{w}$*.*

Exercice 8.5 Montrer que la précision suivante du lemme précédent : Un mot (fini ou infini) $\mathbf{w}$ sur $\{a,b\}$ n'est pas équilibré, si et seulement si, il existe un *palindrome* u tels que aua et bub sont deux facteurs de $\mathbf{w}$.

Lemme 8.6 *Pour* $\alpha \in \{a,b\}$ *lettre, un mot* $\mathbf{w}$ *est équilibré, si et seulement si,* $L_\alpha(\mathbf{w})$ *est équilibré.*

Lemme 8.7 *Pour* $\alpha \in \{a,b\}$ *lettre,*

1. *Si* $\mathbf{w}$ *est un mot équilibré, alors* $R_\alpha(\mathbf{w})$ *est équilibré.*

2. *Si* $R_\alpha(\mathbf{w})$ *est un mot équilibré qui ne commence pas par* α *alors* $\mathbf{w}$ *est équilibré.*

Le résultat précédent n'est pas nécessairement vrai si $R_\alpha(\mathbf{w})$ est un mot sturmien qui commence par α.

Si la proposition 8.1 résume tout l'aspect de désubstitution, elle peut aussi être utilisée pour établir la caractérisation suivante :

Théoreme 8.8 *Les assertions suivantes sont équivalentes.*

1. *Un mot* $\mathbf{w}$ *est sturmien.*

2. *Il existe une suite infinie de mots* $(\mathbf{w}_n)_{n\geq 0}$*, et une suite infinie de morphismes* $(\varphi_n)_{n\geq 0} \in \{L_a, L_b, R_a, R_b\}$ *telles que :*

   (a) $\mathbf{w}_0 = \mathbf{w}$*,*

   (b) $\mathbf{w}_n = \varphi_n(\mathbf{w}_{n+1})$ *pour tout entier* $n \geq 0$*,*

   (c) *un des morphismes* L_a *et* R_a *apparaît infiniment souvent, et un des morphismes* L_b *ou* R_b *apparaît infiniment souvent,*

   (d) *pour un entier* n *et une lettre* α*, si* $\varphi_n = R_\alpha$*, alors* $\varphi_{n+1} \neq L_\alpha$*.*

3. *Il existe deux suites d'entiers* $(a_n)_{n\geq 0}$ *et* $(c_n)_{n\geq 0}$ *telles que :*

   (a) $a_0 \geq 0$ *et* $a_n \geq 1$ *pour* $n \geq 1$*,*

   (b) $0 \leq c_n \leq a_n$ *pour* $n \geq 0$*,*

   (c) $\mathbf{w} = \lim_{n\to\infty} L_a^{c_0} R_a^{a_0-c_0} L_b^{c_1} R_b^{a_1-c_1} \cdots L_a^{c_{2n}} R_a^{a_{2n}-c_{2n}} L_b^{c_{2n+1}} R_b^{a_{2n+1}-c_{2n+1}}(a)$

8.2 Morphismes sturmiens

Nous avons vu dans la partie précédente que les morphismes L_a, R_a, L_b et R_b préservaient les mots sturmiens (c'est-à-dire à la fois les mots infinis apériodiques et les mots équilibrés). Le résultat suivant fournit une caractérisation de l'ensemble des morphismes préservant les mots sturmiens.

Morphisme sturmien. Un morphisme f est *sturmien*, si et seulement si, il préserve les mots sturmiens, i.e., pour tout mot sturmien $\mathbf{w}$, $f(\mathbf{w})$ est un mot sturmien.

Morphisme d'échange. Sur $\{a, b\}$, le *morphisme d'échange* est le morphisme noté E défini par $E(a) = b$, $E(b) = a$. Si au lieu de de prendre l'alphabet $\{a, b\}$, on considère l'alphabet $\{0, 1\}$, l'opération d'échange de lettres correspond à l'opération de complémentation.

Théoreme 8.9 [37] *Un morphisme est sturmien, si et seulement si, il est la composition de morphismes dans* $\{E, L_a, R_a, L_b, R_b\}$*.*

Notons que dans l'article original l'ensemble $\{E, \varphi, \tilde{\varphi}\}$ apparaît à la place de $\{E, L_a, R_a, L_b, R_b\}$ (avec $\tilde{\varphi}(a) = ba$ et $\tilde{\varphi}(b) = a$). Grâce aux relations $L_a = \varphi \circ E$, $L_b = \varphi \circ E$ (relations déjà énoncées dans le chapitre sur les mots finis de Fibonacci), $R_a = \varphi \circ E$, $R_b = \varphi \circ E$.

