



ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
“Antonio Ruberti”
CONSIGLIO NAZIONALE DELLE RICERCHE

G. Pirillo

CONTRIBUTION A LA COMBINATOIRE DES
MOTS

R. 673, Dicembre 2007

Giuseppe Pirillo – Istituto di Analisi dei Sistemi ed Informatica “A. Ruberti”, Unità di
Firenze, Dipartimento di Matematica “U. Dini”, Viale Morgagni 67/A, 50134 Firenze,
Italy, and
Université de Marne-la-Vallée, 5 boulevard Descartes, 77454 Marne-la-Vallée Cedex 2,
France. pirillo@math.unifi.it.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti",
CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161

fax ++39-06-7716461

email: iasi@iasi.cnr.it

URL: <http://www.iasi.cnr.it>

UNIVERSITE PARIS DIDEROT - PARIS 7

Doctorat d'Etat ès Sciences

Giuseppe PIRILLO

CONTRIBUTION A LA COMBINATOIRE DES MOTS

Thèse dirigée par Jacques DESARMENIEN

Soutenue le 17-12-2007

JURY

**M. Jacques DESARMENIEN
M. Christian CHOFFRUT
M. Maxime CROCHEMORE
M. Emanuela FACHINI
M. Jean-Pierre GAZEAU
M. Irène GUESSARIAN
M. Alain LASCoux
M. Pierre-André PICON**

Remerciements

Je remercie Jacques Désarménien pour ses conseils scientifiques et pratiques dans la direction du travail de ma thèse ces dernières années. Il a su maintenir vivant mon rêve de soutenir une thèse d'état. Grazie, Jacques!

Je remercie Jean Berstel. En lisant ses articles sur la combinatoire des mots, toujours clairs et intéressants, j'ai beaucoup appris. Grazie Jean!

Merci aussi à Christian Choffrut. Christian connaissait parfaitement les importants travaux de Justin sur les semigroupes répétitifs. Il m'a toujours prodigué ses précieux conseils. Sans lui, ma thèse de troisième cycle n'aurait pas pu voir le jour. Grazie, Christian!

Je remercie aussi Maxime Crochemore, que je connais depuis de longues années, pour les nombreuses et fructueuses discussions que nous avons eues. Grazie Maxime!

Merci aussi à Emanuela Fachini. J'ai connu Emanuela à Cortone lorsqu'elle et moi étions étudiants de Dominique Perrin. Toute ma gratitude lui est acquise pour avoir accepté de rédiger un rapport sur mon travail. Grazie, Emanuela!

Merci à Jean-Pierre Gazeau de se prêter au rôle de "membre étranger à la spécialité strictement définie du candidat". C'est pour moi un honneur de l'avoir dans mon jury. Grazie Jean-Pierre!

Je remercie aussi Irène Guessarian. J'ai fait sa connaissance à mon arrivée au LITP. Elle a toujours été de bon conseil. Elle a aussi joué un rôle capital dans les démarches qui m'ont permis de soutenir ma thèse. Grazie, Irène!

Des remerciements particulièrement chaleureux à Jacques Justin. J'ai eu le privilège d'être son unique élève. Il m'a fait l'honneur de m'offrir son amitié. Sans son aide patiente et constante et, oserais-je dire, sans sa sollicitude paternelle, la présente thèse n'aurait jamais été écrite. Grazie, Jacques!

À Alain Lascoux, pour m'avoir offert l'hospitalité dans le bureau qu'il partageait avec notre regretté Marcel-Paul Schützenberger, bien cordialement merci. Grazie, Alain!

Toute ma gratitude à Dominique Perrin. J'ai été son étudiant à Cortone et c'est à lui que je dois ma première invitation en France. Il a initié une aventure scientifique qui a changé ma vie professionnelle. Grazie, Dominique!

À Pierre-André Picon, tous mes remerciements pour avoir pris la peine de rédiger un rapport et merci également pour les longues discussions que nous avons eues sur la culture et la civilisation françaises, qui m'ont beaucoup

CONTRIBUTION A LA COMBINATOIRE DES MOTS

appris. Grazie Pierre-André!

Je remercie Stefano Varricchio. Notre amitié s'est renforcé au fil du temps lors de la rédaction de plusieurs articles en collaboration Grazie Stefano!

Je remercie finalement le laboratoire LITP, maintenant LIAFA, et l'IGM pour l'amicale atmosphère dont j'ai profité et qui m'a permis de travailler sereinement. Merci aussi à la "Maison H. Heine (CIUP, Paris)" pour l'accueil et l'environnement culturel qu'elle m'a offerts ces trente dernières années.

Thèse de Doctorat ès Sciences de Giuseppe Pirillo

Table des matières

Remerciements	I
Table des matières	III
Avant-propos	V
Brève histoire de la combinatoire des mots	VII
1 Introduction	1
2 [154] Giuseppe Pirillo, <i>Inequalities characterizing standard Sturmian and episturmian words.</i>	43
3 [91] Jacques Justin, Giuseppe Pirillo, <i>Episturmian words : shifts, morphisms and numeration systems.</i>	61
4 [46] Xavier Droubay, Jacques Justin, Giuseppe Pirillo, <i>Episturmian words and some constructions of de Luca and Rauzy.</i>	81
5 [88] Jacques Justin, Giuseppe Pirillo, <i>Fractional powers in Sturmian words.</i>	97
6 [142] Giuseppe Pirillo, <i>Fibonacci numbers and words.</i>	111
7 [111] Filippo Mignosi, Giuseppe Pirillo, <i>Repetitions in the Fibonacci infinite word.</i>	123
8 [135] Giuseppe Pirillo, <i>Sur les produits directs de semi-groupes répétitifs.</i>	129
9 [92] Jacques Justin, Giuseppe Pirillo, Stefano Varricchio, <i>Unavoidable regularities and finiteness conditions for semigroups.</i>	133
10 Références	139

Avant-propos

Il faut cultiver notre jardin.

Dans *le Petit Prince*, Antoine de Saint-Exupéry dit : *Mais les yeux sont aveugles. Il faut chercher avec le cœur.* Cela est bien vrai de toutes les recherches, qui sont le devoir d'un être humain, y compris les recherches scientifiques.

Nous avons rencontré pour la première fois Marcel-Paul Schützenberger en 1975 à Perugia et Dominique Perrin en 1977 à Cortona quand nous avons été étudiant d'un cours d'été. Peu après notre arrivée à Paris, pendant l'année académique 1977-1978 où nous étions assistant associé chargé de travaux dirigés du cours de Marcel-Paul Schützenberger, nous avons rencontré Jacques Justin.

Pendant une très courte période de temps beaucoup d'événements ont changé notre vie.

Pour ce qui concerne notre travail, nous avons eu le privilège d'être accueilli dans un groupe prestigieux de recherche et nous avons eu en plus un guide personnel, Jacques Justin, qui nous a conduit par la main dans nos premiers pas de chercheur en informatique théorique et nous a toujours fait profiter de ses bons conseils.

Ce fut le début d'une aventure scientifique passionnante.

Dès notre arrivée à Paris, et même avant de bien comprendre ce qu'ils faisaient exactement, nous avons noté la grande passion pour le travail qui régnait parmi les chercheurs du groupe de Marcel-Paul Schützenberger. Nous nous sommes aussi rendu compte du grand nombre de talents qui y appartenaient et du poids remarquable de leurs connaissances scientifiques. Nous aurions dû non seulement garder intacte la passion pour les mathématiques mais aussi élargir, approfondir et élever nos connaissances, c'est-à-dire prendre très sérieusement l'impératif éternel de Voltaire.

Brève histoire de la combinatoire des mots

*Pour le développement de la science logique il sera important,
sans considération pour les applications possibles,
de trouver de grands domaines pour la spéculation
sur des problèmes difficiles.*

(traduction libre d'une proposition d'Axel Thue)

*Même dans les parties les plus élégantes des mathématiques pures abstraites,
la preuve d'un beau théorème se réduit étonnamment souvent
à un lemme combinatoire très terre-à-terre sur les mots.*

(traduction libre d'une proposition de Roger Lyndon,
nous avons traduit *linear arrays of symbols* par *mots*)

La publication de M. Lothaire *Combinatorics on words*, 1983, [103], est l'acte de naissance de la combinatoire de mots qui, avec les publications de M. Lothaire *Algebraic combinatorics on words*, 2002, [106], et *Applied combinatorics on words*, 2005, [107], est devenue une jeune fille splendide.

Bien qu'il y ait déjà eu d'importantes contributions sur des mots (par Hedlund et Morse, [113], [114], par Thue [185], [186] et même par d'autres avant eux) cette discipline mathématique et informatique est reconnue comme telle depuis à peine vingt années. Désormais, dans la dernière classification de *Mathematical Reviews*, la combinatoire des mots a sa propre section, et une conférence internationale *Words* se tient tous les deux ans. En plus, c'est une opinion largement partagée (ou plutôt une évidence) que pendant les deux dernières décennies la recherche sur la combinatoire des mots s'est énormément développée.

Marcel-Paul Schützenberger, *the main architect* de la théorie des codes à longueur variable (Berstel et Perrin, [17]), a, avec Chomsky, *complètement modifié l'approche de la théorie des langages formels* (Berstel et Boasson [22]). La combinatoire des mots est aussi sa création : le "poly rouge" [179], contenant le cours qu'il a donné à l'Université de Paris en 1966 qui fut noté par J.-F. Perrot, est un précurseur du premier livre de Lothaire.

La combinatoire des mots est strictement liée à la théorie des algorithmes, à la dynamique symbolique, à la théorie des nombres,... Aussi les applications à la biologie et à l'étude des langues naturelles sont-elles à souligner. Plusieurs livres (ou chapitres de volumes collectifs) sont parus récemment où tel ou tel aspect de la combinatoire des mots est mis en lumière. Citons, sans prétention

CONTRIBUTION A LA COMBINATOIRE DES MOTS

d'être exhaustif, Berstel et Perrin [17], Allouche et Shallit [2], Pytheas-Fogg (nom de plume du groupe de Marseille) [164], Crochemore et Rytter [32], [33], Choffrut et Karhumäki [28], de Luca et Varricchio [42], [43] et J. Sakarovitch [177].

Thèse de Doctorat ès Sciences de Giuseppe Pirillo

1 Introduction

Nous nous intéressons principalement :

a) à un cas spécial de régularités inévitables, la répétitivité introduite et étudiée par Justin [63], [64], [69], [58], [59], [60], [61], [62], [65], et liée aux résultats de Thue [185], [186], au théorème de Ramsey [165], et au théorème de Van der Waerden [187], [188] ;

b) au théorème de Shirshov [182], à la notion de semi-groupe permutable [170], et à d'autres conditions de finitude des semi-groupes finiment engendrés.

c) aux mots Sturmien (dont l'étude, très ancienne, a récemment progressé surtout par les nombreuses et intéressantes contributions de Berstel, voir, par exemple, [12], [11], [14], [18], [19], [7], [5] et spécialement [106] et sa bibliographie) et aux mots épisturmiens [46].

Nous finissons cette introduction avec une référence très courte à quelques-unes de nos contributions récentes à la biologie théorique. Nous mentionnons également certaines contributions à la didactique de la mathématique et de l'informatique qui a toujours été une de nos préoccupations scientifiques.

Cette thèse n'est pas un survol (tour d'horizon) de l'état de nos recherches, et ne contient pas une analyse détaillée des résultats déjà obtenus, mais c'est un moment de réflexion pour mieux organiser nos recherches futures.

Parmi nos articles, nous en avons choisi environ 60 dont les résumés se trouvent juste après cette introduction. Nous avons choisi aussi 8 travaux dont le texte complet fait partie intégrante de cette thèse. Il s'agit de :

[154] Giuseppe Pirillo, *Inequalities characterizing standard Sturmian and episturmian words.*

[91] Jacques Justin, Giuseppe Pirillo, *Episturmian words : shifts, morphisms and numeration systems.*

[46] Xavier Droubay, Jacques Justin, Giuseppe Pirillo, *Episturmian words and some constructions of de Luca and Rauzy.*

[88] Jacques Justin, Giuseppe Pirillo, *Fractional powers in Sturmian words.*

[142] Giuseppe Pirillo, *Fibonacci numbers and words.*

[111] Filippo Mignosi, Giuseppe Pirillo, *Repetitions in the Fibonacci infinite word.*

[135] Giuseppe Pirillo, *Sur les produits directs de semi-groupes répétitifs.*

[92] Jacques Justin, Giuseppe Pirillo, Stefano Varricchio, *Unavoidable regularities and finiteness conditions for semigroups.*

Définitions préliminaires

En utilisant la terminologie et les notations de Lothaire [107], [106], [105], [104] et [103] (qui sont pratiquement celles du mythique *poly rouge* de Schützenberger [179]) nous notons A^* (resp. A^+) le *monoïde libre* (resp. *semi-groupe libre*) sur l'*alphabet* A . Notons que A est généralement fini et souvent réduit à deux lettres. Un élément de A est une *lettre* et un élément de A^* est un *mot*. Nous notons $|u|$ la *longueur* d'un mot u (il existe un seul mot de A^* de longueur 0, le mot vide noté ϵ).

Un mot u est un *facteur* d'un mot v s'ils existe deux mots u' et u'' tels que $v = u'uu''$. Nous disons que u est un *facteur gauche* (ou *préfixe*) de v si $u' = \epsilon$, et que u est un *facteur droit* (ou *suffixe*) de v si $u'' = \epsilon$.

Pour tout mot $w \in A^*$, de longueur n , notons $w(i)$ la i -ème lettre de w et notons $w(i, j) = w(i) \dots w(j)$, ($1 \leq i, j \leq n$) le facteur de w qui commence au rang i et se termine au rang j . Le couple (i, j) est une *occurrence* du facteur u dans v .

Si $u = u(1)u(2) \dots u(n-1)u(n)$ est un mot alors

$$\tilde{u} = u(n)u(n-1) \dots u(2)u(1)$$

est son *image miroir*. Si un mot u coïncide avec son image miroir, il est un *palindrome*.

Un *mot infini à droite* ou simplement *mot infini* sur un alphabet A est une application de l'ensemble des entiers naturels, N , dans A . Un *mot bi-infini* est une application de l'ensemble des entiers Z dans A .

Nous écrivons selon le cas

$$s = s(0)s(1) \dots s(i) \dots$$

ou

$$s = \dots s(-i) \dots s(-1)s(0)s(1) \dots s(1) \dots$$

où $s(i)$ est la lettre “de rang i ” dans s .

Note. Pour les mots infinis (à droite), le choix du début de la numérotation dépend des auteurs et commence souvent au rang 1.

Soit $s = s(0)s(1) \dots s(i) \dots$ un mot infini, nous notons $s(i, \infty)$ l’unique mot infini tel que $s = us(i, \infty)$ où u est égal à $s(0)s(1) \dots s(i-1)$.

Si x est un mot (fini ou infini), $F(x)$ est l’ensemble de ses facteurs.

Un mot infini s est *ultimement périodique* s’il existe deux entiers i_0 et $p \geq 1$ tels que, pour tout $i \geq i_0$, $s(i+p) = s(i)$. Le mot est *périodique* si $i_0 = 0$.

Un facteur s d’un mot infini r est *spécial à droite* s’il existe deux occurrences de s dans r qui sont suivies l’une d’une lettre et l’autre d’une autre lettre. D’une façon analogue on définit le facteur *spécial à gauche*.

Soit k un entier positif et u un mot. Nous disons que u est une *puissance* k -ième s’il existe un mot v non vide tel que $u = v^k$.

Définition 1 Soit A un alphabet et soit X un sous-ensemble de A^+ . Nous disons que X est un code si, pour tout

$$x_1, \dots, x_n, x'_1, \dots, x'_m \in X, n, m \geq 1,$$

la condition

$$x_1 \cdots x_n = x'_1 \cdots x'_m$$

implique $n = m$, et, pour tout $i = 1, \dots, n$,

$$x_i = x'_i.$$

Définition 2 Un code X est *préfixe* (resp. *suffixe*) si les conditions “ $u, v \in X$ ” et “ u préfixe (resp. suffixe) de v ” impliquent $u = v$. Un code est *bi-préfixe* s’il est préfixe et suffixe.

Définition 3 Soit A un alphabet et soit X un sous-ensemble de A^+ . Nous disons que X est sans ponctuation si, pour tout $y \in X$ et $u, v \in A^*$ tels que $uyv = x_1 \cdots x_n$ avec $x_1, \dots, x_n \in X$, $n \geq 1$, alors $u, v \in X^*$.

Définition 4 Soit A un alphabet et soit X un sous-ensemble de A^+ . Nous disons que X est circulaire si, pour tout

$$x_1, \dots, x_n, x'_1, \dots, x'_m \in X, n, m \geq 1, p \in A^*, s \in A^+,$$

les conditions

$$sx_2 \cdots x_n p = x'_1 \cdots x'_m$$

et

$$x_1 = ps$$

impliquent $n = m$, $p = \epsilon$ (mot vide) et, pour tout $i = 1, \dots, n$,

$$x_i = x'_i.$$

Ordre lexicographique. Supposons que l'alphabet A soit muni d'un ordre total $<$. L'ordre lexicographique sur A^* est ainsi défini : soient $u, v \in A^*$, nous disons que u précède v dans l'ordre lexicographique et nous écrivons $u \leq v$ si ou bien u est préfixe de v ou bien il existe $a, b \in A$, $a < b$, et $r, s, t \in A^*$ tels que $u = ras$ and $v = rbt$.

Définition 5 *Un mot infini est récurrent si et seulement si tous ses facteurs ont une infinité d'occurrences en lui. Un mot infini t est uniformément récurrent si et seulement si pour tout $u \in F(t)$ il existe un entier k tel que tout facteur de t de longueur k contient une occurrence de u .*

Le mot de Fibonacci, le mot de Thue et la majorité des mots étudiés dans cette thèse sont des mots récurrents. Souvent certaines propriétés des mots infinis sont prouvées, explicitement ou implicitement, à l'aide des propriétés des mots récurrents (voir, par exemple, nos preuves du Théorème de Shirshov [141] et [80]).

Résultats classiques.

Les théorèmes de Ramsey et de van der Waerden sont des généralisations profondes du principe des tiroirs.

Un cas particulier simple mais non trivial du théorème de Ramsey est le suivant : *étant données six personnes alors ou bien trois d'entre elles se connaissent l'une l'autre ou bien trois ne se connaissent pas deux à deux.*

La version qui nous sera utile (et qui a plusieurs variantes) est la suivante (nous notons $P_r(X)$ l'ensemble des parties à r éléments de l'ensemble X , c'est-à-dire $\{Y | Y \subset X, |Y| = r\}$) :

Théorème 1 (Ramsey). *Soient r, k, n des entiers positifs tels que $k \geq r \geq 1$. Il existe un entier $R(r, k, n)$, tel que pour tout ensemble X de cardinal $R(r, k, n)$ et toute partition Y_1, Y_2, \dots, Y_n de $P_r(X)$ en n classes, il existe un sous-ensemble Δ de X et une classe Y_i tels que $|\Delta| = k$, $P_r(\Delta) \subset Y_i$.*

La démonstration de ce théorème est très clairement exposée dans l'article original du mathématicien anglais [165] ou, par exemple, dans [176] ou dans [54]. En combinatoire des mots sont souvent utiles plusieurs corollaires comme par exemple le suivant :

Corollaire 1 *Soit*

$$\varphi : A^+ \rightarrow F$$

une application de A^ dans un ensemble fini F de cardinal n . Pour tout k entier positif, tout mot $w \in A^*$ de longueur $R(2, k+1, n)$ contient un facteur $w_1 w_2 \dots w_k$ avec, pour tout i , $1 \leq i \leq k$, $w_i \in A^+$ et pour tout couple $(i, i'), (j, j')$, $1 \leq i \leq i' \leq k$, $1 \leq j \leq j' \leq k$,*

$$\varphi(w_i \dots w_{i'}) = \varphi(w_j \dots w_{j'}).$$

Le Théorème de van der Waerden est prouvé dans [187] et constitue un important chapitre de la *Ramsey theory* (voir [176], [54], [116]). Il a été prouvé en généralisant de 2 à un entier n quelconque une conjecture de Baudet (voir [188] où van der Waerden lui-même donne des explications très intéressantes sur le processus de découverte du théorème qui porte son nom).

Théorème 2 (van der Waerden). *Toute partition de N en un nombre fini de classes comporte une classe qui contient une progression arithmétique de longueur arbitraire.*

Une forme plus adaptée à la combinatoire des mots s'obtient en introduisant la notion de *cadence*. On dira qu'un mot comporte une cadence arithmétique d'ordre k s'il contient une lettre répétée k fois à des rangs équidistants. De façon plus formelle nous pouvons poser la

Définition 6 *On dit qu'un mot $w \in A^+$ contient une cadence d'ordre k s'il existe une factorisation $w = w' a w_1 a w_2 a \dots a w_{k-1} a w''$ avec $a \in A$ et $|w_i| = |w_j|$ (tout mot de longueur supérieure au nombre de lettres de l'alphabet sur lequel il est écrit admet une cadence d'ordre 2 : c'est le principe des tiroirs).*

Le théorème de van der Waerden peut ainsi s'énoncer en termes de cadences

Proposition 1 *Il existe un entier $W(h, k)$ tel que pour tout alphabet A de h lettres tout mot de longueur au moins égale à $W(h, k)$ contient une cadence d'ordre k .*

Le théorème de Van der Waerden a joué un rôle très important dans les premiers travaux de Justin et dans les travaux que nous avons faits ensuite en collaboration avec lui.

Répétitivité.

Etant donnée une application $\varphi : A^+ \rightarrow E$ de A^+ dans un ensemble E nous cherchons dans un mot $w \in A^+$ les facteurs du type $w_1 w_2 \dots w_k$ tels que $\varphi(w_1) = \varphi(w_2) = \dots = \varphi(w_k)$. Si un tel facteur existe dans tout mot suffisamment long, nous dirons que φ est répétitive. Rappelons, plus formellement, les définitions suivantes :

Définition 7 *Etant donné une application $\varphi : A^+ \rightarrow E$ de A^+ dans un ensemble E et un entier positif k , nous dirons qu'un mot $w \in A^+$ est une k -puissance modulo φ s'il existe une factorisation $w = w_1 w_2 \dots w_k$ telle que $\varphi(w_1) = \varphi(w_2) = \dots = \varphi(w_k)$. Une k -puissance modulo φ est dite uniforme si de plus on a $|w_1| = |w_2| = \dots = |w_k|$. Les mots w_1, w_2, \dots, w_k seront dits composants de la k -puissance modulo φ . Nous dirons qu'un mot w contient un k -puissance modulo φ s'il a un facteur qui est une k -puissance modulo φ .*

Définition 8 *Une application $\varphi : A^+ \rightarrow E$ de A^+ dans E est répétitive (resp. uniformément répétitive) si pour tout entier positif k il existe un entier positif l , tel que tout mot $w \in A^+$ de longueur l contient une k -puissance modulo φ (resp. une k -puissance modulo φ uniforme).*

Définition 9 *Un semi-groupe S est dit répétitif (resp. uniformément répétitif) si pour tout alphabet fini A , tout morphisme $\varphi : A^+ \rightarrow S$ est répétitif (resp. uniformément répétitif).*

Le résultat suivant, prouvé à l'aide du Théorème de van der Waerden, est dû à Justin (voir [64]).

Proposition 2 *Le semi-groupe libre A^+ est répétitif si et seulement si $|A| = 1$.*

Comme, à cause de la construction de Thue d'un mot infini sur un alphabet à deux lettres qui est sans cube (voir [185] et [186] et voir aussi après pour des définitions formelles) les semi-groupes libres ne sont pas en général répétitifs, l'étude de ces semi-groupes tient une place importante en combinatoire des mots.

Plusieurs des résultats de Justin (par exemple la répétitivité de tout produit direct du (semi)-groupe des entiers par un semi-groupe fini, et la caractérisation des semi-groupes répétitifs commutatifs) sont des généralisations du théorème de Van der Waerden.

Un résultat de Justin a particulièrement attiré notre attention : un profond théorème portant sur les semi-groupes à générations bornées. En utilisant une propriété de factorisation des mots canoniquement associés aux éléments de ces semi-groupes, nous avons présenté dans [125] une preuve plus simple que l'originale.

Plusieurs variantes de la propriété de répétitivité ont été examinées dans [73]. Dans [72] nous avons considéré une partition en un nombre fini de classes de la suite t de Thue (celle qui est écrite sur un alphabet à deux lettres et qui est sans cube, voir après) et nous avons prouvé que *la suite t ne contient pas trois facteurs de même longueur, consécutifs et dans la même classe*. Dans [135] nous avons prouvé le résultat suivant :

Proposition 3 *Si S et T sont deux semi-groupes finiment engendrés dont le produit direct est répétitif alors ou bien S est fini ou bien T est fini.*

Dans [126] (ou [125]) il est prouvé que tout *semi-groupe fini est uniformément répétitif* et dans [92] il est prouvé que *l'uniforme répétitivité est une condition de finitude pour les semi-groupes finiment engendrés*. D'autres résultats sur les semi-groupes répétitifs sont dans [83] et dans [162].

Deux mots infinis remarquables.

En ce qui concerne le mot de Thue, nous nous limitons pratiquement dans ce paragraphe à la définition. Une présentation détaillée se trouve dans [10] et dans [103] où il y a aussi des remarques historiques très intéressantes.

Mots de Thue. Soit μ l'application

$$\mu : A^+ \rightarrow E$$

définie par

$$\mu(a) = ab \quad \mu(b) = ba.$$

Comme l'image soit de a soit de b n'est pas vide, et comme a (resp. b) est préfixe de $\mu(a) = ab$ (resp. $\mu(b) = ba$) en itérant μ à partir de a (resp. b) on obtiens les mots infinis

$$t = abbabaab\dots$$

$$t' = baababba\dots$$

Dans la littérature le mot t est appelle mot de Thue (et aussi de Thue-Morse). Le mot t est sans cube, c'est-à-dire sans facteurs de la forme uuu . En utilisant le mot t on peut construire un mot sur un alphabet de 3 lettres qui est sans carré, c'est-à-dire sans facteurs de la forme uu . Ce mot, disons t'' , s'obtient plus précisément en utilisant le codage $0 = abb$, $1 = ab$ et $2 = a$ et en regroupant les facteurs de t de façon conséquentes :

$$t = (abb)(ab)(a)(abb)(a)(ab)(abb)(ab)\dots$$

$$t = 012021012102\dots$$

Mot de Fibonacci. En raison de sa facilité de construction et de "l'évolution rapide de ses préfixes" le mot de Fibonacci est une source d'idées pour des résultats valables pour la classe plus générale des mots Sturmien.

Evidemment, toutes les propriétés du mot de Fibonacci ne se prêtent pas aussi bien à une généralisation.

Nous donnons ici quelques exemples de généralisations qui ont abouti à un résultat que nous avons effectivement pu prouver.

Définition et propriétés élémentaires du mot de Fibonacci. Soit (voir [97] et, par exemple, [7]; voir aussi [115]) $\varphi : \{a, b\}^* \rightarrow \{a, b\}^*$ un morphisme ainsi défini sur $\{a, b\}$:

$$\varphi(a) = ab, \varphi(b) = a.$$

Définissons le n -ème mot fini de Fibonacci f_n de la façon suivante : $f_0 = b$ et, pour tout $n \geq 0$,

$$f_{n+1} = \varphi(f_n).$$

En particulier, nous avons : $f_1 = a$, $f_2 = ab$, $f_3 = aba$, $f_4 = abaab$, $f_5 = abaababa$, $f_6 = abaababaabaab$, $f_7 = abaababaabaababaababa \dots$. Il est clair que, pour tout $n \geq 2$, f_n est le produit (concaténation) $f_{n-1}f_{n-2}$ de f_{n-1} et f_{n-2} . Aussi, pour tout $n \geq 0$, $|f_n|$ est le n -ième élément F_n de la suite de nombres de Fibonacci $F_0 = 1$, $F_1 = 1$, $F_2 = 2$, $F_3 = 3$, $F_4 = 5$, $F_5 = 8$, $F_6 = 13$, $F_7 = 21 \dots$. Notons que, pour tout $n \geq 1$, f_n est un préfixe de f_{n+1} .

Il existe un unique mot infini, le mot infini de Fibonacci f , tel que, pour tout $n \geq 1$, f_n est un préfixe de f (voir [7]) et nous avons

$$f = abaababaabaababaabaababaabaababaabaababa \dots$$

Afin de souligner l'importance du mot de Fibonacci nous rappelons que Dominique Perrin présente ce mot déjà dans le chapitre 1 (Words) du livre de Lothaire [115] (en fait le mot de Fibonacci montre que le célèbre résultat [49] de Fine et Wilf est optimal).

Pour tout $n \geq 2$, notons g_n le produit $f_{n-2}f_{n-1}$ et notons h_n le plus long préfixe commun de f_n et g_n . En particulier, nous avons : $g_2 = ba$, $g_3 = aab$, $g_4 = ababa$, $g_5 = abaabaab$, \dots and $h_2 = 1$, $h_3 = a$, $h_4 = aba$, $h_5 = abaaba \dots$. Notons aussi que $bb, aaa \notin F(f)$.

Rappelons ici certains lemmes qui nous ont été très utiles dans notre travail. Le premier appartient au folklore et est très facile à prouver. Le point i) est la “*propriété de quasi-commutativité*” que Knuth Morris et Pratt ont utilisée dans [99].

Lemme 1 *Pour tout $n \geq 2$,*

- i) $f_n = h_nxy$ et $g_n = h_nyx$, où $x, y \in \{a, b\}$, $x \neq y$, et si n est pair alors $xy = ab$, et si n est impair alors $xy = ba$;
- ii) $|h_n| = F_n - 2$;
- iii) h_n est un facteur spécial;
- iv) $h_{n+3} = h_{n+1}xyh_nyxh_{n+1}$, où $x, y \in \{a, b\}$, $x \neq y$;
- v) h_n est un palindrome;
- vi) $h_{n+2} = f_n h_{n+1} = h_{n+1} \tilde{f}_n = h_n \tilde{f}_{n+1} = f_{n+1} h_n$;
- vii) pour tout entier $m \geq 0$, h_n est préfixe et suffixe de h_{n+m} .
- viii) si $v \in F(f)$ alors $\tilde{v} \in F(f)$.

Les lemmes suivants appartiennent aussi au folklore (pour tout $k \geq 1$, notons $s^{[k]}$ l'image miroir du préfixe de f ayant longueur k).

Lemme 2 Pour tout $k \geq 1$:

- i) l'unique facteur spécial de f de longueur k est $s^{[k]}$;
- ii) f a exactement $k + 1$ facteurs de longueur k .

Lemme 3 Si $u, v \in F(f)$ et $|u| = |v|$ alors $||u|_a - |v|_a| \leq 1$.

Lemme 4 Le mot infini de Fibonacci est uniformément récurrent et non ultimement périodique.

Lemme 5 Soit v un préfixe de f . Les conditions suivantes sont équivalentes :
i) v est un palindrome ; ii) $v = h_n$ pour quelque $n \geq 3$.

Les Lemmes 1-5 avec les Lemmes 6-7 que nous rappelons ci-dessous ont été très utiles pour l'étude des puissances fractionnaires dans le mot de Fibonacci ([142], voir paragraphe suivant).

Soient $u, v, w, z, z' \in F(f)$, $v \neq \epsilon$. Nous disons que (u, v, w) est un chevauchement non vide de z et z' si $uv = z$, $vw = z'$ and $uvw \in F(f)$.

Lemme 6 Soit $n \geq 3$. Alors

- i) $(f_{n-1}, f_{n-2}, g_{n-1})$ est l'unique chevauchement non vide de f_n et f_n ;
- ii) $(f_{n-1}, f_{n-2}, f_{n-1})$ est l'unique chevauchement non vide de f_n et g_n ;
- iii) $(f_{n-2}, f_{n-1}, f_{n-2})$ est l'unique chevauchement non vide de g_n et f_n ;
- iv) il n'y a pas de chevauchement non vides de g_n et g_n .

Lemme 7 Soit $n \geq 5$. Il y a exactement deux chevauchements non vides de h_n et h_n , précisément

$$(f_{n-1}, h_{n-2}, \tilde{f}_{n-1})$$

et

$$(f_{n-2}, h_{n-1}, \tilde{f}_{n-2}).$$

Puissances fractionnaires.

A tout mot (fini) non vide w , est naturellement associé un mot infini périodique p_w , ainsi défini : $p_w(1, |w|) = w$ et, pour tout $i \geq 1$, $p_w(i + |w|) =$

$p_w(i)$. Nous disons qu'un mot u est une *puissance fractionnaire* du mot (fini!) w si u est un préfixe de p_w . Nous disons que w est la *base* et $k = |u|/|w|$ l'*exposant* de la puissance fractionnaire et nous écrivons $u = w^k$. En général k est un nombre rationnel, mais si k est un nombre entier nous avons la notion habituelle de puissance d'un mot (voir définitions préliminaires). Nous considérons seulement les puissances d'exposant supérieur ou égal à 1.

Soit t un mot infini, $u \in F(t)$, r un rationnel et $u^r \in F(t)$. Nous disons que la puissance fractionnaire u^r est *maximale* si pour tout mot v tel que $|u| = |v|$ et pour tout rationnel s , si u^r est un facteur de v^s et $v^s \in F(t)$, alors $u = v$ and $r = s$. Nous disons qu'une occurrence (i_0, j_0) de u^r dans t est *maximale* si pour tout $v \in F(t)$ tel que $|u| = |v|$ et pour tous i_1, j_1 tels que $i_1 \leq i_0 \leq j_0 \leq j_1$, si (i_1, j_1) est une occurrence d'une puissance de v dans f alors $u = v$, $i_0 = i_1$ et $j_0 = j_1$. Notons qu'une puissance peut avoir une occurrence maximale bien qu'elle ne soit pas maximale. Par exemple, dans le mot infini de Fibonacci f les facteurs a , $(ab)^{3/2}$, $(aba)^2$, et $(abaab)^{11/5}$ ne sont pas puissances maximales mais ils ont des occurrences maximales dans f .

A l'aide des Lemmes 1-7 déjà rapelés, nous avons caractérisé les bases des puissances fractionnaires maximales ayant s comme exposant dans les cinq cas suivants :

$$(2 + \Phi)/2 < s < 2, \quad s = 2, \quad 2 < s < 3, \quad s = 3, \quad 3 < s.$$

Ces résultats nous ont permis de prouver la Proposition 4 suivante (Proposition 9 de [142]) laquelle, par la Proposition 5 (Proposition 10 dans [142]) est optimale.

Proposition 4 *Soit r un nombre rationnel plus grand que $(2 + \Phi)/2$ et $u \in F(f)$. Si $u^r \in F(f)$, alors il existe $n \geq 1$ tel que u est un conjugué de f_n et, de plus, toute occurrence de u^r est contenue dans une occurrence maximale de $(f_n)^s$ pour quelque $s \in [2, 2 + \Phi[$.*

Proposition 5 *Pour tout $\epsilon > 0$ il existe un nombre rationnel $t \in [(2 + \Phi)/2 - \epsilon, (2 + \Phi)/2[$ et un mot w , tel que w^t a une occurrence maximale dans f et $|w| \neq F_n$ pour tout $n \geq 3$.*

Cette étude nous a permis d'améliorer les résultats suivants qui sont dus à Karhumäki ([95], Proposition 6) et à Séebold ([180], Proposition 7). Signalons

aussi les Propositions 8 et 9 que nous avons obtenues en collaboration avec Mignosi [111].

Proposition 6 *Pour tout $u \in F(f)$ non vide on a $u^4 \notin F(f)$.*

Proposition 7 *Pour tout $u \in F(f)$ non vide, si $u^2 \in F(f)$ alors il existe $n \geq 1$ tel que u est un conjugué de f_n .*

Proposition 8 *Le mot de Fibonacci ne contient pas en préfixe de puissance fractionnaire d'exposant supérieur à $1 + (1 + \Phi)/2$. Plus précisément, si vvu est une puissance fractionnaire et aussi un préfixe de f alors $v = f_n$ pour quelque n et $|vvu| \leq F_n + F_n + F_{n-1} - 2$.*

Proposition 9 *Le mot de Fibonacci ne contient pas de puissances fractionnaires d'exposant supérieur à $2 + (1 + \Phi)/2$ et, pour tout $\epsilon > 0$ il contient une puissance fractionnaire d'exposant supérieur à $2 + (1 + \Phi)/2 - \epsilon$.*

Cette étude sur les puissances fractionnaires du mot de Fibonacci a été aussi utile pour la recherche analogue concernant les mots Sturmien.

Certaines factorisations du mot de Fibonacci ont non seulement un intérêt en elles-même mais aussi suggèrent, comme on le verra, des généralisations. Notons les propositions suivantes

Proposition 10 *Soit*

$$f = u_0 u_1 \dots u_i \dots$$

la factorisation de f tel que $|u_i| = F_{2i+1}$. Alors $\{u_i \mid i \geq 0\}$ est un code préfixe.

Proposition 11 *Soit*

$$f = v_0 v_1 \dots v_i \dots$$

la factorisation de f tel que $|v_i| = F_{2(i+1)}$. Alors $\{v_i \mid i \geq 0\}$ est un code préfixe.

Proposition 12 Soit $n \geq 4$. Soit $f = w_0 w_1 \dots w_i \dots$ la factorisation de f telle que $|w_0| = F_n + F_{n-2} - 1$ et, pour tout $i \geq 1$, $|w_i| = F_{n+2(i-2)-1} + 2F_{n+2(i-1)}$. Alors $\{w_i \mid i \geq 0\}$ est un code bi-préfixe et $w_i \neq w_j$ pour tout entier positif $i \neq j$.

Avant de terminer ce paragraphe rappelons les trois propriétés suivantes du mot de Fibonacci qui ont été très utiles pour l'étude des mots Sturmien :

1) Rappelons que h_n est le plus long préfixe commun de f_n et de g_n . A partir de la propriété de quasi-commutativité (point i) du Lemme 1) on conclut immédiatement que, pour tout $n \geq 2$, $h_n ab$ est conjugué de $h_n ba$ (en formule $h_n ab \sim h_n ba$). Par exemple, $abaab \sim ababa, \dots$. Nous verrons dans le paragraphe concernant les mots Sturmien que les préfixes palindromiques w des mots Sturmien standard sont caractérisés par la relations $wab \sim wba$.

2) Il est clair que la première occurrence du palindrome a est en position centrale dans le préfixe palindromique a de f ! On vérifie simplement que la première occurrence du palindrome b est en position centrale dans le préfixe palindromique $h_4 = aba$ de f . De même la première occurrence de aa est centrale dans $h_5 = abaaba$. Cette simple remarque conduit à la définition d'une propriété qui est à l'origine de l'étude des mots épisturmien [46].

3) Il est encore simple de vérifier que $aa, aab, aaba$ sont respectivement le mot le plus petit dans l'ordre lexicographique parmi les facteurs de f de longueur 2, 3, 4 ... Cette remarque est à l'origine des résultats de [90], [154], [52] sur les mots Sturmien et épisturmien.

Théorème de Shirshov et permutabilité

Théorème de Shirshov. Soit A un alphabet muni d'un ordre total $<$ et soit, encore noté $<$, l'ordre lexicographique induit sur A^* par l'ordre $<$ sur l'alphabet A . Soit n un entier supérieur ou égal à 2 et u un mot de A^* . Nous disons que u est n -divisé s'il existe n mots non vides u_1, u_2, \dots, u_n tels que $u = u_1 u_2 \dots u_n$ et que pour toute permutation non triviale p de $\{1, 2, \dots, n\}$ le mot $u_{p(1)} u_{p(2)} \dots u_{p(n)}$ précède le mot u dans l'ordre lexicographique. Rappelons enfin qu'une puissance k -ième est un mot non vide de la forme $u = v^k$. Shirshov a prouvé le théorème suivant (voir [182] et aussi, par exemple, Lothaire [103]).

Théorème 3 (Shirshov). *Soient k et n des entiers positifs. Tout mot suffisamment long sur un alphabet fini contient un facteur qui est soit une puissance k -ième soit n -divisé.*

Soient A un alphabet, $<$ un ordre total sur A et s un mot infini sur A . Nous disons que s est ultimement ω -divisé pour l'ordre $<$ s'il existe une ω -factorisation

$$u_0 u_1 u_2 \dots u_n \dots$$

de s telle que pour tous les entiers $h \geq 0$, $k \geq 2$ et pour toute permutation non triviale p de $\{1, 2, \dots, k\}$ le mot

$$u_{h+p(1)} u_{h+p(2)} \dots u_{h+p(k)}$$

précède strictement dans l'ordre $<$ le mot

$$u_{h+1} u_{h+2} \dots u_{h+k}.$$

Nous disons que s est ω -divisé si $u_0 = \epsilon$

Par exemple les factorisations

$$\begin{aligned} f &= u_0 u_1 \dots u_i \dots \\ &= (a)(baa)(babaaba)(babaababaabaabaaba) \dots \end{aligned}$$

et

$$\begin{aligned} f &= v_0 v_1 \dots v_i \dots \\ &= (ab)(aabab)(aabaababaabab)(aabaababaabaabaabaabaabaaba) \dots \end{aligned}$$

déjà vues sont des ω -divisions respectivement pour les ordres $b < a$ et $a < b$.

En utilisant une méthode similaire à celle de Reutenauer [173], Varricchio [192] a prouvé le résultat suivant :

Proposition 13 (Varricchio). *Soient A un alphabet fini, $<$ un ordre total sur A et s un mot infini sur A . Alors il existe un mot infini t sur A tel que les facteurs de t sont facteurs de s et tel que t est soit périodique, soit ω -divisé pour l'ordre $<$.*

En introduisant une notion bilatère de ω -division nous avons prouvé avec Justin, voir [80], la proposition suivante

Proposition 14 *Soient A un alphabet fini, $<$ un ordre total sur A et s un mot infini sur A . Alors il existe un mot bi-infini t sur A dont les facteurs sont facteurs de s et tel que t est soit périodique, soit ω -divisé pour l'ordre $<$.*

Nous avons prouvé aussi, voir [134], la proposition suivante

Proposition 15 *Soient A un alphabet fini, $<$ un ordre total sur A . Alors tout mot infini sur A est ou bien ultimement périodique, ou bien ω -divisé pour l'ordre $<$ ou bien ω -divisé pour l'ordre $<^{-1}$ (i.e., l'ordre tel que $a <^{-1} b$ si et seulement si $b < a$).*

Etant donné un mot infini q sur un alphabet totalement ordonné A et un entier k , nous notons $m(q, k)$ le plus petit élément pour l'ordre lexicographique parmi les facteurs de q de longueur k .

Définition 10 *Nous disons qu'un rang i d'un mot infini q est minimal si, pour tout $k \geq 1$, $m(q, k)$ existe et $q(i, i + k - 1) = m(q, k)$.*

Un mot non ultimement périodique a au plus un rang minimal. Cette simple notion nous a conduits à prouver, voir [141], la proposition suivante

Proposition 16 *Soit A un alphabet totalement ordonné, A^* lexicographiquement ordonné, soit q un mot récurrent non ultimement périodique et, s'il existe, soit m son unique rang minimal. Alors*

pour tout i , $i \neq m$, le mot infini $q(i, \infty)$ est ω -divisé

pour $i = m$, le mot infini $q(i, \infty)$ n'est pas ω -divisé.

A l'aide de ce théorème nous avons donné une preuve très simple du théorème de Shirshov (voir [182]).

Définition de semi-groupe permutable. Soit S un semi-groupe. Etant donnés $x_1, x_2, \dots, x_n \in S$ nous disons que le produit $x = x_1 x_2 \dots x_n$ est permutable (resp. faiblement permutable) s'il existe une permutation non triviale σ de $\{1, 2, \dots, n\}$ telle que

$$x_1 x_2 \dots x_n = x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}$$

(resp. s'il existe deux permutations σ et τ de $\{1, 2, \dots, n\}$, $\sigma \neq \tau$, telles que

$$x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)} = x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(n)}).$$

Nous disons que S est n -permutable ou a la propriété P_n (resp. faiblement n -permutable ou a la propriété P_n^*) si tout produit de n éléments de S est permutable (resp. faiblement permutable).

Nous disons que S est permutable ou a la propriété P (resp. faiblement permutable ou a la propriété P^*) s'il existe $n \geq 2$ tel que S est n -permutable (resp. faiblement n -permutable).

La notion de semi-groupe permutable a été introduite par Restivo et Reutenauer dans un célèbre papier [170] où ils ont prouvé la proposition suivante :

Proposition 17 (Restivo et Reutenauer). Soit S un semi-groupe finement engendré. Les conditions suivantes sont équivalentes :

- 1) S est fini ;
- 2) S est périodique et permutable.

Après ce résultat plusieurs variantes de la notion de permutable ont été étudiées. Avec Justin dans [80] nous avons introduit la notion de *two-sided* ω -permutable semi-groupe et nous avons prouvé que

Proposition 18 Un semi-groupe S finement engendré et périodique est *two-sided* ω -permutable si et seulement s'il est fini.

Dans [127] nous avons prouvé l'existence d'un semi-groupe faiblement permutable et non permutable et dans [128] nous avons prouvé l'existence d'un semi-groupe finement engendré faiblement permutable et non permutable.