En particulier, on peut voir que :

Lemme 8.10 *Un mot* $\mathbf{w}$ *est sturmien, si et seulement si,* $E(\mathbf{w})$ *est sturmien.*

Le résultat suivant propose deux autres caractérisations des morphismes sturmiens. La deuxième assertion dans ce théorème permet de tester en temps linéaire si un morphisme est sturmien. Il faut pour cela utiliser un algorithme permettant de détecter en temps linéaire si un mot est équilibré (voir par exemple [31] ou les algorithmes détectant en temps linéaire si un mot représente un segment de droite).

Théoreme 8.11 (voir [25, 8, 37]) *Les assertions suivantes sont équivalentes :*

1. *un morphisme f est sturmien ;*
2. *$f(baabaababaabab)$ est équilibré et f n'est pas périodique ;*
3. *il existe un mot sturmien s dont l'image est un mot sturmien.*

8.3 Mots épisturmiens

La définition des morphismes L et R (donnée au début de ce chapitre) est valable sur des alphabets de cardinaux quelconques. Une des généralisations des mots sturmiens, les mots épisturmiens introduits dans [15, 21] peuvent être caractérisés via ces morphismes. Nous donnons cette caractérisation comme définition :

Mot épisturmien. Un mot $\mathbf{w}$ est épisturmien, si et seulement si, il existe une suite infinie de mots $(\mathbf{w}_n)_{n\geq 0}$, et une suite infinie de morphismes $(\varphi_n)_{n\geq 0} \in \{L_\alpha, R_\alpha \mid \alpha \text{ lettre }\}$ telles que :

1. $\mathbf{w}_0 = \mathbf{w}$,
2. $\mathbf{w}_n = \varphi_n(\mathbf{w}_{n+1})$ pour tout entier $n \geq 0$.

Nous ne rentrerons pas dans le détail des propriétés de ces mots épisturmiens (voir [18] pour plus d'informations). Toutefois une comparaison de cette définition avec le théorème 8.8 amène quelques remarques.

- La définition ne comporte pas d'équivalents des conditions (c) et (d). De fait, il existe des mots épisturmiens qui sont périodiques. Une condition fréquente est de demander que pour chaque lettre α, un des morphismes L_α et R_α apparaissent infiniment souvent. Les mots épisturmiens qui vérifient cette propriété sont appelés les mots d'Arnoux-Rauzy.
- Il existe des mots épisturmiens qui admettent plusieurs désubstitutions infinies différentes. Dès lors, se posent les questions : peut-on savoir si deux désubstitions infinies définissent le même mot ? est-ce qu'une désubstitution infinie définit un unique mot ? (Voir [17] pour un survol des réponses à ces questions).

8.4 Un peu plus sur les mots équilibrés

Quand on s'intéresse aux mots équilibrés sur des alphabets d'au moins trois lettres, il reste une question ouverte appelée la conjecture de Fraenkel.

Le travail le plus récent sur ce sujet semble être l'article de Paquin et Vuillon qui ont caractérisés les suites épisturmiennes qui sont équilibrées. Ils écrivaient entre autres [29] : *This conjecture was first introduced in number theory and has remained unsolved for more than 30 years. It states that for a fixed* $k > 2$*, there is only one way to cover Z by k Beatty sequences with pairwise distinct frequencies. The prob- lem can be translated to combinatorics on words: for a k-letter alphabet, there is only one balanced sequence (up to letter permutation) that has different letter frequencies, which is called Fraenkel's sequence and is denoted by* (Fr_k) *where* $Fr_k = Fr_{k-1}kFr_{k-1}$*, with* $Fr_3 = 1213121$*. The conjecture is verified for k = 3; 4; 5; 6 according to the work of Altman, Gaujal, Hordijk and Tijdeman. The case k = 7 was recently settled by Barát and Varjú. Many cases have been proved by Simpson.*

Pour plus d'informations sur les mots équilibrés, voir l'article de Paquin et Vuillon en 2007 et sa bibliographie.