Nous nous sommes intéressés aussi aux valeurs optimales de l'indice de faible permutable pour certains semi-groupes.

Quotient de Rees. Le complément dans A^+ de l'ensemble des facteurs d'un mot infini s , i.e., $I(s) = A^+ - F(s)$ est un idéal de A^+ . Le quotient de Rees $S = A^+/I(s)$ est défini par la congruence

$$x \sim x' \text{ si et seulement si } x, x' \in I(s) \text{ ou } x = x'.$$

Nous avons obtenu les résultats suivants

Proposition 19 *Le semi-groupe de Thue $A^+/I(t)$, où t est le mot de Thue est un semi-groupe infini, finiment engendré, périodique et a la propriété P_4^* .*

Proposition 20 *Le semi-groupe de Fibonacci $A^+/I(f)$, où f est le mot de Fibonacci est un semi-groupe infini, finiment engendré, périodique et a la propriété P_4^* .*

Proposition 21 *Le quotient de Rees $A^+/I(s)$, où s est un mot Sturmien est un semi-groupe infini, finiment engendré, périodique et a la propriété P_4^* .*

Dans [78] nous avons considéré le mot infini

$$m = aabaabbaabaababb \dots$$

obtenu en itérant le morphisme $\varphi : A^+ \rightarrow A^+$, $\varphi(a) = aab$, $\varphi(b) = abb$, à partir de a et nous avons prouvé que

Proposition 22 *Le quotient de Rees $A^+/I(m)$ (où m est le mot que nous venons de définir) est un semi-groupe infini, finiment engendré, périodique et a la propriété P_3^* .*

Ce résultat prouve que pour tout $n \geq 3$ la périodicité d'un semi-groupe finiment engendré ayant la propriété P_n^* n'entraîne pas sa finitude. En plus, comme un semi-groupe ayant la propriété $P_2^* = P_2$ est abélien et comme un semi-groupe abélien, périodique et finiment engendré est fini, le résultat de la dernière Proposition ne peut pas être amélioré.

Dans [76] nous avons considéré le mot infini

$$m' = x_1x_2y_1x_2x_1y_2x_1y_1y_2x_2x_1y_2x_1x_2y_1x_2y_2y_1 \dots$$

obtenu en itérant le morphisme $\varphi : A^+ \rightarrow A^+$, $\varphi(x_1) = x_1x_2y_1$, $\varphi(x_2) = x_2x_1y_2$, $\varphi(y_1) = x_1y_1y_2$, $\varphi(y_2) = x_2y_2y_1$, à partir de x_1 et nous avons prouvé que

Proposition 23 *Le quotient de Rees $A^+/I(m')$, est un semi-groupe infini, finiment engendré, périodique, a la propriété P_3^* et tous ses éléments sont d'ordre au plus 2.*

Nous avons aussi remarqué qu'il n'est pas possible de faire une construction analogue avec un alphabet de trois lettres. Thue a en effet montré que tout mot sans carré sur trois lettres de longueur supérieure à 30 contient en facteurs les six permutations de ces trois lettres (voir [185] [186] et aussi [57]).

Les propriétés P et P^* , qui sont conservées par passage au sous-semi-groupe et à l'image homomorphe et sont conservées aussi par d'autres constructions voir [75] [79], nous ont conduits à considérer un renforcement naturel, l'intracommutativité (et, aussi, la faible intracommutativité).

Dans la définition suivante S est un semi-groupe et S^1 , comme d'habitude, est S avec une identité ajoutée, si nécessaire. Une n -factorisation est une $(n+2)$ -tuple $(u, x_1, x_2, \dots, x_n, v)$ avec $u, v \in S^1$, $x_i \in S$, $1 \leq i \leq n$.

La k -factorisation $(u', y_1, y_2, \dots, y_k, v')$ est dite une k -factorisation d'une n -factorisation $(u, x_1, x_2, \dots, x_n, v)$ s'il existe $k + 1$ entiers $i_0 < i_1 < \dots < i_k$ dans $[0, n]$ tels que

$$\begin{aligned} u' &= ux_1 \dots x_{i_0} \\ y_{j+1} &= x_{i_j+1} \dots x_{i_{j+1}} \quad (0 \leq j \leq k-1) \\ v' &= x_{i_k+1} \dots x_nv \end{aligned}$$

avec les conventions que $u' = u$ si $i_0 = 0$ et $v' = v$ si $i_k = n$.

Définition 11 *Un n -tuple d'éléments d'un semi-groupe est k -intracommutatif s'il admet une k -factorisation, notée $(u, y_1, y_2, \dots, y_k, v)$, telle que pour tout $\sigma \in \Sigma_k$, nous avons*

$$uy_1y_2 \dots y_kv = uy_{\sigma(1)}y_{\sigma(2)} \dots y_{\sigma(k)}v$$

Un n -tuple d'éléments d'un semi-groupe est faiblement k -intra-commutatif s'il existe $\tau \in \Sigma_n$, telle que

$$x_{\tau(1)}x_{\tau(2)} \cdots x_{\tau(n)}$$

est k -intra-commutatif.

Définition 12 *Un semi-groupe S est k -intra-commutatif (resp. faiblement k -intra-commutatif) ou a la propriété Q (resp. Q^*) si pour tout entier positif k il existe un entier positif n tel que tout n -tuple d'éléments de S est k -intra-commutatif (resp. faiblement k -intra-commutatif).*

Dans [81], nous prouvons certains résultats sur les semi-groupes (parfois à l'aide d'une variante du théorème de Ramsey) et aussi le résultat suivant de la théorie des groupes :

Proposition 24 *Etant donné un groupe G , les conditions suivantes sont équivalentes :*

- (i) G est abélien-par-fini ;*
- (ii) G a la propriété Q , (intra-commutativité) ;*
- (iii) il existe un entier n tel que tout n -tuple d'éléments de G est 2-intra-commutatif.*

Le papier [71] est dédié à l'étude de plusieurs quotients de Rees du semi-groupe libre. Ils ont été étudiés en relation avec les propriétés suivantes : répétitivité, croissance linéaire, permutabilité, faible permutabilité, intra-commutativité, faible intra-commutativité et d'autres variantes concernant des suites infinies d'éléments.

Blyth dans [24] (et déjà dans sa thèse de doctorat, voir aussi [25]) a prouvé que pour un groupe les propriétés P et P^* sont équivalentes.

Nous avons résolu avec Piochi voir [118] une conjecture posée par Curzio. Ce résultat découle aisément d'un théorème plus général de théorie des groupes prouvé aussi dans [118]. Le même problème pour les semi-groupes est beaucoup plus difficile.

Mots Sturmien

Les mots Sturmien, définis ainsi dans [113] et dans [114], interviennent dans de nombreuses théories (dynamique symbolique, fractions continues en mathématiques et quasi-cristaux en physique, ...) où ils jouent un rôle très important. Le mot de Fibonacci est l'exemple le plus remarquable. Ils possèdent plusieurs définitions équivalentes (prouvées telles pour la plupart dans [113] et dans [114]).

Trois de leurs définitions sont les suivantes : les mots Sturmien sont tous les mots sur $\{0, 1\}$ non ultimement périodiques (synonyme : "apériodiques"), qui ont exactement $n + 1$ facteurs de longueur n , pour tout n ; les mots Sturmien sont aussi les mots sur $\{0, 1\}$ non ultimement périodiques, qui ont dans deux facteurs de même longueur des nombres de 1 qui diffèrent au plus de 1 et, finalement

Définition 13 *Un mot infini $s = s(1)s(2)s(3)\cdots, s(i) \in \{a, b\}$, est Sturmien s'il existe deux nombres réels $\alpha, \rho \in [0, 1]$, tels que, pour tout $i \geq 1$,*

$$s(i) = a \text{ si } \lfloor \rho + (n + 1)\alpha \rfloor = \lfloor \rho + n\alpha \rfloor, \quad s(i) = b \text{ sinon}$$

ou pour tout i

$$s(i) = a \text{ si } \lceil \rho + (n + 1)\alpha \rceil = \lceil \rho + n\alpha \rceil, \quad s(i) = b \text{ sinon.}$$

Le mot infini s est Sturmien propre si α est irrationnel, est Sturmien périodique si α est rationnel et est Sturmien standard si $\rho = 0$.

On utilise aussi la définition géométrique suivante : on considère dans le plan euclidien, la grille formée par les droites $x = h, h$ entier, $y = k, k$ entier, on considère aussi une demi-droite de la forme $y = \alpha x, \alpha$, irrationnel et positif; à l'irrationnel α on associe le mot s_α sur l'alphabet $\{0, 1\}$ qui a pour n -ième lettre un 1 si la n -ième intersection de la demi-droite avec la grille (dans le sens des x croissants) est une droite horizontale, et qui a pour n -ième lettre un 0 si la n -ième intersection de la demi-droite avec la grille est une droite verticale. Dans le cas où deux intersections, la n -ième et la $(n + 1)$ -ième, coïncident (c'est à dire que la demi-droite a rencontré un point de coordonnées entières) alors on peut choisir de fixer comme valeur 1 à la n -ième lettre et 0 à la $(n + 1)$ -ième ou 0 à la n -ième lettre et 1 à la $(n + 1)$ -ième.

Récemment plusieurs papiers étudient les relations entre compression de données et mots Sturmien (voir par exemple ceux de Mantaci, Restivo et Sciortino [108] et de Crochemore, Désarménien et Perrin [31]).

Remarquons que dans la plupart des articles le mot “Sturmien” est réservé au cas α irrationnel, mais dans cette thèse “Sturmien” signifie ou bien “propre” ou bien “périodique”. Remarquons aussi que dans les articles pionniers, [113] et [114], tous les *mots équilibrés* (pour la définition, que nous rappelons ci-dessous, voir aussi [21]) sur un alphabet de deux lettres sont appelés “*Sturmian*” et appartiennent à trois classes : celle que nous appelons propres, celle que nous appelons périodiques et les non-récurrents qui sont ultimement périodiques (mais non périodiques) appelés “*skew Sturmian*”.

Les mots Sturmien standard ont une curieuse propriété découverte par Rauzy. Si dans un tel mot on supprime toutes les occurrences de chaque lettre sauf celles de rang multiple d’un entier p donné, alors on retrouve le mot infini initial. On a étudié cette propriété et plusieurs généralisations dans [86]. Dans la dernière partie de ce même papier nous avons donné une présentation autocontenue de la théorie des mots Sturmien qui n’utilise que des arguments combinatoires. Un très court résumé de cette partie de [86] est présenté ici. On reconnaît facilement que plusieurs résultats sont des généralisations des lemmes donnés dans les paragraphes dédiés au mot de Fibonacci.

Définition 14 *Un mot fini ou infini s est équilibré si pour tout $u, v \in F(s)$, $|u| = |v|$ implique $||u|_a - |v|_a| \leq 1$.*

Si $y \in A$, nous notons par \bar{y} l’autre lettre de A . Pour tout mot infini s , nous notons $F_n(s)$ l’ensemble $A^n \cap F(s)$ de tous ses facteurs de longueur n .

Proposition 25 *Soit s un mot infini, alors*

- i) si pour quelque $m \in \mathbb{N}_+$, $|F_m(s)| \leq m$, alors s est ultimement périodique avec période $p \leq m$, et, pour tout $n \geq m$, $|F_n(s)| \leq m$;*
- ii) si s est périodique avec période p alors $|F_n(s)| > n$ pour $n < p$ et $|F_n(s)| = p$ pour $n \geq p$.*

Proposition 26 *Si le mot infini s est équilibré, alors, pour tout $n \in \mathbb{N}_+$, $|F_n(s)| \leq n + 1$.*

Proposition 27 *Pour un mot infini s les conditions suivantes sont équivalentes :*

- i) s a exactement un facteur spécial de longueur n , pour tout $n \in N_+$;*
- ii) $|F_n(s)| = n + 1$, pour tout $n \in N_+$;*
- iii) s est équilibré et non ultimement périodique.*

Notons qu'un mot Sturmien propre s pourrait être défini comme un mot qui satisfait une des conditions (équivalentes) de la Proposition 27 et que toutes ces définitions sont équivalentes.

Proposition 28 *Si s est Sturmien propre, alors $u \in F(s)$ implique $\tilde{u} \in F(s)$.*

Proposition 29 *Si s est Sturmien propre, alors il est uniformément récurrent.*

Proposition 30 *Si les mots infinis s et t sont uniformément récurrents (Sturmiens propres, par exemple), alors ou bien $F(s) = F(t)$ ou bien $F(s) \cap F(t)$ est fini.*

Proposition 31 *Etant donné un mot Sturmien propre s , soit u un mot tel que $u \notin F(s)$ et $F(u) \setminus \{u\} \subset F(s)$. Alors il existe un facteur v de s tel que $|u| = |v|$ et $||u|_a - |v|_a| > 1$.*

Définition 15 *Un mot Sturmien propre s est standard si tout préfixe de s est l'image miroir d'un facteur spécial à droite.*

Proposition 32 *Etant donné un mot Sturmien propre il y a exactement un mot Sturmien standard propre ayant les mêmes facteurs.*

Proposition 33 *Soit s un mot infini non ultimement périodique. Les conditions suivantes sont équivalentes :*

- i) s est un mot Sturmien standard propre ;*
- ii) il y a une suite $(m_i)_{i \in N}$ d'entiers positifs telle que, posant $X_0 = a$, $Y_0 = b$ et, pour tout $i \in N$, $X_{i+1} = X_i^{m_i-1}Y_i$, $Y_{i+1} = X_i^{m_i}Y_i$, nous avons $s = X_1X_2X_3 \dots$.*

Puissances fractionnaires.

Nous nous limitons dans ce paragraphe à un très court résumé de [88]. Soit α un irrationnel dans $]0, 1[$ et s le mot Sturmien standard infini de α . Pour tout m nous calculons la longueur $L(m)$ du facteur maximal de s qui soit un mot périodique de période m . Soit $[0; a_1, a_2, \dots]$ la fraction continue simple de α . Nous calculons $L(m)$ en fonction des a_i . Pour une infinité des valeurs de m l'“exposant” maximal $L(m)/m$ est plus grand que 2 et il est donné par une formule “simple”. Si $L(m)/m$ est plus petit que 2 les formules sont plus compliquées.

Mots épisturmiens.

Les mots infinis épisturmiens sont une généralisation des mots Sturmiens et incluent les suites d'Arnoux-Rauzy. Depuis leur introduction (voir [46]) nos recherches sur les mots Sturmiens et épisturmiens sont conduites en parallèle. Nos principaux articles sur ce sujet sont en annexe. Ceci nous permet de nous limiter aux énoncés des résultats autour desquels nous avons rédigé les articles.

Nous avons déjà vu que les préfixes palindromiques h_n du mot de Fibonacci ont la propriété $h_n ab \sim h_n ba$, i.e., que $h_n ab$ et $h_n ba$ sont conjugués.

Pour les mots Sturmiens, nous avons ([150]) la proposition suivante :

Proposition 34 *Pour un mot w quelconque, les conditions suivantes sont équivalentes :*

- i) w est un préfixe palindromique d'un mot Sturmien standard propre ;*
- ii) $wab \sim wba$.*

Nous avons déjà vu que les premières occurrences des palindromes a , b , $aa \dots$ sont en position centrale dans un préfixe palindromique h_n du mot de Fibonacci f . C'est à l'origine de l'étude des mots épisturmiens [46].

Soit v un mot de A^* . La fermeture palindromique à droite de v est le plus court palindrome w en A^* qui a v comme préfixe. Le mot w est unique. Nous disons qu'un palindrome u est un facteur central du palindrome v si $v = wu\tilde{w}$ pour quelque mot w de A^* .

Définition 16 *Un mot infini s sur un alphabet A est épisturmien si $F(s)$ est fermé par passage au miroir et s a au plus un facteur spécial à droite de longueur n pour tout n .*

Proposition 35 *Un mot infini épisturmien est standard si tous ses facteurs spéciaux à gauche sont aussi préfixes.*

Dans [46] nous avons prouvé la proposition suivante

Proposition 36 *Soit s un mot infini sur un alphabet A . Les conditions suivantes sont équivalentes :*

- i) s est épisturmien standard ;*
- ii) si v est un préfixe de s alors sa fermeture palindromique à droite l'est aussi ;*
- iii) chaque première occurrence en s d'un palindrome est centrale dans un préfixe palindromique de s .*

Pour d'autres résultats sur les mots épisturmiens, nous renvoyons aux articles annexés [91] et [46].

Finalement nous rappelons le résultat suivant de [154]

Proposition 37 *Soit s un mot infini sur $\{a, b\}$. Les conditions suivantes sont équivalentes :*

- i) s est un mot Sturmien standard propre ou un mot Sturmien périodique ;*
- ii) $as \leq \min(s) \leq \max(s) \leq bs$.*

Pour les mots épisturmiens voir [52].

Quelques résultats sur les codes

Nous nous limitons à signaler brièvement ici les articles [87], [139] et [160]. Ce dernière article contient une nouvelle classification des codes circulaires finis.

Biologie théorique

Le code circulaire maximal découvert par Arquès et Michel ([4])

$$\{AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC, GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC\}$$

est un exemple très intéressant d'instrument d'informatique théorique pour l'étude de certains problèmes biologiques.

Nous avons introduit dans [151] une sorte de configuration à éviter, c'est la notion de *collier* (*necklace*, en anglais). Nous avons utilisé cette notion pour l'étude des codes circulaires auto-complémentaires (nous prouvons que parmi eux exactement 528 sont maximaux) dans [161] et pour l'étude des codes sans ponctuation dans [109].

Didactique de la mathématique et de l'informatique.

La didactique des sciences a été depuis longtemps une des nos préoccupations professionnelles (voir, par exemple, [124], [138] et [155] dans la bibliographie générale).

RESUMES DES ARTICLES

Ces résumés ne sont pas nécessairement ceux figurant éventuellement dans les articles.

Sous presse

[160] Giuseppe Pirillo. A hierarchy for circular codes. Fibonacci, Turku, 2006. sous presse

Résumé : *Dans cet article, nous présentons une nouvelle caractérisation des codes circulaires.*

[52] Amy Glen, Jacques Justin et Giuseppe Pirillo. Characterizations of finite and infinite episturmian words via lexicographic orderings. *European Journal of Combinatorics*. sous presse.

Résumé : *Dans cet article, nous caractérisons par l'ordre lexicographique tous les mots Sturmien et épisturmiens finis, c.-à-d., tous les facteurs (finis) de tels mots infinis. En conséquence, nous obtenons une caractérisation des mots épisturmiens infinis au sens large (mots infinis épisturmiens et episkew). C'est-à-dire, nous caractérisons l'ensemble de tous les mots infinis dont les facteurs sont épisturmiens (finis). De même, nous caractérisons par l'ordre lexicographique tous les mots infinis équilibrés sur un alphabet de 2 lettres; en d'autres termes, tous les mots infinis Sturmien et skew.*

[109] Christian Michel, Giuseppe Pirillo et Mario Pirillo. Varieties of comma free codes. *Comput. Math. Appl.* Special Issue, I FIMA Conference, sous presse.

Résumé : *De nouvelles variétés des codes sans ponctuation de longueur 3 sur l'alphabet de 4 lettres sont définies et analysées. De nouvelles propriétés avec des mots de longueur 3, 4, 5 et 6 dans les codes sans ponctuation sont employées pour la détermination des fonctions de croissance dans les variétés de codes étudiées.*

[120] Giuseppe Pirillo. Morse and Hedlund's skew Sturmian words revisited. *Annals of Combinatorics*. sous presse.

Résumé : À n'importe quel mot infini r sur $A = \{a, b\}$ nous associons deux mots infinis $\min(r)$, $\max(r)$ tels que n'importe quel préfixe de $\min(r)$ (resp. $\max(r)$) est lexicographiquement le plus petit (resp. grand) parmi les facteurs de r de même longueur. Nous montrons que $(\min(r), \max(r)) = (as, bs)$ pour un certain mot infini s si et seulement si r est un mot Sturmien propre ou un mot ultimement périodique d'une forme particulière.

2006

[156] Giuseppe Pirillo. Some algebraic remarks on circular codes. *PUMA* vol. 16 (2005), N. 1-2, pp 191–123, 2006.

Résumé : Soit t un trinuéclotide et $C(t)$ son complément. Dans cet article nous prouvons que si X est un code circulaire autocomplémentaire alors X contient une et seulement une des trois paires $(RST, C(RST))$, $(STR, C(STR))$, $(TRS, C(TRS))$, pour chaque trinuéclotide RST différent de AAA , CCC , GGG et TTT .

2005

[154] Giuseppe Pirillo. Inequalities characterizing standard Sturmian and episturmian words. *Theoret. Comput. Sci.*, 341(1-3) :276–292, 2005.

Voir l'article in extenso en annexe.

[161] Giuseppe Pirillo et Mario Pirillo. Growth Function of Self-complementary Circular Codes 2005 Biology Forum 98 97–110.

Résumé : Dans plusieurs articles Arquès et Michel ont étudié les codes circulaires maximaux se composant des mots de longueur 3 (ou trinuéclotides) sur l'alphabet génétique $\{A, C, G, T\}$. Nous présentons ici de l'information additionnelle sur ces codes. Nous étudions la fonction de croissance des codes circulaires auto-complémentaires et nous prouvons que parmi eux exactement 528 sont maximaux.

2004

[91] Jacques Justin et Giuseppe Pirillo. Episturmian words : shifts, mor-

phisms and numeration systems, International Journal on Foundations of Computer Science, Vol. 15, No. 2 (2004) 329-348

Voir l'article in extenso en annexe.

2003

[90] Jacques Justin et Giuseppe Pirillo. On a characteristic property of Arnoux-Rauzy sequences. *Theor. Inform. Appl.*, 36(4) :385–388 (2003), 2002.

Résumé : *Les mots d'Arnoux-Rauzy sont des mots infinis sur un alphabet de k lettres qui ont exactement un facteur spécial à gauche et un facteur spécial à droite pour chaque longueur et sont des généralisations des mots Sturmien. Ici nous donnons une caractérisation des mots d'Arnoux-Rauzy à l'aide des ordres lexicographiques sur leur alphabet.*

[151] Giuseppe Pirillo. A characterization for a set of trinucleotides to be a circular code. pages 241–244, 2002. in : Determinism, Holism, and Complexity (Edited by C. Pellegrini, P. Cerrai, P. Freguglia, V. Benci and G. Israel).

Résumé : *Les codes circulaires auto-complémentaires d'Arquès-Michel sont 216 ensembles de 20 trinucleotides. Ils ont beaucoup de propriétés très intéressantes. Nous avons trouvé le résultat suivant : un ensemble X de trinucleotides est un code circulaire si et seulement s'il n'a aucun 5-collier.*

[148] Giuseppe Pirillo. Remark on the Arquès-Michel code. Biology Forum 2001 327–330 in : Proceedings of the Workshop “Theoretical Biology 2, History and Present Themes”, Arcidosso, September 1-3, 1999.

Résumé : *Arquès et Michel ont découvert par des méthodes statistiques un ensemble de 20 trinucleotides qui a des propriétés remarquables. Nous montrons (en donnant un exemple) qu'il est possible, en employant des arguments théoriques, de trouver des ensembles ayant les propriétés semblables.*

2002

[89] Jacques Justin et Giuseppe Pirillo. Episturmian words and episturmian morphisms. *Theoret. Comput. Sci.*, 276(1-2) :281–313, 2002.

Résumé : *Les mots infinis épisturmiens constituent une généralisation des mots sturmiens qui inclut les suites d'Arnoux-Rauzy. Cette étude, commencée précédemment, concerne maintenant les morphismes épisturmiens et les facteurs palindromiques et périodiques des mots épisturmiens. Nous considérons notamment le cas des mots engendrés par morphisme et introduisons une notion d'“intercept” généralisant celle des mots sturmiens. Enfin nous prouvons l'existence dans un sens fort de fréquences pour les facteurs de ces mots.*

[152] Giuseppe Pirillo. Fibonacci word and Sturmian words. IN MEMORIAM GIAN-CARLO ROTA *Sci. Math. Jpn.*, 55(3) :641–647, 2002.

Résumé : *Si on regarde soigneusement les propriétés du mot de Fibonacci on peut découvrir certaines propriétés générales appropriées des mots Sturmiens standard. Dans deux articles précédents nous avons présenté plusieurs exemples et ici nous en présentons de nouveaux soutenant notre avis.*

2001

[150] Giuseppe Pirillo. A curious characteristic property of standard Sturmian words. In *Algebraic combinatorics and computer science*, pages 541–546. Springer Italia, Milan, 2001.

Résumé : *Une propriété très simple des préfixes palindromiques du mot de Fibonacci (voir Lemme 1, ii)), nous a suggéré le résultat de cet article : un mot w est un préfixe palindromique d'un mot Sturmien standard si et seulement si wab est conjugué de wba .*

[46] Xavier Droubay, Jacques Justin, et Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2) :539–553, 2001.

Voir l'article in extenso en annexe.

[88] Jacques Justin et Giuseppe Pirillo. Fractional powers in Sturmian words. *Theoret. Comput. Sci.*, 255(1-2) :363–376, 2001.

Voir l'article in extenso en annexe.

2000

[149] Giuseppe Pirillo. From the Fibonacci word to Sturmian words. II. *Pure Math. Appl.*, 11(3) :533–537, 2000.

Résumé : *Nous présentons d'autres exemples des propriétés du mot de Fibonacci que l'on peut généraliser aux mots Sturmiens standard.*

1999

[47] Xavier Droubay et Giuseppe Pirillo. Palindromes and Sturmian words. *Theoret. Comput. Sci.*, 223(1-2) :73–85, 1999.

Résumé : *Dans cet article nous présentons une nouvelle caractérisation des mots Sturmiens : un mot infini est Sturmien si et seulement si pour chaque nombre entier n , il contient exactement un palindrome de longueur n si n est pair, et deux palindromes de longueur n si n est impair. La preuve est purement combinatoire. En outre, dans le cas d'un mot Sturmien qui est point fixe d'un morphisme, nous présentons un morphisme qui produit tous les palindromes de ce mot.*

[143] Giuseppe Pirillo. From the Fibonacci word to Sturmian words. *Publ. Math. Debrecen*, 54(suppl.) :961–971, 1999. Automata and formal languages, VIII (Salgótarján, 1996).

Résumé : *Nous présentons plusieurs exemples des propriétés du mot de Fibonacci qui l'on peut généraliser aux mots Sturmiens standard.*

[146] Giuseppe Pirillo. On a conjecture of Erdős. *Publ. Math. Debrecen*, 54(suppl.) :973–978, 1999. Automata and formal languages, VIII (Salgótarján, 1996).

Résumé : *Une conjecture d'Erdős sur les permutations, résolue récemment par Keränen, a une relation étroite avec la théorie de la répétitivité des semi-groupes.*

[147] Giuseppe Pirillo. Some factorizations of the Fibonacci word. *Algebra Colloq.*, 6(4) :361–368, 1999.

Résumé. *Pour tout $n \geq 1$, soit F_n le n -ième nombre de Fibonacci et soit f le mot infini de Fibonacci. Les facteurs de f dans certaines factorisations forment des codes (dans le sens de la théorie de codes de longueur variable). Par exemple, soit $f = u_0u_1u_2\dots$ une factorisation de f comme produit de mots finis et non vides : si la longueur des u_i est F_{2i+1} [resp. $F_{2(i+1)}$], alors les u_i ($i \geq 0$) forment un code préfixe, c.-à-d., aucun de ces mots n'est un préfixe des autres.*

[144] Giuseppe Pirillo. Maximal circular codes and applications to theoretical biology. In *Mathematical and computational biology (Aizu-Wakamatsu City, 1997)*, volume 26 of *Lectures Math. Life Sci.*, pages 187–190. Amer. Math. Soc., Providence, RI, 1999.

Résumé. *Des codes circulaires maximaux intéressants contenant 20 trinuécléotides ont été récemment découverts par l'équipe de biologie théorique de Didier Arquès et de Christian Michel (France). Une application biologique possible de ce résultat est la localisation automatique des gènes dans le génome.*

1998

[44] Pierpaolo Degano, Giuseppe Pirillo et Ugo Vaccaro, editors. *Theoretical computer science*, River Edge, NJ, 1998. World Scientific Publishing Co. Inc.

Résumé. *Cette publication rassemble les papiers choisis présentés lors du colloque de 1998 du chapitre italien d'EATCS.*

1997

[86] Jacques Justin et Giuseppe Pirillo. Decimations and Sturmian words. *RAIRO Inform. Théor. Appl.*, 31(3) :271–290, 1997.

Résumé. *Les mots Sturmiens standard infinis ont une curieuse propriété découverte par G. Rauzy. Si dans un tel mot on supprime toutes les occurrences de chaque lettre sauf celles de rang multiple de p , alors on retrouve le mot infini initial. On étudie cette propriété et plusieurs généralisations. Dans la dernière partie on donne une présentation courte et autocontenue de*

CONTRIBUTION A LA COMBINATOIRE DES MOTS

la théorie des mots Sturmien qui n'utilise que des arguments combinatoires.

[87] Jacques Justin et Giuseppe Pirillo. On some factorizations of infinite words by elements of codes. *Inform. Process. Lett.*, 62(6) :289–294, 1997.

Résumé. *Nous prouvons qu'un mot infini s satisfait $s = u_0u_1u_2\cdots$, où les u_i sont des mots différents non vides dont l'ensemble est un code bi-préfixe, si et seulement si s n'est pas ultimement périodique. Nous donnons aussi des résultats sur les codes arbitraires et nous considérons le cas des mots bi-infinis.*

[142] Giuseppe Pirillo. Fibonacci numbers and words. *Discrete Math.*, 173(1-3) :197–207, 1997. Dedicated to Giovanni Prodi.

Voir l'article in extenso en annexe.

[96] Andrei Kelarev et Giuseppe Pirillo. A note on rewritable products in groups and semigroups. *Boll. Un. Mat. Ital. A (7)*, 11(3) :667–670, 1997.

Résumé. *Nous introduisons la propriété de permutabilité involutionnelle et nous prouvons qu'un groupe a cette propriété si et seulement si il a la propriété de permutabilité.*

1996

[163] Giuseppe Pirillo et Stefano Varricchio. Some combinatorial properties of infinite words and applications to semigroup theory. In *Proceedings of the 5th Conference on Formal Power Series and Algebraic Combinatorics (Florence, 1993)*, Discrete Mathematics, volume 153, pages 239–251, 1996.

Résumé. *Nous présentons un résultat sur les mots infinis et une preuve alternative d'un théorème de Restivo et Reutenauer : un semigroupe finiment engendré est fini si et seulement si il est périodique et permutable. Puis, généralisant des notions étudiées par de Luca, Restivo, Hashiguchi et Varricchio, nous présentons la propriété de ω -itération et nous montrons qu'un semigroupe finiment engendré a la propriété de ω -itération si et seulement s'il est fini.*

Thèse de Doctorat ès Sciences de Giuseppe Pirillo

[85] Jacques Justin et Giuseppe Pirillo. On a combinatorial property of Sturmian words. *Theoret. Comput. Sci.*, 154(2) :387–394, 1996.

Résumé. *Rappelons qu'un semi-groupe a la propriété P_n^* si pour toute suite de n de ses éléments, il y a deux produits de ces n éléments, dans des ordres différents, qui sont égaux. Soit s un mot Sturmien infini (sur un alphabet A à deux lettres). Nous prouvons que le quotient de Rees de A^* par l'ensemble des mots non facteurs de s a la propriété P_4^* et que ce résultat est optimal. Nous prouvons aussi que si St est l'ensemble de tous les mots Sturmiens finis, alors le quotient de Rees $A^*/(A^* - St)$ a la propriété P_8^* .*

[140] Giuseppe Pirillo. On semigroups having the ω -iteration property. *Matematiche (Catania)*, 51(suppl.) :183–187 (1997), 1996. dedicato a Gasapina.

Résumé. *Récemment on a montré qu'un semi-groupe finiment engendré ayant la propriété d'itération est fini. Ici nous montrons que, dans certaines conditions, l'hypothèse que l'ensemble des générateurs soit fini peut être omise. Comme les conditions supplémentaires sont toujours satisfaites par un groupe nous avons qu'un groupe avec la propriété d'itération est fini.*

[141] Giuseppe Pirillo. A proof of Shirshov's theorem. *Adv. Math.*, 124(1) :94–99, 1996. IN MEMORIAM MARCEL-PAUL SCHÜTZENBERGER.

Résumé. *Dans cet article nous montrons que, à l'exception d'au plus un, tous les rangs d'un mot infini à droite qui est récurrent et pas ultimement périodique sont le point de départ d'une ω -division. En utilisant ceci nous fournissons une preuve très simple du Théorème de Shirshov.*

1995

[84] Jacques Justin et Giuseppe Pirillo. On a question about factorization forests. *Theoret. Comput. Sci.*, 141(1-2) :351–355, 1995.

Résumé : *Répondant à une question de I. Simon, nous montrons qu'il existe des fonctions fortement Ramseyennes qui n'admettent aucune forêt de factorisation Ramseyenne de hauteur finie.*

1994

[83] Jacques Justin et Giuseppe Pirillo. On some questions and conjectures in the combinatorial theory of semigroups. *Southeast Asian Bull. Math.*, 18(3) :91–104, 1994.

Résumé. *Nous considérons quelques questions et conjectures liées aux propriétés de répétitivité et de permutabilité des semi-groupes. En particulier, nous discutons deux questions, Question A (existe-t-il un semi-groupe non-répétitif dont tous les sous-semi-groupes propres sont répétitifs ?) et Question B (existe-t-il un groupe non commutatif infini dont tous les sous-groupes propres sont finis ?) et nous montrons, en utilisant un résultat de [135], que la réponse positive à la question B permet de donner également une réponse positive à la question A.*

[162] Giuseppe Pirillo et Stefano Varricchio. On uniformly repetitive semigroups. *Semigroup Forum*, 49(1) :125–129, 1994.

Résumé : *Soit k un entier supérieur à 1. Les propositions suivantes sont équivalentes : 1) le semi-groupe des entiers positifs n'est pas uniformément k -répétitif ; 2) tout semi-groupe finiment engendré et uniformément k -répétitif est fini.*

[139] Giuseppe Pirillo. Infinite words and biprefix codes. *Inform. Process. Lett.*, 50(6) :293–295, 1994. Dédié à A. Saoudi.

Résumé : *Soit A un alphabet et soit f un mot infini à droite sur A . Si f n'est pas ultimement périodique alors il existe un ensemble infini $\{v_i | i \geq 0\}$ de mots (finis) sur A tels que $f = v_0 v_1 \cdots v_i \cdots$, $\{v_i | i \geq 0\}$ est un code bi-préfixe et $v_i \neq v_j$ pour tous entiers positifs $i \neq j$.*

1993

[82] Jacques Justin et Giuseppe Pirillo. Factorial languages and some combinatorial properties of semigroups. *Internat. J. Algebra Comput.*, 3(3) :295–316, 1993.

Résumé : *Nous donnons des caractérisations pour les quotients de Rees des semi-groupes libres qui ont telle ou telle propriété combinatoire : répétiti-*

CONTRIBUTION A LA COMBINATOIRE DES MOTS

vit , croissance lin aire, propri t  de permutation ou une de ses variantes.

[136] Giuseppe Pirillo. On a combinatorial property of Fibonacci semi-group. *Discrete Math.*, 122(1-3) :263–267, 1993.

R sum  : *Nous disons qu'un semi-groupe S a la propri t  P_n^* , $n \geq 2$, si,  tant donn s n  l ments $x_1, \dots, x_n \in S$, au moins deux produits parmi les $n!$ possibles co ncident. Restivo a consid r  le semi-groupe de Fibonacci (c.- -d., le quotient de Rees de $\{a, b\}^+$ par l'ideal des non-facteurs du bien connu mot infini de Fibonacci $abaababaabaab \dots$) et a montr  qu'il a la propri t  P_8^* . Le but de cet article est de montrer qu'il a la propri t  P_4^* . Comme il n'a pas la propri t  P_3^* , ce r sultat est optimal.*

[137] Giuseppe Pirillo. Properties of integers and finiteness conditions for semigroups. *Results Math.*, 24(1-2) :168–173, 1993.

R sum  : *Soient h et k des nombres entiers plus grands que 1 ; nous montrons que les  nonc s suivants sont  quivalents : i) le produit direct des h copies du semi-groupe additif des nombres entiers non n gatifs n'est pas k -r p titif. ii) si le produit direct de h semi-groupes finiment engendr s est k -r p titif alors l'un d'entre eux est fini. En utilisant ceci et quelques r sultats de Dekking et de Pleasants sur des mots infinis, nous montrons que certaines propri t s de r p titivit  sont des conditions de finitude pour des semi-groupes finiment engendr s.*

1992

[111] Filippo Mignosi et Giuseppe Pirillo. Repetitions in the Fibonacci infinite word. *RAIRO Inform. Th or. Appl.*, 26(3) :199–204, 1992.

Voir l'article in extenso en annexe.

[81] Jacques Justin et Giuseppe Pirillo. Intra-commutativity properties for groups and semigroups. *J. Algebra*, 153(2) :424–443, 1992.

R sum  : *Pour un semi-groupe, nous pr sentons la propri t  d'intra-commutativit , Q , comme renforcement naturel de la propri t  P de permutation. Parmi plusieurs r sultats nous prouvons que pour un groupe G les propri t s suivantes sont  quivalentes : i) G est ab lien-par-fini ii) G a Q iii) il existe*

n tel que *n*'importe quel *n*-tuple des éléments de *G* est 2-intracommutatif.

[135] Giuseppe Pirillo. Sur les produits directs de semi-groupes répétitifs. *C. R. Acad. Sci. Paris Sér. I Math.*, 314(8) :645–648, 1992.

Voir l'article in extenso en annexe.

1991

[71] Jacques Justin, Brunetto Piochi et Giuseppe Pirillo. Propriétés de permutabilité avec certaines restrictions. *Arch. Math. (Basel)*, 57(1) :14–20, 1991.

Résumé. *Nous définissons les propriétés de L-permutabilité, R-permutabilité et LR-permutabilité et nous prouvons que, pour un semi-groupe S, les conditions suivantes sont équivalentes : i) S est L-permutable et R-permutable ii) S est LR-permutable iii) S est permutable sans point fixe.*

[80] Jacques Justin et Giuseppe Pirillo. Shirshov's theorem and ω -permutability of semigroups. *Adv. Math.*, 87(2) :151–159, 1991. Dedicated to Marcel-Paul Schützenberger

Résumé. *Reutenauer a fourni une preuve élégante du Théorème de Shirshov. Varricchio l'a prolongée au cas des mots infinis à droite. Dans cet article, nous servant des propriétés des mots uniformément récurrents de Fürstenberg, nous pouvons fournir une preuve très simple et autocontenue du théorème de Shirshov qui est également valide pour les mots bi-infinis.*

[70] Jacques Justin, Brunetto Piochi et Giuseppe Pirillo. On fixed-point-free permutation properties in groups and semigroups. *Ann. Mat. Pura Appl. (4)*, 159 :45–64, 1991.

Résumé : *Nous introduisons la propriété de permutabilité sans points fixes et nous présentons le résultat d'une étude systématique de cette propriété, en précisant les différences principales avec la classique et bien connue propriété de permutation pour les semi-groupes et les groupes (pour fournir un exemple intéressant, nous étudions en détail la propriété de permutabilité sans points fixes des groupes diedraux). Nous prouvons, par un argument combinatoire direct, qu'un semi-groupe finiment engendré et permutable*

sans points fixes est fini et nous prouvons qu'un semi-groupe permutable sans points fixes est fini si et seulement si ses idempotents sont centraux. Les groupes permutable sans points fixes sont caractérisés : un groupe (resp. un groupe finiment engendré) G est permutable sans points fixes si et seulement si son sous-groupe de commutateurs G' est fini (resp. G est centre-par-fini). Enfin nous définissons la propriété de faible permutable sans points fixes qui est équivalente à la propriété de permutable sans points fixes pour des groupes mais pas pour des semi-groupes.

[134] Giuseppe Pirillo. Sur un théorème combinatoire de Shirshov. *C. R. Acad. Sci. Paris Sér. I Math.*, 313(9) :631–634, 1991.

Résumé : *Soit A un alphabet, R un ordre total sur A , R^{-1} l'ordre inverse, $L(R)$ et $L(R^{-1})$ les ordres lexicographiques induits sur le monoïde A^* par les ordres R et R^{-1} respectivement. Nous prouvons, à l'aide du Théorème de Ramsey, que tout mot infini à droite sur A est soit ultimement périodique soit ultimement ω -divisé pour l'ordre $L(R)$, soit ultimement ω -divisé pour l'ordre $L(R^{-1})$.*

[133] Giuseppe Pirillo. On some properties of the Thue infinite word. In *Combinatorics '88, Vol. 2 (Ravello, 1988)*, Res. Lecture Notes Math., pages 325–329. Mediterranean, Rende, 1991.

Résumé : *Nous présentons la notion de l'excès gauche d'un facteur du mot infini de Thue $t = \text{abbabaabbaababba} \dots$, nous montrons que l'ensemble des facteurs de t ayant un excès gauche ambigu est $\{a, b, ab, ba, aba, bab\}$ et nous présentons une preuve très simple du fait que t est sans cube.*

1990

[79] Jacques Justin et Giuseppe Pirillo. Some remarks on the permutation property for semigroups. *European J. Combin.*, 11(2) :151–154, 1990.

Résumé : *La propriété de permutable P pour un semi-groupe a été récemment introduite par Restivo et Reutenauer et étudiée par plusieurs auteurs. Nous montrons que certaines constructions préservent ces propriétés (par exemple, si un semi-groupe S est l'union de ses sous-semi-groupes S_1, S_2, \dots, S_n et si tous les S_i ont la propriété P (resp. P^*) alors S a la propriété P (resp. P^*)).*

[78] Jacques Justin et Giuseppe Pirillo. Infinite words and permutation properties. *Semigroup Forum*, 40(1) :13–22, 1990.

Résumé. *Contrairement au théorème de Restivo et Reutenauer qui dans sa partie non triviale indique que chaque propriété P_n ($n \geq 2$) assure la finitude des semi-groupes périodiques finiment engendrés, nous prouvons qu'aucune propriétés P_n^* ($n \geq 3$) ne peut assurer la même conclusion.*

[77] Jacques Justin et Giuseppe Pirillo. A finiteness condition for semi-groups generated by a finite set of elements of finite order. *Pure Math. Appl. Ser. A*, 1(1) :45–48, 1990.

Résumé : *Soit S un semi-groupe engendré par un ensemble fini d'éléments d'ordre fini. Supposons qu'il existe un nombre entier n tel que chaque produit $x_1 x_2 \dots x_n$ d'éléments de S peut être réobtenu en permutant les x_i de telle manière que x_1 ou x_n ne soit pas un point fixe. Nous montrons que S est fini.*

[57] Gabriele Inglese et Giuseppe Pirillo. A remark on a result by A. Thue. *Pure Math. Appl. Ser. A*, 1(1) :21–28, 1990.

Résumé : *On montre que l'ensemble de nombres entiers P tels que chaque mot sans carré de longueur P sur l'alphabet $\{a, b, c\}$ contient les facteurs abc, acb, bac, bca, cab et cba est non vide et le minimum est 30 (et non 31 comme a été énoncé par Thue).*

[132] Giuseppe Pirillo. On a combinatorial property of Fibonacci infinite word. *Pure Math. Appl. Ser. A*, 1(2) :143–147, 1990.

Résumé : *Karhumäki a montré que dans le mot infini de Fibonacci il n'y a aucune 4-puissance, c.à .d., aucun facteur comme $uuuu$ pour un certain mot non vide u . Nous prouvons ceci d'une manière simple. Dans f il y a des cubes mais aucun cube n'est préfixe de f .*

[93] Jacques Justin, Giuseppe Pirillo et Stefano Varricchio. Some conditions on the product of two elements of a semigroup (Alcune condizioni sul prodotto di due elementi in un semigrupp). *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 1(2) :85–87, 1990.

Résumé *Soit S un semi-groupe finiment engendré. Si pour chaque paire*

x, y d'éléments de S on a que xy ou yx est un idempotent alors S est fini.

1989

[76] Jacques Justin et Giuseppe Pirillo. Mots sans carré et permutabilité de semi-groupes. *C. R. Acad. Sci. Paris Sér. I Math.*, 309(7) :523–525, 1989.

Résumé. *Un semi-groupe S a la propriété P_n^* , $n \geq 2$, si et seulement si étant donnés n éléments quelconques de S , deux au moins, parmi les $n!$ produits qu'on peut faire avec ces éléments, sont égaux. Nous présentons ici un semi-groupe engendré par quatre éléments, infini, dont tous les éléments sont d'ordre au plus 2 et qui a P_3^* . Ce résultat améliore ceux déjà connus.*

[75] Jacques Justin et Giuseppe Pirillo. Comments on the permutation property for semigroups. *Semigroup Forum*, 39(1) :109–112, 1989.

Résumé. *Nous prouvons que le semi-groupe bi-cyclique a la propriété de faible permutation P_5^* et nous posons la question s'il a la propriété P_4^* .*

[119] Brunetto Piochi et Giuseppe Pirillo. On a constant word in finite semigroups. *European J. Combin.*, 10(5) :457–467, 1989.

Résumé : *Nous prouvons, au moyen de techniques de la théorie des semi-groupes, que pour chaque semi-groupe fini il existe des éléments $a_1, a_2, \dots, a_k, a_{k+1}$ et des nombres entiers i_1, i_2, \dots, i_k tels que le mot $a_1 x^{i_1} a_2 x^{i_2} \dots a_k x^{i_k} a_{k+1}$ est constant. Nous caractérisons également les semi-groupes où le mot prend une forme courte et particulièrement simple.*

[131] Giuseppe Pirillo. The van der Waerden theorem and the Burnside problem for semigroups. *Arch. Math. (Basel)*, 53(1) :1–3, 1989.

Résumé : *Le théorème de van der Waerden sur les progressions arithmétiques est un cas particulier de la proposition suivante : tout semi-groupe fini est uniformément répétitif [103]. Pendant notre travail afin de donner une réponse à une question naturelle (les semi-groupes finiment engendrés et uniformément répétitifs sont-ils finis ?) nous avons trouvé le résultat suivant, qui est le résultat principal de cet article : n'importe quel semi-groupe uniformément répétitif est périodique. Nous employons également un résultat de Restivo et de Reutenauer sur le problème de Burnside pour les semi-groupes*

et nous avons le résultat suivant : un semi-groupe finiment engendré et n -permutable est fini si et seulement s'il est uniformément répétitif.

[129] Giuseppe Pirillo. A finiteness condition for finitely generated semi-groups. (Su una condizione di finitezza per semigrupperi finitamente generati). *Rend. Circ. Mat. Palermo (2)*, 38(2) :212–216, 1989.

Résumé : Nous fournissons une preuve plus simple du résultat principal de [74] et nous montrons également qu'un groupe G est fini si et seulement s'il existe un nombre entier n tel que pour chaque suite x_1, x_2, \dots, x_n de n éléments de S il existe i, j , $1 \leq i \leq j \leq n$ tels que $x_i \cdots x_j$ est l'identité de G . Nous prouvons ce résultat en utilisant seulement des faits élémentaires sur la forme canonique d'un élément d'un (semi-)groupe finiment engendré.

[130] Giuseppe Pirillo. Two finiteness conditions for finitely generated and periodic semigroups. *Acta Sci. Math. (Szeged)*, 53(3-4) :233–236, 1989.

Résumé : Dans cet article nous présentons deux conditions de finitude pour un semi-groupe finiment engendré et périodique. La première condition exige que la fonction qui compte le nombre d'éléments des i premières générations croît moins rapidement que $\frac{i(i+1)}{2}$. La seconde exige que le semi-groupe soit répétitif et qu'il existe un nombre entier positif p tels que chaque élément du semi-groupe est d'ordre plus petit que p .

[92] Jacques Justin, Giuseppe Pirillo et Stefano Varricchio. Unavoidable regularities and finiteness conditions for semigroups. Proceedings of the Third Italian Conference THEORETICAL COMPUTER SCIENCE, 1989.

Voir l'article in extenso en annexe.

1988

[74] Jacques Justin et Giuseppe Pirillo. Une condition de finitude pour les semi-groupes finiment engendrés. *C. R. Acad. Sci. Paris Sér. I Math.*, 306(2) :63–65, 1988.

Résumé : Nous démontrons qu'un semi-groupe finiment engendré S est fini si, et seulement si, il existe un entier n , tel que dans toute suite de n éléments de S il existe deux facteurs distincts non vides ayant la même valeur

dans S .

[128] Giuseppe Pirillo. On permutation properties for finitely generated semigroups. In *Combinatorics '86 (Trento, 1986)*, volume 37 of *Ann. Discrete Math.*, pages 375–376. North-Holland, Amsterdam, 1988.

Résumé. *Nous présentons un semi-groupe ayant la propriété de faible permutation P^* mais pas la propriété de permutation P et qui est également finiment engendré.*

[118] Brunetto Piochi et Giuseppe Pirillo. Sur une propriété de permutableté des groupes finis. *C. R. Acad. Sci. Paris Sér. I Math.*, 307(3) :115–117, 1988.

Résumé : *Soit $k(n)$ le plus petit entier tel que, pour tout groupe d'ordre n , tout produit $x_1 \cdots x_{k(n)}$ d'éléments du groupe peut être réordonné d'au moins une manière sans que sa valeur change. Nous prouvons que $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0$.*

[117] Brunetto Piochi et Giuseppe Pirillo. A proof of a theorem by G. Higman. In *Notes on algebraic systems, V*, volume 88, pages 7–14. Budapest, 1988.

Résumé : *Nous montrons que si le sous-groupe G' des commutateurs d'un groupe G est le groupe non cyclique K d'ordre quatre, alors G est 4-permutable. Notre preuve de ce résultat (d'abord annoncé par G. Higman) emploie seulement des notions élémentaires de théorie des groupes.*

1987

[127] Giuseppe Pirillo. On permutation properties for semigroups. In *Group theory (Bressanone, 1986)*, volume 1281 of *Lecture Notes in Math.*, pages 118–119. Springer, Berlin, 1987.

Résumé : *Nous présentons un semi-groupe qui n'a pas la propriété P définie par Restivo et Reutenauer, mais satisfait une propriété de permutation plus faible, P^* .*

1986

CONTRIBUTION A LA COMBINATOIRE DES MOTS

[73] Jacques Justin et Giuseppe Pirillo. On a natural extension of Jacob's ranks. *J. Combin. Theory Ser. A*, 43(2) :205–218, 1986.

Résumé : *Cet article est consacré à plusieurs variantes de la propriété de répétitivité pour des semi-groupes. En particulier, nous prouvons le résultat suivant : si S est un semi-groupe finiment engendré alors les conditions suivantes sont équivalentes : (i) S est fini; (ii) S est Ramseyen; (iii) S est fortement répétitif; (iv) S est fortement Ramseyen; (v) S est ω -répétitif.*

1984

[72] Jacques Justin et Giuseppe Pirillo. Two combinatorial properties of partitions of the free semigroup into finitely many parts. *Discrete Math.*, 52(2-3) :299–303, 1984.

Résumé : *Soit t le mot infini sans cubes de Thue. Nous prouvons ici qu'il y a une partition finie de l'ensemble de tous les facteurs de t tels que jamais trois d'entre eux sont de même longueur, consécutifs et dans la même partie. Nous présentons également une autre propriété d'irrégularité pour un mot infini similaire.*

1983

[103] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1983.

Résumé : *C'est le premier livre sur la combinatoire des mots. Nous sommes responsables du chapitre "Repetitive mappings and morphisms" qui contient essentiellement les résultats de Jacques Justin.*

[124] Giuseppe Pirillo. Codages et transductions. Consiglio Nazionale delle Ricerche, Istituto di Analisi Globale, 1979.

Résumé : *Ce volume contient les actes du premier colloque que nous avons organisé.*



Inequalities characterizing standard Sturmian and episturmian words

Giuseppe Pirillo^{a, b, *}

^aIASI CNR, Unità di Firenze, Dipartimento di Matematica “U. Dini”, Viale Morgagni 67/A 50134 Firenze, Italia

^bUniversité de Marne-la-Vallée 5, boulevard Descartes Champs sur Marne 77454 Marne-la-Vallée Cedex 2, France

Received 12 December 2003; received in revised form 13 August 2004; accepted 11 April 2005

Communicated by D. Perrin

To Franco Migliorini for his 65th birthday

Abstract

Considering the smallest and the greatest factors with respect to the lexicographic order we associate to each infinite word r two other infinite words $\min(r)$ and $\max(r)$. In this paper we prove that the inequalities $as \leq \min(s) \leq \max(s) \leq bs$ characterize standard Sturmian words (proper ones and periodic ones) and that the condition “for any $x \in A$ and lexicographic order $<$ satisfying $x = \min(A)$ one has $xs \leq \min(s)$ ” characterizes standard episturmian words.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Words; Lexicographic order; Infinite words; Sturmian words; Episturmian words

1. Introduction

We use notions and terminology of theoretical computer science (see [11,12]).

Sturmian words have been extensively studied for a long time (see, for example [2,14,15]). They are infinite words over a two letter alphabet ($\{a, b\}$ in this paper) and have several characterizations [12]. A recent one (*the set of the factors of an infinite word contains*

* IASI CNR, Unità di Firenze, Dipartimento di Matematica “U. Dini”, Viale Morgagni 67/A 50134 Firenze, Italia.

E-mail address: pirillo@math.unifi.it.

exactly one palindrome of even length and two palindromes of odd length if and only if it is Sturmian) is in [5]. In [4] we began to study the episturmian words, an apparently interesting generalization to k letter alphabet A , with $k \geq 2$.

In this paper a standard Sturmian word is presented as a *cutting sequence* of a half-line $y = \alpha x$. We consider the two possible cases: α irrational (*proper* standard Sturmian words) and α rational (*periodic* standard Sturmian words). So our terminology is different from the most usual one where standard Sturmian word refers only to the proper standard Sturmian case (α irrational).

Now, let A be a finite alphabet and let $<$ be a lexicographic order on the free monoid generated by A . Given an infinite word r over the alphabet A we denote by $\min(r)$ (resp. $\max(r)$) the infinite word having for each integer $k \geq 0$ its prefix of length k equal to the smallest (resp. greatest) factor of r of length k with respect to the lexicographic order $<$. With this notation and with the order $a < b$, we have the following property: *an infinite word r satisfies $\min(r) = ar$ and $\max(r) = br$ if and only if it is proper standard Sturmian* (see [20] and, in a different form, [6]). This result has been generalized to episturmian words in [9]: *an infinite word s over a finite alphabet A is a standard Arnoux-Rauzy sequence if and only if for any $x \in A$ and for any lexicographic order $<$ for which x is the smallest letter of A we have $xs = \min(s)$.*

Proposition 3.2 hereafter (*an infinite word s over a finite alphabet A is standard episturmian if and only if for any $x \in A$ and lexicographic order $<$ satisfying $x = \min(A)$ we have $xs \leq \min(s)$) is an extension of the just recalled result of [9]. Proposition 3.1, which has been announced without proof in [18] and which is a corollary of Proposition 3.2, has an interest in itself (preliminary lemmas are descriptions of $\min(r)$ and $\max(r)$). It may be remarked that our Propositions 3.1 and 3.2, which seem slight variations of [20,9] have fairly more complicated proofs than these two ones.*

This paper is organized as follows.

In Section “2 Preliminaries” we recall the definitions of (proper and periodic) Sturmian words, we make some comments on the Fibonacci word (a very particular case of Sturmian word), we recall the definition of episturmian words, we recall the notion of extremal words and, finally, we prove some preliminary lemmas which will be very useful in the next section.

In Section “3 Results” we prove the characterization of standard Sturmian words, Proposition 3.1, and the characterization of standard episturmian words, Proposition 3.2.

We seize the occasion to mention an error in [8]. Theorem 5.1 there (not used elsewhere) is false and should be ignored. This has been corrected in [10] where complete results are given.

2. Preliminaries

2.1. Words

We refer to [11,12] for the notions of *semigroup* and *monoid*, for the notions of *word*, *factor*, *prefix*, *suffix*, *free semigroup* and *free monoid* over a finite *alphabet* and for the notion of *morphism*.

When A is a finite alphabet, we denote by A^* the free monoid over A and by A^+ the free semigroup over A . We denote by ε the *empty word*, which has length 0, and we have $A^+ = A^* \setminus \{\varepsilon\}$. In order to make clear the sequence of *letters* which gives a word u we write $u = u(1)u(2) \dots u(n)$, $n \geq 1$. By definition, n is the *length* of u and it is denoted by $|u|$. We denote by A^n the set of words of length n over the alphabet A . A *right infinite word* w is a function from $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ into A . We write $w = w(1)w(2) \dots w(i) \dots$. We denote by A^ω the set of the right infinite words over A . If x is a letter and w is an infinite word then the infinite word $w' = xw$ is defined as follows: $w' = w'(1)w'(2) \dots w'(i) \dots$, where $w'(1) = x$ and, for $i > 1$, $w'(i) = w(i-1)$.

The *reversal* of a finite word $u = u(1)u(2) \dots u(m)$ is $u(m)u(m-1) \dots u(1)$ denoted by \tilde{u} if u is non-empty and is ε itself if $u = \varepsilon$. A word u is a *palindrome* (or a *palindromic word*) if and only if $u = \tilde{u}$. If $u \in A^*$, the *right palindromic closure* of u is the shortest palindrome $u^{(+)}$ having u as a prefix. We have $u^{(+)} = u'v\tilde{u}'$ where v is the longest palindromic suffix of u and $u = u'v$.

For a finite or infinite word m , we denote by $m(i, j)$ the factor $v = m(i)m(i+1) \dots m(j)$ of m . The set of *factors* of m is denoted by $F(m)$. The subset of the alphabet consisting of the letters which have at least one occurrence in m is denoted by $Alph(m)$. A factor u of an infinite word m is *right* (resp. *left*) *special* in m if there exist two letters a, b such that ua and ub (resp. au and bu) are also factors of m .

An infinite word p is *periodic* (resp. *ultimately periodic*) if there exists $k \geq 1$ such that $p(j+k) = p(j)$ for each $j \geq 1$ (resp. for each $j \geq i$ for some $i \geq 1$). If u is a finite word then u^ω is the periodic infinite word $uuu \dots u \dots$ having $|u|$ as a period.

Morphisms of A^* extend to infinite words. In particular, if $b = x(1) \dots, x(i) \in A$, and $\varphi : A^* \rightarrow A^*$ is a morphism then $\varphi(b) = \varphi(x(1)) \dots$.

2.2. Fibonacci word

Before giving a general definition of Sturmian words we present the most famous of them, the Fibonacci word (see for instance [3,7,11–13]), which we describe now. Let $\varphi : \{a, b\}^* \rightarrow \{a, b\}^*$ be the morphism given by $\varphi(a) = ab$, $\varphi(b) = a$. Let $f_0 = b$ and, for $n \geq 0$,

$$f_{n+1} = \varphi(f_n).$$

Notice that, for $n \geq 0$, $|f_n|$ is the n th element of the sequence of Fibonacci numbers F_n . For each $n \geq 2$, $f_n = f_{n-1}f_{n-2}$ implies that, for each $n \geq 1$, f_n is a prefix of f_{n+1} .

Hence there exists a unique infinite word, called the Fibonacci word, denoted by f such that, for each $n \geq 1$, f_n is a prefix of f . The first letters of f are

$$f = abaababaabaababaabaababaabaababaabaababa \dots$$

We are convinced that looking *carefully* at the properties of the Fibonacci word one can discover interesting properties of Sturmian (and episturmian!) words. Any case, the results of this paper were discovered in this way.

2.3. Sturmian words

There are many definitions of Sturmian words, see [12, Chap. 2]. In particular among them there is the following one.

Definition 2.1. An infinite word $s = s(1)s(2)s(3) \dots, s(i) \in \{a, b\}$, is *Sturmian* if there exist reals $\alpha, \rho \in [0, 1]$, such that either for all i

$$s(i) = a \text{ if } \lfloor \rho + (i + 1)\alpha \rfloor = \lfloor \rho + i\alpha \rfloor, \quad s(i) = b \text{ otherwise}$$

or for all i

$$s(i) = a \text{ if } \lceil \rho + (i + 1)\alpha \rceil = \lceil \rho + i\alpha \rceil, \quad s(i) = b \text{ otherwise.}$$

The infinite word s is proper Sturmian if α is irrational, it is periodic Sturmian if α is rational and it is standard Sturmian if $\rho = 0$.

Consider the grid g of the half-lines

$$c_k = \{(x, y) \mid x = k, y > 0, k \text{ positive integer}\}$$

and

$$l_h = \{(x, y) \mid y = h, x > 0, h \text{ positive integer}\}$$

and consider the half-line l which has the form

$$\{(x, y) \mid y = \beta x, x > 0, \beta \in (0, \infty)\}.$$

In this paper standard Sturmian words are obtained as the *cutting sequence* of a half-line l with respect to the grid g , that is the sequences like those of Fig. 1, which we will describe more precisely hereafter.

Put in a sequence, by increasing distance from the origin, the intersection points of l with the half-lines of the grid. Let $I(1), I(2), I(3), \dots, I(n), I(n + 1), \dots$ be this sequence.

If β is irrational, then for no $I(j)$ its two coordinates are both integers. If β is rational, then there are infinitely many $I(j)$ with the two coordinates integer.

In the first case, the cutting sequence of l is the infinite word $s = s(1)s(2) \dots s(j)s(j + 1) \dots$, where $s(j) = a$ if $I(j)$ is on an horizontal half-line l_h of g and $s(j) = b$ if $I(j)$ is on a vertical half-line c_k of g .

When β is rational, let $I(n)$ be the intersection point which has integer coordinate and is at the shortest distance from the origin. Note that there exist suitable relatively prime integers p, q such that $I(n) = (p, q)$ and $\beta = q/p$. If $I(n) \neq (1, 1)$ we proceed on $I(1), I(2), \dots, I(n - 1)$ in the same way used for the β irrational case and we obtain a word $w = s(1)s(2) \dots s(n - 1)$. If $I(n) = (1, 1)$ we put $w = \varepsilon$, ε the empty word.

Note that w , which is a prefix of our cutting sequence in construction, is restricted to the interior of the rectangle R of vertices $(0, 0), (p, 0), (p, q), (0, q)$ and, by geometrical

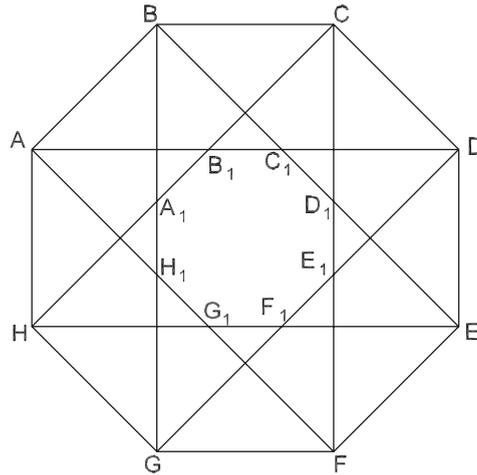


Fig. 2. Regular octagons with incommensurable sides.

are presented: for example, see Fig. 2, BC and B_1C_1 are incommensurable segments (hint: $AD = 2BC + B_1C_1$ and $BC = A_1D_1$).

2.4. Episturmian words

We follow the notations and terminology of [4,9].

Definition 2.2. A right infinite word is *episturmian* if the set of its factors is closed under reversal and it has at most one right (or equivalently left) special factor of each length.

A right infinite word is *standard episturmian* if all its left special factors are prefixes of it. Standard words (also called characteristic in the Sturmian case) are characterized in [4] using the right palindromic closure. More precisely s , right infinite, is standard episturmian if and only if there exists a right infinite word $\Delta = x(1)x(2)\dots, x(i) \in A$, called its *directive word* such that the sequence $u_1 = \varepsilon, u_2, u_3, \dots$ of palindromic prefixes of s (which are also right special factors of s) is given by $u_{n+1} = (u_n x(n))^{(+)}$.

Consider for any $a \in A$ the morphisms ψ_a given by $\psi_a(a) = a, \psi_a(x) = ax$ and $\bar{\psi}_a$ given by $\bar{\psi}_a(a) = a, \bar{\psi}_a(x) = xa$ for $x \in A, x \neq a$.

A right infinite word s is standard episturmian if there exist Δ as before and an infinite sequence of right infinite words $s^{(0)} = s, s^{(1)}, \dots$ given by $s^{(i-1)} = \psi_{x(i)}(s^{(i)})$; moreover Δ is the directive word of s .

Consider the prefix $x(1)x(2)\dots x(n)$ of the directive word $\Delta = x(1)x(2)\dots, x(i) \in A$, we write $\mu_n = \psi_{x(1)} \dots \psi_{x(n)}$. The standard episturmian word directed by Δ has the $h_n = \mu_n(x(n+1))$ as prefixes and we have the important relation

$$u_{n+1} = h_{n-1}u_n.$$

2.5. Lexicographic order

Let $<$ be a *total order* over the alphabet A . The *lexicographic order* on A^+ is defined as follows: for all $u, v \in A^+$

$$u < v \text{ if and only if } \begin{cases} \text{either } u \text{ is a proper prefix of } v \\ \text{or } u=ras, v=rbt \text{ for some } a, b \in A, a < b, \text{ and } r, s, t \in A^*. \end{cases}$$

The lexicographic order of infinite words arises from that of their prefixes. For example $f < (abaab)^\omega$, where f is the Fibonacci word, because $abaababaabaa < abaababaabab$.

2.6. Extremal words

Let $r \in \{a, b\}^\omega$ and let k be a positive integer. We denote by $\min(r|k)$ and by $\max(r|k)$, respectively, the smallest and the greatest factor of r of length k in the lexicographic order.

Definition 2.3. Let $r \in \{a, b\}^\omega$. We say that the pair

$$(\min(r|k), \max(r|k))$$

is *k extremal* for r .

It is clear that $\min(r|k)$ (resp. $\max(r|k)$) is a prefix of $\min(r|k+1)$ (resp. $\max(r|k+1)$). So we can define, by taking the limits, the infinite words

$$\min(r) = \lim_{k \rightarrow \infty} \min(r|k) \quad \text{and} \quad \max(r) = \lim_{k \rightarrow \infty} \max(r|k).$$

Remark 2.3. For a^ω and b^ω , we have $a^\omega = \min(a^\omega) = \max(a^\omega)$ and $b^\omega = \min(b^\omega) = \max(b^\omega)$; for any other $r \in \{a, b\}^\omega$, we have

$$\min(r) < \max(r).$$

Definition 2.4. Let $r \in \{a, b\}^\omega$. We say that the pair

$$(\min(r), \max(r))$$

of infinite words is *extremal* for r .

Remark 2.4. Given an infinite word r , if there is a positive integer i such that, for each positive integer k , $r(i, i+k-1) = \min(r|k)$ then i is called a *minimal rank* for r . This definition was given in [16] for obtaining a very short proof of Shirshov's Theorem.

2.7. Some preliminary lemmas

It is useful to note the following

Fact. Let r be in $\{a, b\}^\omega$ such that $(\min(r), \max(r)) = (ar, br)$. If k is a positive integer and u is the prefix of r of length $k - 1$ then

$$(\min(r \mid k), \max(r \mid k)) = (au, bu).$$

Example 2.1. The following pairs

$$(a, b), (aa, ba), (aab, bab), (aaba, baba), (aabaa, babaa)$$

are respectively 1, 2, 3, 4 and 5-extremal for f , the Fibonacci word. Note that if we cancel the first letter in the first or in the second component of each pair the remaining suffix is always a prefix of f .

This phenomenon is very well known and holds not only for Fibonacci word but also for any proper standard Sturmian word s . In fact, for this class of words we always have

$$(\min(s), \max(s)) = (as, bs).$$

This result is proved in [1] (where the terminology of Lyndon words and infinite Christoffel words is used) and it is also mentioned in [12, Exercise 2.2.13].

We proved (see [20]) the following.

Proposition 2.1. Let r be in $\{a, b\}^\omega$. The following conditions are equivalent:

- (a) r satisfies $(\min(r), \max(r)) = (ar, br)$;
- (b) r is proper standard Sturmian.

Remark 2.5. The proof of Proposition 2.1 is actually in this paper and it is “contained in” that of Proposition 3.1. Indeed, for sake of completeness and as [20] is unpublished, we repeat in the appropriate parts of Proposition 3.1 the arguments of the proof of Proposition 2.1.

Now before proving Lemmas 2.2–2.6, needed for proving our hereafter Proposition 3.1, we give an example of infinite words satisfying the hypotheses of these lemmas.

Example 2.2. An infinite word satisfying the conditions of the following lemmas is $s = (wab)^\omega$, for $w = aba$, i.e. the word $s = (abaab)^\omega$. We have $\min(s) = \min(s') = (awb)^\omega$ and $\max(s) = \max(s') = (bwa)^\omega$, where $s' = (wba)^\omega$.

Definition 2.5. Let $s \in \{a, b\}^\omega$ and $w \in F(s)$. We say that the property $P(s, w)$ holds if (awb, bwa) is $(|w| + 2)$ -extremal for s .

The previous example shows that $P((wab)^\omega, aba)$ holds. Similarly $P((aabaaab)^\omega, aabaa)$ holds. The aim of the following lemmas is precisely to show that if $P(s, w)$ holds then wab is a conjugate of wba and $s = (wab)^\omega$ or $s = (wba)^\omega$. In other words, if $P(s, w)$ holds then s is periodic Sturmian having form $s = (wab)^\omega$ or $s = (wba)^\omega$.

Lemma 2.2. *Let $s \in \{a, b\}^\omega$ and $w \in F(s)$ such that $P(s, w)$ holds. If $|w| = 0$ then s has no left special factors, no right special factors and $s = (ab)^\omega$ or $s = (ba)^\omega$. If $|w| > 0$ then for each integer i , $1 \leq i \leq |w|$, we have:*

- (a) *the prefix $w(1, i)$ is the unique left special factor of s of length i ;*
- (b) *the suffix $w(|w| - i + 1, |w|)$ is the unique right special factor of s of length i ;*
- (c) *$w(1, i) = w(|w| - i + 1, |w|)$.*

Proof. The case $|w| = 0$ is immediate since aa and bb are not factors of s .

Now suppose $|w| > 0$. We proceed by induction. Let us prove that (a)–(c) hold for $i = 1$.

Clearly, $w(1)$ is a left special factor of s for length 1. Suppose that $y \in \{a, b\}$, $y \neq w(1)$, is also left special. Then $aa, bb \in F(s)$ and this contradicts that $(aw(1), bw(1))$ is 2-extremal for s . Then only the first letter of w is a left special factor of s of length 1 and (a) holds for $i = 1$.

Clearly, the last letter of w , $w(|w|)$, is a right special factor of s for length 1. Suppose that $y \in \{a, b\}$, $y \neq w(|w|)$ is also right special. Then $aa, bb \in F(s)$ and this contradicts that $(aw(1), bw(1))$ is 2-extremal for s . Then only the last letter of w is a right special factor of s of length 1 and (b) holds for $i = 1$.

Moreover, the first and the last letter of w must coincide (in the opposite case $aa, bb \in F(s)$ and this contradicts that $(aw(1), bw(1))$ is 2-extremal for s) and clearly $\widetilde{w(1)} = w(|w|)$. So (c) holds for $i = 1$.

This completes the proof that (a)–(c) hold for $i = 1$.

Now, we can suppose that (a)–(c) hold for i , $1 \leq i < |w|$, and we will prove that they hold also for $i + 1$.

Let $u = w(1, i)$ and ux , $x \in \{a, b\}$, be prefixes of w , $|u| < |w|$. By induction, u is the unique left special factor of s of length i . Also ux is left special and we have to prove that it is the unique left special factor of s of length $i + 1$. Suppose that v of length $i + 1$ is also left special. The prefix of v of length i must be u because it is the unique left special factor of s of length i and so the last letter of v must be y , $y \neq x$. Then $aua, bub \in F(s)$ and this contradicts that (aux, bux) is $(i + 2)$ -extremal for s . Then only ux is a left special factor of s of length $i + 1$. So (a) holds for $i + 1$.

Now, suppose that $u = w(|w| - i + 1, |w|)$ and xu , $x \in \{a, b\}$, are suffixes of w , $|u| < |w|$. By induction, u is the unique right special factor of s of length i and \tilde{u} is a prefix of w . Clearly, xu is right special and we have to prove that it is the unique right special factor of length $i + 1$. Suppose that v of length $i + 1$ is also right special. The suffix of v of length i must be u because it is the unique right special factor of s of length i and so the first letter of v must be y , $y \neq x$. So $xu, yu \in F(s)$ with $x \neq y$ and consequently u is left special and $u = \tilde{u}$, a prefix of w . Then

$$aua = a\tilde{u}a \in F(s)$$

and

$$bub = b\tilde{u}b \in F(s)$$

and this contradicts that $(auw(i + 1), buw(i + 1))$ is $(i + 2)$ -extremal for s . Then only xu is a right special factor of s of length $|u| + 1$. So (b) holds for $i + 1$.

Finally, by way of contradiction, suppose that $ux = w(1, i)x$ is a prefix of w and that $yw(|w| - i + 1, |w|) = y\tilde{u}$ is a suffix of w with $x \neq y$. In all the possible cases (u is a palindrome or u is not a palindrome) there exists a prefix v of u such that ava and bvb belong to $F(s)$ and we are in contradiction with the extremality of $(avw(|v| + 1), bvw(|v| + 1))$. So $w(i + 1) = w(|w| - i)$ and consequently $w(1, i + 1) = w(|w| - i, |w|)$. So (c) holds for $i + 1$. This completes the proof that (a)–(c) hold for $i + 1$. \square

Example 2.2 (continued). Let $s = (abaab)^\omega$. Then a (resp. ab , resp. aba) is the unique left special factor of s of length 1 (resp. 2, resp. 3) and a (resp. ba , resp. aba) is the unique right special factor of s of length 1 (resp. 2, resp. 3).

Lemma 2.3. Let $s \in \{a, b\}^\omega$ and $w \in F(s)$ such that $P(s, w)$ holds. Then w is a palindrome.

Proof. Clearly, w is a palindrome in case $|w| = 0$ and, by previous lemma (point (c)), this also holds in the case $|w| > 0$. \square

Lemma 2.4. Let $s \in \{a, b\}^\omega$ and $w \in F(s)$ such that $P(s, w)$ holds. Then, for each $i \geq |w| + 1$,

- (a) no factor of s of length i is left special;
- (b) no factor of s of length i is right special.

Proof. (a) Suppose that the factor v of s of length $|v| = |w| + 1$ is left special. Then $v = wx$ for some letter x . So $xwx \in F(s)$. If $x = a$ then $awa \in F(s)$. As $awa < awb$, we are in contradiction with $awb = \min(r \mid |w| + 2)$. If $x = b$ then $bwb \in F(s)$. As $bwa < bwb$, we are in contradiction with $bwa = \max(r \mid |w| + 2)$. As in each of the two possible cases we are in contradiction, v is not left special and s has no left special factor of length $|w| + 1$. Since a prefix of a left special factor must also be left special, s has no left special factor of length $i > |w|$.

(b) Similar to part (a). \square

Example 2.2 (continued). Let $s = (abaab)^\omega$. No factor of s of length equal or greater than 4 is left special or right special.

Lemma 2.5. Let $s \in \{a, b\}^\omega$ and $w \in F(s)$ such that $P(s, w)$ holds. Then s has exactly $|w| + 2$ distinct factors of length $|w| + 2$.

Proof. The argument is based on the previous lemmas. When $|w| = 0$, by Lemma 2.2, s has exactly two factors of length 2: ab, ba . Let $|w| > 0$. By Lemma 2.2, s has exactly one right special factor of length i , $1 \leq i \leq |w|$. It follows that $|F(s) \cap A^{i+1}| - |F(s) \cap A^i| = 1$ for $1 \leq i \leq |w|$. As s has clearly two factors of length 1, namely a, b , it follows that s has $|w| + 1$ factors of length $|w|$. As w is right special, s has $|w| + 2$ factors of length $|w| + 1$. As s has no right special factors of length $|w| + 1$, the number of factors of s of length $|w| + 2$ remains $|w| + 2$. \square

Example 2.2 (continued). Let $s = (abaab)^\omega$. The factors of s of length 5 are $abaab$, $baaba$, $aabab$, $ababa$, $babaa$. They are the conjugates of $abaab$.

The next lemma is an “ad hoc” variation of a classical well-known result: *an infinite word over a finite alphabet whose right special factors are of bounded length is ultimately periodic*.

Lemma 2.6. *Let $s \in \{a, b\}^\omega$ and $w \in F(s)$ such that $P(s, w)$ holds. Then w is a palindromic prefix of a proper standard Sturmian word and s is periodic Sturmian with least period equal to $|w| + 2$. Moreover, $\min(s) = (awb)^\omega$ and $\max(s) = (bwa)^\omega$.*

Proof. Denote by u_i , $1 \leq i \leq |w| + 3$, the factor of s beginning at rank i and having length $|w| + 2$.

By Pigeon Hole Principle and Lemma 2.5 there exist i, j , $i \neq j$, such that $u_i = u_j$ and, for some prefixes v' and v'' of s , we have:

$$s = v'u_i s'$$

and

$$s = v''u_j s''$$

where s' and s'' are infinite words which are equal because in s , by Lemma 2.4, there is no right special factor of length $|w| + 2$.

So s is ultimately periodic. Moreover, it is periodic (otherwise there will be factors of length greater than $|w|$ which should be left special, in contradiction with Lemma 2.4). The period of s cannot be i with $i \leq |w|$ (otherwise s could not have special factors of length i , in contradiction with Lemma 2.2). The period of s cannot be $|w| + 1$ (otherwise $awa, bwb \in F(s)$ which is impossible by extremality of (awb, bwa)).

So necessarily the period of s is at least $|w| + 2$. On the other hand, by the above construction, the period of s is at most $|w| + 2$. Consequently, the least period of s is exactly $|w| + 2$. Hence s can be presented under the form:

$$s = u'awbawb\dots awb\dots$$

or the form

$$s = u''bwabwa\dots bwa\dots$$

for some u' and u'' , respectively, suffixes of awb and bwa .

Since awb and bwa are factors of s , we have that either awb is a factor of $(bwa)^2$ or bwa is a factor of $(awb)^2$. Consequently, awb and bwa are *conjugate*, i.e. there exist words w_1 and w_2 such that $awb = w_1w_2$ and $bwa = w_2w_1$.

We proved in [17] that *if awb and bwa are conjugate then w is a palindromic prefix of a proper standard Sturmian word*. Applying this here we get that s is a periodic Sturmian word.

Finally, being (awb, bwa) an extremal pair, we have $\min(s) = (awb)^\omega$ and $\max(s) = (bwa)^\omega$. \square

Example 2.2 (continued). Let $s = (abaab)^\omega$. Note that aba is a palindromic prefix of the Fibonacci word.

3. Results

Now we prove our

Proposition 3.1. *Let s be an infinite word on $\{a, b\}$. The following conditions are equivalent:*

- (i) s is a proper standard Sturmian word or a periodic standard Sturmian word;
- (ii) $as \leq \min(s) \leq \max(s) \leq bs$.

Proof. (i) \Rightarrow (ii). We have to prove that a *proper* standard Sturmian word as well as a *periodic* standard Sturmian word verifies the inequalities in (ii). Let us distinguish the two possible cases.

Proper. Let $s = s(1)s(2) \dots s(i) \dots$ be a proper standard Sturmian word. The pair (a, b) is 1-extremal for s .

Now, let $u = s(1)s(2) \dots s(i)$ be the prefix of length i of s . Put $c = s(i+1)$ and suppose by induction hypothesis, that

$$(au, bu)$$

is $(i+1)$ -extremal for s . We will prove that

$$(auc, buc)$$

is $(i+2)$ -extremal for s . By contradiction, assume that

$$(aud, bue)$$

is $(i+2)$ -extremal for s . Suppose, for instance, $d \neq c$. As $aud < auc$ we have $d = a$ and $c = b$, thus au is right special. As in a proper standard Sturmian word the mirror image of a right special factor is a prefix we have that \widetilde{au} is a prefix of s . Hence $a = c = b$. Contradiction.

Similarly, we reach a contradiction if we suppose that $e \neq c$.

So the minimal (resp. maximal) element in the lexicographic order of length n is av (resp. bv) with v prefix of s of length $n-1$. Hence $as = \min(s)$ and $bs = \max(s)$ and in conclusion (ii) is satisfied in the stronger form $as = \min(s) < \max(s) = bs$.

Periodic. Let us consider first the case $|Alph(s)| = 1$. If $s = a^\omega$ then $as = \min(s) = \max(s) < bs$. If $s = b^\omega$ then $as < \min(s) = \max(s) = bs$. In both cases (ii) is satisfied. So we can suppose that s is a periodic standard Sturmian word with $|Alph(s)| = 2$, i.e., different from a^ω and from b^ω .

As seen in Section 2.3, s has form $(wab)^\omega$ or $(wba)^\omega$ where w is a palindrome corresponding to the cutting points of the half-line $y = (q/p)x$ with the lines of the grid in the strict interior of the rectangle of vertices $(0, 0)$, $(p, 0)$, (p, q) , $(0, q)$, for suitable relatively primes p, q .

Suppose for instance $(wab)^\omega$ (the other case is similar).

Now, note that w is also a palindromic prefix of infinitely many proper standard Sturmian words. In particular when α is suitably near to q/p , the cutting sequence of $y = \alpha x$ contains the factors of any large required length which are in the cutting sequence of $y = (q/p)x$. More precisely, the cutting sequence, say s' , of $y = \alpha x$ with $\alpha = q/p + \varepsilon$, for a suitable small positive ε , begins with $(wab)^2$ and consequently s' contains as factors all the factors of s having length at most $|w| + 2$ and in particular those having length at most $|w| + 1$.

It is also well known that wab is a conjugate of wba (i.e. there exist words w_1 and w_2 such that $awb = w_1w_2$ and $bwa = w_2w_1$, see, for example, [11,12,17]) and consequently $(wab)^\omega = u'(wba)^\omega$, for some prefix u' of wab , and so there exist prefixes u'' and u''' of wab such that $s = u''(awb)^\omega = u'''(bwa)^\omega$. Consequently $awb, bwa \in F(s)$ and $aw, bw \in F(s)$.

As $aw, bw \in F(s) \cap \{a, b\}^{|w|+1} = F(s') \cap \{a, b\}^{|w|+1}$ and as the pair (aw, bw) is $(|w| + 1)$ -extremal for s' (see the proof given just above for the proper case), then (aw, bw) is also $(|w| + 1)$ -extremal for s .

Now, as in s two factors of length $|w| + 2$ must contain the same number of occurrences of a and as $awb, bwa \in F(s)$ we have $awa, bwb \notin F(s)$ and consequently (awb, bwa) is $(|w| + 2)$ -extremal for s .

As awb is minimal and it is a factor of the periodic word s we have that $\min(s) = (awb)^\omega$. Similarly considering bwa which is maximal, we have that $\max(s) = (bwa)^\omega$.

As awa is a prefix of as and awb is a prefix of $\min(s)$ we have $as < \min(s)$. On the other hand, $\max(s) = (bwa)^\omega = b(wab)^\omega = bs$. So (ii) is satisfied in the stronger form $as < \min(s) < \max(s) = bs$.

The case $s = (wba)^\omega$ leads to $as = \min(s) < \max(s) < bs$.

This ends the proof of (i) \Rightarrow (ii).

(ii) \Rightarrow (i). Let us distinguish the following cases:

- (1) $as = \min(s)$ and $\max(s) = bs$;
- (2) $as = \min(s)$ and $\max(s) < bs$;
- (3) $as < \min(s)$ and $\max(s) = bs$;
- (4) $as < \min(s)$ and $\max(s) < bs$.

Case 1: $as = \min(s)$ and $\max(s) = bs$.

For any $i \geq 1$, the prefix u of length i of s is clearly a left special factor of s . We will prove that u is the unique left special factor of s of length i .

By contradiction, suppose that $v \in F(s)$, $|v| = |u|$, $v \neq u$ and v is left special. Since (au, bu) is $(i + 1)$ -extremal for s , we have both $au \leq av$ and $bu \geq bv$. So $u \leq v$ and $u \geq v$ and $u = v$. Contradiction.

So for each i , s has a unique left special factor of length i which is its prefix of length i . So s is proper standard Sturmian.

Case 2: $as = \min(s)$ and $\max(s) < bs$.

First, note that from $\max(s) < bs$ it follows that letter a occurs in s .

If $s = a^\omega$ then it verifies Case 2 ($a \cdot a^\omega = a^\omega = \min(a^\omega) = \max(a^\omega) < b \cdot a^\omega$) and it is clearly periodic standard Sturmian.

If s contains also an occurrence of b then $\max(s) = bs'$ for some $s' \in \{a, b\}^\omega$.

Let w be the longest common prefix of s and s' . Then, for some $x, y \in \{a, b\}$, $y \neq x$, we have

awx is a prefix of $as = \min(s)$,

bwx is a prefix of bs ,

$bw y$ is a prefix of $bs' = \max(s)$.

As $\max(s) = bs' < bs$ we have $y < x$ and consequently $y = a$ and $x = b$. Then (awb, bwa) is $(|w| + 2)$ -extremal for s . By Lemma 2.6, w is a palindromic prefix of a proper standard Sturmian word and s is a periodic Sturmian word of period $|w| + 2$. Being (awb, bwa) an extremal pair, we have also $\min(s) = (awb)^\omega$ and $\max(s) = (bwa)^\omega$.

As $as = \min(s) = (awb)^\omega$, we have that wb is a prefix of s and consequently s is the periodic standard Sturmian word $(wba)^\omega$.

Case 3: $as < \min(s)$ and $\max(s) = bs$.

Similarly to Case 2 we first consider the case $s = b^\omega$ and then we prove that the other words satisfying this case are the periodic standard Sturmian words of the form $(wab)^\omega$.

Case 4: $as < \min(s)$ and $\max(s) < bs$.

We will prove that this case is impossible. The two inequalities imply $s \in \{a, b\}^\omega - \{a^\omega, b^\omega\}$. There exist two infinite words s' and s'' satisfying $\min(s) = as'$ and $\max(s) = bs''$ and two finite words w' and w'' which are respectively the longest common prefix of s and s' and longest common prefix of s and s'' . So:

$aw'a$ is a prefix of as ,

$aw'b$ is a prefix of $as' = \min(s)$,

$bw''a$ is a prefix of $bs'' = \max(s)$,

$bw''b$ is a prefix of bs .

If $|w'| < |w''|$ then $w'a$ is a prefix of w'' and consequently $bw'a$ is prefix of $\max(s)$. As $aw'b$ is a prefix of $\min(s)$ then $P(s, w')$.

If $|w'| = |w''|$ then $w' = w''$. As $aw'b$ is a prefix of $\min(s)$ and $bw'a$ is a prefix of $\max(s)$ then $P(s, w')$. If $|w''| < |w'|$ then $w''b$ is a prefix of w' and consequently $aw''b$ is prefix of $\min(s)$. As $bw''a$ is a prefix of $\max(s)$ then $P(s, w'')$.

In any case there exists a prefix w of s (indeed the shorter between w' and w'') such that $P(s, w)$.

From Lemma 2.6, we deduce that $as < (awb)^\omega < (bwa)^\omega < bs$. As w is a prefix of s , the prefixes of as and $(awb)^\omega$ of length $|w| + 1$, as well as the prefixes of bs and $(bwa)^\omega$ of the same length, coincide. Now, whatever happens for the value a or b of $s(|w| + 1)$, one of the two inequalities $as < (awb)^\omega$ or $(bwa)^\omega < bs$ holds and the other fails. Contradiction. This ends the examination of Case 4. \square

Following the previous proposition, all the words over $\{a, b\}$ satisfying $as \leq \min(s) \leq \max(s) \leq bs$ are proper or periodic standard Sturmian words. More precisely we proved the following: *let s be an infinite word on $\{a, b\}$. Then:* (i) $as = \min(s) < \max(s) = bs$ if and only if s is a proper Standard Sturmian word; (ii) $as = \min(s) < \max(s) < bs$ if and only if s is a periodic Sturmian word having the form $(wba)^\omega$ for some palindromic word w such that wab is a conjugate of wba ; (iii) $as < \min(s) < \max(s) = bs$ if and only if s is a periodic Sturmian word having the form $(wab)^\omega$ for some palindromic word w such that wab is a conjugate of wba ; (iv) it is impossible that $as < \min(s) \leq \max(s) < bs$.

Remark 3.1. Proposition 3.1 is a particular case of the next proposition as easily seen if we observe that on a two letter alphabet proper and periodic episturmian words are exactly the proper and periodic Sturmian words and that when A is a two letter alphabet there are only two lexicographic orders over A^* .

Proposition 3.2. *For an infinite word s over a finite alphabet A the following properties are equivalent:*

- (i) s is standard episturmian;
- (ii) for any $x \in A$ and lexicographic order $<$ satisfying $x = \min(A)$ we have $xs \leq \min(s)$.

Proof. Denote by s_i the prefix of length i of the infinite word s .

(i) \Rightarrow (ii). Here s is standard episturmian. With $x = \min(A)$ suppose by contradiction $xs > \min(s)$ and let n be minimal such that $xs_{n-1} > \min(s | n)$. Then $xs_{n-2} \leq \min(s | n-1)$. Set $s_{n-1} = s_{n-2}y$, $y \in A$, and $w = \min(s | n-1)$. Then $\min(s | n) = wz$ for some $z \in A$. From $xs_{n-2} \leq w$ and $xs_{n-2}y > wz$ it follows $w = xs_{n-2}$ and $y > z$.

As $s_{n-2}y$ and $s_{n-2}z$ are factors of s , s_{n-2} is right special, hence is a palindromic prefix, u_p say, of s . Here the notations for s , its directive word, and so on, are as in the Preliminaries.