Pour terminer sur cette parenthèse, signalons que toutes les suites épisturmiennes ne sont pas équilibrées. Une notion d'équilibre local, généralisant la notion d'équilibre sur deux lettres a été proposée dans [34]. Cette propriété est caractéristique des mots épisturmiens.

8.5 Complexité

Fonction de complexité. La *fonction de complexité* ou *complexité* d'un mot infini $\mathbf{w}$ est la fonction notée $p_{\mathbf{w}}$ qui associe à tout entier n, le nombre de facteurs de longueur n de $\mathbf{w}$:

$$p_{\mathbf{w}}(n) = \#\mathcal{F}_n(\mathbf{w}).$$

Lemme 8.12 *[13] La complexité d'un mot infini non ultimement périodique est strictement croissante. Plus précisément, sont équivalents pour un mot* $\mathbf{w}$ *infini :*

1. $\mathbf{w}$ *est ultimement périodique ;*
2. *la fonction de complexité de* $\mathbf{w}$ *est bornée ;*
3. *il existe* n *tel que* $p_{\mathbf{w}}(n) \leq n$ *;*
4. *il existe* n *tel que* $p_{\mathbf{w}}(n) = p_{\mathbf{w}}(n+1)$ *;*
5. *il existe* n_0 *tel que pour tout entier* $n \geq n_0$*,* $p_{\mathbf{w}}(n) = p_{\mathbf{w}}(n+1)$*.*

Propriété 8.13 *Dans le lemme précédent, la condition* $p_{\mathbf{w}}(n) = p_{\mathbf{w}}(n+1)$ *est équivalente au fait que* $\mathbf{w}$ *n'a pas de facteur spécial droit de longueur* n*.*

Lemme 8.14 *Un mot* $\mathbf{w}$ *est de complexité* $n+1$ *pour tout* n *si et seulement si* $\mathbf{w}$ *est écrit sur un alphabet de deux lettres et possède un et un seul facteur spécial droit de longueur* n *pour chaque entier* n*.*

Théoreme 8.15 *[13] Les assertions suivantes sont équivalentes pour un mot infini* $\mathbf{w}$ *sur* $\{a, b\}$ *:*

1. $\mathbf{w}$ *est sturmien ;*
2. $\mathbf{w}$ *a exactement un mot spécial droit de chaque longueur ;*
3. $p_{\mathbf{w}}(n) = n + 1$ *pour tout entier* $n \geq 0$.

Ce résultat peut être démontré en utilisant le graphe des facteurs. Soit x un mot et n un entier.

Le graphe $G_n(x)$ des facteurs de longueur n d'un mot x est le graphe $(\mathcal{F}_n(x), E)$ avec : $E = \{(bs, a, sa) \mid a, b \in A, bsa \in \mathcal{F}_{n+1}(\mathbf{w})\}$

A noter qu'un facteur p d'un mot x est *spécial droit* s'il y a au moins deux arcs issus de p dans le graphe $G_{|p|}(x)$.

Une autre démarche pour démontrer ce résultat peut être d'utiliser le théorème 8.8 et la démarche pour le démontrer. Pour cela les résultats suivants peuvent être utiles (on pourra noter leur très forte similarité avec des résultats de la partie 8.1). Pour simplifier l'écriture, nous dirons qu'un mot est de complexité $n+1$ si son nombre de facteurs est $n+1$ pour tout entier n.

Proposition 8.16 *Soit* $\mathbf{w}$ *un mot sur* $\{a, b\}$, α *sa première lettre et* β *la lettre telle que* $\{a, b\} = \{\alpha, \beta\}$. *Ce mot* $\mathbf{w}$ *est de complexité* $n + 1$, *si et seulement si, il existe un mot* $\mathbf{w}'$ *de complexité* $n + 1$ *tel que :*
- $\mathbf{w} = L_\alpha(\mathbf{w}')$ *si* $\mathbf{w}$ *contient* $\alpha\alpha$,
- $\mathbf{w} = R_\beta(\mathbf{w}')$ *sinon.*

Lemme 8.17 *Un mot (fini ou infini)* $\mathbf{w}$ *sur* $\{a, b\}$ *n'est pas de complexité* $n + 1$, *si et seulement si, il existe un mot u tel que au et bu sont deux facteurs spéciaux droits de* $\mathbf{w}$.