We can assume that among all standard episturmian words s and lexicographic orders $<$ such that $xs > \min(s)$ where $x = \min(A)$ we have chosen one with minimal value of p . Consider $s^{(1)}$ given by $\psi_{x_1}(s^{(1)}) = s$ and the sequence $u_1^{(1)} = \varepsilon, u_2^{(1)}, \dots, u_i^{(1)}, \dots$ of its palindromic prefixes. It easily follows from the definition of ψ_{x_1} that if dx_1ex_1 is a prefix of s then uv is a prefix of $s^{(1)}$ where $d = \psi_{x_1}(u)$ and $x_1e = \psi_{x_1}(v)$. By [4, Lemma 3] we know that $u_p = \psi_{x_1}(u_{p-1}^{(1)})x_1$. It follows that $u_{p-1}^{(1)}y$ is a prefix of $s^{(1)}$ and that $xu_{p-1}^{(1)}z \in F(s^{(1)})$. Indeed consider $xu_{p-1}^{(1)}z$ for instance. If $x \neq x_1$ then $x_1xu_pz = x_1xs_{n-2}z \in F(s)$. If $z = x_1$ then $x_1xu_p = \psi_{x_1}(v)$ for some factor v of $s^{(1)}$ and $v = xu_{p-1}^{(1)}x_1$. If $z \neq x_1$ then $x_1xu_pz = \psi_{x_1}(v)$ for some factor v of $s^{(1)}$, and $v = xu_{p-1}^{(1)}z$. Case $x = x_1$ has similar proof.

Now, as $xu_{p-1}^{(1)}y > xu_{p-1}^{(1)}z$ because $y > z$, we get $xu_{p-1}^{(1)}y > \min(s^{(1)} | (2 + |u_{p-1}^{(1)}|))$ and this contradicts the minimality of p .

Thus $xs \leq \min(s)$.

(ii) \Rightarrow (i). We will show first that s satisfying (ii) is episturmian. If, for some $n \in \mathbb{N}_+$, s has at least two different left special factors u, v of length $n-1$, take n minimal with this property. Set $u = u'y, v = v'z, y, z \in A$. Then u', v' are left special whence, by minimality of n , $u' = v'$, whence $y \neq z$. We have $au'y, bu'y, cu'z, du'z \in F(s)$, for some letters $a, b, c, d \in \text{Alph}(s)$, $a \neq b, c \neq d$. If $u' \neq s_{n-2}$ write $u' = weu'', s_{n-2} = wfs'', e, f \in A, e \neq f$. As $a \neq b$ suppose for instance $f \neq a$ and consider an order $<$ of A such that $a = \min(A), e < f$. Thus $as_{n-2} > au'$, a contradiction. Thus $u' = s_{n-2}$. Now set $s_{n-1} = s_{n-2}x, x \in A$. As $y \neq z$ we can suppose $x \neq y$, for instance. Consider an order $<$ of A such that $a = \min(A)$ and $y < x$. Then $as_{n-1} = as_{n-2}x > as_{n-2}y = au'y$ in contradiction with $as \leq \min(s)$.

Consequently, $u = v$. Thus for all n , s has at most one left special factor of length n . Moreover, the argument above shows that all left special factors of s are prefixes of s .

Let us show now that $F(s)$ is closed under reversal. Suppose by contradiction that for some $u = xvy$, $x, y \in A$, $u \in F(s)$, $\tilde{u} = y\tilde{v}x \notin F(s)$ and suppose $|u|$ minimal. Then $y\tilde{v}, \tilde{v}x \in F(s)$ whence, for some letter $t \neq x$, $y\tilde{v}t \in F(s)$. Thus by minimality of $|u|$, $tv \in F(s)$. As $xv \in F(s)$, v is left special hence is a prefix of s .

Now suppose that $\tilde{v}x$ is not a prefix of s . As it is a factor of s , there exists $m \in A$ such that $m\tilde{v}x \in F(s)$, and $m \neq y$. Thus as $m\tilde{v}, y\tilde{v} \in F(s)$, \tilde{v} is left special hence is a prefix of s . Thus $\tilde{v} = v$ whence in particular $x \neq y$. Thus, as we have supposed that $\tilde{v}x$ is not a prefix of s , va is a prefix of s for some letter $a \neq x$. Thus $yvt, mvx, xvy, va \in F(s)$ and $yvx \notin F(s)$. Recall $t \neq x, m \neq y, a \neq x, y \neq x$. At least one of the two following cases is realized.

(c) $a \neq m$. Consider an order $<$ such that $m = \min(A)$, $x < a$. Then $mva > mvx$ contradicting $ms \leq \min(s)$.

(c') $a \neq y$. Then with order $<$ such that $x = \min(A)$, $y < a$ we get a similar contradiction $xva > xvy$.

Consequently, $\tilde{v}x$ is a prefix of s . As v is left special, hence is a prefix of s , $v = \tilde{v}$. Thus $xvy, yvt \in F(s)$. Consider an order $<$ such that $\min(A) = y$ and $t < x$ (possible as $y \neq x$ and $t \neq x$), we get $yvx > yvt$ contradicting $ys \leq \min(s)$. Consequently $F(s)$ is closed under reversal and s satisfies the definition of episturmian words. Moreover, as remarked above all its left special factors are prefixes of it, hence it is standard episturmian. \square

Acknowledgements

We thank Dipartimento di Matematica “U. Dini” for the friendly hospitality. We thank J. Justin for the very helpful conversations. We thank also the referees for their constructive remarks.

References

- [1] J.-P. Borel, F. Laubie, Quelques mots sur la droite projective réelle, *J. Théor. Nombres Bordeaux* 5 (1) (1993) 23–51.
- [2] C.B. Christoffel, *Observatio arithmetica*, *Math. Annal.* 6 (1875) 145–152.
- [3] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theoret. Comput. Sci.* 183 (1) (1997) 45–82.
- [4] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* 255 (1–2) (2001) 539–553.
- [5] X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* 223 (1–2) (1999) 73–85.
- [6] S. Gan, Sturmian sequences and the lexicographic world, *Proc. Amer. Math. Soc.* 129 (5) (2001) 1445–1451 (electronic).
- [7] J. Justin, G. Pirillo, Decimations and Sturmian words, *Theoret. Inform. Appl.* 31 (3) (1997) 271–290.
- [8] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* 276 (1–2) (2002) 281–313.
- [9] J. Justin, G. Pirillo, On a characteristic property of Arnoux–Rauzy sequences, *Theoret. Inform. Appl.* 36 (4) (2002) 385–388.
- [10] J. Justin, G. Pirillo, Episturmian words: shifts, morphisms and numeration systems, *Internat. J. Found. Comput. Sci.* 15 (2) (2004) 329–348.
- [11] M. Lothaire, *Combinatorics on words*, *Encyclopedia of Mathematics and its Applications*, Vol. 17, Addison-Wesley Publishing Co., Reading, MA, 1983.

- [12] M. Lothaire, Algebraic combinatorics on words, *Encyclopedia of Mathematics and its Applications*, Vol. 90, Cambridge University Press, Cambridge, 2002.
- [13] F. Mignosi, G. Pirillo, Repetitions in the Fibonacci infinite word, *Theoret. Inform. Appl.* 26 (3) (1992) 199–204.
- [14] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories, *Amer. J. Math.* 62 (1940) 1–42.
- [15] M. Morse, G.A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, *Duke Math. J.* 11 (1944) 1–7.
- [16] G. Pirillo, A proof of Shirshov’s theorem, *Adv. Math.* 124 (1) (1996) 94–99.
- [17] G. Pirillo, A curious characteristic property of standard Sturmian words, in: *Algebraic Combinatorics and Computer Science—A Tribute to Gian-Carlo Rota*, Springer, Italia, Milan, 2001, pp. 541–546.
- [18] G. Pirillo, Inequalities characterizing standard Sturmian words, *Pure Math. Appl.* 14 (1–2) (2003) 141–144 Dedicated to the 65th birthday of Professor Franco Migliorini.
- [19] G. Pirillo, Numeri irrazionali e segmenti incommensurabili, *Nuova Secondaria* 7 (2005) 87–91.
- [20] G. Pirillo, Characterization of (proper) standard Sturmian words, unpublished manuscript.

EPISTURMIAN WORDS: SHIFTS, MORPHISMS AND NUMERATION SYSTEMS

JACQUES JUSTIN*[†] and GIUSEPPE PIRILLO[‡]

[†]*LIAFA, ERS 586, Université Paris VII, case 7014,
2 place Jussieu, 75251 Paris Cedex 5, France*

[‡]*IASI CNR Viale Morgagni 67/A 50134 Firenze,
Italy and Université de Marne-la-Vallée,
5 boulevard Descartes Champs-sur-Marne,
77454 Marne-la-Vallée Cedex 2, France*

[†]*justin@liafa.jussieu.fr*

[‡]*pirillo@math.unifi.it*

Received 12 February 2003

Accepted 4 October 2003

Communicated by Juhani Karhumäki

Abstract: Episturmian words, which include the Arnoux-Rauzy sequences, are infinite words on a finite alphabet generalizing the Sturmian words and sharing many of their same properties. This was studied in previous papers. Here we gain a deeper insight into these properties. This leads in particular to consider numerations systems similar to the Ostrowski ones and to give a matrix formula for computing the number of representations of an integer in such a system. We also obtain a complete answer to the question: if an episturmian word is morphic, which shifts of it, if any, also are morphic ?

Introduction

Sturmian words on a 2-letter alphabet were explicitly introduced in [12] with some approaches going back to Bernoulli and other independent approaches such as [7]. They have been extensively studied, see survey in [11, Chap. 2 by J. Berstel and P. Séébold]. An interesting generalization to a 3-letter alphabet (and even to any finite alphabet) is known as the Arnoux-Rauzy sequences [13, 14, 1, 16] and even a slightly wider class of infinite words, the episturmian words, is considered in [8, 10, 3]. These words have many properties extending those of Sturmian words.

Our aim is to study more in detail some properties considered in [10]. As in [8, 10] our study is purely combinatorial. We make use of biinfinite words which allow more natural formulations. In [10] it was shown that any episturmian word could be obtained or “directed” using an infinite “spinned word”, which was a way

*Present mailing address: J. JUSTIN, 19 rue de Bagneux 92330 SCEAUX France.

of representing the infinite sequence of elementary morphisms “generating” it (episturmian words are “ S -adic”, see for instance [16]). The relation between episturmian words and “spinned directive words” is not one-to-one. This is studied in Section 3 while in Section 2 we study the same problem for finite spinned words. In Section 2 we also see that there are numeration systems, resembling the Ostrowski systems [5] associated to Sturmian words and we find results for the number of representations of integers in such systems similar to those for Sturmian words [6, 3]. This section can also be viewed as a study of the monoid of pure episturmian morphisms as the quotient of the free monoid of spinned finite words by the “block-equivalence” relation. This has some relation with the study of episturmian morphisms made by G. Richomme in [15].

In Section 3 we extend the “block-equivalence” to infinite spinned words and show that, roughly speaking, two infinite spinned words direct the same episturmian word if and only if they are block-equivalent. We also describe the particular case of periodic episturmian words.

In Section 4 we apply results of Section 3 to (infinite) episturmian words which are morphic (i.e., fixed points of morphisms) and this allows to correct an error in [10]. It was said there (Theorem 5.1 of Subsection 5.1) that if an episturmian word is morphic then any shift of this word also is morphic. This is false and indeed I. Fagnot has shown [9] that if s is a morphic standard Sturmian word on $\{a, b\}$ then as, bs, abs, bas (which are morphic [4]) are the only morphic Sturmian words related to s by a shift. Here we give complete results for all episturmian words.

1. Preliminaries

1.1. Words

The alphabet A is finite, ε is the empty word, $A^+ = A^* \setminus \{\varepsilon\}$ is the set of (finite) non-empty words. A right infinite (resp. left infinite, biinfinite) word is a function from $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ (resp. $\mathbb{Z} \setminus \mathbb{N}_+, \mathbb{Z}$) into A . For instance a biinfinite word b is represented by a sequence $(b(i))_{i \in \mathbb{Z}} = \cdots b(-1)b(0).b(1)b(2)\cdots$, $b(i) \in A$, with a dot written between $b(0)$ and $b(1)$ for avoiding ambiguity. The set of right infinite words on A is denoted by A^ω . The shift operator T is defined for a biinfinite word $b = (x_i)_{i \in \mathbb{Z}}$ by $T(b) = (x_{i+1})_{i \in \mathbb{Z}}$, T^k is defined by iteration. This extends to right infinite words for $k \geq 0$ and left infinite words for $k \leq 0$. For a finite word the shift T will be here a circular one, more precisely if $w = xv$, $x \in A, v \in A^*$ then $T(w) = vx$.

The *reversal* of a finite word $u = u(1)u(2)\cdots u(m)$ is $u(m)u(m-1)\cdots u(1)$ denoted by \tilde{u} . This extends to infinite words, for instance the reversal of the biinfinite word $l.r$, l left infinite, r right infinite is $\tilde{r}.\tilde{l}$.

If $u \in A^*$, the *right palindromic closure* of u is the shortest palindrome $u^{(+)}$ having u as a prefix. We have $u^{(+)} = uv^{-1}\tilde{u}$, where v is the longest palindromic suffix of u .

For a finite or infinite word m , if $v = m(i)m(i+1)\cdots m(j)$, $i \leq j$, then v is a *factor* of m denoted by $m(i, j)$. Here we often consider an *occurrence* of a factor, that

is, this factor together with its position. The terminology will be rather informal and clear from the context, for example if $m = m'vm''$ this defines a particular occurrence of v in m . Factors $u = m(i, j)$ and $v = m(k, l)$, with $i \leq j$, $k \leq l$ and $i \leq k$ of m are *disjoint* if $j < k$ (in particular *adjacent* if $j = k - 1$) and *overlap* if $j \geq k$. The set of all infinite words having the same factors as the infinite word m will be denoted by $\mathcal{C}(m)$.

A factor u of an infinite word m is *right* (resp. *left*) *special* if for at least two different letters x, y , ux and uy (resp. xu and yu) are factors of m .

The set of letters occurring in a word t is denoted by $Alph(t)$ and, if t is right infinite, the set of letters occurring infinitely many times in it is $Ult(t)$. Morphisms of A^* (endomorphism here) extend to infinite words. In particular if $b = \cdots x_{-1}x_0.x_1 \cdots$, $x_i \in A$, is biinfinite then $\varphi(b) = \cdots \varphi(x_{-1})\varphi(x_0).\varphi(x_1) \cdots$. An infinite word is *morphic* if it is a fixed point of some non-trivial morphism.

Some basic notions, not recalled here, such as (uniformly) recurrent, primitive, periodic can be found in [11, Chap. 1].

1.2. Episturmian words

We follow the notations and terminology and shortly recall basic results of [8, 10].

Definition 1.1. A right infinite word is *episturmian* if the set of its factors is closed under reversal and it has at most one right (or equivalently left) special factor of each length.

When $|A| = 2$ this gives the (proper as well as periodic) Sturmian words.

This definition extends to left infinite and biinfinite words.

A right infinite episturmian word is *standard* if all its left special factors are prefixes of it. Standard words (also called characteristic in the Sturmian case) are characterized in [8, 10] using the right palindromic closure. More precisely s , right infinite, is standard episturmian if and only if there exists a right infinite word $\Delta = x_1x_2 \cdots$, $x_i \in A$, called its *directive word*, such that the sequence $u_1 = \varepsilon, u_2, u_3, \dots$ of the palindromic prefixes of s is given by $u_{n+1} = (u_n x_n)^{(+)}$.

Consider for any $a \in A$ the morphisms ψ_a given by $\psi_a(a) = a$, $\psi_a(x) = ax$ and $\bar{\psi}_a$ given by $\bar{\psi}_a(a) = a$, $\bar{\psi}_a(x) = xa$, for $x \in A$, $x \neq a$. These “elementary” morphisms generate by composition the monoid of *pure episturmian morphisms* (more generally a morphism is *episturmian* if it is obtained by composition of a pure episturmian morphism and a permutation of the alphabet). A right infinite word s is standard episturmian if there exist $\Delta = (x_i)_{i \geq 1}$, $x_i \in A$, and an infinite sequence of right infinite words $s^{(0)} = s, s^{(1)}, \dots$ such that $s^{(i-1)} = \psi_{x_i}(s^{(i)})$. In [10] this is extended as follows.

A right infinite word t is episturmian if there exist Δ as before, a *spinned word* $\check{\Delta} = \check{x}_1\check{x}_2 \cdots$ where \check{x}_i is either x_i (spin 0) or \bar{x}_i (spin 1) and an infinite sequence of recurrent right infinite words $t^{(0)} = t, t^{(1)}, \dots$ given by $t^{(i-1)} = \psi_{x_i}(t^{(i)})$ if \check{x}_i has spin 0 and $t^{(i-1)} = \bar{\psi}_{x_i}(t^{(i)})$ if \check{x}_i has spin 1. We say that $\check{\Delta}$ *directs* (or is a *spinned directive word* for) t .

We can even extend this to finite words. We will say that a spinned finite word $\check{w} = \check{y}_1\check{y}_2 \cdots \check{y}_n$, $y_i \in A$, directs the morphism $\check{\psi}_{y_1} \cdots \check{\psi}_{y_n}$ where $\check{\psi}_{y_i} = \psi_{y_i}$ or $\bar{\psi}_{y_i}$ according to the spin of \check{y}_i and we represent this morphism by $\mu_{\check{w}}$ (in particular, for x a letter, $\mu_x = \psi_x$). In this case “directs” is a morphism from the free monoid $(A \cup \bar{A})^*$ where $\bar{A} = \{\bar{x} \mid x \in A\}$ onto the monoid of pure episturmian morphisms.

In particular considering the prefix $x_1x_2 \cdots x_n$ of the directive word $\Delta = x_1x_2 \cdots$, $x_i \in A$, we write μ_n for $\mu_{x_1x_2 \cdots x_n}$ and we have $\mu_n = \psi_{x_1} \cdots \psi_{x_n}$. The standard episturmian word directed by Δ has the $h_n = \mu_n(x_{n+1})$ as prefixes and we have the important relation

$$u_{n+1} = h_{n-1}u_n. \tag{1}$$

The above characterization of right episturmian words (in short rew) by a sequence $(t^{(i)})$ extends to biinfinite episturmian words (in short bew), and all $t^{(i)}$ are now bew. Moreover if $t = l.r$ is directed by $\check{\Delta}$ with associated bew $t^{(i)} = l^{(i)}.r^{(i)}$ then r is directed by $\check{\Delta}$ with associated rew $r^{(i)}$.

Consider a rew, $t \in A^\omega$. If it is standard directed by Δ then $at \in \mathcal{C}(t)$, $a \in A$, if and only if $a \in \text{Ult}(\Delta)$. If t is not standard then $at \in \mathcal{C}(t)$ for only one letter $a \in A$. It easily follows that if $t = T^k(s)$, $k \geq 0$, with s a standard rew directed by Δ then t has exactly $|\text{Ult}(\Delta)|$ left extensions giving bew in $\mathcal{C}(t)$ and that otherwise t has exactly one such extension.

Finally, with Δ as above and $\check{\Delta}$ a spinned version of Δ , μ_n directed by $x_1 \cdots x_n$ as above and $\check{\mu}_n$ directed by $\check{x}_1 \cdots \check{x}_n$, we recall the proposition.

Proposition 1.1. [10, prop. 3.15] *For any $u \in A^*$ (resp. A^ω) we have $f\check{\mu}_n(u) = \mu_n(u)f$ (resp. $f\check{\mu}_n(u) = \mu_n(u)$) with*

$$f = h_{n-1}^{\epsilon_n} h_{n-2}^{\epsilon_{n-1}} \cdots h_0^{\epsilon_1} \tag{2}$$

where ϵ_i is the spin (0 or 1) of \check{x}_i in $\check{\Delta}$.

Thus $\check{\mu}_n(u) = T^{|f|}(\mu_n(u))$. Putting $\sigma_n = |h_{n-1}|$, σ_n is the elementary shift associated to position n in Δ , and $|f| = \epsilon_n\sigma_n + \cdots + \epsilon_1\sigma_1$. Hereafter we will say that $|f|$ is the shift induced by the prefix $\check{w} = \check{\Delta}(1, n)$ of $\check{\Delta}$ and will denote it by $S(\check{w})$. The word f will be called the shifting factor of $\check{\mu}_n$.

2. Finite Spinned Words

Here we study the structure of finite spinned words in relation with the morphism they direct. This is applied for studying a numeration system and later in Sections 3 and 4 where infinite words are considered.

2.1. Block-equivalence

Let us introduce a convention for spins. If $v \in A^+$ then \bar{v} has all its spins 1; also $\bar{\bar{\epsilon}} = \epsilon$.

Definition 2.1. A (x -based) *block* is a word of the form xvx , $x \in A$, $v \in (A \setminus \{x\})^*$. A (x -based) *multiblock* is a word xvx , $x \in A$, $v \in A^*$.

Definition 2.2. A (x -based) *block-transformation* is the replacement in a spinned word of an occurrence of $xv\bar{x}$ (where xvx is a block) by $\bar{x}v\bar{x}$ or vice-versa. A *multiblock-transformation* (mbt) is the result of x -based block-transformations successively acting on a same x -based multiblock.

Example 2.1.

$$ab\bar{a}\bar{a}bc\bar{a} \rightarrow \bar{a}b\bar{a}\bar{a}bc\bar{a} \rightarrow \bar{a}\bar{b}\bar{a}abc\bar{a} \rightarrow \bar{a}\bar{b}\bar{a}\bar{a}bc\bar{a}.$$

Definition 2.3. Two spinned versions \check{w} , \hat{w} of $w \in A^*$ are *block-equivalent*, written $\check{w} \equiv \hat{w}$, if we can pass from one to the other by a (possibly empty) chain of block-transformations.

In particular $x\bar{x} \equiv x\bar{x}$. This allows permutations in sequences of the same letter with various spins. On the contrary $xv\bar{x} \not\equiv \bar{x}v\bar{x}$ for $v \neq \varepsilon$, showing the dissymmetry of Definition 2.3.

Lemma 2.1.

- 1) Let (if it exists) z be the leftmost letter in \check{u} having spin 0. Then $\mu_{\check{u}}(t)$ begins with z for any finite or right infinite word t
- 2) let m_1, m_2 be finite or infinite words and φ be an episturmian morphism. Then $\varphi(m_1) = \varphi(m_2)$ if and only if $m_1 = m_2$
- 3) $\mu_{\check{w}} = \mu_{\hat{w}}$ if and only if $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$ with \check{w}, \hat{w} two spinned versions of $w \in A^*$
- 4) if $\mu_{u\check{w}} = \mu_{u\hat{w}}$ then $\mu_{\check{w}} = \mu_{\hat{w}}$, with u any spinned word.

Proof. Properties 1) and 2) follow at once from basic properties of episturmian morphisms.

For 3), let t be a biinfinite non-periodic word. As $\mu_{\check{w}}(t) = T^{\mathcal{S}(\check{w})}(\mu_w(t))$ and similarly for \hat{w} and as $\mu_w(t)$ is not periodic (easily seen by induction on $|w|$), we have $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$ if and only if $\mu_{\check{w}} = \mu_{\hat{w}}$.

For 4), for any letter x , $\mu_{u\check{w}}(x) = \mu_u(\mu_{\check{w}}(x))$ and similarly for \hat{w} , thus by 2) $\mu_{\check{w}}(x) = \mu_{\hat{w}}(x)$ whence $\mu_{\check{w}} = \mu_{\hat{w}}$. □

Theorem 2.2. Let \check{w}, \hat{w} be two spinned versions of $w \in A^*$. Then \check{w} and \hat{w} direct the same morphism (or equivalently induce the same shift) if and only if they are block-equivalent.

Proof. Equivalence between “direct the same morphism” and “induce the same morphism” is Lemma 2.1, 3).

For the “if” part, $\check{w} \equiv \hat{w}$. We pass from \check{w} to \hat{w} by block-transformations. Consider w as a prefix of the infinite directive word $\Delta = x_1x_2 \cdots$, $x_i \in A$. When a block

$x_i x_{i+1} \cdots x_{j-1} \bar{x}_j$, $x_i = x_j$, $x_{i+1}, \dots, x_{j-1} \neq x_i$, is replaced by $\bar{x}_i \bar{x}_{i+1} \cdots \bar{x}_{j-1} x_j$ or vice-versa, the induced shift remains unchanged because $\sigma_i + \cdots + \sigma_{j-1} = \sigma_j$ (see [10, Prop. 3.16]). Thus $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$ whence by Lemma 2.1, 3), $\mu_{\check{w}} = \mu_{\hat{w}}$.

For the “only if” part, $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$. The result $\check{w} \equiv \hat{w}$ is true for $|w| = 0$. Suppose it is true for $|w| < n$ and assume $|w| = n$. We have $\mu_{\check{w}} = \mu_{\hat{w}}$. Let $w = yw'$, $y \in A$. If $\check{y} = \hat{y}$ then $\mu_{\check{y}\check{w}'} = \mu_{\hat{y}\hat{w}'}$, whence $\mu_{\check{w}'} = \mu_{\hat{w}'}$ by Lemma 2.1, whence $\mathcal{S}(\check{w}') = \mathcal{S}(\hat{w}')$ whence $\check{w}' \equiv \hat{w}'$ by inductive hypothesis, whence $\check{w} \equiv \hat{w}$.

Suppose now $\check{y} \neq \hat{y}$ and $\hat{y} = \bar{y}$ for instance and let z be the leftmost letter of w having spin 0 in \hat{w} . If this letter does not exist, then $\hat{w} = \bar{w}$ and the unique way of having $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$ is $\check{w} = \bar{w}$, a contradiction. Then by Lemma 2.1 $\mu_{\hat{w}}(t)$ begins with z for any t in A^+ , thus, as $\mu_{\check{w}}(t) = \mu_{\hat{w}}(t)$, $y = z$. Thus \hat{w} begins with some $\bar{y}\bar{v}y$, $v \in A^*$, but this is a y -based multiblock which can be transformed by some mbt into $y\check{v}\bar{y}$ for suitable spins in \check{v} , giving, say \check{w} . As both \check{w} and \hat{w} begin with y , $\check{w} \equiv \hat{w}$ whence $\hat{w} \equiv \check{w}$. □

Corollary 2.3. *For any $u, w \in A^*$ and \check{w}, \hat{w} spinned versions of w , $u\check{w}$ and $u\hat{w}$ induce the same shift if and only if $\check{w} \equiv \hat{w}$.*

Proof. “If” part is trivial by Theorem 2.2. Now if $\mathcal{S}(u\check{w}) = \mathcal{S}(u\hat{w})$ then $\mu_{u\check{w}} = \mu_{u\hat{w}}$ whence $\mu_{\check{w}} = \mu_{\hat{w}}$ whence $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$. Thus by the same theorem $\check{w} \equiv \hat{w}$. □

Remark 2.1. Theorem 2.2 could also be proved using Theorem 7.1 of [15]. Indeed this one gives a representation of the monoid of episturmian morphisms by generators and relations and the last of these relations simply expresses a block-transformation.

Now we make precise the structure of finite spinned words.

Definition 2.4. Let $\check{w} = \check{x}\check{v}_1\check{x}\check{v}_2\check{x}\cdots\check{x}\check{v}_n\check{x}$, $x \in A$, $v_i \in (A \setminus \{x\})^*$, be a x -based spinned multiblock. **Note that different occurrences of \check{x} may have different spins.** Then

1) \check{w} is a *semigood* multiblock (in short *smb*) if a) x and \bar{x} occur in it, b) each \check{v}_i is v_i or \bar{v}_i , c) for any non-empty v_i write $w = ev_i f$, then if $\check{v}_i = v_i$ (resp. $\check{v}_i = \bar{v}_i$) then x (resp. \bar{x}) occurs in \check{e} and \check{f} begins with a prefix $g\bar{x}$ (resp. $\bar{g}x$), $g \in A^*$

2) the *smb* \check{w} is a *good* multiblock (in short *gmb*) if its non-empty \check{v}_i all are v_i (*low form*) or all are \bar{v}_i (*high form*). When $n = 1$ (i.e., $\check{w} = xv_1\bar{x}$ or $\bar{x}\bar{v}_1x$) we simply say “good block”.

Example 2.2. Example 2.1 shows block-equivalent a -based *smb*, the first one in good low form, the last one in good high form.

If \check{w} is good of one form then, as easily seen, there exists a block-equivalent *gmb* of the other form. This can be generalized.

Proposition 2.4. *If \check{w} is a *smb* then there exists a *gmb* \hat{w} in low (or similarly high) form such that $\check{w} \equiv \hat{w}$.*

Proof. With \check{w} as in Definition 2.4, 1), if all non-empty \check{v}_i are v_i then \check{w} is already a gmb in low form. Otherwise let p , $1 \leq p \leq n$ be maximal such that $\check{v}_p = \bar{v}_p \neq \varepsilon$. Write $\check{w} = \check{e}\bar{v}_p\check{f}$. Then \check{f} begins with $\check{g}x\check{h}$, $g, h \in \{A\}^*$. But $g = xv_{p+1}x \cdots xv_q$ for some $q \geq p$ ($g = \varepsilon$ if $q = p$). Thus the v_i , $p+1 \leq i \leq q$ are empty, i.e., \check{f} begins with $\bar{x}^{q-p}x$.

Now we can write $\check{e}\bar{v}_p = \check{e}_1\bar{x}\check{v}_r x \cdots x\check{v}_{p-1}x\bar{v}_p$ where the rightmost occurrence of \bar{x} in \check{e} just precedes \check{v}_r . If some non-empty \check{v}_j , $r \leq j \leq p-1$, is v_j , then we are in contradiction with Definition 2.4, 1) c), as all \check{x} until \bar{v}_p are x . Thus

$$\check{w} = \check{e}_1\bar{x}\check{v}_r x \cdots x\bar{v}_{p-1}x\bar{v}_p\bar{x}^{q-p}x\check{h}.$$

By mbt (starting with $\bar{x}\bar{v}_r x$) we transform \check{w} into

$$\check{w} = \check{e}_1xv_r x \cdots xv_{p-1}xv_p\bar{x}^{q-p+1}\check{h}.$$

Then $\check{w} \equiv \check{w}$ is semigood and the number of non-empty $\check{v}_i = \bar{v}_i$ has decreased as at least \bar{v}_p is not empty. Continuing we get a good gmb (in low form). \square

Lemma 2.5. *Let $u = \check{e}\check{w}\check{f}$, $e, w, f \in A^*$ with \check{w} a x -based smb. If a good block $\bar{y}s\bar{y}$ or $\bar{y}s\bar{y}$, $y \in A$, $s \in (A \setminus \{y\})^*$ occurs in \check{w} or overlaps it in u then $y = x$. Moreover any $\hat{w} \equiv \check{w}$ is a x -based smb.*

Proof. With \check{w} the smb of Definition 2.4, 1), if the good block $\bar{y}s\bar{y}$ for instance, $y \neq x$, occurs in \check{w} then \bar{y} occurs in some non-empty \bar{v}_i . Thus the first letter with spin 0 on the right of this \bar{v}_i is an x , a contradiction because it is y . If $\bar{y}s\bar{y}$ overlaps \check{w} on the left of this one, then y occurs in some non-empty v_i . By Definition 2.4, 1) c) some x occurs in \check{w} on the left of y , a contradiction. If $\bar{y}s\bar{y}$ overlaps \check{w} on the right of this one then there is a \bar{y} within some \bar{v}_i in \check{w} and then an x before the end of \check{w} hence before the end of $\bar{y}s\bar{y}$, a contradiction. It follows that $x = y$, thus any block-transformation acting on \check{w} must be x -based.

Clearly \check{w} remains a x -based smb when modified by a block-transformation. Thus $\check{w} \equiv \hat{w}$ implies that \hat{w} is a x -based smb. \square

Lemma 2.6. *Let \check{w} be a x -based smb, factor of \check{u} . Then*

- 1) *if a smb \check{w}_1 occurs in \check{w} then it is x -based*
- 2) *if a smb \check{w}_1 overlaps \check{w} then it is x -based and both together form a x -based smb*
- 3) *if \check{w}_1 is a x -based smb adjacent to \check{w} then they together form a smb.*

Proof. For 1), the proof of Proposition 2.4 shows that if \check{w}_1 is a y -based smb then it contains a y -based good block. This one is a factor of \check{w} , hence by Lemma 2.5 $y = x$.

For 2) suppose by contradiction that \check{w}_1 is y -based, $y \neq x$. Write $\check{w}_1 = \check{w}'_1\check{x}\check{v}\check{y}$, $\check{w} = \check{x}\check{v}\check{y}\check{w}'$, $v \in A^*$. Suppose for instance $\check{x} = x$. Then as \check{w}_1 is a y -based smb, $\check{x}\check{v}\check{y}$ has a prefix $xg\bar{y}$, $g \in A^*$. As \check{w} is a x -based smb, \bar{x} must occur in the prefix xg of \check{w} , a contradiction. Thus $y = x$, i.e., \check{w}_1 is x -based.

Now the fact that $\check{w}_1\check{w}'$ is a x -based smb follows by verifying that Definition 2.4 holds for this word. The same reason also proves 3). □

Now let $\check{u} = \check{e}\check{w}\check{f}$ with \check{w} a smb. Then \check{w} is a maximal smb (in \check{u}) if $\check{u} = \check{e}'\check{w}'\check{f}'$, $|\check{e}'| \leq |\check{e}|$, $|\check{f}'| \leq |\check{f}|$, \check{w}' a smb, imply $\check{w}' = \check{w}$.

Theorem 2.7. *A spinned word \check{u} , has a unique factorization*

$$\check{u} = \check{e}_1\check{f}_1\check{e}_2\check{f}_2 \cdots \check{e}_n\check{f}_n\check{e}_{n+1}$$

where the \check{f}_i are all the maximal smb of \check{u} . Moreover any $\hat{u} \equiv \check{u}$, has form

$$\hat{u} = \check{e}_1\hat{f}_1\check{e}_2\hat{f}_2 \cdots \check{e}_n\hat{f}_n\check{e}_{n+1}$$

with $\hat{f}_i \equiv \check{f}_i$ for all i .

Proof. By Lemma 2.6 two maximal smb are disjoint whence the factorization as claimed. Moreover as a block, like $\bar{y}\bar{s}\bar{y}$ or $y\bar{s}\bar{y}$, is a smb, it is contained in some maximal \check{f}_i . Thus all block-transformations act on the \check{f}_i letting the \check{e}_i unchanged, whence the result. □

It is not difficult to design a finite automaton recognizing smb (resp. gmb). Thus

Theorem 2.8. *Given a finite alphabet the language of semigood (resp. good) multi-blocks on it is rational.*

2.2. A numeration system

Let $\Delta = x_1x_2 \cdots$ direct the standard rew s , with u_n the palindromic prefixes of s and σ_n , $n \geq 1$, as in the Preliminaries. Put $d_n = \Delta(1, n)$. Then as already said in [10] all shifts $\mathcal{S}(\check{d}_n)$ from 0 to $|u_{n+1}|$ can be obtained by suitable spins in \check{d}_n . Indeed this is easily shown by induction on n , using $\sigma_n \leq |u_n| + 1$, which itself follows from $|u_{n+1}| \leq 2|u_n| + 1$ and Equation (1).

Thus $(\sigma_n)_{n \in \mathbb{N}_+}$ can be considered as a numeration system for positive integers, with digits 0 and 1. However if $x_i = x_{i+1}$ then $\sigma_i = \sigma_{i+1}$, thus it is natural to ignore permutations among adjacent \check{x} in \check{d}_n . Then we represent Δ by $\Delta = z_1^{e_1}z_2^{e_2} \cdots$, $z_i \in A$, $z_i \neq z_{i+1}$, $e_i > 0$, and we can associate to Δ the numeration system $(\tau_n)_{n \in \mathbb{N}_+}$ where the τ_n are the different σ_i in increasing value order. Now any integer $K \geq 0$ can be represented as $\sum_{i=1}^n k_i\tau_i$, $0 \leq k_i \leq e_i$ for all i . We get a numeration system which can be called a generalized Ostrowski system as the Ostrowski systems [5] correspond to the Sturmian case.

The relation $\tau_{i+1} = e_i\tau_i + \tau_{i-1}$ valuable in the Sturmian case is rather simple because z_1, z_2, \dots are alternatively the two letters of the alphabet. In the general episturmian case more complex similar relations follow from the relations between the σ_i given in [10, Prop. 3.16].

We are interested in the number of representations of an integer $K \geq 0$ in the system above. With $d_n = \Delta(1, n)$, n large enough and \check{d}_n such that $\mathcal{S}(\check{d}_n) = K$, it

suffices to count the number of $\hat{d}_n \equiv \check{d}_n$ up to permutations of adjacent z and \bar{z} , $z \in A$. In view of Theorem 2.7 this number is the product of analogous numbers for the maximal smb of \check{d}_n .

Thus it suffices to make the count for an arbitrary smb. A lemma will be useful.

Lemma 2.9. *With $w = xv_1xv_2 \cdots xv_nx$, $x \in A$, $v_i \in (A \setminus \{x\})^*$, let \check{w} , \hat{w} be two x -based block-equivalent smb. Write, for any non-empty v_h , $w = ev_hf$. Then*

- 1) *the numbers of \bar{x} (resp. x) in \check{w} and \hat{w} are equal*
- 2) *if $\check{v}_h = \hat{v}_h$ then $\check{e} \equiv \hat{e}$ and $\check{f} \equiv \hat{f}$*
- 3) *if $\check{v}_h = v_h$ and $\hat{v}_h = \bar{v}_h$ then \hat{e} (resp. \check{f}) contains one more \bar{x} than \check{e} (resp. \hat{f}).*

Proof. 1) is clear because each block-transformation from \check{w} to \hat{w} lets the number of \bar{x} invariant. For 2) and 3) a block-transformation from \check{w} to \hat{w} has form $xv_j\bar{x} \rightarrow \bar{x}\bar{v}_jx$ or the inverse, thus if $\hat{v}_h \neq \check{v}_h$ the number of such transformations such that $j = h$ is odd whence the result for 3). Now if $\hat{v}_h = \check{v}_h$ the number of block-transformations such that $j = h$ is even. We have to show that in this case $\check{e} \equiv \hat{e}$. Let $e = e'x$ and $f = xf'$. Without loss of generality suppose that $\check{e} = \check{e}'x$, $\check{v}_h = v_h$, $\check{f} = \bar{x}\check{f}'$ and that the first block-transformation is $xv_h\bar{x} \rightarrow \bar{x}\bar{v}_hx$ giving $\check{w} = \check{e}\bar{v}_h\check{f} = \check{e}'\bar{x}\bar{v}_hx\check{f}'$. Thus $\mathcal{S}(\check{e}) = \mathcal{S}(\check{e}') + \sigma$ where σ is the elementary shift corresponding to the x before v_h . Thus the shift of \check{e} has increased by σ . The following block-transformations may act on \check{e} and \check{f} replacing them by block-equivalent words, \check{e}^+ , \check{f}^+ , say. They are followed by a block-transformation $\bar{x}\bar{v}_hx \rightarrow xv_h\bar{x}$ which decreases the shift of \check{e}^+ by σ , giving without loss of generality $\hat{e} \equiv \check{e}$ by Theorem 2.2. \square

Now consider a y -based gmb in low form $\check{B} = \check{y}_1^{a_1}w_1\check{y}_2^{a_2}w_2 \cdots \check{y}_n^{a_n}$ with, for all i , $y_i = y \in A$, $w_i \in (A \setminus \{y\})^+$, $a_i > 0$ and denote by $q_1 < q_2 \cdots$ the indices j such that $\check{y}_j^{a_j}$ contains \bar{y} , and put $p_1 = q_1$, $p_i = q_i - q_{i-1}$. Also let $c_i = 1$ if $\check{y}_{q_i}^{a_{q_i}}$ contains y , $c_i = 0$ otherwise. As \check{B} is in low form $\check{y}_n^{a_n}$ must contain \bar{y} , thus $n = q_k$ for some k . We will show that the number of representations of $\mathcal{S}(\check{B})$ in the above numeration system, i.e., the number of $\hat{B} \equiv \check{B}$ (up to permutations within the $y_i^{a_i}$) only depends on $(p_1, c_1, p_2, c_2, \dots, p_k, c_k)$ and can be represented by $\mathcal{R}(p_1, c_1, p_2, c_2, \dots, p_k, c_k)$. We have the theorem

Theorem 2.10. *Let \check{B} be a good multiblock in low form, with sequence $(p_1, c_1, p_2, c_2, \dots, p_k, c_k)$ as defined above. Then the number of representations of $\mathcal{S}(\check{B})$ in the numeration system described above satisfies*

$$\mathcal{R}(p_1, 0, p_2, c_2, \dots, p_k, c_k) = \mathcal{R}(p_2, c_2, \dots, p_k, c_k) + (p_1 - 1)\mathcal{R}(p_2 + 1, c_2, \dots, p_k, c_k)$$

$$\mathcal{R}(p_1, 1, p_2, c_2, \dots, p_k, c_k) = p_1\mathcal{R}(p_2 + 1, c_2, \dots, p_k, c_k).$$

Proof. With $\check{B} = \check{y}_1^{a_1}w_1\check{y}_2^{a_2}w_2 \cdots \check{y}_n^{a_n}$ a gmb in low form as previously write $\check{B} = f w_{p_1-1} \check{y}_{p_1}^{a_{p_1}} \check{g}$. If $\hat{B} \equiv \check{B}$ let $\hat{B} = \hat{f} \hat{w}_{p_1-1} \hat{y}_{p_1}^{a_{p_1}} \hat{g}$.

Suppose firstly $c_1 = 0$. If $\hat{w}_{p_1-1} = w_{p_1-1}$ then $\hat{f} = f$ by Lemma 2.9, 2). In this case we also have $\hat{y}_{p_1}^{a_{p_1}} \hat{g} \equiv \bar{y}_{p_1}^{a_{p_1}} \check{g}$. As $\hat{y}_{p_1}^{a_{p_1}}$ may contain at most one \bar{y} more than $\bar{y}_{p_1}^{a_{p_1}}$ we have exactly $\hat{y}_{p_1}^{a_{p_1}} = \bar{y}_{p_1}^{a_{p_1}}$ and $\check{g} \equiv \hat{g}$, giving $\mathcal{R}(p_2, c_2, \dots, p_k, c_k)$ possibilities for \hat{g} . If $\hat{w}_{p_1-1} = \bar{w}_{p_1-1}$ then there is exactly one \bar{y} in \hat{f} giving $p_1 - 1$ possibilities and now the possibilities for $\hat{y}_{p_1}^{a_{p_1}} \hat{g}$ are the same as for $\bar{y}_{p_1}^{a_{p_1}-1} \hat{y} \hat{g}$ whence for $y \hat{g}$, giving $\mathcal{R}(p_2 + 1, c_2, \dots, p_k, c_k)$ possibilities. The first formula follows.

Suppose now $c_1 = 1$, then $\hat{y}_{p_1}^{a_{p_1}}$ contains y . Thus up to permutations of adjacent spinned y we can write $\check{B} = f w_{p_1-1} \bar{y} \hat{y}_{p_1}^{a_{p_1}-2} y \check{g}$ and $\hat{B} = \hat{f} \hat{w}_{p_1-1} \hat{y}_{p_1}^{a_{p_1}-2} \hat{y} \hat{g}$ (the two \hat{y} may have different spins). Using Lemma 2.9 we get $f w_{p_1-1} \bar{y} \equiv \hat{f} \hat{w}_{p_1-1} \hat{y}$ giving p_1 possibilities. As $\check{B} \equiv \hat{B}$ it follows $y \check{g} \equiv \hat{y} \hat{g}$, giving $\mathcal{R}(p_2 + 1, c_2, \dots, p_k, c_k)$ possibilities for $\hat{y} \hat{g}$ whence the second formula. \square

Formulas of Theorem 2.10 can be collected into one:

$$\mathcal{R}(p_1, c_1, p_2, c_2, \dots, p_k, c_k) = (1 - c_1) \mathcal{R}(p_2, c_2, \dots, p_k, c_k) + (p_1 + c_1 - 1) \times \mathcal{R}(p_2 + 1, c_2, \dots, p_k, c_k)$$

whence, as $\mathcal{R}(p, 0) = \mathcal{R}(p, 1) = p$ for any $p > 0$, the following corollary.

Corollary 2.11. *The number of representations of $\mathcal{S}(\check{B})$ in the numeration system is given by*

$$(1, 0) \begin{pmatrix} \mathcal{R}(p_1, c_1, p_2, c_2, \dots, p_k, c_k) \\ \mathcal{R}(p_1 + 1, c_1, p_2, c_2, \dots, p_k, c_k) \end{pmatrix} = (1, 0) M_1 M_2 \dots M_k \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

where $M_i = \begin{pmatrix} 1 - c_i & p_i + c_i - 1 \\ 1 - c_i & p_i + c_i \end{pmatrix}$.

Example 2.3. We take the example of [3] to which we refer for details. Sequence (2,2,3,2,2,2,3) there corresponds to the prefix $\Delta(1, 16) = a^2 b^2 a^3 b^2 a^2 b^2 a^3$ of a directive word on $\{a, b\}$. Number 660 has the Zeckendorf representation $2000200 (= 2\tau_7 + 2\tau_3 = 2 \times 323 + 2 \times 7)$ in the corresponding Ostrowski system (the Zeckendorf representation $\sum k_i \tau_i$, $0 \leq k_i \leq e_i$, of a number is the one where, for $i \geq 2$, $k_i = e_i$ implies $k_{i-1} = 0$). This corresponds to the spinned word $\check{\Delta}(1, 16) = aabba\bar{a}\bar{a}bbaabba\bar{a}\bar{a}$. In this particular example $\check{\Delta}(1, 16)$ is its own unique maximal smb (it is even a gmb: observe that in general the \check{d}_n where all maximal smb are in good low form correspond to Zeckendorf representations in reverse order of digits). Thus if we write $\Delta(1, 16)$ as $y_1^2 w_1 y_2^3 w_2 y_3^2 w_3 y_4^3$, where $y_1 = y_2 = y_3 = y_4 = a$, we have $q_1 = 2$, $q_2 = 4$ whence $p_1 = p_2 = 2$ and $c_1 = c_2 = 1$. We get $M_1 = M_2 = \begin{pmatrix} 0 & 2 \\ 0 & 3 \end{pmatrix}$ whence $\mathcal{R}(2, 1, 2, 1) = 6$ as in [3].

Remark 2.2. It is possible to make a theory of generalized Ostrowski systems (and in particular to show existence and unicity of generalized Zeckendorf representations) without any explicit reference to episturmian words and spinned words.

3. Infinite Spinned Words

In this section we examine in which conditions two spinned versions of a right infinite word direct a same bew.

3.1. Non-periodic episturmian words

We will consider directive (right infinite) words $\Delta \in A^\omega$ such that $|Ult(\Delta)| > 1$. The case $|Ult(\Delta)| = 1$ which corresponds to periodic episturmian words will be shortly examined after in Subsection 3.2.

The duality between spins 0 and 1 is made precise by the following lemma. If \check{w} is a finite or infinite spinned word its *opposite* is \hat{w} obtained by changing all spins in \check{w} .

Lemma 3.1. *Let \check{w} be a finite (resp. $\check{\Delta}$ be an infinite) spinned word. Then if \hat{w} (resp. $\hat{\Delta}$) is its opposite we have*

- 1) for all $u \in A^*$, $\mu_{\hat{w}}(\tilde{u}) = \widetilde{\mu_{\check{w}}(u)}$
- 2) if $\check{\Delta}$ directs a bew b then $\hat{\Delta}$ directs \tilde{b} .

Proof. When u is a letter 1) is easy by induction on $|w|$ and the formula for arbitrary u follows. For 2) observe that if $X.Y$ is a bew and $a \in A$ then $\bar{\psi}_a(\tilde{Y}.\tilde{X}) = \psi_a(\widetilde{X.Y})$. Thus if $(b^{(i)})_{i \in \mathbb{N}}$ is the sequence of the bew associated with b (see Preliminaries) then $b^{(i-1)} = \mu_{\check{x}_i}(b^{(i)})$ implies $\tilde{b}^{(i-1)} = \mu_{\hat{x}_i}(\tilde{b}^{(i)})$, thus the sequence of the $\tilde{b}^{(i)}$ corresponds to \tilde{b} hence this one is directed by $\hat{\Delta}$. \square

A spinned version $\check{\Delta}$ of Δ will be called *wavy* if it contains infinitely many spins 0 and spins 1. When $\check{\Delta}$ is wavy it directs exactly one bew ([10, Prop.3.2]). When the spins are ultimately 0 (resp. 1) $\check{\Delta}$ directs $|Ult(\Delta)|$ bew which are shifts of bew $m.s$ (resp. $\tilde{s}.\tilde{m}$) with s the standard rew directed by Δ .

Now we give a continuity theorem extending Theorem 3.20 of [10].

Theorem 3.2. *Let $\check{\Delta}_i, i \in \mathbb{N}$ be wavy spinned versions of $\Delta \in A^\omega, |Ult(\Delta)| > 1$, and suppose the $\check{\Delta}_i$ have a limit $\check{\Delta}$ say. Let b_i be the bew directed by the $\check{\Delta}_i$. Then the b_i have a limit b which is directed by $\check{\Delta}$.*

Proof. Let $\check{v}_n = \check{\Delta}(1, n)$. Then \check{v}_n is a prefix of all $\check{\Delta}_i$ for i large enough (depending of n). Since $|Ult(\Delta)| > 1$ there exist infinitely many values of n , say $n_1 < n_2 < \dots$ such that for some $x \in A, x_{n_j+1} = x$ and the x_{n_j+1} have same spin in $\check{\Delta}$. Suppose first this one is 0. Then $\check{x}_{n_j+1} = x$ in all $\check{\Delta}_i$ for i large enough (depending of n_j). Denote $\mu_{\check{v}_{n_j}}$ by $\check{\mu}_{n_j}$. Then $\check{\mu}_{n_j}(x)$ is a prefix of the right part of b_i for i large enough (depending of n_j). As $|Ult(\Delta)| > 1$ the $|\check{\mu}_{n_j}(x)|$ increase with j , hence the b_i have a limit, b say. Now if r is a rew directed by $\check{\Delta}$ then by the Preliminaries $r^{(n_j)}$ begins with x , whence $\check{\mu}_{n_j}(x)$ is a prefix of r . Thus the right part of b , hence b itself, is directed by $\check{\Delta}$.

If on the contrary the \check{x}_{n_j+1} have spin 1 we consider the opposites of $\check{\Delta}$ and the $\check{\Delta}_i$ and we apply what precedes whence by Lemma 3.1 the result. \square

Lemma 3.3. *Let $\check{\Delta}$, wavy, and $\hat{\Delta}$ be spinned versions of $\Delta \in A^\omega$, $|Ult(\Delta)| > 1$, directing a same bew, with $\check{\Delta}$ beginning with \bar{x} (resp. x) and $\hat{\Delta}$ beginning with x (resp. \bar{x}), $x \in A$. Then the leftmost letter of Δ having spin 0 (resp. spin 1) in $\check{\Delta}$ is an occurrence of x .*

Proof. Suppose for instance $\check{\Delta}$ begins with \bar{x} . Let r be the right part of the bew directed by both $\check{\Delta}$, $\hat{\Delta}$. As $\hat{\Delta}$ begins with x , r also begins with x . By Lemma 2.1, 1), the first letter with spin 0 in $\check{\Delta}$ is x (this letter exists as $\check{\Delta}$ is wavy), thus $\check{\Delta}$ begins as claimed. If on the contrary $\check{\Delta}$ begins with x we obtain the result by considering the opposites of $\check{\Delta}$, $\hat{\Delta}$. \square

Now we extend Definition 2.3 to right infinite words.

Definition 3.1. Let $\check{\Delta}$, $\hat{\Delta}$ be spinned versions of Δ , then

1) $\check{\Delta}$ leads to $\hat{\Delta}$ (written $\check{\Delta} \rightsquigarrow \hat{\Delta}$) if there exist infinitely many prefixes \check{f}_i of $\check{\Delta}$, \hat{g}_i of $\hat{\Delta}$, $|g_i| \leq |f_i|$, with the g_i of strictly increasing lengths, such that, with $f_i = g_i c_i$, we have $\check{f}_i \equiv \hat{g}_i \check{c}_i$ for suitable spins in the \check{c}_i

2) $\check{\Delta}$ and $\hat{\Delta}$ are block-equivalent (written $\check{\Delta} \equiv \hat{\Delta}$) if $\check{\Delta} \rightsquigarrow \hat{\Delta}$ and $\hat{\Delta} \rightsquigarrow \check{\Delta}$.

Example 3.1. With alphabet $\{a, b\}$ let $\check{\Delta} = \bar{a}(\bar{b}a)^\omega$. Easily $\bar{a}(\bar{b}a)^n \equiv (ab)^n \bar{a}$. As $(ab)^n$ is a prefix of Δ we have $\check{\Delta} \rightsquigarrow \Delta$.

It is easily verified that \equiv is an equivalence.

In the “wavy case” we have a rather simple result.

Theorem 3.4. *Let $\check{\Delta}$ and $\hat{\Delta}$ be wavy spinned versions of $\Delta \in A^\omega$, $|Ult(\Delta)| > 1$. Then $\check{\Delta}$ and $\hat{\Delta}$ direct the same (unique) bew if and only if $\check{\Delta} \equiv \hat{\Delta}$.*

Proof. For the “if” part let $\check{\Delta} = \check{f}_i \check{\Delta}_i$, $\hat{\Delta} = \hat{g}_i \hat{\Delta}'_i$, with $\check{f}_i \equiv \hat{g}_i \check{c}_i$ as in the first part of Definition 3.1. Then the $\hat{g}_i \check{c}_i \check{\Delta}_i$ direct the same bew b as $\check{\Delta}$, thus as these infinite words have limit $\hat{\Delta}$, by Theorem 3.2, $\hat{\Delta}$ also directs b .

For the “only if” part, $\check{\Delta}$ and $\hat{\Delta}$ are wavy and direct the same bew, b . If $\Delta = w\Delta'$, $w \in A^+$, $\Delta' \in A^\omega$ and $\check{w} \equiv \hat{w}$ then $\check{\Delta}'$, $\hat{\Delta}'$ are wavy and direct a same bew (which is $b^{(|w|)}$, see Preliminaries). If we can continue this way indefinitely then clearly $\check{\Delta} \equiv \hat{\Delta}$. Otherwise $\Delta = w\Delta'$ with $|w|$ maximal such that $\check{w} \equiv \hat{w}$. Then $\check{\Delta}'$, $\hat{\Delta}'$ direct the same bew and have no “common modulo \equiv ” non-empty prefix. Without loss of generality we can suppose that this is the case for $\check{\Delta}$, $\hat{\Delta}$. Then for instance $\check{\Delta}$ begins with $x \in A$ and $\hat{\Delta}$ with \bar{x} . By Lemma 3.3 $\check{\Delta}$ begins with $\check{f}_0 = xv_1xv_2x \cdots xv_p\bar{x}$, $p \geq 1$, $v_i \in (A \setminus x)^*$. Write $\Delta = f_0\Delta_0$, and replace \check{f}_0 in $\check{\Delta}$ by the block-equivalent $\hat{f}_0 = \bar{x}\bar{v}_1\bar{x}\bar{v}_2\bar{x} \cdots \bar{x}\bar{v}_p\bar{x}$, obtaining $\check{\Delta} = \hat{f}_0\check{\Delta}_0$.

Clearly $\check{\Delta}$ also directs b . Let y be the leftmost letter in Δ having different spins in $\check{\Delta}$ and $\hat{\Delta}$. This spin disagreement must occur within f_0 because otherwise we would have $\check{f}_0 \equiv \hat{f}_0 = \hat{f}_0$, a contradiction.

Thus we can write $\check{f}_0 = \check{g}_0\check{c}_0$ with $\check{g}_0 = \hat{g}_0$ and the first letter y of c_0 having different spins in \check{c}_0, \hat{c}_0 . If $y \neq x$, it occurs in some $v_r, 1 \leq r \leq p$ and is \bar{y} in \check{f}_0 . Thus by Lemma 3.3 the first letter with spin 0 in $\check{c}_0\check{\Delta}_0$ must be y , and this is impossible as x immediately follows \bar{v}_r in \check{c}_0 . Thus $y = x$, say the x following v_r and $g_0 = xv_1 \cdots xv_r$ for some $r, 1 \leq r \leq p$. Thus $\check{c}_0\check{\Delta}_0$ begins with x and $\hat{c}_0\hat{\Delta}_0$ with \bar{x} . By Lemma 3.3 $\check{c}_0\check{\Delta}_0$ begins with $xv_{r+1} \cdots xv_q\bar{x}, q > p$. Put $xv_{r+1} \cdots v_qx = d_0$ and $f_1 = f_0d_0 = xv_1 \cdots xv_qx$ and $c_0\Delta_0 = c_0d_0\Delta_1$ whence $\Delta_0 = d_0\Delta_1$. Replacing $\check{c}_0\check{d}_0$ in $\check{c}_0\check{\Delta}_0$ by the equivalent $\bar{x}\bar{v}_{r+1}x \cdots \bar{x}\bar{v}_qx$ we get, say, $\check{c}_0^+\check{d}_0^+\check{\Delta}_1$. As previously $\check{c}_0^+\check{d}_0^+\check{\Delta}_1$ and $\hat{c}_0\hat{d}_0\hat{\Delta}_1$ have a maximal common prefix of the form $\bar{x}\bar{v}_{r+1}x \cdots \bar{x}\bar{v}_m, m \leq q$. Let $e_0 = xv_{r+1} \cdots v_m$ and $g_1 = g_0e_1 = xv_1 \cdots v_m$ and $c_1 = xv_{m+1} \cdots v_qx$ whence $f_1 = g_1c_1$. We have $\check{c}_0^+\check{d}_0^+ = \check{e}_0^+\check{c}_1^+ = \hat{e}_0^+\hat{c}_1^+$ whence

$$\check{f}_1 = \check{f}_0\check{d}_0 \equiv \check{f}_0\check{d}_0^+ \equiv \hat{g}_0\check{c}_0\check{d}_0^+ \equiv \hat{g}_0\check{c}_0^+\check{d}_0^+ = \hat{g}_0\hat{e}_0\hat{c}_1^+ = \hat{g}_1\hat{c}_1^+.$$

Thus $\check{\Delta}$ has a prefix \check{f}_1 and $\hat{\Delta}$ has a prefix \hat{g}_1 such that $\check{f}_1 \equiv \hat{g}_1\hat{c}_1^+$ for suitable spins in c_1 . Continuing this way, we get prefixes \check{f}_i of $\check{\Delta}$ and \hat{g}_i of $\hat{\Delta}$ such that $\check{f}_i \equiv \hat{g}_i\hat{c}_i^+$ for suitable spins in the c_i . Clearly the f_i and g_i have increasing lengths. Thus $\check{\Delta} \rightsquigarrow \hat{\Delta}$ by Definition 3.1. Similarly $\hat{\Delta} \rightsquigarrow \check{\Delta}$ whence $\check{\Delta} \equiv \hat{\Delta}$. □

Corollary 3.5. *If $\check{\Delta} \equiv \hat{\Delta}$ where $\check{\Delta}, \hat{\Delta}$, wavy and without common modulo \equiv non-empty prefix, begin with x, \bar{x} respectively, then $\check{\Delta} = xv_1\check{x}\bar{v}_2\check{x} \cdots$ with infinitely many spins 1 among the \check{x} and $\hat{\Delta} = \bar{x}\bar{v}_1\hat{x}\bar{v}_2\hat{x} \cdots$ with infinitely many spins 0 among the \hat{x} .*

Proof. Immediate from the proof of Theorem 3.4. □

Remark 3.1. In this case $\check{\Delta}$ and $\hat{\Delta}$ can be considered as “good infinite multi-blocks” with a natural extension of definitions given for finite words.

Now let us consider all cases not covered by Theorem 3.4.

Proposition 3.6. *If, with Δ as previously, $\check{\Delta}$ wavy and $\hat{\Delta}$ with all spins ultimately 0 (resp. 1) direct a same bew, then $\check{\Delta} \rightsquigarrow \hat{\Delta}$. Moreover there is a factorization $\Delta = w\Delta'$ such that $\check{w} \equiv \hat{w}, \hat{\Delta}' = \Delta'$ for instance (case $\hat{\Delta}' = \bar{\Delta}'$ is obtained by taking the opposites) and $\check{\Delta}' = \bar{x}\bar{v}_1x\bar{v}_2x \cdots, x \in A, v_i \in (A \setminus \{x\})^*$.*

Proof. We prove $\check{\Delta} \rightsquigarrow \hat{\Delta}$ exactly as in proof of Theorem 3.4. Now supposing spins of $\hat{\Delta}$ are ultimately 0 let $\Delta = w\Delta'$ with $|w|$ maximal such that $\check{w} \equiv \hat{w}$. Then $\check{\Delta}'$ and $\hat{\Delta}'$ direct a same bew and begin with different spins. If $\check{\Delta}'$ begins with x and $\hat{\Delta}'$ with \bar{x} then $\check{\Delta}'$ begins with $\check{f} = xv_1xv_2 \cdots xv_q\bar{x}, q \geq 1, v_i \in (A \setminus \{x\})^*$, which can be replaced by $\bar{x}\bar{v}_1x\bar{v}_2x \cdots \bar{x}\bar{v}_qx = \check{f}$. As in the proof of Theorem 3.4 a new disagreement must occur on some x in \check{f} whence a \bar{x} in $\hat{\Delta}'$. Continuing we get infinitely many occurrences of \bar{x} in $\hat{\Delta}'$ and this impossible as all spins of $\hat{\Delta}$

are ultimately 0. Thus $\hat{\Delta}' = \Delta'$ and easily, using Lemma 3.3, $\check{\Delta}' = \bar{x}\bar{v}_1x\bar{v}_2x \cdots$ as claimed. \square

Proposition 3.7. *With $|\text{Ult}(\Delta)| > 1$, if $\check{\Delta} = \check{w}\Delta'$ and $\hat{\Delta} = \hat{w}\Delta'$ (resp. $\check{w}\bar{\Delta}'$ and $\hat{w}\bar{\Delta}'$) then $\check{\Delta}, \hat{\Delta}$ direct a same bew if and only if $\check{w} \equiv \hat{w}$.*

Proof. “If” part follows from the fact that $\mu_{\check{w}} = \mu_{\hat{w}}$. For the “only if” part let r be the right part of a bew directed by both $\check{\Delta}$ and $\hat{\Delta}$. Then by the Preliminaries we have $r = \mu_{\check{w}}(s) = \mu_{\hat{w}}(s)$, i.e., $\text{T}^{\mathcal{S}(\check{w})} \mu_w(s) = \text{T}^{\mathcal{S}(\hat{w})} \mu_w(s)$, where s is the (standard) rew directed by Δ' . Thus as $\mu_w(s)$ is not periodic, $\mathcal{S}(\check{w}) = \mathcal{S}(\hat{w})$, i.e., $\check{w} \equiv \hat{w}$. \square

The last case shows that a well-known property of Sturmian words is exceptional. A lemma will be useful.

Lemma 3.8. *Let s be a standard non-periodic rew. Then*

- 1) $\tilde{s}.s$ does not belong to $\mathcal{C}(s)$.
- 2) for any letter y , $\tilde{s}y.s$ does not belong to $\mathcal{C}(s)$.

Proof. 1) Let s , a standard rew, have directive word Δ , with $\Delta = x_1x_2 \cdots, u_n, h_n$ as in the Preliminaries. If x_1 does not occur in $x_2x_3 \cdots$ then x_1x_1 does not occur in s . As s begins with x_1 , x_1x_1 occurs in $\tilde{s}.s$ whence $\tilde{s}.s \notin \mathcal{C}(s)$.

Otherwise let x_n be the second occurrence of x_1 in Δ . Then $u_{n+1} = (u_nx_1)^{(+)} = u_nu_n$ whence by Equation (1) $u_n = h_{n-1}$. Let p be maximal such that h_{n-1}^p is a prefix of s . Then, by [10, Prop. 2.10], h_{n-1}^{p+2} does not occur in s . But $h_{n-1}^{2p} = u_n^{2p}$ clearly occurs in $\tilde{s}.s$ whence if $\tilde{s}.s \in \mathcal{C}(s)$, $2p < p + 2$, whence $p < 2$. But clearly $h_{2n-1} = u_{n+1}$ is a prefix of s , a contradiction.

2) With the same notations let s begin with x_1^p , $p \geq 1$. The maximal power of x_1 in s is $p + 1$. If $y = x_1$ then x_1^{2p+1} occurs in $\tilde{s}.s$. As $2p + 1 > p + 1$ this one does not belong to $\mathcal{C}(s)$. If $y \neq x_1$, write $s = \psi_{x_1}(t)$, t the standard rew directed by $x_2x_3 \cdots$. We easily get $\tilde{s}y.s = \psi_{x_1}(\tilde{t}y.t)$. If $\tilde{s}y.s \in \mathcal{C}(s)$ then $\tilde{t}y.t \in \mathcal{C}(t)$. If $y = x_2$ this impossible as previously. Otherwise we continue until $y = x_i$ for some i whence impossibility. If y never occurs in Δ then $\tilde{s}y.s \notin \mathcal{C}(s)$. \square

Proposition 3.9. *With $|\text{Ult}(\Delta)| > 1$, Δ and $\bar{\Delta}$ cannot direct two bew b_1 and b_2 which differ only by a shift, unless $\text{Ult}(\Delta)$ is a 2-letter alphabet, $\{a, b\}$ say. In this case $\Delta = w\Delta'$ with $\Delta' \in \{a, b\}^\omega$, $w \in A^*$, and Δ' directs $\bar{t}ab.t$ and $\bar{t}ba.t$, while $\bar{\Delta}'$ directs $\bar{t}.bat$ and $\bar{t}.abt$, with t a standard Sturmian word on $\{a, b\}$.*

Proof. Let $\Delta = x_1x_2 \cdots$ and morphisms $\mu_n, \bar{\mu}_n$ be directed by $x_1x_2 \cdots x_n, \bar{x}_1\bar{x}_2 \cdots \bar{x}_n$ and s be the standard rew directed by Δ . Then b_1 has right part s and b_2 has left part \tilde{s} . As they differ only by a shift we easily see that for some $w \in A^+$, either $b_1 = \tilde{s}w.s$ and $b_2 = \tilde{s}.ws$ or, with $s = ws'$, $b_1 = \tilde{s}'.ws' = \tilde{s}'.s$ and $b_2 = \tilde{s}'w.s' = \tilde{s}'.s'$.

In the second case, w is a prefix of s and a suffix of \tilde{s} , hence is a palindromic prefix of s , u_{n+1} , say. Consider the $b_1^{(n)}$ and $b_2^{(n)}$ in the sequences of the bew associated to b_1 and b_2 . We have $b_1 = \mu_n(b_1^{(n)})$ and $b_2 = \bar{\mu}_n(b_2^{(n)})$. Let $b_1^{(n)} = \tilde{m}.s^{(n)}$, with $s^{(n)}$ the standard rew directed by $x_{n+1}x_{n+2}\dots$. Then $b_1 = \mu_n(\tilde{m}.s^{(n)}) = \mu_n(\tilde{m}).s$. Thus we have $\mu_n(\tilde{m}) = \tilde{s}'$. Thus, recalling that the shifting factor of $\bar{\mu}_n$ is $\overline{u_{n+1}}$ (see [10, end of Prop. 3.10]) we get $\tilde{s} = \mu_n(\tilde{m})w = \mu_n(\tilde{m})u_{n+1} = \bar{\mu}_n(\tilde{m}) = \mu_n(m)$ whence $s = \mu_n(m)$ whence $m = s^{(n)}$. Thus $b_1^{(n)} = \tilde{s}^{(n)}.s^{(n)}$. By Lemma 3.8 this is impossible.

Consequently we are in the first case: $b_1 = \tilde{s}w.s$. We have $b_1 = \psi_{x_1}(b_1^{(1)})$. Thus, writing $b_1^{(1)} = \tilde{m}.s^{(1)}$, we have $\tilde{s}w. = \psi_{x_1}(\tilde{m})$ whence $\tilde{s}wx_1. = \bar{\psi}_{x_1}(\tilde{m})$ whence $x_1\tilde{w}s = \psi_{x_1}(m)$, whence, as $s = \psi_{x_1}(s^{(1)})$, $m = v_1s^{(1)}$ with $\psi_{x_1}(v_1) = x_1\tilde{w}$. Thus $b_1^{(1)} = \tilde{s}^{(1)}\tilde{v}_1.s^{(1)}$ and more generally $b_1^{(h)} = \tilde{s}^{(h)}\tilde{v}_h.s^{(h)}$ with $x_h\tilde{v}_{h-1} = \psi_{x_h}(v_h)$.

Among all $b_1^{(h)}$ consider one with $|v_h|$ minimal and without loss of generality suppose it is $b_1^{(0)}$, i.e., b_1 itself. Then $|w| \leq |v_1|$. Thus, as $x_1w = \psi_{x_1}(v_1)$, at most one letter different from x_1 occurs in v_1 . Thus $v_1 = x_1^p$ or $v_1 = x_1^p y x_1^q$, $y \in A \setminus \{x_1\}$.

If $v_1 = x_1^p$ then $w = x_1^{p-1}$, hence $p \geq 2$ because both $w = \varepsilon$ and $w = x_1$ are forbidden by Lemma 3.8. Thus as x_2 is the separating letter of $b_1^{(1)}$ (i.e., occurs in any 2-letter factor of $b_1^{(1)}$) we must have $x_1 = x_2$. Continuing, all letters x_i are x_1 , a contradiction with $|Ult(\Delta)| > 1$.

Consequently $v_1 = x_1^p y x_1^q$ for some letter $y \neq x_1$. If $x_2 \neq x_1$ and $x_2 \neq y$ then as x_2 is separating in v_1 , $p = q = 0$ which is impossible by Lemma 3.8. Thus $x_2 \in \{x_1, y\}$. If $x_2 = x_1$ we get, by $x_2\tilde{v}_1 = \psi_{x_2}(v_2)$, $v_2 = \tilde{v}_1$, and we continue. As $|Ult(\Delta)| > 1$ we arrive to $x_i = y$. Suppose for instance, without loss of generality, $x_2 = y$. As y is separating in v_1 this gives $p, q \in \{0, 1\}$. Thus v_1 is $x_1 y x_1$ or $x_1 y$ or $y x_1$. Case $v_1 = x_1 y x_1$ is impossible as it gives $v_2 = x_1 x_1$ which is impossible as seen previously. Cases $v_1 = x_1 y$, $v_1 = y x_1$ give respectively $v_2 = y x_1$ $v_2 = x_1 y$ which allow continuation. Thus $b_1^{(1)}$ is Sturmian on $\{x_1, y\}$ and v_1 is $x_1 y$ or $y x_1$ as claimed.

Now if $\Delta' \in \{x_1, y\}^\omega$ directs $b_1^{(1)} = \tilde{s}^{(1)}x_1 y.s^{(1)}$ for instance then it also directs $\tilde{s}^{(1)}y x_1.s^{(1)}$ and $\bar{\Delta}'$ directs $\tilde{b}_1^{(1)} = \tilde{s}^{(1)}.y x_1 s^{(1)}$ and $\tilde{s}^{(1)}.x_1 y s^{(1)}$ which achieves the proof. □

3.2. The periodic case

Now $|Ult(\Delta)| = 1$, giving periodic episturmian words.

If $\Delta = a^\omega$, $a \in A$ then the unique rew (resp. bew) directed by $\check{\Delta}$ is a^ω itself (resp. $a^{\mathbb{Z}}$). Thus we can ignore spins in \check{a}^ω .

If $\Delta = x_1 x_2 \dots = w y a^\omega$ with $|w| = n - 1$, $y = x_n$, $a = x_{n+1}$, then Δ directs the standard rew $s = (\mu_{wy}(a))^\omega = (\mu_n(a))^\omega = h_n^\omega$ (and the bew $b = h_n^{\mathbb{Z}}$ with the dot between two h_n).

All rew or bew directed by $\check{\Delta} = \check{w}\check{y}a^\omega$ are shifts of s or b and their number is $|\mathcal{C}(s)| = |\mathcal{C}(b)| = |h_n| = \sigma_{n+1}$ because h_n is primitive ([10, Prop. 2.8]).

Now observe that $h_n = \mu_{n-1}(\psi_y(a)) = h_{n-1}\mu_{n-1}(a)$ and consider $\Delta' = way^\omega$. Thus Δ' directs the standard rew $h'_n{}^\omega$ where $h'_n = \mu_{n-1}(\psi_a(y)) = \mu_{n-1}(a)h_{n-1} = T^{\sigma_n}(h_n)$ as $|h_{n-1}| = \sigma_n$. Thus Δ' directs the standard rew $s' = T^{\sigma_n}(s)$. Thus $\mathcal{C}(s)$ contains two standard words. We can show that this is exactly two. Indeed let f be the longest left special factor of s . Definition 1.1 shows that f is also right special hence is a palindromic prefix of s . Now u_n is right special as two letters, y and a , occur in $x_n x_{n+1} \cdots = ya^\omega$ ([8, Th. 6]). For the same reason u_{n+1} is not right special. Thus u_n is the longest right special factor of s , and also of s' . Thus only two standard rew can begin with u_n , the first one beginning with $u_n y$, the second with $u_n a$.

Finally $|h_n u_{n+1}| = |u_{n+2}| \leq 2|u_{n+1}| + 1$ whence $|h_n| - 1 \leq |u_{n+1}|$. Thus as $\mathcal{S}(\bar{w}\bar{y}) = |u_{n+1}|$ all shifts from 0 to $|h_n| - 1$ can be obtained by suitable spins in $\check{w}\check{y}$.

Also as $|u_n| < |h_n|$ by Prop. 2.11 of [10] we have $|u_{n+1}| \leq 2|u_n| + 1 < 2|h_n|$. Thus if $\mathcal{S}(\check{w}\check{y}) \equiv \mathcal{S}(\hat{w}\hat{y}) \pmod{|h_n|}$ then $|\mathcal{S}(\check{w}\check{y}) - \mathcal{S}(\hat{w}\hat{y})| \in \{0, |h_n|\}$. As $|h_n| = \sigma_{n+1}$ is the elementary shift associated to x_{n+1} we get the following proposition.

Proposition 3.10. *With $\Delta = wya^\omega$, $\check{\Delta} = \check{w}\check{y}a^\omega$ and $\hat{\Delta} = \hat{w}\hat{y}a^\omega$ direct the same rew (or bew) if and only if $\check{w}\check{y}\check{a} \equiv \hat{w}\hat{y}\hat{a}$ for suitable spins in \check{a} and \hat{a} .*

4. The Morphic Case

A non-periodic bew b is fixed by a morphism φ if and only if it is directed by some periodic spinned word $\check{\Delta} = \check{f}^\omega$, $f \in A^+$, with $\varphi = \mu_{\check{f}}$ ([10]). Note that in \check{f}^ω all factors \check{f} are identical, including spins. As $Alph(\Delta) = Ult(\Delta)$, b is “strict” (i.e., is an Arnoux-Rauzy sequence). Our aim is to search the conditions for another bew, c , also fixed point of a morphism, to be in the shift-orbit of b . Clearly c is directed by some $\hat{\Delta} = \hat{F}^\omega$ and without loss of generality we can take for F a power of f .

We first examine the particular cases where $\check{\Delta}$ or $\hat{\Delta}$ is not wavy.

Proposition 4.1. *If $\check{\Delta} = \check{f}^\omega$, wavy, directs the bew b then $c = T^k(b)$ is directed by Δ (resp. $\bar{\Delta}$) if and only if $k = 1$ (resp. $k = -1$) and $\check{\Delta}$ has all its spins 1 (resp. 0) except for the occurrences of some letter x all of whose spins are 0 (resp. 1).*

Proof. Consider $c = T^k(b)$ directed by Δ for instance. For large enough p and suitable spins and some g, e we have $\check{f}^p = \check{g}\check{e} \equiv \check{g}\check{e}$, $\mathcal{S}(\check{g}) - \mathcal{S}(\check{e}) = k$. Then $\check{g}\check{e}\check{\Delta}$ directs $T^k(b) = c$, thus $\check{g}\check{e}\check{\Delta} \rightsquigarrow \Delta$. Let w be the longest common prefix of $\check{g}\check{e}\check{\Delta}$ and Δ . Then by Proposition 3.6 $T^{|\check{w}|}(\check{g}\check{e}\check{\Delta}) = \bar{x}\bar{v}_1\bar{x}\bar{v}_2\bar{x}\cdots$, $x \in A$, $v_i \in (A \setminus \{x\})^*$. Thus by periodicity all \check{x} in $\check{\Delta}$ have spin 0 and all other spins are 1. Then by [10, Th. 3.17], the bew b directed by $\check{\Delta}$ has right part xs where s is the standard rew directed by Δ . Thus $T(b)$ has right part s as c does, hence $k = 1$. This proves the “only if” part.

For the “if” part let $k = 1$ for instance. Then b has right part xs for some $x \in Alph(\Delta)$ and s the standard rew directed by Δ . But $\check{\Delta}$ where all x have spin

0 and all other spins are 1 is wavy and directs xs . Thus $\check{\Delta} \equiv \check{\Delta}$. As no good block occurs in $\check{\Delta}$, we have $\check{\Delta} = \check{\Delta}$. □

The other case shows an exceptional property of Sturmian words.

Proposition 4.2. Δ directs b and $\bar{\Delta}$ directs $c = T^k(b)$ if and only if $k = -2$ and the right part of b , is standard Sturmian.

Proof. This immediately follows from Proposition 3.9. □

Now we pass to the “general case” of wavy $\check{\Delta}$ and $\hat{\Delta}$. A theorem gives the answer and a corollary specifies all the possible values for the shift.

Theorem 4.3. Let b be directed by $\check{\Delta} = \check{f}^\omega$ with $\check{\Delta}$ wavy. Then $c = T^k(b)$, $k \in \mathbb{Z} \setminus \{0\}$ is directed by some periodic wavy $\hat{\Delta}$ if and only if for some power F of f we have $\check{F} = \check{F}_1\check{t}_1 = \check{t}_2\check{F}_2$, $t_1, t_2 \neq \varepsilon$, $F_1, F_2 \in A^*$, such that $\check{t} = \check{t}_1\check{t}_2$ is a maximal smb of $\check{F}^2 = \check{F}_1\check{t}\check{F}_2$ and, for suitable spins in \check{t} , $\check{t}_1\check{t}_2 \equiv \check{t}_1\check{t}_2$ with $\mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = k$.

Proof. For the “only if” part the bew $c = T^k(b)$ is directed by $\hat{\Delta}$, periodic and wavy. Let $v_n = \Delta(1, n)$. We claim that, for large enough n and for suitable spins in \check{v}_n , $\mathcal{S}(\check{v}_n) = \mathcal{S}(\check{v}_n) + k$. Indeed $0 \leq \mathcal{S}(\check{v}_n) \leq |u_{n+1}|$ (see Proposition 1.1 and Equation (1)). As $\check{\Delta}$ contains infinitely many spins 0 and 1, both $\mathcal{S}(\check{v}_n)$ and $|u_{n+1}| - \mathcal{S}(\check{v}_n)$ grow to infinity with n . Thus for large enough n , $0 \leq \mathcal{S}(\check{v}_n) + k \leq |u_{n+1}|$ and then there is a spinning \check{v}_n of v_n such that $\mathcal{S}(\check{v}_n) = \mathcal{S}(\check{v}_n) + k$ as claimed.

Now, without loss of generality, take for v_n a power, G say, of f . Then \check{G}^ω directs b whence \check{G}^ω directs $T^k(b)$, i.e., c .

Thus, by Theorem 3.4 and Definition 3.1, $\check{G}^\omega \equiv \hat{\Delta}$ and thus Δ has a prefix $F = GH$ such that, for suitable spins in \check{H} , $\hat{F} \equiv \check{H}\check{H}$.

Without loss of generality we can take for F a power of f . Then both $\hat{\Delta} = \hat{F}^\omega \equiv (\check{H}\check{H})^\omega = \check{G}(\check{H}\check{G})^\omega$ and $\check{G}(\check{H}\check{G})^\omega$ direct c . Thus both $(\check{H}\check{G})^\omega$ and $(\check{H}\check{G})^\omega$ direct a same bew (which is $c^{(|G|)}$, see Preliminaries). Thus $\mu_{\check{H}\check{G}}^\times = \mu_{\check{H}\check{G}}$, i.e., $\check{H}\check{G} \equiv \check{H}\check{G}$.

Now as $\mathcal{S}(\check{G}) = \mathcal{S}(\check{G}) + k$ we have $\check{G} \not\equiv \check{G}$. Thus, by Theorem 2.7, there exists in $\check{H}\check{G}$ a maximal smb $\check{s} = \check{s}_1\check{s}_2$, $s_1, s_2 \neq \varepsilon$, such that $\check{H} = \check{H}_1\check{s}_1$, $\check{G} = \check{s}_2\check{G}_2$ for some H_1, G_2 in A^* and that, for suitable spins, $\check{s}_1\check{s}_2 \equiv \check{s}_1\check{s}_2$ and $\check{H}\check{G} = \check{H}_1\check{s}_1\check{s}_2\check{G}_2$. As $\check{s}_1\check{s}_2$ is a maximal smb we even have $\check{H}_1 \equiv \check{H}_1$ and $\check{G}_2 \equiv \check{G}_2$ whence $\check{G} \equiv \check{s}_2\check{G}_2$. Thus $k = \mathcal{S}(\check{G}) - \mathcal{S}(\check{G}) = \mathcal{S}(\check{s}_2\check{G}_2) - \mathcal{S}(\check{s}_2\check{G}_2) = \mathcal{S}(\check{s}_2) - \mathcal{S}(\check{s}_2)$.

It remains to pass to \check{F}^2 . We have $\check{F}^2 = \check{G}\check{H}\check{G}\check{H}$. The maximal smb \check{s} in $\check{H}\check{G}$ is part of some maximal smb \check{t} of \check{F}^2 and we have $\check{F} = \check{F}_1\check{t}_1 = \check{t}_2\check{F}_2$, with $F_1, F_2 \in A^*$ and for some $s', s'' \in A^*$ $t_1 = s's_1$, $t_2 = s_2s''$ whence $\check{t} = \check{s}'\check{s}_1\check{s}_2\check{s}''$. Writing $\check{t}_1 = \check{s}'\check{s}_1$ and $\check{t}_2 = \check{s}_2\check{s}''$, we get $\check{t} = \check{t}_1\check{t}_2 \equiv \check{t}_1\check{t}_2$ and $\mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = k$.

For the “if” part we have, for some power F of f , $\check{F} = \check{F}_1\check{t}_1 = \check{t}_2\check{F}_2$, $F_1, F_2 \in A^*$, $t_1, t_2 \neq \varepsilon$ and $\check{t}_1\check{t}_2 \equiv \check{t}_1\check{t}_2$ for some spins in $\check{t}_1\check{t}_2$. Thus $\check{t}_2\check{F}_2\check{\Delta}_2$ directs b , with $\Delta_2 = T^{|t_2F_2|}(\Delta)$, whence $\check{t}_2\check{F}_2\check{\Delta}_2$ directs $c = T^k(b)$ for $k = \mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2)$.

But $\check{t}_2\check{F}_2\check{\Delta}_2 = \check{t}_2\check{F}_2(\check{F}_1\check{t}_1\check{t}_2\check{F}_2)^\omega \equiv \check{t}_2\check{F}_2(\check{F}_1\check{t}_1\check{t}_2\check{F}_2)^\omega = (\check{t}_2\check{F}_2\check{F}_1\check{t}_1)^\omega$. Thus c is morphic. □

Corollary 4.4. *It is sufficient in Theorem 4.3 to take $F = f^2$. With the notations and in the conditions of this theorem let b be directed by $\check{\Delta} = \check{f}^\omega$, wavy. Then*

1) *if no maximal smb of $\check{F}^2 = \check{F}.\check{F}$ (from here on the dot denotes the separation between the two factors F) overlaps both factors F then there is no morphic bew, directed by some wavy $\check{\Delta}$, in the shift-orbit of b apart from b itself. However recall that if $\check{\Delta}$ has form given in Proposition 4.1 another morphic shift of b is given by this proposition.*

Otherwise, for parts 2) and 3), let $\check{t} = \check{t}_1\check{t}_2$ be the maximal smb of F^2 (unique by Theorem 2.7) such that $F = F_1t_1 = t_2F_2$, $|t_1|, |t_2| > 0$. Let also F have suffix zx^q and prefix x^py , $x, y, z \in A$, $x \neq y$, $x \neq z$, $p > 0$, $q \geq 0$. Then

2) *if \check{t} is not x -based there are exactly two morphic bew in the shift-orbit of b , b itself and either $\Gamma(b)$ or $\Gamma^{-1}(b)$*

3) *if $\check{t} = \check{t}_1\check{t}_2$ is x -based let $\check{t} \equiv \check{t}$ be a gmb in low form. Let $r = |\check{x}^q.\check{x}^p|_{\check{x}}$. Suppose without loss of generality that the r letters \check{x} in $\check{x}^q.\check{x}^p$ are the leftmost ones, i.e., $\check{x}^q\check{x}^p = \check{x}^r x^{p+q-r}$. Let d be the morphic bew corresponding to \check{t} as in the theorem, i.e., d is directed by $\check{t}_2\check{F}_2\check{F}_1\check{t}_1 \equiv (\check{t}_2\check{F}_2\check{F}_1\check{t}_1)^\omega$. Then $c = \Gamma^k(d)$ is morphic if and only if $k \in [\eta, \min(p, q, r, p + q - r)]$ where $\eta = -1$ if $|t_1| \leq q$ and $r < q$, or $|t_2| \leq p$ and $r > q$, or $|t_1| > q$ and $|t_2| > p$, otherwise $\eta = 0$.*

Proof. For the moment we assume F is a sufficiently large power of f , as in Theorem 4.3, $F = f^e$ say. Case 1) was proved in the theorem.

Consider case 2). As \check{t} is not x -based there exists a letter $a \neq x$ such that $t_1 = t'aw_1$, $t_2 = w_2at''$ and $w_1w_2 \in (A \setminus \{a\})^+$ and all spins in $\check{w}_1\check{w}_2$ are equal. For instance $\check{w}_1\check{w}_2 = \bar{w}_1\bar{w}_2$ (case w_1w_2 is similar). As in the theorem $c = \Gamma^k(b)$ is morphic if and only if there exists $\check{t} \equiv \check{t}$ such that $\mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = k$.

Using Lemma 2.9, if $\check{w}_1\check{w}_2 = \bar{w}_1\bar{w}_2$ then $\check{t}_2 \equiv \check{t}_2$ whence $k = 0$. Otherwise we can suppose without loss of generality that $\check{t} = \check{t}'\bar{a}\bar{w}_1\bar{w}_2\bar{a}\check{t}''$ has been replaced by $\check{t} = \check{t}'aw_1w_2\bar{a}\check{t}''$ whence $\check{t}_2 = \bar{w}_2\bar{a}\check{t}''$, $\check{t}_2 = w_2\bar{a}\check{t}''$. Thus $\mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = \mathcal{S}(\bar{w}_2a) - \mathcal{S}(w_2\bar{a}) = -1$ because a does not occur in w_2 (see [10, Prop. 3.16]).

Now consider case 3) and firstly suppose $|t_1| > q$, $|t_2| > p$. We have $t_1 = t'zx^q$, $t_2 = x^pyt''$, for some $t', t'' \in A^*$. By the theorem, $c = \Gamma^k(d)$ is morphic if and only if there exists $\check{t} \equiv \check{t}$ such that $\mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = k$. We have $\check{t} = \check{t}'\check{z}\check{x}^q.\check{x}^p\check{y}\check{t}''$ and $(\check{z}, \check{y}) \in \{(z, y), (\bar{z}, y), (z, \bar{y}), (\bar{z}, \bar{y})\}$, giving four subcases. Recall that \check{t} is a gmb in low form whence $(\check{z}, \check{y}) = (z, y)$.

In the first subcase, $(\check{z}, \check{y}) = (z, y)$, we have by Lemma 2.9, $\check{t}'' \equiv \check{t}''$ and $|\check{x}^q\check{x}^p|_{\check{x}} = r$. Thus $k = \mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = |\check{x}^p|_{\check{x}} - |\check{x}^p|_{\check{x}} = |\check{x}^p|_{\check{x}} - \max(0, r - q)$. As $|\check{x}^p|_{\check{x}}$ can take any value in $[\max(0, r - q), \min(p, r)]$ we get that $\Gamma^k(d)$ is morphic if and only if $k \in [\max(0, r - q), \min(p, r)] - \max(0, r - q) = [0, \min(p, q, r, p + q - r)]$ (equality is easily verified).

Subcase $(\check{z}, \check{y}) = (\bar{z}, y)$ can be realized if and only if $r > 0$. We have again $\check{t}'' \equiv \check{t}''$. Now $|\check{x}^q \check{x}^p|_{\bar{x}} = r - 1$. It follows as previously $k \in [\max(0, r - 1 - q), \min(p, r - 1)] - \max(0, r - q)$. This introduces a new possible value for k , namely $k = -1$, if and only if $\max(0, r - 1 - q) < \max(0, r - q)$, i.e., $r > q$.

Subcase $(\check{z}, \check{y}) = (z, \bar{y})$ is slightly different. It can be realized if and only if x occurs in $\check{x}^q \check{x}^p$, i.e., $r < p + q$. Then $\check{t} = \check{t}'' z \check{x}^q \check{x}^p \check{y} \check{t}''$. If \bar{x} occurs in \check{x}^p we can write $\check{t}_2 \equiv \check{x}^{p-1} \bar{x} \check{y} \check{t}''$ and get by a mbt $\check{t}_2 \equiv \check{x}^{p-1} x y \check{t}''$ for suitable spins in \check{t}'' . Thus $\check{t} \equiv \check{t}_1 \check{x}^{p-1} x y \check{t}''$. As $\check{t} \equiv \check{t}$, Lemma 2.9 gives $\check{t}'' \equiv \check{t}''$. It follows $k = \mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2) = |\check{x}^p|_{\bar{x}} - |\check{x}^p|_{\bar{x}} - 1$. Using the fact that \bar{x} occurs in $\check{x}^q \check{x}^p$ it can be verified that this relation holds even if \bar{x} does not occur in \check{x}^p .

Consequently as $|\check{x}^q \check{x}^p|_{\bar{x}} = r + 1$ we get for k the interval $[\max(0, r + 1 - q), \min(p, r + 1)] - 1 - \max(0, r - q)$. This introduces for k the new value -1 not already given by the first subcase, if and only if $r - q < 0$.