Lemme 8.18 *Pour* $\alpha \in \{a, b\}$ *lettre, un mot* $\mathbf{w}$ *est de complexité* $n + 1$, *si et seulement si,* $L_\alpha(\mathbf{w})$ *est de complexité* $n + 1$.

Lemme 8.19 *Pour* $\alpha \in \{a, b\}$ *lettre,*

1. *Si* $\mathbf{w}$ *est un mot de complexité* $n + 1$, *alors* $R_\alpha(\mathbf{w})$ *est de complexité* $n + 1$.
2. *Si* $R_\alpha(\mathbf{w})$ *est un mot de complexité* $n + 1$ *qui ne commence pas par* α *alors* $\mathbf{w}$ *est de complexité* $n + 1$.

Le résultat précédent n'est pas nécessairement vrai si $R_\alpha(\mathbf{w})$ est un mot de complexité $n+1$ qui commence par α.

8.6 Palindromes

Théoreme 8.20 (voir [16]) *Les assertions suivantes sont équivalentes pour un mot infini* $\mathbf{w}$ *sur* $\{a, b\}$ *:*

1. $\mathbf{w}$ *est sturmien ;*

2. $\mathbf{w}$ *a exactement un facteur palindromique pour chaque longueur paire et deux facteurs palindromiques pour chaque longueur impaire.*

8.7 Codage de droite

Dans cette partie, nous travaillons par confort sur l'alphabet $\{0, 1\}$.

Mots mécaniques. Un mot $\mathbf{w}$ est dit *mécanique* s'il existe un réel α entre 0 et 1 et un réel ρ tel qu'une des deux situations suivantes existe :

- (mot mécanique inférieur) $\forall n \geq 0$, $\mathbf{w}[n] = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor$;
- (mot mécanique supérieur) $\forall n \geq 0$, $\mathbf{w}[n] = \lceil \alpha(n+1) + \rho \rceil - \lceil \alpha n + \rho \rceil$.

Un tel mot est dit mécaniquement irrationnel si α est irrationnel. Le réel α est appelé la pente du mot.

Théoreme 8.21 (voir [25] pour plus de détails) *Les assertions suivantes sont équivalentes pour un mot infini* $\mathbf{w}$ *sur* $\{a, b\}$ *:*

1. $\mathbf{w}$ *est sturmien ;*

2. $\mathbf{w}$ *est mécaniquement irrationnel.*

À noter que si $\mathbf{w}$ est un mot mécanique de pente α, alors $\alpha = \lim_{n\to\infty} \frac{|\mathbf{w}[0..n]|_1}{|\mathbf{w}[0..n]|}$. Si par ailleurs, $(a_n)_{n\geq 0}$ est la suite qui apparaît dans la troisième assertion du théorème 8.8, alors la fraction continue de α vaut $[0; a_1, a_2, \ldots]$.

References

[1] J.-P. Allouche and T. Jonhson. Finite automata and morphisms in assisted musical composition. *Journal of New Music Research*, 24:97–108, 1995. voir http://www.lri.fr/∼allouche.

[2] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, 2003.

[3] E. Altman, B. Gaujal, and A. Hordjik. Balanced sequences and optimal routing. *Journal of the ACM*, 47:752–775, 2000.

[4] J. Berstel. Tracé de droites, fractions continues et morphismes itérés. In M. Lothaire, editor, *Mots*, pages 298–309. Hermès, Paris, 1990.

[5] J. Berstel and J. Karhumäki. Combinatorics on Words – A Tutorial. *Bull. Eur. Assoc. Theor. Comput. Sci.*, 78:178–228, 2003.

[6] J. Berstel, A. Lauve, C. Reutenauer, and F. Saliola. *Combinatorics on Words: Christoffel Words and Repetitions in Words*, volume 27 of *CRM Monograph Series*. American Mathematical Society, 2008.

[7] J. Berstel and D. Perrin. The origins of combinatorics on words. *European J. Combin.*, 28:996–1022, 2007.

[8] Jean Berstel and Patrice Séébold. A remark on morphic Sturmian words. *RAIRO Theoretical Informatics and Applications*, 28:255–263, 1994.

[9] V. Berthé, C. Holton, and L. Q. Zamboni. Initial powers of sturmian sequences. *Acta Arith.*, 122:315–347, 2006.