Subcase $(\check{z}, \check{y}) = (\bar{z}, \bar{y})$ can always be realized. Now $\mathcal{S}(\check{t}_2) - \mathcal{S}(\check{t}_2)$ satisfies the same relation as in subcase $(\check{z}, \check{y}) = (z, \bar{y})$, and also $|\check{x}^q \check{x}^p|_{\bar{x}} = r$, whence $k \in [\max(0, r - q), \min(p, r)] - 1 - \max(0, r - q)$. Thus we always have the value -1 in this subcase.

The cases $|t_1| \leq q, |t_2| \leq p$ follow from what precedes. For instance $|t_2| \leq p$ corresponds to subcases (z, y) and (\bar{z}, y) because in these subcases $y \check{t}'' \equiv y \check{t}''$ and thus suffix $y \check{t}''$ of t_2 has no influence.

It remains to show that we can take $F = f^2$. If $F = f^e, e \geq 2$ and if $\check{t}_1 \check{t}_2$ exists then in view of the periodicity of \check{F} there exists in $\check{f} \check{f}$ a maximal smb $\check{s}_1 \check{s}_2$ such that $t_1 t_2 = s' s_1 s_2 s''$ and $|s'| = |s''| \equiv 0 \pmod{|f|}$. Then clearly all arguments can be repeated with $F = f^2$, and $s_1 s_2$ instead of $t_1 t_2$. □

Example 4.1. In several of the following examples it suffices to take $F = f$, we do so when possible. Here the alphabet is $\{x, y\}$ and b is directed by \check{f}^ω .

1) $\check{f} = x \bar{x} y \bar{y}$. No smb $\check{t}_1 \check{t}_2, t_1, t_2 \neq \varepsilon$ occurs in $\check{f} \check{f}$ and \check{f}^ω has not form of Proposition 4.1. Thus no non-trivial shift of b is morphic.

2) $\check{f} = x \bar{y} y$. Then $\check{t}_1 \check{t}_2 = \bar{y} y . x \bar{y} y$ occurs in $\check{f} \check{f}$ and is not x -based. As $\check{t}_1 \check{t}_2 \equiv \bar{y} \bar{y} . \bar{x} y y$ and as $\mathcal{S}(\bar{x} y) - \mathcal{S}(x \bar{y}) = -1, \Gamma^{-1}(b)$ directed by $(\bar{x} y y x \bar{y} \bar{y})^\omega$ is the unique non-trivial morphic shift of b . Here $F = f^2$ is necessary for the construction of Theorem 4.3 (but observe that $(x y \bar{y})^\omega$ also directs $\Gamma^{-1}(b)$).

3) $\check{f} = x^2 y \bar{x} y \bar{y} x y \bar{x} 3$. Then in $\check{f}^2, \check{t}_1 \check{t}_2 = x y \bar{x}^3 . x^2 y \bar{x}$ is the maximal smb overlapping both \check{f} . It is x -based. We have $|t_1| > q = 3, |t_2| > p = 2, r = 3$. As $\check{t}_1 \check{t}_2$ is already a gmb in low form, we have that $\Gamma^k(b)$ is morphic if and only if $k \in [-1, \min(2, 3, 3, 2)] = [-1, 2]$, and this gives the four bew directed by \check{f}^ω where \check{f} is respectively $x^2 \bar{y} x y \bar{y} \bar{x} y \bar{x} 3, \check{f}$ itself, $x \bar{x} y \bar{x} y \bar{y} x y \bar{x} 2x$ and $\bar{x} 2 y \bar{x} y \bar{y} x y \bar{x} 2$.

4) $\check{f} = x^p y \bar{x}^p$ (simplest example). Then we have $|t_1|, |t_2| > p = q = r$ whence $\eta = -1$. For any $0 \leq k \leq p, (\bar{x}^k x^{p-k} y \bar{x}^k \bar{x}^{p-k})^\omega$ directs $\Gamma^k(b)$ and also $(x^p \bar{y} \bar{x}^p)^\omega$ directs $\Gamma^{-1}(b)$.

Acknowledgement. We thank P. Arnoux, I. Fagnot, an anonymous referee and, all especially, G. Richomme for very useful remarks.

References

1. P. Arnoux, G. Rauzy, *Représentation géométrique de suites de complexité $2n+1$* , Bull. Soc. Math. France 119 (1991) 199–215.
2. J. Berstel, *An exercise on Fibonacci representations*, Theoret. Informatics Appl. 35 (2001) 491–498.
3. J. Berstel, *Recent results on extensions of Sturmian words*, Internat. J. Algebra Comput. 12, 1-2 (2002) 371–385.
4. J. Berstel, P. Séébold, *A remark on morphic Sturmian words*, Theoret. Informatics Appl. 28, 3-4 (1993) 255–263.
5. V. Berthé, *Autour du système de numération d’Ostrowski*, Bull. Belg. Math. Soc. 8 (2001) 209–239.
6. L. Carlitz, *Fibonacci representations*, The Fibonacci Quater. 6(4) (1968) 193–220.
7. C. B. Christoffel, *Observatio arithmetica*, Math. Annal., 6 (1875) 145–152.
8. X. Droubay, J. Justin, G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci. 255, 1-2 (2001) 539–553.
9. I. Fagnot, *A little more about morphic Sturmian words*, preprint (2002).
10. J. Justin, G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci. 276 (2002) 281–313.
11. M. Lothaire, *Algebraic Combinatorics on Words*, M. Lothaire ed., vol. 90 of Encyclopedia of Mathematics, Cambridge University Press (2002).
12. M. Morse, G.A. Hedlund, *Symbolic dynamics II: Sturmian trajectories*, Amer. J. Math. 62 (1940) 1–42.
13. G. Rauzy, *Nombres algébriques et substitutions*, Bull. Soc. Math. France 110 (1982) 147–178.
14. G. Rauzy, *Mots infinis en arithmétique*, in Automata on infinite words, M. Nivat, D. Perrin eds., Lect. Notes Comp. Sci. 192 (1985) 165–171.
15. G. Richomme, *Conjugacy and episturmian morphisms*, Theoret. Comput. Sci. 302, 1-3 (2003) 1–34.
16. R.N. Risley, L.Q. Zamoni, *A generalization of Sturmian sequences, combinatorial structure and transcendence*, Acta Arithmetica 95 (2000) 167–184.

Episturmian words and some constructions of de Luca and Rauzy

Xavier Droubay^a, Jacques Justin^{b,*}, Giuseppe Pirillo^{c,d}

^a*LaRIA, Université de Picardie Jules Verne, CURI, 5, rue du Moulin Neuf, 80000 Amiens, France*

^b*LIAFA, 2 place Jussieu, F-75251 Paris, Cedex 05, France*

^c*IAMI CNR, Viale Morgagni 67/A, 50134 Firenze, Italy*

^d*Université de Marne-la-Vallée, 5, boulevard Descartes, Champs sur Marne,
77454 Marne-la-Vallée Cedex 2, France*

Received May 1999; revised June 1999

Communicated by D. Perrin

Abstract

In this paper we study infinite episturmian words which are a natural generalization of Sturmian words to an arbitrary alphabet. A characteristic property is: they are closed under reversal and have at most one right special factor of each length. They are first obtained by a construction due to de LUCA which utilizes the palindrome closure. They can also be obtained by the way of extended RAUZY rules. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Sturmian words; Palindrome factors; Rauzy rules; Episturmian words

0. Introduction

In [5] de Luca makes a deep study of finite Sturmian words and gives a characterization of standard Sturmian (infinite) words, namely if u_n is a palindrome prefix of the infinite word, then $u_{n+1} = (u_n x_n)^{+}$ is also a palindrome prefix for some $x_n \in \{a, b\}$, where v^{+} denotes the palindrome right closure of v , i.e. the shortest palindrome having prefix v . As we shall see this property is equivalent to the following one, *Al*: an infinite word s has *Al* if for any prefix v of s , v^{+} is also a prefix of s .

On the other hand, we have observed another characteristic property of standard Sturmian words, *Pi*: an infinite word s has *Pi* if any leftmost occurrence of a palindrome

* Correspondence address: J. Justin, 19 rue de Bagneux, 92330 Sceaux, France.

E-mail addresses: droubay@laria.u-picardie.fr (X. Droubay), justin@liafa.jussieu.fr (J. Justin), pirillo@udini.math.unifi.it (G. Pirillo).

in s is a central factor of a prefix palindrome of s . Finite Sturmian words have also the following property, Ju : a word has Ju if it has a palindrome suffix with no other occurrence in it.

In Section 2, we study the relations between these properties in the case of an arbitrary alphabet. In particular, we prove the non-trivial equivalence between Al and Pi . In Section 3, we apply this to a two letter alphabet and obtain a simple and short proof of de Luca's characterization.

The infinite words considered in Section 2 are a natural and promising generalization to an arbitrary alphabet of the standard Sturmian words. So in Section 4 we call them standard episturmian and study some of their properties. In particular (Theorem 5), an infinite word is episturmian if and only if it is closed under reversal and has at most one right special factor of each length. It appears that, at least in the case that we call strict episturmian and with a three letter alphabet, they have been evoked by Rauzy [11] and, together with Arnoux, studied in [1], so it could perhaps be better to call them Rauzy words. We prove the equivalence of their various definitions, we begin the study of morphisms preserving these words and we also mention some open problems.

Let us also recall that generalizations of Sturmian words in other directions have been considered by several authors (see notes of Berstel and Seebold [2] for a bibliography), in particular the billiard sequences which generalize the definition of Sturmian words by cutting sequences [12].

1. Preliminaries

1.1. Words

Given a set A (*alphabet*) whose elements are called *letters*, the free monoid A^* generated by A is the set of the (finite) words on A . The *empty word* is denoted by ε . If $u = u(1)u(2) \cdots u(m)$, $u(i) \in A$, is a word, its *length* is $|u| = m$. Also $|u|_x$, $x \in A$, is the number of occurrences of x in u . Lastly \tilde{u} denotes the *reversal* of u , i.e. the word $u(m)u(m-1) \cdots u(2)u(1)$. A word equal to its reversal is a *palindrome*.

In the same way, an *infinite word* s is an infinite sequence $s = s(1)s(2) \cdots s(i) \cdots, s(i) \in A$, $i \in \mathbb{N}_+ = \mathbb{N} \setminus \{0\}$. The set of infinite words on A is A^ω . For a finite or infinite word t , the word $w = t(i)t(i+1) \cdots t(j)$ is a *factor* of t (*proper factor* if $w \neq t$). It is a *prefix* of t if $i = 1$. If $t \in A^*$ and $|t| = m$, then w is a *suffix* of t if $j = m$. Also we speak of an *occurrence* of w in t if we consider w together with its position in t defined, for instance, by i or by a prefix uw , $u \in A^*$ of t . The factor w of t is *unioccurrent* in t if it has exactly one occurrence in t . Lastly, given two occurrences of w in t given by the prefixes uw and $u'w$ of t , the *shift* between them is $\|u| - |u'|\|$. Let u, v be two palindromes, then u is a *central factor* of v if $v = wu\tilde{w}$ for some $w \in A^*$.

Let $m \in A^* \cup A^\omega$. Then $F(m)$ denotes the set of the factors of m and the *alphabet* of m is $Alph(m) = F(m) \cap A$. Also m is closed under reversal if $\tilde{u} \in F(m)$ whenever $u \in F(m)$.

Given, $u \in A^*$, its *palindrome right-* (resp. *left-*) *closure* is the (unique) shortest palindrome $w = u^{(+)}$ (resp. $w = u^{(-)}$) which has prefix (resp. suffix) u [5].

A word is *primitive* if it is not a power of some shorter word. An infinite word s is *ultimately periodic* if it can be written $s = uvvv \cdots = uv^\omega$ for $u \in A^*$ and v a primitive word (*periodic* if $u = \varepsilon$).

Let s be an infinite word, then a factor u of s is *right* (resp. *left*) *special* in s if there exist $x, y \in A$, $x \neq y$, such that $ux, uy \in F(s)$ (resp. $xu, yu \in F(s)$). Clearly, if s is closed under reversal, its right special factors are exactly the reversals of its left special factors.

Finally, an infinite word s is *uniformly recurrent* if for each $u \in F(s)$ there exists k such that $u \in F(w)$ for any $w \in F(s)$ with $|w| = k$.

1.2. Sturmian words

Sturmian words, whose theory was founded by Morse and Hedlund 60 years ago in the frame of symbolic dynamics, are extensively studied (see [2] for a survey).

Definition 1 (*Berstel and Seebold* [2]; *Crisp et al.* [3]; *de Luca* [5]; *Justin and Pirillo* [7]). Let ρ, α be real numbers in $[0,1]$ with α irrational. A *Sturmian word* on $\{a, b\}$ is an infinite word s given either by

$$s(n) = \begin{cases} a & \text{if } \lfloor \rho + (n+1)\alpha \rfloor - \lfloor \rho + n\alpha \rfloor = 0, \\ b & \text{otherwise} \end{cases}$$

or by

$$s(n) = \begin{cases} a & \text{if } \lceil \rho + (n+1)\alpha \rceil - \lceil \rho + n\alpha \rceil = 0, \\ b & \text{otherwise.} \end{cases}$$

Some interesting geometrical equivalent definitions are known, cutting sequences in particular.

Definition 2. A Sturmian word is *standard* if $\rho = 0$ in Definition 1.

Definition 3. A word is a *finite Sturmian word* if it is a factor of some (infinite) Sturmian word.

Some useful properties [2] are recalled hereafter.

- An infinite word is Sturmian if and only if for every n it has exactly $n + 1$ factors of length n .
- A finite or infinite non-ultimately-periodic word u is Sturmian if and only if it is *balanced*, that is whenever w, w' are factors of u with $|w| = |w'|$ we have $\|w\|_a - \|w'\|_a \leq 1$.
- An infinite word on two letters is standard Sturmian if and only if its prefixes are exactly its left special factors.

2. Palindrome factors

In this section the alphabet A is arbitrary.

Definition 4. A word w has property Ju (resp. LJu) if there exists a palindrome suffix (resp. prefix) of w which is unioccurrent in w .

Clearly, if w has Ju it has exactly one unioccurrent palindrome suffix v and this one is the longest palindrome suffix of w , moreover if $w = uv$, then $w^{(+)} = uv\tilde{u}$.

Now, for $u \in A^*$, let $P(u)$ denote the number of different palindrome factors of u . We have

Proposition 1. $P(w)$ is the number of prefixes (resp. suffixes) of w which have Ju (resp. LJu).

Proof. For any $u \in A^*$ and $x \in A$, we have:

$P(ux) = P(u)$ if ux has not Ju ,

$P(ux) = P(u) + 1$ if ux has Ju .

The result for Ju follows by induction as $P(\varepsilon) = 1$. \square

In particular, we have the trivial but seemingly not widely known fact

Proposition 2. A word w has at most $|w| + 1$ different palindrome factors.

Also

Proposition 3. A word w has exactly $|w| + 1$ palindrome factors if and only if all its prefixes (resp. suffixes) have Ju (resp. LJu).

Corollary 1. If $P(w) = |w| + 1$ then for any factor u of w , $P(u) = |u| + 1$. In other words, the language of all such w is factorial.

Proof. Let $w = w'uw''$. Then by Proposition 3 all prefixes of $w'u$ have Ju whence, again by Proposition 3, all suffixes of u have LJu , that is $P(u) = |u| + 1$. \square

Definition 5. An infinite word s has property Al if, for any prefix u of s , $u^{(+)}$ is also a prefix of s .

Definition 6. An infinite word s has property Pi if every leftmost occurrence of a palindrome in s is a central factor of a palindrome prefix of s .

Theorem 1. For an infinite word $s \in A^\omega$ the following conditions are equivalent

- (i) s has Al ,
- (ii) s has Pi ,

(iii) there exist an infinite sequence $u_1 = \varepsilon, u_2, u_3, \dots$ of palindromes and an infinite word $\Delta(s) = x_1 x_2 x_3 \dots$, $x_i \in A$, such that $u_{n+1} = (u_n x_n)^{(+)}$ for all $n \geq 1$ and that all the u_n are prefixes of s .

Remark. When s satisfies (iii), u_1, u_2, u_3, \dots are all the palindrome prefixes of s in increasing length order, because if some palindrome prefix u of s would satisfy $|u_n| < |u| < |u_{n+1}|$ we would have $|(u_n x_n)^{(+)}| \leq |u| < |u_{n+1}|$, a contradiction.

Proof. (iii) \Rightarrow (i): With notations as in (iii) let u be any non-empty prefix of s . There exists n such that $|u_n| < |u| \leq |u_{n+1}|$. As $u_n x_n$ is a prefix of u , $|(u_n x_n)^{(+)}| \leq |u^{(+)}|$; as u_{n+1} is a palindrome $|u^{(+)}| \leq |u_{n+1}|$. So $|u^{(+)}| = |u_{n+1}|$ and by the unicity of the palindrome right closure of $u_n x_n$ we get $u^{(+)} = u_{n+1}$. Consequently s has Al .

(i) \Rightarrow (ii): Let v be any palindrome factor of s with its leftmost occurrence given by $s = uvs'$, $u \in A^*$, $s' \in A^\omega$. If v is not the longest palindrome suffix of uv we have $uv = u'wv$ with $w \neq \varepsilon$ and wv a palindrome, whence $uv = u'v\tilde{w}$ and this is a contradiction as $|u'| < |u|$. So v is the longest palindrome suffix of uv , whence $(uv)^{(+)} = uv\tilde{u}$. As s has Al , $uv\tilde{u}$ is a prefix of s and so v is a central factor of a palindrome prefix of s .

For the more difficult proof of (ii) \Rightarrow (iii) we require a lemma

Lemma 1. *If an infinite word s on an alphabet A has property Pi , then all its prefixes have Ju .*

Proof. By induction on $n = |w|$, where w is prefix of s , we exhibit a palindrome w_2 suffix of w and unioccurrent in it, which is trivial for $n = 1$. Let $w = vx$, where $x \in A$, $v \in A^n$, ($|w| = n + 1$) and $s = wt$ with $t \in A^\omega$.

The induction allows to set $v_1 \in A^*$, $v_2 \in A^+$ with $v = v_1 v_2$ and v_2 palindrome unioccurrent in v . Then Pi involves that $v_1 v_2 \tilde{v}_1$ is a prefix of s . If $v_1 \neq \varepsilon$, x is prefix of \tilde{v}_1 and $w_2 = xv_2 x$ is the palindrome suffix of w which is unioccurrent in it, since v_2 is unioccurrent in v .

Let us suppose now that $v_1 = \varepsilon$, so $v = v_2$ is a palindrome. Either x is unioccurrent in w and $w_2 = x$ is obviously the unique palindrome suffix of w , or, let ux be the prefix of v such that x is unioccurrent in ux , and $|u| < |vx| = n + 1$. We first prove that u is a palindrome: the case x prefix of v being clear since $u = \varepsilon$ is clearly a palindrome, we now suppose $u \neq \varepsilon$. By induction we can set $u_1 \in A^*$, $u_2 \in A^+$ with $u = u_1 u_2$ and u_2 a palindrome unioccurrent in u . Again, the assumption $u_1 \neq \varepsilon$ with Pi would involve that $u_1 u_2 \tilde{u}_1$ is a prefix of s , and would contradict the unioccurrence of x in ux . So $u_1 = \varepsilon$, and $u = u_2$ is a palindrome. Now we can choose qx , the longest prefix of v with q a palindrome. As v is a palindrome, xqx is clearly a palindrome suffix of $w = vx$, and the proof is complete if we show that $w_2 = xqx$ is unioccurrent in w . If it is not true, we set $v = oxqxp$ with $o, p \in A^*$ such that xqx is unioccurrent in $oxqx$. Let us prove that $l = oxq$ is a palindrome. Otherwise there exists by induction $l_1, l_2 \in A^+$ with $l = l_1 l_2$ and l_2 a palindrome unioccurrent in l . Moreover $l_2 \neq q$ (q is a prefix of l since it is a prefix of v and thus q is not unioccurrent in l), so $l_1 l_2 \tilde{l}_1$ is, from Pi , a prefix of s

and x is a prefix of \tilde{l}_1 . Thus xl_2x and xqx would be two distinct palindrome suffixes of lx , both unioccurrent in it! So l is a palindrome with lx a prefix of v , but then, as $|l| > |q|$, this contradicts the choice of q , and thus w_2 is unioccurrent in w . \square

(ii) \Rightarrow (iii): Now s has Pi . Let us build the sequences $(u_n)_{n \geq 1}$ and $\Delta(s)$ as denoted in (iii). Clearly, $u_1 = \varepsilon$ and x_1 is the letter prefix of s . Then we can set $u_2 = (u_1x_1)^{(+)} = x_1$ and x_2 is the letter such that u_2x_2 is a prefix of s , with $|u_2| \geq |u_1|$. By induction on n , the above Lemma 1 ensures the existence of a unioccurrent palindrome suffix v_n of u_nx_n , and the leftmost occurrence of v_n in s is given by $s = wv_ns' = u_nx_ns'$, for some $w \in A^*$, $s' \in A^\omega$. So with Pi , v_n is a central factor of a palindrome prefix of s , say u' . Moreover, v_n is the longest palindrome suffix of u_nx_n , so that it is exactly a central factor of $u' = (u_nx_n)^{(+)}$. Setting $u_{n+1} = (u_nx_n)^{(+)}$ which is a palindrome prefix of s , and x_{n+1} the letter such that $u_{n+1}x_{n+1}$ is a prefix of s , the proof is over. \square

From now on, $\Delta(s)$ will be called the *directive word* of s .

Corollary 2. *If an infinite word satisfies the (equivalent) conditions of Theorem 1, then any factor w of it has exactly $|w| + 1$ palindrome factors.*

Proof. By Lemma 1 all prefixes have Ju , so by Proposition 3 any factor w has exactly $|w| + 1$ palindrome factors. \square

Remark. The infinite words having Al are, by Corollary 2, “rich” in palindrome factors. For example, by Corollary 4 hereafter, the well-known [2] Fibonacci word $abaababa \dots$ has this property. On the contrary, it is easy to find infinite words which are “poor” in palindrome factors, for example $(abc)^\omega$ whose unique palindromes are a, b, c , and for a non-periodic example, the image of the Fibonacci word by the morphism $a \mapsto ac$, $b \mapsto b$, whose unique palindromes are a, b, c and aca .

Proposition 4. *If an infinite word s satisfies the (equivalent) conditions of Theorem 1, then it is closed under reversal.*

Proof. This is trivial as any factor of s is a factor of some palindrome prefix of s . \square

Theorem 2. *If an infinite word s satisfies the (equivalent) conditions of Theorem 1, then it is uniformly recurrent.*

Proof. Following the notations of (iii) of Theorem 1, the palindrome prefixes of s , $u_1 = \varepsilon$, u_2, u_3, \dots satisfy $u_{i+1} = (u_ix_i)^{(+)}$, $x_i \in A$. It suffices to show, given any factor v of s , that v has infinitely many occurrences in s : this is trivial, and that the shift between successive occurrences of v is bounded. Without loss of generality, we may suppose that v is a palindrome prefix of s , that is $v = u_m$ for some m . For $n > m$, v is a prefix and a suffix of u_n . Let k_n be the maximum shift between successive occurrences of v in u_n . As u_n is prefix and suffix of u_{n+1} and $|u_{n+1}| \leq 2|u_n| + 1$, we get $k_{n+1} \leq \max(k_n, |v| + 1)$. For the same reason $k_{m+1} \leq |v| + 1$. Hence for all $i \geq m$, we have $k_i \leq |v| + 1$. \square

Theorem 3. *An infinite word s having AI is ultimately periodic if and only if its directive word $\Delta(s)$ has the form gx^ω , $g \in A^*$, $x \in A$. In this case, s is even periodic.*

Proof. If s is ultimately periodic, let $s = fv^\omega$, $f \in A^*$ and v a primitive word. There exists n such that for $i \geq n$ the palindrome prefix u_i defined in (iii) of Theorem 1 satisfies $|u_i| \geq |fv|$. Consider two such palindromes u_i, u_j , $n \leq i < j$. Then $u_i = fv^p v'$, $u_j = fv^q v''$, p, q positive integers and v', v'' proper prefixes of v . As u_i is a suffix of u_j , $v^p v'$ is a suffix of $v^q v''$ whence, as v is primitive, $v' = v''$. Consequently, the letters x_i, x_j following respectively u_i, u_j are the same, x say, whence $\Delta(s) = x_1 x_2 \cdots x_{n-1} x^\omega$.

Conversely, let $\Delta(s) = x_1 x_2 \cdots x_{n-1} x^\omega$. As s satisfies condition (iii) of Theorem 1, $u_{i+1} = (u_i x)^{(+)}$ for $i \geq n$. Now, for $i \geq n$, $xu_i x$ is the longest palindrome suffix of $u_{i+1} x$ because u_i is the longest palindrome proper prefix (and suffix) of u_{i+1} . Let $u_{n+1} = cxu_n$, $c \in A^*$. Then $u_{n+2} = (cxu_n x)^{(+)} = cxu_n x \tilde{c} = cxu_{n+1}$, and continuing in the same way, $u_{n+j} = cxu_{n+j-1}$, for $j \geq 1$. It follows $u_{n+j} = (cx)^j u_n$, for $j \geq 0$ whence $s = (cx)^\omega$, that is s is periodic. \square

Corollary 3. *If the infinite word s has AI and is not periodic, then all its prefixes are left special factors.*

Proof. Let u be a prefix of s . Then \tilde{u} is a suffix of all palindrome prefixes u_i of s such that $|u_i| \geq |u|$. As s is not periodic, by Theorem 3, two letters at least occur infinitely many times in $\Delta(s) = x_1 x_2 \cdots$. So the occurrences of \tilde{u} in s are followed at least by two different letters, and \tilde{u} is right special in s . \square

In order to prove Proposition 5 hereafter which completes Corollary 3, we require a lemma

Lemma 2. *Let s be an infinite word having AI and u be a factor of s . With the notations of (iii) in Theorem 1, let r be such that $u \in F(u_{r+1}) \setminus F(u_r)$, and write $u_{r+1} = ded\tilde{d}$ with $u_r x_r = de$, $d, e \in A^*$, and e being the unioccurrent palindrome suffix of $u_r x_r$. Then*

- (a) *the leftmost occurrence of u in the prefix u_{r+1} of s is given by $u_{r+1} = ded\tilde{d} = cuc' = cfef'c'$, with $c, c', f, f' \in A^*$. In particular, e is a factor of $u = fef'$*
- (b) *if u is a palindrome, then $\tilde{f} = f'$ and e is a central factor of u*
- (c) *if $u = xvy$, with $x, y \in A$ and v a palindrome prefix of s , then u is either a suffix of $u_r x_r$ or a prefix of $x_r u_r$.*

Proof. (a) As $u \in F(u_{r+1})$, putting into evidence the leftmost occurrence of u , we have $u_{r+1} = cuc'$ for some $c, c' \in A^*$. As $u \notin F(u_r)$ and $u_{r+1} = u_r x_r \tilde{d}$ we have $|cu| \geq |u_r x_r|$ whence $|c'| \leq |d|$. In the same way, $|c| \leq |d|$. So writing $d = cf$, $\tilde{d} = f'c'$, $f, f' \in A^*$, we get $u_{r+1} = cuc' = cfef'c'$, whence $u = fef'$.

(b) As u_{r+1} and u are palindromes we have $u_{r+1} = cuc' = \tilde{c}'u\tilde{c}$. So $|c| \leq |c'|$ because u is a unioccurrent suffix of cu . As $u_{r+1} = cfef'c' = c\tilde{f}'e\tilde{f}c'$ and as e is a unioccurrent suffix of $cfef'$, we have $|cf| \leq |c\tilde{f}'|$ whence $|f| \leq |f'|$. As $|cf| = |c'f'| = |d|$ we get $|c| = |c'|$, $|f| = |f'|$. So e is a central factor of u .

(c) We have $u = xvy = fef'$. If $|f| \geq 1$ and $|f'| \geq 1$, then e is a factor of v . As v is a palindrome prefix of s , shorter than u_{r+1} , e is a factor of u_r , a contradiction. So we have, for instance, $f' = \varepsilon$ whence $u_r x_r = cfef' = cu$. Similarly if $f = \varepsilon$, then $x_r u_r = uc'$. \square

Proposition 5. *If the infinite word s has Al , then all its left special factors are prefixes of it.*

Proof. If this is false, let u be a shortest left special factor of s which is not a prefix of s . Let $u = vx$, $x \in A$. Then v is left special, hence is a prefix of s . So vz is a prefix of s for some $z \in A$, $z \neq x$. So v is right special. So \tilde{v} is a prefix of s whence $v = \tilde{v}$. As u is left special we have $yvx, y'vx \in F(s)$ for some $y, y' \in A$, $y \neq y'$. Applying Lemma 2(c) to yvx we get that yvx or xvy is a prefix of $x_r u_r$ for some r . Clearly $|u_r| > |v|$, so vz is a prefix of u_r whence $z \in \{x, y\}$. As $z \neq x$, we get $z = y$. In the same way $z = y'$ whence $y = y'$, a contradiction. \square

3. Case of a two letter alphabet

Now the alphabet is $A = \{a, b\}$. If $x \in A$ we denote by \bar{x} the other letter of A . The infinite words satisfying Theorem 1 can be completely characterized: the equivalence between (iii) and (iv) in Theorem 4 hereafter is due to de Luca [5, Theorem 5 and Proposition 11]. Indeed his formulation of (iii) makes use of palindrome left closure (-) but this amounts to the same thing. Here we prove in a simple way de Luca's result with the help of Section 2.

Theorem 4. *For an infinite non-ultimately-periodic word s on the two letter alphabet $A = \{a, b\}$, the following conditions are equivalent*

- (i) s has Al ,
- (ii) s has Pi ,
- (iii) *there exist an infinite sequence of palindromes $u_1 = \varepsilon, u_2, u_3, \dots$ and an infinite word $\Delta(s) = x_1 x_2 x_3 \dots$, $x_i \in A$, $\Delta(s) \in A^\omega \setminus (A^* a^\omega \cup A^* b^\omega)$ such that the u_i are prefixes of s and $u_{i+1} = (u_i x_i)^{(+)}$ for all $i \geq 1$,*
- (iv) *the infinite word s is standard Sturmian.*

Proof. The equivalence among (i), (ii) and (iii) follows from Theorems 1 and 3.

(iv) \Rightarrow (ii): Here s is standard Sturmian. Suppose it has not Pi . Then there exist a shortest $u \in A^*$, a palindrome v unioccurrent in uv and $s' \in A^\omega$ such that $s = uvs'$ and \tilde{u} is not a prefix of s' . Clearly $u \neq \varepsilon$ so putting $u = u'x$, $u' \in A^*$, $x \in A$

and $s' = ys''$, $y \in A$, $s'' \in A^\omega$ we have $s = u'xvys''$. If $x = y$, $xvx = v'$ is unioccurrent in $u'v'$ and $u'v'u' = uv\tilde{u}$ is not a prefix of s . As this contradicts the minimality of $|u|$ we have $x \neq y$. So $s = u'xv\bar{x}s''$. So xv , $v\bar{x}$ and their reversals $v\bar{x}$, $\bar{x}v$ are factors of s . So v is left special and is a prefix of s , a contradiction with (iv).

(i) \Rightarrow (iv): Now s satisfies Theorem 1. So by Proposition 5 it has at most one left special factor of each length and as it is not ultimately periodic it has exactly one. As by Proposition 5 these left special factors are prefixes, we get that s is standard Sturmian. \square

Corollary 4. Any finite Sturmian word u has Ju and has exactly $|u| + 1$ different palindrome factors.

Proof. This follows from Corollary 2 and Theorem 4. \square

4. Episturmian words

In this section the alphabet A is finite.

4.1.

The words satisfying Theorem 1 are very similar to standard Sturmian words and as we will see have several similar properties. So we introduce the following definition

Definition 7. An infinite word s is *standard episturmian* if it satisfies the equivalent conditions of Theorem 1.

An infinite word t is *episturmian* if $F(t) = F(s)$ for some standard episturmian word s .

The second part of this definition is justified by the uniform recurrence property (Theorem 2) and the analogy with Sturmian words.

An episturmian word is periodic (resp. aperiodic) if the corresponding standard episturmian word is periodic (resp. aperiodic).

The case of periodicity is given by Theorem 3. The next theorem gives a characterization of episturmian words.

Theorem 5. Let s be an infinite word on the finite alphabet A , then the following conditions are equivalent:

- (i) s is episturmian,
- (ii) s is closed under reversal and has at most one right special factor of each length.

Proof. (i) \Rightarrow (ii): It suffices to prove this for standard episturmian words, and this follows immediately from Propositions 4 and 5.

(ii) \Rightarrow (i): Remark that as s is closed under reversal, then any factor of s has infinitely many occurrences in s . Construct a “standard” infinite word $\Sigma(s)$ as follows:

- (a) if s has exactly one left special factor of each length, then $\Sigma(s)$ is the infinite word whose prefixes are those factors;
- (b) if not, s is ultimately periodic and even is periodic as each factor occurs infinitely many times. Let u be the longest left special factor of s and for one of its occurrences, write $s = fus'$, $f \in A^*$, $s' \in A^\omega$. Then $\Sigma(s) = us'$.

We prove first that $\Sigma(s)$ has *Pi*. If not there is a leftmost occurrence of a palindrome w in $\Sigma(s)$ such that $\Sigma(s) = uxvw\tilde{v}ys'$ for some $u, v \in A^*$, $s' \in A^\omega$, $x, y \in A$, $x \neq y$. As $yvw\tilde{v}x \in F(s)$, $vw\tilde{v}$ is left special in s , whence $\Sigma(s) = vw\tilde{v}s''$, $s'' \in A^\omega$ and this contradicts the fact that the occurrence of w considered previously was the leftmost one. Consequently, $\Sigma(s)$ is standard episturmian.

Let us show now that $F(s) = F(\Sigma(s))$. If s is periodic this is evident by the definition of $\Sigma(s)$. If not, let u be any factor of s and let $s(i)s(i+1)\cdots s(i+k-1)$ and $s(j)s(j+1)\cdots s(j+k-1)$ be two occurrences of \tilde{u} in s with $k = |u|$ and $j > i$. If \tilde{u} is not right special we have $s(i+k) = s(j+k)$. Continuing this way we see that either $s(i+q) = s(j+q)$ for all $q \geq 0$ or \tilde{u} is a prefix of some right special factor of s . The first case is impossible as s is not periodic. So u is a factor of some left special factor of s , whence $u \in F(\Sigma(s))$. So $F(s) = F(\Sigma(s))$ and s is episturmian. \square

Remark. The condition of closure under reversal cannot be deleted from (ii). Indeed replace in the Fibonacci word $abaab\cdots$ each b by bc , then we get the infinite word $abcaabcabc\cdots$ which has exactly one left (resp. right) special factor of each length but is not episturmian because it is not closed under reversal.

4.2. Complexity of episturmian words

Recall that the complexity function p of an infinite word s is given by $p(n) = |A^n \cap F(s)|$ for all $n \in \mathbb{N}$.

Theorem 6. *Let s be a standard episturmian word with sequence of palindrome prefixes $u_1 = \varepsilon, u_2, \dots$ and $\Delta(s) = x_1x_2\cdots$ as in (iii) of Theorem 1. Then for $n \in \mathbb{N}_+$ and $x \in A$, $u_nx \in F(s)$ if and only if x occurs in $x_nx_{n+1}\cdots$.*

Proof. *If part:* If $x = x_q$ for some $q \geq n$, then u_qx_q is a prefix of s , hence its suffix u_nx_q is a factor of s .

Only if part: Let $u_nx \in F(s)$, $x \in A$, and consider a factor yu_nx , $y \in A$, of s . By Lemma 2(c) yu_nx or xu_ny is a suffix of u_rx_r for some r . So $|yu_nx| \leq |u_rx_r|$ whence $n < r$. So xu_n is a suffix of u_r . So $x = x_r$ or $x = x_n$, so x occurs in $x_nx_{n+1}\cdots$. \square

Theorem 7. *Let t be an aperiodic episturmian word, s be the corresponding standard episturmian word, with directive word $\Delta(s) = x_1x_2x_3\cdots$. Then, for n large enough, t has complexity $p(n) = (h-1)n + q$ for some $q \in \mathbb{N}_+$ and with h the number of letters that occur infinitely many times in $\Delta(s)$.*

Proof. For any right special factor u of t let h_u be its *order*, that is the number of letters x such that $ux \in F(t)$. Then h_u is a decreasing function of $|u|$ and for $|u|$ large enough $h_u = h$ where, by Theorem 6, h is the number of the letters occurring infinitely many times in $A(s)$. As $p(|u| + 1) - p(|u|) = h_u - 1$, the result follows. \square

The case where any right special factor of t has order $k = |A|$ gives the maximum complexity $p(n) = (k - 1)n + 1$. Such episturmian words will be called A -strict according to the following definition.

Definition 8. Let $t \in A^\omega$ be an episturmian word and $B = \text{Alph}(t)$. Then t is *strict* if for any right special factor u of t , $uB \subset F(t)$. Also, t is A -*strict* if t is strict and $B = A$.

In [1] uniformly recurrent infinite words of complexity $p(n) = (k - 1)n + 1$ with exactly one right and one left special factor of each length have been studied (with emphasis on the case $k = 3$). We claim that they are exactly the strict episturmian words. This follows from Section 4.3 hereafter but we can sketch a more direct proof. Let T be such an infinite word with alphabet A . In [1] it is studied by the mean of its factor graphs Γ_n . The directed graph Γ_n has for vertice set $A^n \cap F(T)$ and has an arrow from u to v if $uy = xv \in F(T)$ for some $x, y \in A$. It consists of k simple paths from D_n to G_n and one simple path from G_n to D_n , with G_n (resp. D_n) the left (resp. right) special factor of length n . The evolution of these graphs when n grows shows that T has infinitely many factors w_1, w_2, \dots which are both left and right special and that $|w_{n+1}| \leq 2|w_n| + 1$. Consequently if w_n is a palindrome, then w_{n+1} is also a palindrome. Consequently, T has infinitely many palindrome factors hence is closed under reversal. So, by Theorem 5, T is episturmian.

4.3. Rauzy rules

The construction of standard Sturmian words by Rauzy pairs and rules is well known [8]. The extension to alphabets of at least three letters is evoked in [11] and implicitly made in [1].

Let $A = \{a_1, a_2, \dots, a_k\}$ be a k -letter alphabet, $k \geq 2$. We construct a sequence $(R_n)_{n \in \mathbb{N}}$ of Rauzy k -uples $R_n = (A_n^{(1)}, A_n^{(2)}, \dots, A_n^{(k)})$ as follows: $R_0 = (a_1, a_2, \dots, a_k)$, R_{n+1} is obtained from R_n by applying one of the Rauzy rules, labelled $1, 2, \dots, k$, with the rule $i \in [1, k]$ defined by

$$A_{n+1}^{(i)} = A_n^{(i)},$$

$$A_{n+1}^{(j)} = A_n^{(i)} A_n^{(j)} \quad \text{for } j \in [1, k] \setminus \{i\}.$$

We impose no restriction on the choice of the rule at each stage. There exists a unique infinite word t such that every prefix of t is a prefix of infinitely many of the $A_n^{(q)}$, $n \in \mathbb{N}$, $q \in [1, k]$. So any Rauzy sequence $(R_n)_{n \in \mathbb{N}}$ defines an infinite word.

Theorem 8. *An infinite word s on the finite alphabet $A = \{a_i; 1 \leq i \leq k\}$ is standard episturmian if and only if it can be obtained by the Rauzy rules for A . Moreover, if $\Delta(s) = a_{i_1} a_{i_2} a_{i_3} \dots$ then the sequence of the applied Rauzy rules is i_1, i_2, i_3, \dots .*

Proof. (*Only if*): Let s be the standard episturmian with directive word $\Delta(s) = x_1 x_2 \dots$ and corresponding palindrome prefixes $u_1 = \varepsilon, u_2, \dots$. For any n define v_n, v'_n and w_n by $u_{n+1} = v_n u_n = u_n \tilde{v}_n, v_n = v'_n x_n, u_n x_n = v'_n w_n$. Then w_n is the unioccurrent palindrome suffix of $u_n x_n$. We have also $u_{n+1} = v_n v_{n-1} \dots v_p u_p$ for any $p \in [1, n]$. Now define k -uples $R_n = (A_n^{(1)}, A_n^{(2)}, \dots, A_n^{(k)})$ by $R_0 = (a_1, a_2, \dots, a_k)$ and, for $i \in [1, k]$,

- (a) $A_n^{(i)} = v_{n-1} v_{n-2} \dots v_1 a_i = u_n a_i$ if a_i does not occur in $x_1 x_2 \dots x_{n-1}$
- (b) $A_n^{(i)} = v_{n-1} v_{n-2} \dots v_p$ with p maximal in $[1, n-1]$ such that $x_p = a_i$ if a_i occurs in $x_1 x_2 \dots x_{n-1}$.

We will show that R_{n+1} is obtained from R_n by applying a Rauzy rule and more precisely the rule i such that $x_n = a_i$.

- (a) Suppose we are in the first case for the definition of $A_n^{(i)}$, that is $A_n^{(i)} = u_n a_i$. Then as a_i does not occur in u_n , we have $u_{n+1} = u_n a_i u_n$ whence $v_n = u_n a_i = A_n^{(i)}$. We have also $A_{n+1}^{(i)} = v_n$ because $x_n = a_i$, so $A_{n+1}^{(i)} = A_n^{(i)}$.
- (b) Suppose we are in the second case for the definition of $A_n^{(i)}$, that is $A_n^{(i)} = v_{n-1} \dots v_p$ with $x_p = a_i$ and $x_{n-1}, \dots, x_{p+1} \neq a_i$. Then we have $u_n = v_{n-1} \dots v_{p+1} v'_p a_i u_p$, hence u_p is the longest palindrome suffix of u_n which is preceded by a_i in u_n . So $a_i u_p a_i = w_n$. It follows that $u_n x_n = v_{n-1} \dots v_{p+1} v'_p w_n$ whence $v_{n-1} \dots v_{p+1} v'_p = v'_n$. As $A_{n+1}^{(i)} = v_n$ we have again $A_{n+1}^{(i)} = A_n^{(i)}$.

So in both cases $A_{n+1}^{(i)} = A_n^{(i)} = v_n$.

Now for $j \neq i$, by the definition of $A_n^{(j)}$ and $A_{n+1}^{(j)}$ we have $A_{n+1}^{(j)} = v_n A_n^{(j)}$ because the occurrences, if any, of a_j in $x_1 x_2 \dots x_{n-1}$ and in $x_1 x_2 \dots x_n$ are the same. Consequently $A_{n+1}^{(j)} = A_n^{(j)} A_n^{(j)}$. So R_{n+1} can be obtained from R_n by applying the Rauzy rule i .

(*If*): It follows from the only if part that any sequence of Rauzy rules can be obtained with a suitable standard episturmian word, and so generates it.

4.4. Morphisms

The theory of Sturmian morphisms [2] on a 2-letter alphabet has a natural extension for any k -letter alphabet $A = \{a_1, \dots, a_k\}$ with $k \geq 2$. Hereafter we sketch some results and proofs, at least for standard episturmian words.

We define endomorphisms ψ_a, θ_{ab} of A^* , $a, b \in A, a \neq b$ by

$$\psi_a(a) = a, \quad \psi_a(x) = ax, \quad x \in A, \quad x \neq a,$$

$$\theta_{ab}(a) = b, \quad \theta_{ab}(b) = a, \quad \theta_{ab}(x) = x, \quad x \in A, \quad x \neq a, b.$$

These morphisms satisfy relations such as $\theta_{ab} \circ \psi_a = \psi_b \circ \theta_{ab}$. Endomorphisms of A^* are naturally extended to A^ω . The following lemmas have easy proofs.

Lemma 3. *Let $u \in A^*$ and $a \in A$. Then*

- (a) $\psi_a(\tilde{u})a$ is the reversal of $\psi_a(u)a$,

- (b) u is a palindrome if and only if $\psi_a(u)a$ is a palindrome,
- (c) u is a right (resp. left) special factor of an infinite word t if and only if $\psi_a(u)a$ is a right (resp. left) special factor of $\psi_a(t)$.

For any $m \in A^* \cup A^\omega$ let us say that $a \in A$ is *separating* for m if for any $x, y \in A$, $xy \in F(m)$ implies $a \in \{x, y\}$.

Lemma 4. (a) *If s is a standard episturmian word with first letter a , then a is separating for s and its factors.*

(b) *If two different letters a, b are separating for $m \in A^* \cup A^\omega$, then m is $(ab)^\omega$ or $(ba)^\omega$ or $m \in F((ab)^\omega)$.*

Theorem 9. *The infinite word $s \in A^\omega$ is standard episturmian if and only if there exist a standard episturmian word t and $a \in A$ such that $s = \psi_a(t)$. Moreover the first letter of s is a , t is unique, and the directive words satisfy $\Delta(s) = a\Delta(t)$.*

Proof. (Only if): If s is standard episturmian its first letter, a say, is separating, so $s = \psi_a(t)$ for some $t \in A^\omega$. By Lemma 3 and easy arguments, t is closed under reversal and has at most one left special factor of each length, and this one is a prefix of t , hence by Theorem 5, t is standard episturmian. Clearly t is unique. At last if $u'_1 = \varepsilon, u'_2, \dots$ are the palindrome prefixes of t and $\Delta(t) = x'_1 x'_2 \dots, x'_n \in A$ then by Lemma 3, $\psi_a(u'_n)a$ is a palindrome prefix of s and even is u_{n+1} from the remark after Theorem 1, with $u_1 = \varepsilon, u_2, \dots$ being the sequence of all palindrome prefixes of s . Moreover $x_{n+1} = x'_n$, so $\Delta(s) = x_1 \Delta(t) = a\Delta(t)$.

(If): Follows from Lemma 3 in the same way. \square

Now let M be the monoid of morphisms generated by the ψ_a and the θ_{ab} , $a, b \in A$. We have

Theorem 10. (a) *If $\varphi \in M$, then for any standard episturmian word $t \in A^\omega$, $\varphi(t)$ is standard episturmian.*

(b) *If φ is an endomorphism of A^* such that for any A -strict standard episturmian word t the infinite word $\varphi(t)$ is also A -strict standard episturmian, then $\varphi \in M$.*

Proof. Part (a) follows by induction from Theorem 9. For part (b) suppose by contradiction that there exists a morphism $\varphi \notin M$ such that $\varphi(t)$ is A -strict standard episturmian whenever t is. Choose φ such that $|\varphi| = \sum_{x \in A} |\varphi(x)|$ is minimal. Let C be the set of the first letter of each $\varphi(x) \neq \varepsilon, x \in A$. For any $x \in A$ such that $\varphi(x) \neq \varepsilon$ let t be an A -strict standard episturmian word with first letter x . Then $\varphi(x)$ is a prefix of $\varphi(t)$, so the first letter of $\varphi(x)$ is separating for $\varphi(t)$ hence for all $\varphi(y) \neq \varepsilon, y \in A$. So if $|C| \geq 3$, three different letters are separating for all non-empty $\varphi(y), y \in A$. By Lemma 4(b) this is impossible unless $|\varphi(x)| \leq 1$ for all $x \in A$. In this case we must have $A = \{\varphi(x); x \in A\}$, so φ is a permutation of A , hence $\varphi \in M$.

If $|C|=2$, say $C = \{a, b\}$, $a \neq b$, then by Lemma 4(b), $|\varphi(x)| \geq 2$, $x \in A$ gives $\varphi(x) \in F((ab)^\omega)$. Also $|\varphi(x)| = 1$ gives $\varphi(x) \in C$. So only a and b occur in $\varphi(x)$, $x \in A$. So $A = C$ and the A -strict standard episturmian words are the standard Sturmian words on A . Then $\varphi(x) \in M$ by the theory of standard Sturmian morphisms [2,4].

Finally if $C = \{a\}$ for some $a \in A$, a is separating for all non-empty $\varphi(x)$, $x \in A$ and we see easily that $\varphi(x) = \psi_a(u_x)$, $u_x \in A^*$, $x \in A$. Hence $\varphi(x) = \psi_a \circ \varphi_1$ where φ_1 is defined by $\varphi_1(x) = u_x$, $x \in A$. Also if $s = \varphi(t)$ with t A -strict standard episturmian, then by Theorem 9, $s = \psi_a(s')$ with s' A -strict standard episturmian. So $\varphi_1(t) = s'$ whence φ_1 satisfies the hypotheses for φ . By the minimality of $|\varphi|$ we have either $\varphi_1 \in M$, whence $\varphi \in M$, or $|\varphi_1| \geq |\varphi|$ whence $\varphi(x) \in a^*$ for all $x \in A$, which is impossible as $|A| \geq 2$. \square

The next theorem is a stronger version of part (b) of Theorem 10.

Theorem 11. *Let $\varphi \in \text{Endo}(A^*)$ be such that $\varphi(t) = s$ for some A -strict standard episturmian words $s, t \in A^\omega$. Then $\varphi \in M$.*

Proof (sketch). If this is false then there exist φ, s, t with $|\varphi|$ minimal satisfying the hypotheses and such that $\varphi \notin M$.

Let x be the first letter of t . If $\varphi(x) = \varepsilon$, we have $s = \varphi(t')$ with $t' = \psi_x^{-1}(t)$. If y is the first letter of t' , either $\varphi(y) \neq \varepsilon$ and we consider t' instead of t , or $\varphi(y) = \varepsilon$ and we repeat the argument. So, without loss of generality, we assume $\varphi(x) \neq \varepsilon$. Let a be the first letter of $\varphi(x)$, hence of s . Put $s_1 = \psi_a^{-1}(s)$. If all non-empty $\varphi(z)$, $z \in A$ have a for their first letter we can write $\varphi(z) = \psi_a(u_z)$, $u_z \in A^*$, $z \in A$. Let $\varphi_1 : A^* \rightarrow A^*$ be given by $\varphi_1(z) = u_z$ for all $z \in A$. Then $\varphi = \psi_a \circ \varphi_1$ and $s_1 = \varphi_1(t)$. Clearly $\varphi_1 \notin M$ and $|\varphi_1| \leq |\varphi|$ so, by the minimality of $|\varphi|$, $|\varphi_1| = |\varphi|$ whence $\varphi(z) \in a^*$ for all $z \in A$. This is impossible.

So at least two different letters are prefixes of the $\varphi(z)$, $z \in A$. It can be shown then that all $\varphi(z)$, $z \in A$ are palindromes. Let $\varphi' = \varphi \circ \psi_x$ and $t' = \psi_x^{-1}(t)$. Then $s = \varphi'(t')$. As all the $\varphi'(z)$, $z \in A$ have the same first letter, a , and with $s_1 = \psi_a^{-1}(s)$ as previously we have $s_1 = \varphi_1(t')$ with φ_1 being given by $\varphi' = \psi_a \circ \varphi_1$. In the particular case where $|\varphi(x)| = 1$, i.e. $\varphi(x) = a$ it is easy to see that $|\varphi_1| \leq |\varphi|$, whence by the minimality of $|\varphi|$, $\varphi_1 \in M$ or $|\varphi_1| = |\varphi|$ and then the restriction of φ_1 to A is a permutation of A . Hence $\varphi_1 \in M$, whence $\varphi \in M$. If $|\varphi(x)| > 1$, and if b is the first letter of s_1 we repeat the argument with $s_2 = \psi_b^{-1}(s_1)$, φ_2 given by $\varphi_1 = \psi_b \circ \varphi_2$ and $s_2 = \varphi_2(t')$. Continuing this way as far as possible and with a careful analysis of the process, we arrive to some φ_n such that $|\varphi_n| \leq |\varphi|$ and we conclude that $\varphi \in M$. \square

From Theorem 11 we deduce a result for standard episturmian words which are fixed points of morphisms.

Theorem 12. *Let s be a strict standard episturmian word. Then s is the fixed point of a morphism if and only if its directive word $\Delta(s)$ is periodic.*

Proof. (If): Let $\Delta(s) = (x_1 x_2 \cdots x_n)^\omega$. It follows from Theorem 9 that $s = (\psi_{x_1} \circ \psi_{x_2} \circ \cdots \circ \psi_{x_n})(s)$.

(Only if): Without loss of generality, assume s is A -strict. If $s = \varphi(s)$ then, by Theorem 11, $\varphi \in M$. So $\varphi = \psi_{y_1} \circ \psi_{y_2} \circ \cdots \circ \psi_{y_n} \circ \sigma$, $y_i \in A$, with σ a permutation of A . It follows by Theorem 9 that $\Delta(s) = y_1 y_2 \cdots y_n \sigma(\Delta(s))$ whence, easily, $\Delta(s)$ is periodic. \square

4.5. Other problems

Let us mention three among the many problems about episturmian words.

First, do episturmian words satisfy a “balance” condition which extends the one for Sturmian words?

Another problem is: is there a number theoretical definition of episturmian words on A recalling the one for Sturmian words and is there some kind of continued fraction expansion that could play a role there as for Sturmian words? Some particular cases have been examined. In [10] the infinite word on $\{a, b, c\}$ which is the fixed point of the morphism φ given by $\varphi(a) = ab$, $\varphi(b) = ac$, $\varphi(c) = a$, that is the standard episturmian word with directive word $(abc)^\omega$, is studied in this spirit. In [1] any A -strict episturmian word on $\{a, b, c\}$ is characterized by partitioning the circle of length 1 into 3 intervals of lengths α , β , γ where α , β , γ are the frequencies of a, b, c in the infinite word, and performing a transformation on those intervals.

Lastly, it has been shown (Corollary 2) that any factor w of an episturmian word has exactly $|w| + 1$ palindrome factors. However, the converse is false: if in the Fibonacci word $abaab \cdots$ we replace a by aa and b by bb , we get $aabbaaaabb \cdots$ whose any factor of length n has exactly $n + 1$ palindrome factors. So the problem of characterizing such words is open.

References

- [1] P. Arnoux, G. Rauzy, Représentation géométrique des suites de complexité $2n + 1$, Bull. Soc. Math. France 119 (1991) 199–215.
- [2] J. Berstel, P. Seebold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Preliminary Version (February 16, 1999), Institut Gaspard Monge, Univ. Marne-la-Vallée (Chapter 2).
- [3] D. Crisp, W. Moran, A. Pollington, P. Shiue, Substitution invariant cutting sequences, J. Théorie Nombres Bordeaux 5 (1993) 123–138.
- [4] A. de Luca, On standard Sturmian morphisms, Theoret. Comput. Sci. 178 (1–2) (1997) 205–224.
- [5] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, Theoret. Comput. Sci. 183 (1997) 45–82.
- [6] X. Droubay, G. Pirillo, Palindromes and Sturmian words, Theoret. Comput. Sci. 223 (1999) 73–85.
- [7] J. Justin, G. Pirillo, Decimations and Sturmian words, Inform. Théor. Appl. 31 (1997) 271–290.
- [8] F. Mignosi, P. Seebold, Morphismes Sturmiens et règles de Rauzy, J. Théor. Nombres Bordeaux 5 (1993) 221–233.
- [9] G. Rauzy, Suites à termes dans un alphabet fini, Sém. Théor. Nombres Bordeaux 25 (1982–1983) 2501–2516.
- [10] G. Rauzy, Nombres algébriques et substitutions, Bull. Soc. Math. France 110 (1982) 147–178.
- [11] G. Rauzy, Mots infinis en arithmétique, in: M. Nivat, D. Perrin (Eds.), Automata on Infinite Words, Lecture Notes in Computer Science, Vol. 192, 1985, pp. 167–171.
- [12] S. Tabachnikov, Billiards, Panoramas et Synthèses 1, Soc. Math. France, 1995.



Fractional powers in Sturmian words

Jacques Justin^{a,*}, Giuseppe Pirillo^{b,c}

^a*LIAFA, 2 place Jussieu, F-75251 Paris Cedex 05*, France*

^b*IAMI CNR, Viale Morgagni 67/A, 50134 FIRENZE, Italy*

^c*Institut Gaspard Monge, IFI Université de Marne-la-Vallée, 2 rue de la Butte Verte,
93160 NOISY-LE-GRAND, France*

Received January 1999; revised April 1999

Communicated by D. Perrin

Résumé

Etant donné un mot infini Sturmian s , nous calculons la fonction $L(m)$ qui donne la longueur du plus long facteur de s ayant pour période m . L'expression de $L(m)$ fait intervenir le développement en fraction continue de l'irrationnel α associé à s . © 2001 Elsevier Science B.V. All rights reserved.

Abstract

Given an infinite Sturmian word s , we calculate the function $L(m)$ which gives the length of the longest factor of s having period m . The expression of $L(m)$ makes use of the continued fraction of the irrational α associated with s . © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Sturmian word; Continued fraction; 3-distance theorem; Periodic factor

1. Introduction

The study of the existence of factors of the form $uu \cdots u = u^n$ in long words over a finite alphabet has been initiated for a long time and is the subject of many publications. It is a part of the theory of (un)avoidable patterns in words. The first nontrivial result seems to be due to Thue [13]. Another interesting result of Karhumäki [3] says that the Fibonacci word, which trivially contains 3 th-powers, does not contain 4 th-powers. This result involving a particular Sturmian word was later developed and extended by several authors.

* Correspondence address. J. Justin, 19 rue de Bagneux 92330 SCEAUX France.

E-mail addresses: justin@liafa.jussieu.fr (J. Justin), pirillo@udini.math.unifi.it (G. Pirillo).

In particular, in [6] the Sturmian words where the exponent of the powers are bounded are characterized, and in [7] it is proved that the exponents of fractional powers in the Fibonacci word have least upper bound $(5 + \sqrt{5})/2$ (a periodic word w with period m is also called a fractional power of its prefix u of length m , with exponent $e = |w|/m$, and is written $w = u^e$).

The fractional powers with exponent $e \geq 2$ in the Fibonacci word are described in [9].

Recently, a generalization of these results has been given in the case of Sturmian words which are fixed points of morphisms [14]. The existence of a least upper bound which is also a point of accumulation for the exponents is proved and its value given.

In the present paper (whose results were announced in [10]), given any Sturmian word s , we compute the maximal length $L(m)$ of a periodic factor of period m , at least in the more interesting case $L(m) \geq 2$. To this aim we give and use a combinatorial presentation of the well-known three distance theorem (Theorem 2). The method, inspired by [6], makes use of a bijection [4] between the factors of length m and a partition of the circle into $m + 1$ intervals. The expressions for $L(m)$ are obtained in terms of denominators of the convergents in the representation of the “slope” of s by a continued fraction (Theorem 4).

As this result gives a precise expression for the maximal exponent $L(m)/m$ in function of m it allows to recover all the aforementioned results. This will be precised in Section 4.2 where some further properties of maximal fractional powers and examples are also given.

2. Preliminaries

2.1. Sturmian words

Let $A = \{0, 1\}$ be a two letter alphabet. *Words* over A are finite (possibly empty) sequences $u = u(1)u(2) \cdots u(m)$, $m \in \mathbb{N}_+$, (the set of positive integers), $u(i) \in A$. The *length* of u is $|u| = m$. Similarly, *infinite words* are infinite sequences $s = s(1)(2) \cdots s(i) \cdots$, $i \in \mathbb{N}_+$, $s(i) \in A$. A word u is a *factor* of a finite or infinite word t if $t = t'ut''$, where t' is a word and t'' is a finite or infinite word according to the case. When t' is the empty word, u is a *prefix* of t . Also if $u = t(i)t(i+1) \cdots t(j)$ we write $u = t(i, j)$. The set of the factors of t is denoted by $F(t)$.

Let $u = u(1)u(2) \cdots u(n)$ be a word of length n and let $m \in \mathbb{N}_+$, $m \leq n$. Then u is periodic with period m if $u(i+m) = u(i)$ for $1 \leq i \leq n-m$. In this case we can equivalently say that u is a *fractional power* of $v = u(1)u(2) \cdots u(m)$, with *exponent* $e = n/m$, and write $u = v^e$. When e is an integer, we find the classical definition of powers.

Let $\alpha \in [0, 1]$ be an irrational number. Then the *standard Sturmian word of slope α* (also called the *characteristic sequence of α*) is the infinite word s given by

$$s(n) = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor, \quad n \in \mathbb{N}_+$$

or equivalently by

$$\begin{aligned} s(n) &= 0 && \text{if } \{n\alpha\} \in [0, 1 - \alpha[\\ s(n) &= 1 && \text{if } \{n\alpha\} \in [1 - \alpha, 1[\end{aligned}$$

where as usual $\{x\} = x - [x]$ denotes the fractional part of the real x .

An infinite word t is *Sturmian* if $F(t) = F(s)$ for some standard Sturmian word s . Among the many equivalent algebraic or combinatorial definitions of Sturmian words, let us recall the following one which is of concern for our purpose. An infinite word is Sturmian if and only if for all $n \in \mathbb{N}$ it has exactly $n + 1$ factors of length n .

Hereafter, as we are interested only in factors, we limit ourselves to the consideration of standard Sturmian words.

2.2. Simple continued fractions [2, Chap. X]

Let $\alpha \in [0, 1]$ be an irrational number and $[a_0, a_1, \dots]$ be its representation by a simple continued fraction. As $\alpha \in [0, 1]$, we have $a_0 = 0$. Also a_1, a_2, \dots are positive integers. The convergents are fractions p_n/q_n , $n \in \mathbb{N}$ with p_n, q_n defined by the recurrence relations

$$\begin{aligned} p_0 &= a_0, & p_1 &= a_1 a_0 + 1, & p_{n+1} &= a_{n+1} p_n + p_{n-1}, & (n \geq 1) \\ q_0 &= 1, & q_1 &= a_1, & q_{n+1} &= a_{n+1} q_n + q_{n-1}, & (n \geq 1) \end{aligned}$$

As customary we extend the general formulas to $n = 0$ by putting $p_{-1} = 1$, $q_{-1} = 0$. By Theorem 164 of [2] we have, for n even for instance

$$1 > q_n \alpha - p_n > q_{n+2} \alpha - p_{n+2} > 0 \tag{1}$$

whence, as $q_{n+2} = q_n + a_{n+2} q_{n+1}$ and $p_{n+2} = p_n + a_{n+2} p_{n+1}$,

$$\begin{aligned} 1 > q_n \alpha - p_n &> (q_n + q_{n+1}) \alpha - (p_n + p_{n+1}) > \dots > ((q_n + (a_{n+2} - 1) q_{n+1}) \alpha \\ &- (p_n + (a_{n+2} - 1) p_{n+1})) > q_{n+2} \alpha - p_{n+2} > 0 \end{aligned} \tag{2}$$

It also follows from (1) and from Theorem 164 of [2]

$$q_{n+2} \alpha - p_{n+2} + (-q_{n+1} \alpha + p_{n+1}) > -q_{n+1} \alpha + p_{n+1} > 0 > q_{n+2} \alpha - p_{n+2},$$

that is

$$\begin{aligned} (q_n + (a_{n+2} - 1) q_{n+1}) \alpha - (p_n + (a_{n+2} - 1) p_{n+1}) &> -q_{n+1} \alpha + p_{n+1} \\ &> (q_n + (a_{n+2} + 1) q_{n+1}) \alpha - (p_n + (a_{n+2} + 1) p_{n+1}) > 0. \end{aligned} \tag{3}$$

Using (2) and (3) and observing that positive members belong to $]0, 1[$ we get, for even $n \geq 2$

$$\begin{aligned} 1 > \{q_n \alpha\} > \{(q_n + q_{n+1}) \alpha\} > \{(q_n + 2q_{n+1}) \alpha\} \\ &> \dots > \{(q_n + (a_{n+2} - 1) q_{n+1}) \alpha\} > \{-q_{n+1} \alpha\} > 0 \end{aligned} \tag{4}$$

In the same way, for odd $n \geq 1$ we have

$$\begin{aligned} 1 &> \{-q_n \alpha\} > \{-(q_n + q_{n+1}) \alpha\} > \{-(q_n + 2q_{n+1}) \alpha\} \\ &> \cdots > \{-(q_n + (a_{n+2} - 1)q_{n+1}) \alpha\} > \{q_{n+1} \alpha\} > 0. \end{aligned} \quad (5)$$

3. Factors of Sturmian words

Let s be a Sturmian word. Then for any $m \in \mathbb{N}_+$ there exists a well-known bijection between $F_m(s) = A^m \cap F(s)$ and some partition of the circle of length 1 into $m + 1$ intervals [8, p. 26, 4, 6]. As in [6] instead of considering the directed circle of length 1 we will consider the interval $[0, 1[$ of \mathbb{R} with some conventions.

(a) Let $\lambda, \mu \in [0, 1[$, then the (*generalized*) interval $[\lambda, \mu[$ is the customary interval $[\lambda, \mu[$ if $0 \leq \lambda < \mu < 1$ and is $[\lambda, 1[\cup [0, \mu[$ if $0 \leq \mu < \lambda < 1$. Also $[\lambda, \lambda[$ is the empty set. Accordingly, the *amplitude* $|[\lambda, \mu[|$ of $[\lambda, \mu[$ is $\{\mu - \lambda\}$ that is $\mu - \lambda$ if $\lambda \leq \mu$ and $\mu + 1 - \lambda$ if $\mu < \lambda$.

(b) If $\mu, \lambda, \nu \in [0, 1[$ and $\mu \in [\lambda, \nu[$ we write $\lambda \prec \mu \prec \nu$. This means that when starting from λ in the positive direction on the circle of length 1 we meet μ before ν . More generally $\lambda \prec \mu_1 \prec \mu_2 \prec \cdots \prec \mu_n \prec \nu$ means that $[\lambda, \mu_1[, \dots, [\mu_n, \nu[$ are disjoint and $|[\lambda, \nu[| = |[\lambda, \mu_1[| + \cdots + |[\mu_n, \nu[|$. We will often make use of evident properties of “translations modulo 1”. For instance $\mu \in [\lambda, \nu[\Leftrightarrow \{\mu + \rho\} \in [[\lambda + \rho], \{\nu + \rho\}[$ for $\lambda, \mu, \nu \in [0, 1[$ and any real ρ .

Now, considering the standard Sturmian word $s = s_\alpha$ with slope α and $m \in \mathbb{N}_+$, denote by c_i , $0 \leq i \leq m$ the $\{-j\alpha\}$, $0 \leq j \leq m$ in increasing order: $0 = c_0 < c_1 < \cdots < c_m < 1$ and put $I_i = [c_i, c_{i+1}[$, $0 \leq i < m$, and $I_m = [c_m, 0[$. We recall

Theorem 1 (Knuth [4], Mignosi [6]). *There exists a bijection between $F_m(s)$ and the set of the I_k . More precisely the value of $s(n, n + m - 1)$, $n \in \mathbb{N}_+$ is determined by the I_k containing $\{n\alpha\}$.*

Consequently we are led to a study of the I_k . Theorem 2 hereafter is essentially the 3-distance theorem [11, 12, 1], but with a presentation suited to our purpose.

Let $[a_0, a_1, a_2, \dots]$ be the simple continued fraction representing α with notations for the convergents as in the Preliminaries (here $a_0 = 0$ as $0 < \alpha < 1$). For $m \in \mathbb{N}_+$ there exists a unique $i \in \mathbb{N}$ such that $q_{i-1} + q_i \leq m < q_i + q_{i+1}$. We define $b, r \in \mathbb{N}$ by putting

$$m = q_{i-1} + bq_i + r, \quad 1 \leq b \leq a_{i+1}, \quad 0 \leq r < q_i \quad (6)$$

Define a circular permutation φ_m of $\{0, 1, \dots, m\}$ by

if $0 \leq x \leq m - q_i$ then $\varphi_m(x) = x + q_i$

if $m - q_i < x < m - r$ then $\varphi_m(x) = x + q_i + r - m$

if $m - r \leq x \leq m$ then $\varphi_m(x) = x + r - m$.

Putting $\psi_m = \varphi_m^{-1}$ we also have:

if $0 \leq x \leq r$ then $\psi_m(x) = x + m - r$

if $r < x < q_i$ then $\psi_m(x) = x + m - r - q_i$

if $q_i \leq x \leq m$ then $\psi_m(x) = x - q_i$.

Lastly for any $n, n' \in \mathbb{Z}$ we write for shortness $\langle n, n' \rangle = [\{-n\alpha\}, \{-n'\alpha\}]$.

All those notations and conventions will be valuable throughout this paper.

Theorem 2. *Let $m \in \mathbb{N}_+$, with notations as above. Then the corresponding I_k are given by*

$$I_k = \langle \varphi_m^k(0), \varphi_m^{k+1}(0) \rangle \text{ if } i \text{ is odd}$$

$$\text{(resp. } I_k = \langle \psi_m^k(0), \psi_m^{k+1}(0) \rangle \text{ if } i \text{ is even)}$$

The amplitudes of these intervals have at most three values, $l_1, l_2 = l_1 + l_3, l_3$ for, respectively, $m - q_i - 1, q_i - r - 1$ and $r + 1$ intervals. In particular there is no interval of amplitude l_2 if $r = q_i - 1$. These amplitudes are $l_1 = \{-q_i\alpha\}, l_2 = \{(m - q_i - r)\alpha\}, l_3 = \{(m - r)\alpha\}$ if i is odd (resp. $l_1 = \{q_i\alpha\}, l_2 = \{-(m - q_i - r)\alpha\}, l_3 = \{-(m - r)\alpha\}$ if i is even).

Lastly, $l_1 < l_3$ if $b < a_{i+1}$, i.e. $m < q_{i+1}$ and $l_3 < l_1$ if $b = a_{i+1}$, i.e. $q_{i+1} \leq m$.

Proof. We proceed by induction on m . When $m = 1$, as $q_{-1} + q_0 = 1$ we have $i = 0, b = 1, \psi_1(0) = 1, \psi_1(1) = 0, I_0 = \{0, -\alpha\}, I_1 = \{-\alpha, 0\}$ hence $I_0 = \langle \psi_1^0(0), \psi_1(0) \rangle, I_1 = \langle \psi_1(0), \psi_1^2(0) \rangle, l_1 = \{\alpha\}, l_3 = \{-\alpha\}$. At last if $b < a_1$ we have $a_1 > 1$ whence $\alpha < 1/2$ whence $l_1 < l_3$, and if $b = a_1$, then $a_1 = 1$ whence $\alpha > 1/2$ whence $l_3 < l_1$. So the theorem is verified for $m = 1$. Now supposing the theorem true for some $m \in \mathbb{N}_+$, let us prove it for $m' = m + 1$. With the same notations as previously, let $m = q_{i-1} + bq_i + r$. Suppose for instance i is odd (the case i even is similar). For simplicity of notation we write now φ instead of φ_m . The intervals of $[0, 1[$ corresponding to m' are the same as those, $I_k, k \in \{0, 1, \dots, m\}$, corresponding to m , except that the one containing $\{-m'\alpha\}$ is divided in two parts. As $\varphi(m - q_i) = m$ and $\varphi(m) = r$ we have

$$\{-m\alpha\} \in [(-m - q_i)\alpha, -m\alpha[\cup [-m\alpha, -r\alpha[= \langle m - q_i, r \rangle.$$

By translation modulo 1 of value $-\alpha$ we get

$$\{-m'\alpha\} \in J = \langle m' - q_i, r' \rangle$$

with $r' = r + 1$.

Consider three cases:

(a) *Case 1:* $r < q_i - 1$. Then $m' = q_{i-1} + bq_i + r', r' < q_i$. In this case, $\varphi(m' - q_i) = r'$ whence J is one of the I_k, I_h say, and has amplitude l_2 . Put $J_1 = \langle m' - q_i, m' \rangle, J_2 = \langle m', r' \rangle$ and define a permutation φ' of $\{0, 1, \dots, m'\}$ by $\varphi'(x) = \varphi(x)$ for $x \neq m' - q_i, \varphi'(m' - q_i) = m', \varphi'(m') = r'$. An easy verification shows that φ' is $\varphi_{m'}$, and that the intervals, I'_t say, corresponding to m' are $\langle \varphi_{m'}^t(0), \varphi_{m'}^{t+1}(0) \rangle, 0 \leq t \leq m'$ and are related to the I_k by $I'_t = I_t, 0 \leq t < h, I'_h = J_1, I'_{h+1} = J_2, I'_t = I_{t-1}, h + 1 < t \leq m'$.

As J_1 and J_2 have amplitude l_1 and l_3 , respectively, we gain one interval of each amplitude l_1 , l_3 and loss one interval of amplitude l_2 when passing from m to m' . Lastly, as m and m' have the same value for b , the condition $l_1 < l_3 \Leftrightarrow b < a_{i+1}$ is satisfied for m' .

(b) *Case 2:* $r = q_i - 1$ and $b < a_{i+1}$. We have now $m' = q_{i-1} + b'q_i \leq q_{i+1} < q_i + q_{i+1}$, with $b' = b + 1$. In this case $r' = r + 1 = q_i$ and $\varphi(m' - q_i) = \varphi(q_{i-1} + bq_i) = 0$ and $\varphi(0) = r'$. So $J = J_1 \cup J_2$ where $J_1 = \langle m' - q_i, 0 \rangle$ and $J_2 = \langle 0, q_i \rangle$ are two intervals I_k of amplitude, respectively, l_3 and l_1 . As $b < a_{i+1}$ and the theorem is true for m , we have $l_1 < l_3$. As $[\{-(m' - q_i)\alpha\}, \{-m'\alpha\}] = \langle m' - q_i, m' \rangle$ has amplitude l_1 it follows $\{-m'\alpha\} \in J_1$. Define a permutation φ' of $\{0, 1, \dots, m'\}$ by $\varphi'(x) = \varphi(x)$ for $x \neq m' - q_i$, $\varphi'(m' - q_i) = m'$, $\varphi'(m') = 0$. Then it is easily verified that $\varphi' = \varphi_{m+1} = \varphi_{m'}$, and that the intervals I'_i corresponding to m' are the $\langle \varphi'_{m'}(0), \varphi'^{t+1}_{m'}(0) \rangle$.

The amplitudes of the I'_i are $l'_1 = l_1$, $l'_2 = l_3 = l'_1 + l'_3$, $l'_3 = l_3 - l_1 = \{(q_{i-1} + b'q_i)\alpha\}$. These amplitudes and the number of occurrences of each of them satisfy the theorem as it is easily verified. It remains to compare l'_1 and l'_3 . It follows from formula (4) in the Preliminaries that if $b' < a_{i+1}$ then $\{(q_{i-1} + b'q_i)\alpha\} > \{-q_i\alpha\}$, that is $l'_3 > l'_1$. Also if $b' = a_{i+1}$ then by formula (5) $l'_3 = \{q_{i+1}\alpha\} < \{-q_i\alpha\} = l'_1$. So this part of the theorem is verified for m' .

(c) *Case 3:* $r = q_i - 1$ and $b = a_{i+1}$. Then $m' = q_{i-1} + a_{i+1}q_i + q_i = q_i + q_{i+1}$, whence $q_{i'-1} + q_{i'} \leq m' < q_{i'} + q_{i'+1}$ with $i' = i + 1$. As in the second case, we have $\{-m'\alpha\} \in J = J_1 \cup J_2$ with J_1 and J_2 being intervals I_k of respective amplitude l_3 and l_1 . But now $l_3 < l_1$ whence $\{-m'\alpha\} \in J_2 = \langle 0, q_i \rangle$. We define a permutation φ' by $\varphi'(x) = \varphi^{-1}(x)$ for $0 < x \leq m$, $\varphi'(0) = m'$, $\varphi'(m') = q_i$. We easily verify that φ' is $\psi_{m'}$, and that the intervals I'_j , $j = 0, 1, \dots, m'$, corresponding to m' are the $\langle \psi_{m'}^j(0), \psi_{m'}^{j+1}(0) \rangle$.

The amplitudes of the I'_j are $l'_1 = l_3$, $l'_2 = l_1 = l'_1 + l'_3$, $l'_3 = l_1 - l_3$ and their numbers of occurrences are, respectively, $q_i + 1$, $q_{i+1} - 1$ and 1, that is $m' - q_{i'-1} + 1$, $q_{i'} - 1$ and $0 + 1$ as stated for m' by the theorem. It remains to compare l'_1 and l'_3 . As $m' = q_i + q_{i+1}$ we have $l'_1 = \{q_{i+1}\alpha\}$ and $l'_3 = \{-(q_i + q_{i+1})\alpha\}$. Also $q_{i+2} = q_i + a_{i+2}q_{i+1}$. So if $a_{i+2} > 1$ it follows from formula (5) that $l'_3 > l'_1$, while if $a_{i+2} = 1$, by formula (4), $l'_1 < l'_3$. These two cases correspond respectively to $m' < q_{i+2}$ and $m' = q_{i+2}$. \square

This theorem gives a comparison between l_1 and l_3 , however in view of Section 4, a more precise comparison is wanted.

Theorem 3. (i) *With notations as previously,*

$$(a_{i+1} - b)l_1 < l_3 < (a_{i+1} + 1 - b)l_1. \quad (7)$$

(ii) *If $b = a_{i+1}$ (or equivalently $m \geq q_{i+1}$) then*

$$a_{i+2}l_3 < l_1 < (a_{i+2} + 1)l_3. \quad (8)$$

Proof. With i odd for instance, we have by the Preliminaries, formula (2), $\{q_{i-1}\alpha\} > h\{-q_i\alpha\} > 0$ for $1 \leq h \leq a_{i+1}$. It follows $\{q_{i-1}\alpha\} - b\{-q_i\alpha\} = \{(q_{i-1} + bq_i)\alpha\}$ whence

$\{q_{i-1}\alpha\} = l_3 + bl_1$. By the Preliminaries, formulas (2, 3), we have $a_{i+1}l_1 < \{q_{i-1}\alpha\} < (a_{i+1} + 1)l_1$ and (7) follows.

Now if $b = a_{i+1}$ we have $q_{i-1} + bq_i = q_{i+1}$ whence $l_3 = \{q_{i+1}\alpha\}$. Using the dual for n odd of formulas (2, 3) of Preliminaries we get (8). \square

4. Periodic factors of Sturmian words

4.1. Calculation of $L(m)$

Given the standard Sturmian word s of slope α we define, for $m \in \mathbb{N}_+$, $L(m)$ by $L(m) = \text{Sup}\{|w|; w \in F(s) \text{ and } w \text{ has period } m\}$.

We will compute $L(m)$, but as the case $L(m) < 2m$ leads to complicated formulas, we limit ourselves here to the case $L(m) \geq 2m$.

The notations are those of Section 3, in particular (6) holds and the I_k are those of Theorem 2. For simplicity, we write φ, ψ instead of φ_m, ψ_m .

Part A of Theorem 4 gives the cases where $L(m) \geq 2m$ and Part B gives the value of $L(m)$. There are five cases (and even seven as cases (iii) and (v) contain a disjunction in their formulation). This is due to the intervention of some relative values of the parameters. For instance when $r = q_i - 1$ there is no interval I_k of size l_2 . However only cases (i) and (ii) are general, the other ones imply $i \in \{0, 1, 2\}$ hence concern small values of m . An acceptable assumption $\alpha < 1/2$ or $\alpha > 1/2$ could also eliminate case (iv) or (v) but here we have preferred uniformity.

In order to calculate $L(m)$ we need a lemma.

Lemma 1. *Let λ, μ, v_0, ρ be real numbers and $h \in \mathbb{N}_+$ such that putting $v_j = v_0 + j\rho$, $0 \leq j \leq h$ we have $\lambda, \mu \in [0, 1[$ and λ, μ and all $\{v_j\}$ are different and $\{v_j\} \in [\lambda, \mu[$ for $0 \leq j \leq h$. Then*

- (a) *if $\{\mu - \lambda\} + \{\rho\} \leq 1$ then $\lambda < v_0 < v_1 < \dots < v_h < \mu$ and $h\{\rho\} < \{\mu - \lambda\}$*
- (b) *if $\{\mu - \lambda\} + \{-\rho\} \leq 1$ then $\lambda < v_h < v_{h-1} < \dots < v_0 < \mu$ and $h\{-\rho\} < \{\mu - \lambda\}$.*

Proof. Consider case (a) for instance. Without loss of generality we can suppose $\lambda = 0$ because this amounts to replacing $[\lambda, \mu[$ by $[0, \{\mu - \lambda\}[$ and $\{v_j\}$ by $\{v_j - \lambda\}$. We have then to show that $0 < \{v_0\} < \{v_1\} < \dots < \{v_h\} < \mu < 1$.

If this is false we have, for some $j \in [1, h]$, $\{v_{j-1}\} > \{v_j\}$, that is $\{v_0 + (j-1)\rho\} > \{v_0 + j\rho\}$. It follows $\{v_0 + j\rho\} = \{v_0 + (j-1)\rho\} + \{\rho\} - 1$ whence $\{v_0 + (j-1)\rho\} + \{\rho\} \geq 1$. As $\{v_0 + (j-1)\rho\} < \mu$ we get $\mu + \rho > 1$ contrarily to the hypothesis. \square

Theorem 4. *Let s be a standard Sturmian word with slope $\alpha = [a_0 = 0, a_1, a_2, \dots]$. For any $m \in \mathbb{N}_+$ define integers i, r, b by*

$$q_{i-1} + q_i \leq m = q_{i-1} + bq_i + r < q_i + q_{i+1},$$

with $1 \leq b \leq a_{i+1}$, $0 \leq r < q_i$, where $q_{-1} = 0$ and $q_0 = 1$, q_1, q_2, \dots are the denominators of the convergents of α .

Let $L(m)$ be the maximal length of the factors of s having period m . Then

(A) We have $L(m) \geq 2m$ if and only if one of the following conditions is satisfied

- (i) $0 = r \neq q_i - q_{i-1}$ and $q_i > 1$
- (ii) $0 < r = q_i - q_{i-1}$ and $q_{i-1} > 1$ and $2b \leq a_{i+1}$
- (iii) $0 = r \neq q_i - q_{i-1}$ and $q_i = 1$ and either $b = a_{i+1}$ or $2b \leq a_{i+1}$
- (iv) $0 < r = q_i - q_{i-1}$ and $q_{i-1} = 1$ and $2b < a_{i+1}$
- (v) $0 = r = q_i - q_{i-1}$ and either $b = a_{i+1}$ or $2b < a_{i+1}$.

(B). When $L(m) \geq 2m$ its value is as follows where the cases are those of Part A.

- (i) $L(m) = 2m + q_i - 2$ if $b < a_{i+1}$ and $L(m) = (2 + a_{i+2})m + q_i - 2 = q_{i+2} + 2q_{i+1} - 2$ if $b = a_{i+1}$
- (ii) $L(m) = q_{i+1} + 2q_i - 2$
- (iii) $L(m) = q_2 + 2q_1 - 2$ if $b = a_1$ and $L(m) = q_1$ if $2b \leq a_1$
- (iv) $L(m) = q_{i+1} + 2q_i - 2$
- (v) $L(m) = q_3 + 2q_2 - 2$ if $b = a_2$ and $L(m) = q_2$ if $2b < a_2$.

Proof of Part A. Clearly, $L(m) \geq 2m$ if and only if $s(n, n+m-1) = s(n+m, n+2m-1)$ for some $n \in \mathbb{N}_+$, that is if $\{n\alpha\}$ and $\{(n+m)\alpha\}$ belong to the same interval I_k . As the sequence $\{j\alpha\}$, $j \in \mathbb{N}_+$ is dense in $[0, 1[$ this will occur if and only if $\{m\alpha\} < M$ or $\{-m\alpha\} < M$, where $M = \text{Sup}\{|I_k|; 0 \leq k \leq m\}$. As $\{m\alpha\} = |\langle m, 0 \rangle|$ and $\{-m\alpha\} = |\langle 0, m \rangle|$ we will study these two intervals. Both are union of successive I_k . More precisely, with i odd for instance, define t in $[0, m]$ by $m = \varphi^t(0)$. Then $\langle 0, m \rangle = I_0 \cup I_1 \cup \dots \cup I_{t-1}$ and $\langle m, 0 \rangle = I_t \cup I_{t+1} \cup \dots \cup I_m$. Hence $|\langle 0, m \rangle| \geq l_1 + l_3$, unless either $r = q_i - q_{i-1}$, which gives $m = (b+1)q_i$ and $|\langle 0, m \rangle| = (b+1)l_1$, or $r = q_{i-1} = 0$ which gives $i = 0$, $m = b$, $|\langle 0, m \rangle| = ml_1$. In the same way $|\langle m, 0 \rangle| \geq l_1 + l_3$, unless $r = 0$ which gives $|\langle m, 0 \rangle| = l_3$. Also observing that $M = l_2$ if $r < q_i - 1$ and that $M = \text{Sup}\{l_1, l_3\}$ if $r = q_i - 1$, we are led to the cases of the theorem.

If $0 = r \neq q_i - q_{i-1}$ and $q_i > 1$ then $M = l_2 > l_3$, so $L(m) \geq 2$. This is case i).

If $0 < r = q_i - q_{i-1}$ and $q_{i-1} > 1$ then $L(m) \geq 2m$ if and only if $(b+1)l_1 < M = l_2 = l_1 + l_3$. Using formula (7) this gives $2b \leq a_{i+1}$, this is case (ii).

If $0 = r \neq q_i - q_{i-1}$ and $q_i = 1$ then $i = 0$ and $q_{i-1} = 0$ and $M = \text{Sup}\{l_1, l_3\}$. So $L(m) \geq 2m$ if and only if either $l_3 < l_1$, that is $b = a_{i+1}$, or $bl_1 < l_3$, that is $2b \leq a_{i+1}$ (case iii)).

If $0 < r = q_i - q_{i-1}$ and $q_{i-1} = 1$ then in the same way $L(m) \geq 2m$ if and only if $2b < a_{i+1}$ (case iv)).

If $0 = r = q_i - q_{i-1}$ then $i = 1$, $q_i = q_{i-1} = 1$, $M = \text{Sup}\{l_1, l_3\}$. So $L(m) \geq 2m$ if and only if either $l_3 < l_1$, that is $b = a_{i+1}$, or $(b+1)l_1 < l_3$, that is $2b < a_{i+1}$ (case v)).

Proof of Part B. Let $v = s(n, n+L(m)-1)$ be a factor of s of length $L(m)$ and period m . Put $u = s(n, n+m-1)$ and $L(m) = em + z$, $e, z \in \mathbb{N}$ and $0 \leq z < m$. Then $v = u^e w$ with w a prefix of u of length z . Consequently there exists an I_k containing $\{n\alpha\}$, $\{n\alpha + m\alpha\}$, ..., $\{n\alpha + (e-1)m\alpha\}$ but containing neither $\{n\alpha - m\alpha\}$ nor $\{n\alpha + em\alpha\}$ as $u^{e+1} \notin F(s)$.

Now, for any $y \in \mathbb{Z}$, let $S(y) = \{ \{ nx + y\alpha + jm\alpha \}; j = 0, 1, \dots, e - 1 \}$. If $0 \leq y \leq z$, then $s(n+y, n+y+em-1)$ is an e th power of some conjugate of u whence $S(y) \subset I(y)$, where $I(y)$ is one of the I_k . On the contrary $S(-1)$ and $S(z+1)$ are not included in any I_k . If $I(0) = [\lambda, \mu[$, put for simplicity $T(y) = [\{ \lambda + y\alpha \}, \{ \mu + y\alpha \}[$ for any $y \in \mathbb{Z}$. It follows $S(y) \subset T(y)$. We will consider the five cases corresponding to $L(m) \geq 2$. In each of them we will make use of Lemma 1 applied to $S(0)$ and $I(0)$. The verification that the hypotheses of this lemma are satisfied becomes trivial in view of the following remarks.

Let $\Delta = \text{Inf} \{ \{ m\alpha \}, \{ -m\alpha \} \}$ and as previously $M = \text{Sup} \{ |I_k|; 0 \leq k \leq m \}$. Then the hypothesis $\{ \mu - \lambda \} + \{ \rho \} \leq 1$ or $\{ \mu - \lambda \} + \{ -\rho \} \leq 1$ will be satisfied if $M + \Delta \leq 1$, hence in particular if $M < 1/2$. By Theorem 2 we have $(m - q_i + 1)l_1 + (q_i - r - 1)l_2 + (r + 1)l_3 = 1$. If $q_i \neq r + 1$ it follows easily $M = l_2 < 1/2$. If $q_i = r + 1$ then $M = \text{Sup} \{ l_1, l_3 \}$ and $(m - r)l_1 + q_i l_3 = 1$. Suppose firstly that $M = l_3$. Then if $M > 1/2$ we must have $q_i = 1$ whence $r = 0$, $\Delta = ml_1$ and $M + \Delta = 1$. Secondly suppose $M = l_1$. Then if $M > 1/2$ we must have $m - r = 1$ whence $m = q_i$ whence $q_{i-1} = 0$, $i = 0$, $m = q_0 = 1$ and $\Delta = l_3$ whence $M + \Delta = 1$.

The following lemma will simplify the proof of some cases.

Lemma 2. *If $I(0) > l_1$ and $q_i | m$ then $L(m) = L(q_i)$.*

Proof. With $v = s(n, n + L(m) - 1)$ and $I(0) = [\lambda, \mu[$ as previously we have for instance $\lambda \prec \{ nx \} \prec \{ (n + m)\alpha \} \prec \mu$ (the case $\lambda \prec \{ (n + m)\alpha \} \prec \{ nx \} \prec \mu$ is similar).

As $\{ j\alpha; j \in \mathbb{N}_+ \}$ is dense in $[0, 1[$ we can choose n in such a way that $\mu - \{ (n + m)\alpha \}$ is arbitrarily small positive. As $l_1 = \{ -q_i\alpha \} < |I(0)|$ we get that $\{ (n + m - q_i)\alpha \} \in I(0)$ whence $s(n + m - q_i, n + 2m - q_i - 1) = s(n + m, n + 2m - 1) = u$. So u has period q_i . Moreover as $q_i | m$, v also has period q_i whence $L(m) \leq L(q_i)$. As trivially any word with period q_i also has period m , $L(m) \geq L(q_i)$ and the result follows. \square

Now successively consider cases (i) to (v) of Part A.