[10] E. Bombieri and J.E. Taylor. Which distributions of matter diffract ? An initial investigation. *Journal de Physique*, Colloque C3, supplément au n^{o}7, Tome 47(juillet):19–28, 1986.

[11] J. Cassaigne. On extremal properties of the Fibonacci word. *RAIRO Theoretical Informatics and Applications*, 42:701–715, 2008.

[12] M.G. Castelli, F. Mignosi, and A. Restivo. Fine and Wilf's theorem for three periods and a generalization of Sturmian words. *Theor. Comput. Sci.*, 218(1):83–94, 1999.

[13] E. M. Coven and G. A. Hedlund. Sequences with minimal block growth. *Mathematical Systems Theory*, 7:138–153, 1973.

[14] X. Droubay. Palindromes in the fibonacci word. *Inform. Process. Lett.*, 55:217–221, 1995.

[15] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theor. Comput. Sci.*, 255:539–553, 2001.

[16] X. Droubay and G. Pirillo. Palindromes and Sturmian words. *Theor. Comput. Sci.*, 223:73–85, 1999.

[17] A. Glen, F. Levé, and G. Richomme. Directive words of episturmian words. *RAIRO Theoretical Informatics and Applications*, 43(299-319), 2009.

[18] Amy Glen and Jacques Justin. Episturmian words: A survey. *RAIRO - Theoretical Informatics and Applications*, 43:403–442, 2009.

[19] R.L. Graham, D.E. Knuth, and O. Patashnik. *Mathématiques concrètes.* Vuibert, 2003 (deuxième édition).

[20] J. Justin. On a paper by Castelli, Mignosi, Restivo. *RAIRO Theoretical Informatics and Applications*, 34:373–377, 2000.

[21] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theor. Comput. Sci.*, 276(1-2):281–313, 2002.

[22] J. Karhumäki. On cube-free ω-words generated by binary morphisms. *Discret. Appl. Math.*, 5:279–297, 1983.

[23] F. Levé and G. Richomme. Quasiperiodic infinite words: some answers (column: Formal language theory). *Bull. Europ. Assoc. Theoret. Comput. Sci.*, 84:128–238, 2004.

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

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

[26] M. Lothaire. *Applied Combinatorics on Words*, volume 105 of *Encyclopedia of Mathematics.* Cambridge University Press, Cambridge, UK, 2005.

[27] F. Mignosi, A. Restivo, and S. Salemi. Periodicity and the golden ratio. *Theor. Comput. Sci.*, 204:153–167, 1998.

[28] M. Morse and G.A. Hedlund. Symbolic Dynamics II: Sturmian trajectories. *Amer. J. Math.*, 62(1):1–42, 1940.

[29] G. Paquin and L. Vuillon. A characterization of balanced episturmian sequences. *The Electronic Journal of Combinatorics*, 14:#R33, 2007.

[30] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics.* Springer, 2002. (V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel, editors).

[31] G. Richomme. Another characterization of Sturmian words (one more). *Bulletin of the European Association of Theoretical Computer Science*, 67:173–175, 1999.

[32] G. Richomme. Lyndon morphisms. *Bulletin of the Belgian Mathematical Society*, 10(5):761–786, 2003.

[33] G. Richomme. Conjugacy of morphisms and Lyndon decomposition of standard sturmian words. *Theoretical Computer Science*, 380 (numéro spécial Words'2005)(3):393–400, 2007. Version journal de l'article présenté à la conférence Words'2005.

[34] G. Richomme. A local balance property of episturmian words. In T. Harju, J. Karhumäki, and A. Lepistö, editors, *Developments in Language Theory, 11th International Conference, DLT 2007, Turku, Finland, July 3-6, 2007*, volume 4588 of *Lecture Notes in Computer Science*, pages 371–381. Springer, 2007.

[35] G. Richomme. SudoLyndon. *Bulletin of the European Assocciation of Theoretical Computer Sicence*, 92:143–149, 2007.

[36] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages*. Springer, 1997.

[37] P. Séébold. Fibonacci morphisms and Sturmian words. *Theor. Comput. Sci.*, 88:365–384, 1991.

[38] Zhi-Xiong Wen and Zhi-Ying Wen. Some properties of the singular words of the Fibonacci word. *European J. Combin.*, 15:587–598, 1994.