(i) Case $0 = r \neq q_i - q_{i-1}$ and $q_i > 1$. First suppose i is odd. We have $I(0) = \langle x, \varphi(x) \rangle$ for some $x \in \{ 0, 1, \dots, m \}$. Then $\{ m\alpha \} = |\langle m, 0 \rangle| = l_3$. As $S(0) \subset I(0)$ we have $S(-1) \subset T(-1) = \langle x + 1, \varphi(x) + 1 \rangle$. As this one is not an I_k we get $x = m$ or $\varphi(x) = m$ or $0 \leq x \leq m$ and $\varphi(x) + 1 \neq \varphi(x + 1)$. This gives for $I(0)$ the three possibilities $J_1 = \langle m, 0 \rangle$, $J_2 = \langle m - q_i, m \rangle$, $J_3 = \langle m - 1, q_i - 1 \rangle$. As $|J_1| = l_3$, J_1 is excluded. If $I(0) = J_2$ then $|I(0)| = l_1$ and $T(-1) = \langle m - q_i + 1, m + 1 \rangle$. As $\langle m - q_i + 1, 1 \rangle$ is an I_k of amplitude l_2 , $S(-1) \subset T(-1) \subset \langle m - q_i + 1, 1 \rangle$ and this is a contradiction. Consequently $I(0) = \langle m - 1, q_i - 1 \rangle$ and has amplitude l_2 .

Observing now that for $0 \leq y \leq q_i - 2$, $S(y) \subset T(y) = \langle m - 1 - y, q_i - 1 - y \rangle$ which is an I_k we get $z \geq q_i - 2$. Also $S(q_i - 1) \subset \langle m - q_i, 0 \rangle = \langle m - q_i, m \rangle \cup \langle m, 0 \rangle$. If $S(q_i - 1)$ were included in $\langle m - q_i, m \rangle$ which is an I_k of amplitude l_1 we would have $(e - 1)l_3 < l_1$ whence $el_3 < l_2$ whence s would contain an $(e + 1)$ th power of some conjugate of u , a contradiction. As $S(q_i - 1)$ is not included in $\langle m, 0 \rangle$ which is an I_k of amplitude l_3 , $S(q_i - 1)$ is not included in any I_k . Consequently $z = q_i - 2$.

If we suppose i even a similar argument gives the same conclusion.

Lastly, using Lemma 1, $e - 1 = \lfloor l_2/l_3 \rfloor = 1 + \lfloor l_1/l_3 \rfloor$ whence by formulas (7) and (8) $e = 2$ if $b < a_{i+1}$ and $e = 2 + a_{i+2}$ if $b = a_{i+1}$, whence the values as claimed for $L(m) = em + z$.

(ii) Case $0 < r = q_i - q_{i-1}$ and $q_{i-1} > 1$ and $2b \leq a_{i+1}$. Suppose i is odd for instance. Now $m = (b+1)q_i$ and $\{-m\alpha\} = |\langle 0, m \rangle| = (b+1)l_1$. Writing as previously that $T(-1)$ is not an I_k we get for $I(0)$ the possible values $J_1 = \langle m, r \rangle$, $J_2 = \langle m - q_i, m \rangle$, $J_3 = \langle m - r - 1, q_i - 1 \rangle$. As $|J_2| = l_1$, J_2 is excluded. If $I(0) = J_1$ then $T(-1) = \langle m+1, r+1 \rangle$ has amplitude l_3 . As $\langle bq_i+1, r+1 \rangle$ is an I_k of amplitude l_2 , $S(-1) \subset T(-1) \subset \langle bq_i+1, r+1 \rangle$, a contradiction. Consequently $I(0) = \langle m - r - 1, q_i - 1 \rangle$ of amplitude l_2 .

As $I(0) > l_1$ and $q_i | m$, by Lemma 2, $L(m) = L(q_i)$. Now putting $q_i = m'$, $i - 1 = i'$ we have $m' = q_{i'-1} + b'q_{i'} + r'$ with $b' = a_{i'+1}$, $r' = 0$.

So we are in case (i) if $q_{i'} > 1$ and in case (iii) if $q_{i'} = 1$. In case (i) we get $L(q_i) = q_{i'+2} + 2q_{i'+1} - 2 = q_{i+1} + 2q_i - 2$. In case (iii), noting that $i' = 0$, we get $L(q_i) = q_2 + 2q_1 - 2 = q_{i+1} + 2q_i - 2$.

(iii) Case $0 = r \neq q_i - q_{i-1}$ and $q_i = 1$ and $b = a_{i+1}$ or $2b \leq a_{i+1}$. This implies $q_{i-1} = 0$, $i = 0$, $m = b$, $M = \text{Sup}\{l_1, l_3\}$, $l_3 + ml_1 = 1$, $\{-m\alpha\} = l_3$, $\{m\alpha\} = bl_1$. As i is even, we have $I(0) = \langle x, \psi(x) \rangle$. The possibility for $I(0)$ are now $J_1 = \langle m, m - 1 \rangle$, $J_2 = \langle 0, m \rangle$. Consider two subcases

(a) $b = a_{i+1} = a_1$, whence $l_3 < l_1 = M$. As J_2 has amplitude l_3 , J_2 is excluded. So $I(0) = \langle m, m - 1 \rangle$ of amplitude l_1 . So for $0 \leq y \leq m - 1$, $S(y) \subset T(y) = \langle m - y, m - 1 - y \rangle$ which is an I_k . So $z \geq m - 1$. Also $S(m) \subset T(m) = \langle 0, -1 \rangle \subset \langle 0, m \rangle \cup \langle m, m - 1 \rangle = I_0 \cup I_1$. We have $|I_0| = l_3$ and $|I_1| = l_1$. Clearly, $S(m) \not\subset I_0$. If $S(m) \subset I_1$ we get $S(m) \subset T(m) \cap I_1 = \langle m, -1 \rangle$. As $|\langle m, -1 \rangle| = l_1 - l_3$ we get $(e - 1)l_3 < l_1 - l_3$ whence $el_3 < l_1$ and then s contains an $(e + 1)$ th power of some conjugate of u . As this is impossible $S(m)$ is not included in any I_k whence $z = m - 1 = q_1 - 1$.

As $e - 1 = \lfloor l_1/l_3 \rfloor = a_2$ we get $L(m) = (a_2 + 1)q_1 + q_1 - 1 = q_2 + 2q_1 - 2$. (b) $2b \leq a_{i+1} = a_1$ whence $l_1 < l_3 = M$. As $\{m\alpha\} \geq l_1$ and $\{-m\alpha\} > l_1$ the possibility J_1 for $I(0)$ is excluded, so $I(0) = J_2 = \langle 0, m \rangle$ of amplitude l_3 . As $i = 0$, $q_i = 1$, so by Lemma 2, $L(m) = L(1)$. Also as $2b \leq a_1$, $a_1 > 1$. Hence as it is well known the maximal power of a letter in s is a_1 , so $L(m) = a_1 = q_1$.

(iv) Case $0 < r = q_i - q_{i-1}$ and $q_{i-1} = 1$ and $2b < a_{i+1}$. This implies $i = 1$ or $i = 2$, $m = (b+1)q_i$, $\{m\alpha\} \geq l_1 + l_3$, $\{-m\alpha\} = (b+1)l_1$. This case recalls case (ii) but as $r = q_i - 1$ there is no I_k of amplitude l_2 . We have for $I(0)$ the possibilities $J_1 = \langle m, r \rangle$, $J_2 = \langle m - q_i, m \rangle$ if i is odd (resp. $J_1 = \langle r, m \rangle$, $J_2 = \langle m, m - q_i \rangle$ if i is even) so as $|J_2| = l_1 < l_3$, $I(0) = J_1$ of amplitude l_3 . By Lemma 2 we have then $L(m) = L(q_i)$. Defining $m' = q_i$, $i' = i - 1$, b' as above we are for m'

in case (i) if $q_{i'} > 1$ which implies $i' = 1$

in case (iii) if $q_{i'} = 1 > q_{i'-1}$ which implies $i' = 0$

in case (v) if $q_{i'} = q_{i'-1} = 1$ which implies $i' = 1$.

Using the formulas for these cases with $b' = a_{i'+1}$ we always get $L(m) = L(q_i) = q_{i+1} + 2q_i - 2$.

(v) Case $0 = r = q_i - q_{i-1}$ and either $b = a_{i+1}$ or $2b < a_{i+1}$. This implies $q_i = q_{i-1} = 1$,

$i = 1$, $m = 1 + b$, $\{m\alpha\} = l_3$, $\{-m\alpha\} = (b + 1)l_1$ and the possibilities for $I(0)$ are $J_1 = \langle m, 0 \rangle$ of amplitude l_3 , $J_2 = \langle m - 1, m \rangle$ of amplitude l_1 . We have two subcases

(a) $b = a_{i+1}$. Then $l_3 < l_1$ whence $I(0) = J_2 = \langle m - 1, m \rangle$, also $m = q_{i+1}$. Reasoning as in subcase a) of case (iii) we obtain $L(m) = (a_{i+2} + 1)q_{i+1} + q_{i+1} - 1 = q_3 + 2q_2 - 2$.

(b) $2b < a_{i+1}$. Then $l_1 < l_3$ whence $I(0) = \langle m, 0 \rangle$ of amplitude l_3 . By Lemma 2, $L(m) = L(1)$. As $a_1 = 1$ it is well known that $L(1) = 1 + a_2 = q_2$. \square

4.2. Complements and examples

Some known notions used here can be found in [5, Ch. 2] for instance. As previously $L(m)$ is the maximal length of factors having period m . We give two propositions about them.

Proposition 1. *When $L(m) \geq 2m$ all factors of s having period m and length $L(m)$ are equal. Their common value v is a palindrome and $v(1, L(m) - m)$ is left special.*

Proof. Let $s(n, n + L(m) - 1) = v$ have period m . Then $s(n - 1) \neq s(n + m - 1)$ so $s(n, n + L(m) - m - 1)$ is left special and is unique. So v is unique. As \tilde{v} occurs in s , $v = \tilde{v}$, v is a palindrome. \square

Now define $\Lambda(m)$ to be the length of the longest prefix of s having period m .

Proposition 2. (a) *if $u = s(1, m)$ is a primitive word then $\Lambda(m) = L(m) - m$ and $s(1, \Lambda(m))$ is a palindrome.*

(b) *if $u = g^e$, $e \in \mathbb{N}_+$, g primitive, then $L(m) = L(m/e)$ and $\Lambda(m) = \Lambda(m/e) = L(m) - m/e$.*

Proof. (a) Let $w = s(1, L(m) - m)$. As s is standard Sturmian w is a prefix and also a suffix of $v = s(n, n + L(m) - 1)$ of Proposition 1. So $\Lambda(m) \geq L(m) - m$. Now let w_1 be the longest left special suffix of v . Then $w_1 = u_1 w$ for some u_1 , $0 \leq |u_1| < m$. If $|u_1| > 0$, as both w and w_1 are prefixes of s we get that w_1 and also v have period $|u_1|$. So by the Theorem of Fine and Wilf v has period $p = (m, |u_1|)$ and u is an integral power of a shorter word, a contradiction. So w is the longest left special suffix of v .

Now let $s(t, t + \Lambda(m) - 1)$ be an occurrence in s of $s(1, \Lambda(m))$ such that $s(t - 1) = s(t + m - 1) = u(m)$. As $s(t - 1, t + L(m) - m - 1)$ is not special, $s(t - 2) = s(t + m - 2) = u(m - 1)$. Continuing this way we see that $s(t, t + \Lambda(m) - 1)$ is preceded by u . So $|s(t - m, t + \Lambda(m) - 1)| \leq L(m)$ whence $\Lambda(m) = L(m) - m$.

(b) If $u = g^e$, g primitive, $L(m) = L(m/e)$ as in the proof of Lemma 2 and similarly $\Lambda(m) = \Lambda(m/e)$ whence by Part (a) $\Lambda(m) = L(m) - m/e$.

Remark. $s(1, \Lambda(m))$ is a palindrome prefix of s . The structure of such words is known. Let $s_n = L(1, q_n)$ and $h_n = L(1, q_n - 2)$ with the q_n and a_n as previously. Then ignoring the shortest ones for simplicity, the palindrome prefixes of s are $s_n^x h_{n-1}$, $n \geq 2$, $1 \leq x \leq a_{n+1}$. This word has period q_n and length $q_{n+1} - 2$ for $x = a_{n+1}$. So $\Lambda(q_n) \geq$

$q_{n+1} - 2$. We are not far from the result for case (i) of Theorem 4. It is our opinion that, with some work, a substantial part of Theorem 4 could be deduced this way.

Now let $\xi(m) = L(m)/m$ be the maximal fractional exponent corresponding to period m . Assume for simplicity $\alpha < 1/2$ i.e. $a_1 > 1$. Then

Proposition 3. *The maximum of $\xi(m)$ when $q_{i-1} + q_i \leq m < q_i + q_{i+1}$, $i \geq 1$, is $\xi(q_{i+1}) = a_{i+2} + 2 + (q_i - 2)/q_{i+1}$.*

Proof. We are in case (i) of Theorem 4. If $b < a_{i+1}$ then $\xi(m) = 2 + (q_i - 2)/m < 3$. If $b = a_{i+1}$ then $m = q_{i+1}$ and $L(m) = q_{i+2} + 2q_{i+1} - 2 = (a_{i+1} + 2)q_{i+1} + q_i - 2$, whence the result. \square

This allows to find again all known results about maximal exponents. When the partial quotients a_n are bounded, the exponents are bounded [6] and the maximum integer exponent is $2 + \sup\{a_1 - 2, a_2, a_3, \dots\}$.

When the law of the a_n is suitable it is possible to calculate, using (9) hereafter, bounds and limits for the exponents. In particular let $\alpha = [0, a_1, a_2, \dots, a_h, \overline{a_{h+1}, \dots, a_{h+p}}]$ be a quadratic irrational. Recall the general formula (easily deducible from $q_n/q_{n-1} = a_n + q_{n-2}/q_{n-1}$)

$$q_n/q_{n-1} = [a_n, a_{n-1}, \dots, a_2, a_1] \quad (9)$$

Let

$$\lambda_t = [\overline{a_{t-1}, a_{t-2}, \dots, a_{h+1}, a_{h+p}, \dots, a_t}]$$

for $t \in [h+1, h+p]$.

Then

$$\lim_{x \rightarrow \infty} ((q_t + xp - 1)/(q_t + xp - 2)) = \lambda_t.$$

For $n > h$, $[\xi(q_{n-1})]$ is maximal when $a_n = \sup\{a_{h+1}, \dots, a_{h+p}\} = M$.

It follows that the greatest point of accumulation of the $\xi(m)$ is

$$\Omega = M + 2 + \max\{\lambda_t^{-1}; a_t = M, t \in [a_{h+1}, \dots, a_{h+p}]\}$$

It can also be shown that $\xi(m) < \Omega$ except possibly for small values of m . This is the result of [14] in a slightly more general context (s is only the morphic image of a fixed point).

Example 1. With $\alpha = (0, 2, \bar{1}) = (3 - \sqrt{5})/2$, s is the Fibonacci word 01001010... and q_0, q_1, q_2, \dots are the Fibonacci numbers 1, 2, 3, Equation (6) becomes

$$q_{i+1} \leq m = q_{i-1} + bq_i + r < q_{i+2}, \quad 1 \leq b \leq a_{i+1}, \quad 0 \leq r < q_i$$

If $i=0$ then $r=0$, $m=b \leq 2$. We are in case (iii) of Theorem 4 whence if $m=2$ $L(m)=3+2 \times 2-2=5$, corresponding to factors 01010, while if $m=1$ $L(m)=2$ corresponding to factor 00.

If $i > 0$ then $a_{i+1} = 1$, whence $b = 1 = a_{i+1}$ and $m = q_{i+1} + r$, $0 \leq r < q_i$. If $r = 0$ we are in case (i) whence $L(m) = L(q_{i+1}) = q_{i+2} + 2q_{i+1} - 2$. If $r \neq 0$ we are in no case whence $L(m) < 2m$.

For instance $L(5) = 8 + 2 \times 5 - 2 = 16$, corresponding to $(01001)^{30}$, while $L(6) < 12$ and indeed as can be verified $L(6) = 9$, corresponding to $(010010)010$.

Now by Proposition 3

$$\zeta(q_n) = 3 + q_{n-1}/q_n - 2/q_n.$$

As q_{n-1}/q_n has limit $\theta = [0, \bar{1}] = (\sqrt{5} - 1)/2$ we get the limit $3 + (\sqrt{5} - 1)/2$ for the maximal fractionary exponents. Moreover, observing that the q_{n-1}/q_n are the convergents of θ , we have $|q_{n-1}/q_n - \theta| < 1/q_n^2$ whence easily $\zeta(q_n) < 3 + \theta$. These are the results of [7].

Example 2. With $\alpha = (\sqrt{15} - 3)/2 = [0, \overline{2, 3}]$ we have $q_1 = 2$, $q_2 = 7$, $q_3 = 16$, $q_4 = 55$, ... and $s = 01010100101010010101 \dots$

Here are some cases.

If $m = 2$ then $i = 0$, $b = 2 = a_1$ we are in case (iii) whence $L(m) = q_2 + 2q_1 - 2 = 9$, corresponding to $(01)^40$.

If $m = 3$ then $i = 1$, $b = 1$, $r = 0$, we are in case (i) whence $L(m) = 2m + q_i - 2 = 6$, corresponding to $(010)^2$.

If $m = 4$ then $i = 1$, $b = 1$, $r = 1$ we are in case (iv) and also find $L(4) = 9$.

If $m = 5$ then $i = 1$, $b = 2$, $r = 0$, we are in case (i) whence $L(m) = 2m + q_i - 2 = 2m = 10$, corresponding to $(01010)^2$.

If $m = 14$ then $i = 2$, $b = 1$, $r = 5$, we are in case (ii) whence $L(m) = q_3 + 2q_2 - 2 = 28 = L(7)$.

If $m = 6, 8, 10, 11$ for instance we are in no case, $L(m) < 2m$.

Now the exponents $\zeta(m)$ to consider for getting the upper bound are the $\zeta(q_{2n+1}) = 5 + (q_{2n} - 2)/q_{2n+1}$. Their limit is $5 + [0, \overline{2, 3}] = 5 + (\sqrt{15} - 3)/2$ and by using explicit formulas for the q_n it is possible to see that all $\zeta(m)$ are less than this value.

Acknowledgements

We thank the referee for constructive comments and signalling reference [14].

References

- [1] P. Alessandri, V. Berthe, Three distance theorems and combinatorics on words, *L'enseignement Math.* 44 (1–2) (1998) 103–132.
- [2] G.H. Hardy, E.M. Wright, *An introduction to the Theory of Numbers*, Oxford University Press, London 1971.
- [3] X. Karhumaki, On cube-free ω -words generated by binary morphism, *Discrete Appl. Math.* 5 (1985) 279–297.

- [4] D.E. Knuth, Sequences with precisely $k + 1$ k -blocks, solution of problem E2307, *Amer. Math. Monthly* 79 (1972) 773–774.
- [5] M. Lothaire, *Algebraic Combinatorics on Words*, Preliminary Version, February 16, 1999, Institut Gaspard Monge, Univ. Marne-la-Vallée, 1999.
- [6] F. Mignosi, Infinite words with linear subword complexity, *Theoret. Comput. Sci.* 65 (1989) 221–242.
- [7] F. Mignosi, G. Pirillo, Repetitions in the Fibonacci infinite word, *Inform. Theorique. Appl.* 26 (1992) 199–204.
- [8] M. Morse, G.A. Hedlund, Symbolic dynamics II: Sturmian trajectories, *Amer. J. Math.* 62 (1940) 1–42.
- [9] G. Pirillo, Fibonacci numbers and words, *Discrete Math.* 173 (1997) 197–207.
- [10] G. Pirillo, From Fibonacci word to Sturmian words, *Publ. Mat. Debrecen* 54, supplement (1999) 961–971.
- [11] T. van Ravenstein, The three gap theorem (Steinhaus conjecture), *J. Austral. Math. Soc. Ser A* 5 (1988) 360–370.
- [12] N.B. Slatter, Gaps and steps for the sequence $n\theta \bmod 1$, *Proc. Cambridge Philos. Soc* 63 (1967) 1115–1123.
- [13] A. Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen* *Skr. Vid. Kristiania I. Mat. Naturv. Klasse I* (1912) 1–67.
- [14] D. Vandeth, Sturmian words and words with a critical exponent, *Theoret. Comput. Sci.*, in press.

Fibonacci numbers and words

Giuseppe Pirillo

IAMI CNR viale Morgagni 67/A, I-50134 Firenze, Italy

Received 7 April 1994

Dedicated to Giovanni Prodi.

Abstract

Let Φ be the golden ratio $(\sqrt{5} + 1)/2$, f_n the n th Fibonacci finite word and f the Fibonacci infinite word. Let r be a rational number greater than $(2 + \Phi)/2$ and u a non-empty word. If u^r is a factor of f , then there exists $n \geq 1$ such that u is a conjugate of f_n and, moreover, each occurrence of u^r is contained in a maximal one of $(f_n)^s$ for some $s \in [2, 2 + \Phi)$. Several known results on the Fibonacci infinite word follow from this.

1. Introduction

In analogy with the definition of the Fibonacci numbers, one sets $f_0 = b$, $f_1 = a$, and, for $n \geq 2$, one defines the n th Fibonacci finite word as the product $f_{n-1}f_{n-2}$ of the words f_{n-1} and f_{n-2} (see [4, 7]). The two products $f_{n-1}f_{n-2}$ and $f_{n-2}f_{n-1}$ are almost the same, being different only on the last two letters. This is the amusing, very simple but very interesting ‘near-commutative property’ used in [6] to study concrete algorithms. It plays an important role also in this paper.

The Fibonacci infinite word f is the *Sturmian word* associated with the golden ratio $\Phi = (\sqrt{5} + 1)/2$ and can also be defined as the unique infinite word having, for each $n \geq 1$, f_n as a left factor.

If a power u^r has an occurrence in f , we try to extend it to the left and to the right as far as possible preserving periodicity. We call *maximal* the occurrences of the powers that cannot be locally extended and we prove that we always reach one of them, call it v^s , where v is a conjugate of u . The main result of present paper says that: if $r > (2 + \Phi)/2$ then $v = f_n$ for some $n \geq 1$ and $s \in [2, 2 + \Phi)$. Several known results on f are consequences of this.

2. Definitions and preliminary results

This paper is organized so as to be self-contained; terminology and notations are those currently used in theoretical computer science [4, 7].

We consider only the two-letter alphabet $\{a, b\}$ and we call (finite) words the elements of the free monoid $\{a, b\}^*$; we denote by 1 the empty word and by $|u|$ the length of a word u . We consider a word u of length $k \geq 1$ as a map $u: \{0, 1, \dots, k-1\} \rightarrow \{a, b\}$; we write $u = u(0) \dots u(i) \dots u(k-1)$ where $u(0)$, $u(i)$ and $u(k-1)$ are, respectively, the first, the i th and the last letter of u .

A word u is a factor of a word v if there exist two words $u', u'' \in \{a, b\}^*$ such that $v = u'uu''$. When $u' = 1$ (resp. $u'' = 1$) we say that u is a left factor (resp. right factor) of v . A proper factor, (resp. proper left factor, proper right factor) u of v is a factor (resp. left factor, right factor) u of v such that $|u| < |v|$.

A (right) infinite word on $\{a, b\}$ is a map q from the set of non-negative integers into $\{a, b\}$. We write $q = q(0)q(1) \dots q(i) \dots$. A word u is a factor of q if there exist a word u' and an infinite word q' such that $q = u'uq'$. If $u' = 1$ we say that u is a left factor of q .

A non-empty word u may be a factor of another (finite or infinite) word w in several ways. So it is useful to speak about occurrences. For this reason, let i, j be integers such that $0 \leq i \leq j$ (and that $j < |w|$ if w is a finite word) and let us denote by $w(i, j)$ the word $w(i) \dots w(j)$. We say that the pair of integers (i, j) is an occurrence of the factor u in the word w if $u = w(i, j)$. We say that an occurrence (i_0, j_0) of u in w is contained in an occurrence (i_1, j_1) of v in w if $i_1 \leq i_0 \leq j_0 \leq j_1$. We only speak about occurrences of non-empty words.

Now, let $\varphi: \{a, b\}^* \rightarrow \{a, b\}^*$ be the morphism whose restriction to $\{a, b\}$ is given by $\varphi(a) = ab$, $\varphi(b) = a$. Remark that φ is injective. Let us define the n th Fibonacci finite words f_n in the following way: $f_0 = b$ and, for each $n \geq 0$,

$$f_{n+1} = \varphi(f_n).$$

In particular, we have: $f_1 = a$, $f_2 = ab$, $f_3 = aba$, $f_4 = abaab$, $f_5 = abaababa$, $f_6 = abaababaabaab$, $f_7 = abaababaabaababaababa \dots$. It is clear that, for each $n \geq 2$, f_n is the product (juxtaposition) $f_{n-1}f_{n-2}$ of f_{n-1} and f_{n-2} . Also, for each $n \geq 0$, $|f_n|$ is the n th element F_n of the sequence of Fibonacci numbers 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377 ...

Remark now that, for each $n \geq 1$, f_n is a left factor of f_{n+1} . So there exists a unique infinite word, namely the Fibonacci infinite word f , such that, for each $n \geq 1$, f_n is a left factor of f (see, [4, 7]) and we have

$$f = abaababaabaababaabaababaabaababaabaababa \dots$$

We denote by $F(f)$ the set of the non empty factors of f and by $LF(f)$ its subset containing the non empty left factors of f .

For each $n \geq 2$, we denote by g_n the product $f_{n-2}f_{n-1}$ and by h_n the common longest left factor of f_n and g_n . In particular, we have: $g_2 = ba$, $g_3 = aab$, $g_4 = ababa$, $g_5 = abaabaab$, ... and $h_2 = 1$, $h_3 = a$, $h_4 = aba$, $h_5 = abaaba \dots$.

Remark that a non-empty factor of an element of $F(f)$ is again in $F(f)$; for each $n \geq 2$, $g_n \in F(f)$, $\varphi(g_n) = g_{n+1}$ and $h_{n+1} \in LF(f)$; if $f(i) = b$ then $i > 0$ and $f(i-1) = f(i+1) = a$ and if $f(i, i+1) = aa$ then $i > 0$ and $f(i-1) = f(i+2) = b$ (i.e., bb , $aaa \notin F(f)$).

A factor v of f is *special* if $va, vb \in F(f)$. Let $k \geq 1$; we denote by \tilde{u} the mirror image $u(k-1)u(k-2)\dots u(1)u(0)$ of the word $u = u(0)u(1)\dots u(k-2)u(k-1)$. We say that a non-empty word v is a *palindrome* if $v = \tilde{v}$.

Lemma 1 belongs to the folklore (see e.g. [1, 2, 4, 5, 6]) and is very easy; so we can give just an hint of its proof. The point (i) is the ‘near-commutative property’ quoted in the introduction.

Lemma 1. For each $n \geq 2$,

(i) $f_{n+2} = f_{n+1}f_n = f_n g_{n+1} = h_{n+2}xy$ and $g_{n+2} = f_n f_{n+1} = f_{n+1}g_n = h_{n+2}yx$, where $x, y \in \{a, b\}$, $x \neq y$ and if n is even then $xy = ab$ and if n is odd then $xy = ba$;

(ii) $|h_n| = F_n - 2$;

(iii) h_n is a special factor;

(iv) $f_{n+3} = h_{n+1}xyh_nyxh_{n+1}xy$, where $x, y \in \{a, b\}$, $x \neq y$;

(v) h_n is a palindrome;

(vi) $h_{n+2} = f_n h_{n+1} = h_{n+1} \tilde{f}_n = h_n \tilde{f}_{n+1} = f_{n+1} h_n$;

(vii) for each integer $m \geq 0$, h_n is a left and a right factor of h_{n+m} .

Moreover, (viii) if $v \in F(f)$ then $\tilde{v} \in F(f)$.

Proof. One can prove (i) easily by induction; (ii)–(iv) are consequence of (i); one can prove (v) by induction using (iv); (vi) is a consequence of (v); (vii) is a corollary of (vi) and finally one can prove (viii) using (v). \square

Lemma 2. For each $n \geq 2$,

(i) $f_n(0) = a (= f_1)$; $g_{n+1}(0) = a$;

(ii) if $f(i, j) = f_{n+2}$ then $f(j+1) = a$; if $f(i, j) = g_{n+1}$ then $f(j+1) = a$;

(iii) if $f(i, j) = h_{n+1}$ then $f(i, j+2) = f_{n+1}$ or $f(i, j+2) = g_{n+1}$.

Proof. (i) is trivial; to prove (ii) and (iii) use the fact that aba is a right factor of each h_{n+2} (point (vii) of Lemma 1) and the fact that $bb, aaa \notin F(f)$. \square

Lemma 3. Let $u \in F(f)$. Then $\varphi^{-1}(u)$ exists and belongs to $F(f)$ if and only if one of these two conditions holds: (i) $u(0) = a$ and $u(|u|-1) = b$; (ii) $u(0) = a$, $u(|u|-1) = a$ and $ua \in F(f)$.

Proof. By induction on $|u|$. \square

Remark. We have $\varphi^{-1}(aa) = bb \notin F(f)$. We often use Lemma 3 together with points (i) and (ii) of Lemma 2 in order to prove the existence of $\varphi^{-1}(u)$ in $F(f)$ for suitable $u \in F(f)$.

The following Lemma 4 says that no occurrence of g_n is too close to the left of f .

Lemma 4. For each $n \geq 2$, if $f(i, j) = g_n$ then $i \geq F_{n-1}$ and $f(i - F_{n-1}, i - 1) = f_{n-1}$.

Proof. By induction. Let $n = 2$. If $f(i, i + 1) = ba = g_2$ then $i \neq 0$ and $f(i - 1) = a = f_1$. Now, let $n \geq 3$ and $f(i, j) = g_n$. As $f(0, F_n - 1) = f_n$, $i \neq 0$ and so $w = f(0, i - 1)$ is non empty. By Lemmas 2 and 3, one has $\varphi^{-1}(wg_n) = w'g_{n-1} \in F(f)$ for some non-empty $w' \in LF(f)$. By induction hypothesis f_{n-2} is a right factor of w' ; so $i \geq F_{n-1}$ and $f(i - F_{n-1}, i - 1) = f_{n-1}$. \square

The following Lemma 5 belongs to the folklore. Point (i) is proved, in [1] for example, using an auxiliary morphism which is not necessary here. Point (ii) is a particular case of a more general result on sturmian words (see [4, 5]). For each $k \geq 1$, let us denote by $s^{[k]}$ the mirror image of the left factor of f having length k .

Lemma 5. For each $k \geq 1$, (i) the unique special factor of length k is $s^{[k]}$; (ii) in $F(f)$ there are exactly $k + 1$ elements of length k .

Proof. (i) Remark that $s^{[k]}$ is a right factor of h_m for each m such that $F_m > k + 1$ and so, by point (iii) of Lemma 1, $s^{[k]}$ is special. Suppose now that for a given k there is a special factor v of length k which is different from $s^{[k]}$. The last letter of v is necessarily a , hence the greatest integer, such that h_n is a common right factor of $s^{[k]}$ and v , is greater than or equal to 3. Let also k' be the greatest integer such that $s^{[k]}(k') \neq v(k')$. We have $s^{[k]} = uxu'h_n$ and $v = u''yu'h_n$, for some $u, u', u'' \in \{a, b\}^*$ and for some $x, y \in \{a, b\}$, $x \neq y$. We have also $|u'| < F_{n-1}$ otherwise, by point (vi) and (vii) of Lemma 1, we have a contradiction with the maximality of n . Being a right factor of a special factor, $yu'h_n$ is special and so $yu'h_n a, yu'h_n b \in F(f)$. In both cases, n even or odd, $yu'g_n \in F(f)$ and, by Lemma 4, yu' is a right factor of f_{n-1} . Hence, xu' is not a right factor of f_{n-1} . Contradiction. (ii) It is an easy consequence of (i). \square

Let us recall that an infinite word p is *periodic* (resp. *ultimately periodic*) if there exists $k \geq 1$ such that $p(j + k) = p(j)$ for each $j \geq 0$ (resp. for each $j \geq i$ for some $i \geq 0$).

Lemma 6. The Fibonacci infinite word is not ultimately periodic.

Proof. This is easy by point (ii) of Lemma 5. \square

Let $u, v, w, z, z' \in F(f)$. We say that (u, v, w) is a *non empty overlap* of z and z' if $uv = z$, $vw = z'$ and $uvw \in F(f)$. The possible non empty overlaps concerning f_n and g_n are considered in the following Lemma.

Lemma 7. Let $n \geq 3$. Then

- (i) $(f_{n-1}, f_{n-2}, g_{n-1})$ is the unique non-empty overlap of f_n and f_n ;
- (ii) $(f_{n-1}, f_{n-2}, f_{n-1})$ is the unique non-empty overlap of f_n and g_n ;
- (iii) $(f_{n-2}, f_{n-1}, f_{n-2})$ is the unique non-empty overlap of g_n and f_n ;
- (iv) there is no non-empty overlap of g_n and g_n .

Proof. By induction. (i) Let $n = 3$. As a is the unique word which is a proper non empty right and also left factor of aba , $(ab, a, ba) = (f_2, f_1, g_2)$ is the unique non-empty overlap of f_3 and f_3 ; hence the statement is true for $n = 3$. Now, let $n \geq 4$. Clearly, $(f_{n-1}, f_{n-2}, g_{n-1})$ is a non-empty overlap of f_n and f_n . Now, let (u, v, w) be a non-empty overlap of f_n and f_n . By Lemmas 2 and 3 there exist $u' = \varphi^{-1}(u)$, $v' = \varphi^{-1}(v)$, $w' = \varphi^{-1}(w)$ in $F(f)$ such that (u', v', w') is a non-empty overlap of f_{n-1} and f_{n-1} . By induction hypothesis, $u' = f_{n-2}$, $v' = f_{n-3}$, $w' = g_{n-2}$. Hence $u = \varphi(f_{n-2}) = f_{n-1}$, $v = \varphi(f_{n-3}) = f_{n-2}$, $w = \varphi(g_{n-2}) = g_{n-1}$; (ii) the argument is analogous, but starting with the fact that a is the unique word which is a proper non-empty right factor of aba and also a proper non-empty left factor of aab ; (iii) ab is the unique word which is a proper non-empty right factor of aab and also a proper non-empty left factor of aba ; (iv) no word is a proper non empty right and also left factor of aab . \square

Lemma 8. Let $n \geq 5$. There are exactly two non-empty overlaps of h_n and h_n , namely just

$$(f_{n-1}, h_{n-2}, \tilde{f}_{n-1})$$

and

$$(f_{n-2}, h_{n-1}, \tilde{f}_{n-2}).$$

Proof. This is an easy consequence of point (iii) of Lemma 2, Lemma 7 and point (vi) of Lemma 1. \square

Remark. In some sense, $(aba)(aba)$ can be considered as an ‘overlap’ of h_4 and h_4 but we have chosen to consider only non-empty overlaps. So there is a unique non-empty overlap of h_4 and h_4 and this is (ab, a, ba) , in accordance with point (i) of Lemma 7.

Lemma 9. Let $v \in LF(f)$. Then the following two conditions are equivalent: (i) v is a palindrome and (ii) $v = h_n$ for some $n \geq 3$.

Proof. (ii) \rightarrow (i) is point (v) of Lemma 1. We prove (i) \rightarrow (ii) by induction. The palindromes of $LF(f)$, having length less than or equal to $F_5 - 2$, are $a = h_3$, $aba = h_4$ and $abaaba = h_5$. Let $n \geq 4$ and suppose, by induction hypothesis, that h_3, \dots, h_n and h_{n+1} are all the palindromes of $LF(f)$ having length less than or equal to $F_{n+1} - 2$. Suppose also, by way of contradiction, that there exists a proper left factor u of h_{n+2} such that h_{n+1} is a proper left factor of u with u a palindrome. By points (i) and (vii) of Lemma 1, we can write $h_{n+2} = f_n h_{n-1} c d h'_n b a$, where h'_n is the left factor of h_n of length $F_n - 4$ and $c, d \in \{a, b\}$, $c \neq d$. We can see that $|u| \notin \{F_{n+1} - 1, F_{n+1}, F_{n+2} - 3\}$. So there exist a non-empty left factor u' of h'_n , $x, y \in \{a, b\}$ and $u'' \in \{a, b\}^*$ such that $f_n = u' x y u'' d c$ and $u = u' x y u'' d c h_{n-1} c d u'$. Being $u'' d c h_{n-1} c d u'$ a right factor of u , u a palindrome and a left factor of h_{n+2} , one has that $u'' d c h_{n-1} c d u'$ is a right factor of h_{n+2} . As $|u'' d c h_{n-1} c d u'| = F_{n+1} - 2$, one has that $u'' d c h_{n-1} c d u' = h_{n+1}$. By Lemma 8 and by $|u' x y| < F_n$, we have $u' x y = f_{n-1}$ and so $x y = c d$. Then $u = u' c d u'' d c h_{n-1} c d u'$ is not a palindrome. Contradiction. \square

Lemma 10. For each $n \geq 2$,

- (i) $f_n g_n \notin F(f)$;
- (ii) if $f(i, j) = g_{n+1} h_{n+1}$ then $f(i - F_n, j) = f_{n+1} f_{n+1} h_n$;
- (iii) if $z, f_{n+1} z g_{n+1} \in F(f)$ then $|z| \geq F_{n-2}$;
- (iv) if $z, g_{n+1} z f_{n+1} \in F(f)$ then $|z| \geq F_{n-1}$.

Proof. (i) follows from point (i) of Lemma 1 and Lemma 4. (ii) Let $f(i, j) = g_{n+1} h_{n+1}$. By Lemma 4, we have $f(i - F_n, i - 1) = f_n$. Hence, $f(i - F_n, j) = f_{n+1} f_{n+1} h_n$. The proofs of (iii) and (iv) are similar to that of Lemma 4. \square

For each non empty finite word w , there exists a naturally associated periodic infinite word p_w , defined as follows: $p_w(0, |w| - 1) = w$ and, for each $i \geq 0$, $p_w(i + |w|) = p_w(i)$. We say that a word u is a *power* of the (finite!) word w if u is a left factor of p_w . We say that w is the *base* and $k = |u|/|w|$ is the *exponent* of this power and we write $u = w^k$. In general k is rational, but if k is an integer we obtain the usual notion of power. We consider only powers with exponent greater than or equal to 1. We say that two words u and v are *conjugate* if there exist two words u' and u'' such that $u = u'u''$ and $v = u''u'$. The following lemma is trivial.

Lemma 11. Let $r, s \geq 1$ and u, v be non-empty words of equal length. If u^r is a factor of v^s then u is a conjugate of v .

3. Maximal powers in the Fibonacci infinite word

Let $u \in F(f)$, r rational and $u^r \in F(f)$. We say that the power u^r is *maximal* if for each word v such that $|u| = |v|$ and for each rational s , if u^r is a factor of $v^s \in F(f)$, then $u = v$ and $r = s$. We say that an occurrence (i_0, j_0) of u^r in f is *maximal* if for each $v \in F(f)$ such that $|u| = |v|$ and for each i_1, j_1 such that $i_0 \leq i_1 \leq j_0 \leq j_1$, if (i_1, j_1) is an occurrence of some power of v in f then $u = v$, $i_0 = i_1$ and $j_0 = j_1$. A power can have maximal occurrences even if it is not maximal. For example, a , $(ab)^{3/2}$, $(aba)^2$, and $(abaab)^{1/5}$ are not maximal powers but the pairs $(5, 5)$, $(8, 10)$, $(13, 18)$ and $(21, 31)$ are, respectively, maximal occurrences of them in f .

Proposition 1. Let $u \in F(f)$ and $r \geq 1$. If (i_0, j_0) is an occurrence of u^r in f then there exist a conjugate v of u and $s \geq r$ such that (i_0, j_0) is contained in a maximal occurrence of v^s in f .

Proof. Let I be the set of all $i \geq 0$ such that (i, j_0) is an occurrence of some $(u')^r$ such that $|u| = |u'|$. Since I contains at least i_0 , it is non empty. Let i_1 be its minimum. Now, let J be the set of all $j \geq j_0$ such that (i_1, j) is an occurrence of some $(u'')^r$ such that $|u| = |u''|$. Since J contains at least j_0 , it is non empty. By Lemma 6, there is a maximum in J , say j_1 . Clearly, (i_1, j_1) is a maximal occurrence of some v^s such that $|v| = |u|$ and $s \geq r$. By Lemma 11, u is a conjugate of v . \square

Hereafter we denote by Φ the golden ratio $(\sqrt{5} + 1)/2$. One of the arguments used proving the following Propositions 2–6 consists in reading some words in both directions, left–right and right–left.

Proposition 2. *Let $v \in F(f)$ and s be a rational number such that $(2 + \Phi)/2 < s < 2$. Then no occurrence of v^s in f is a maximal one.*

Proof. By way of contradiction, suppose that (i_1, j_1) is a maximal occurrence of v^s . There exist a positive integer k , a proper non-empty left factor v' of v and a proper non-empty right factor v'' of v such that $|v'| = k$, $v = v'v''$ and $v^s = v'v''v'$. We pose $x = v''(0)$ and $x' = v''(|v'| - 1)$ and we have $v'' = xv''' = v''''x'$ for some $v''', v'''' \in \{a, b\}^*$.

(ia) $i_1 = 0$. Let $f(j_1 + 1) = y$. By maximality of v^s we have $x \neq y$, and so v' is special. By Lemma 5, $v' = s^{[k]}$ and, by definition of $s^{[k]}$, $v' = \tilde{v}'$. As $v' \in LF(f)$ and v' is a palindrome, we have, by Lemma 9, $v' = h_n$ for some $n \geq 3$.

(ib) $i_1 \neq 0$. Let $f(i_1 - 1) = y'$, $f(j_1 + 1) = y$ and consider the words $v^s y = v'xv''v'y$ and $y'v^s = y'v'v''''x'v'$. Remark that, by point (viii) of Lemma 1, $\tilde{v}'x'\tilde{v}''''\tilde{v}'y' \in F(f)$. By maximality of v^s we have $x \neq y$ and $x' \neq y'$ and from this v' and \tilde{v}' are both special factors. Then $v' = \tilde{v}' = s^{[k]}$ and so $v' \in LF(f)$ and v' is a palindrome. Hence, by Lemma 9, $v' = h_n$ for some $n \geq 3$.

Thus in both cases (ia) and (ib) we have $v' = h_n$ for some $n \geq 3$.

Moreover, $n > 3$, otherwise we would have $v'v''v' = (av'')a = (av'')^s$ and $s = (2 + |v''|)/(|v''| + 1) \leq 3/2 < (2 + \Phi)/2$ which is a contradiction. So $v' = h_n$ for $n \geq 4$.

(i1) $|v''| = 1$. By point (vii) of Lemma 1, v' begins and ends with aba . As $aaa, abababa \notin F(f)$ we reach a contradiction in the case $v'' = a$ as well as in the case $v'' = b$.

(i2) $|v''| = 2$. By maximality of v^s we have $v'xyv'yx \in F(f)$ where $x, y \in \{a, b\}$, $x \neq y$ and $v'' = xy$. By point (iii) of Lemma 2 we have to consider two cases: $v'xy = f_n$ and $v'xy = g_n$.

(i2a) If $v'xy = f_n$ then we have $v'xyv'yx = f_n g_n \in F(f)$ which is impossible by point (i) of Lemma 10.

(i2b) If $v'xy = g_n$ then we have $v'xyv' = g_n h_n$ which contradicts, by point (ii) of Lemma 10, the maximality of v^s (more precisely, $g_n h_n$ is a fractional power of g_n with exponent $(2F_n - 2)/F_n < 2$ and each of its occurrence is contained, as right factor, in an occurrence of $f_n f_n h_{n-1}$ which is a fractional power of f_n with exponent $(2F_n + F_{n-1} - 2)/F_n > 2$).

(i3) $|v''| \geq 3$. We have $v'xyzv'yx \in F(f)$ where x, y are letters and z is a word. Again we have to consider two cases: $v'xy = f_n$ and $v'xy = g_n$.

(i3a) If $v'xy = f_n$ then, by point (iii) of Lemma 10, $|z| \geq F_{n-3}$.

(i3b) If $v'xy = g_n$ then, by point (iv) of Lemma 10, $|z| \geq F_{n-2}$.

So in both (i3a) and (i3b) we have $|z| \geq \min\{F_{n-2}, F_{n-3}\} = F_{n-3}$. But then we have

$$s = (|v'| + |xyz| + |v'|)/(|v'| + |xyz|) = 1 + (F_n - 2)/(F_n + |z|) < 1 + (F_n - 2)/(2F_{n-1}) < (2 + \Phi)/2,$$

i.e., a contradiction. \square

Proposition 3. Let $v \in F(f)$. If v^2 has a maximal occurrence in f then $v = a$ or $v = aba$.

Proof. As in the previous case we first prove that v is a palindrome and $v \in LF(f)$ and so that $v = h_n$ for some $n \geq 3$. By point (vii) of Lemma 1, $abaaba$ is a right and a left factor of h_n for each $n \geq 5$. Since $abaabaabaaba \notin F(f)$, the only two possibilities are $v = h_3 = a = f_1$ or $v = h_4 = aba = f_3$. \square

Proposition 4. Let $v \in F(f)$ and s be a rational such that $2 < s < 3$. If v^s has a maximal occurrence in f then either

$$v = f_2 = ab \quad \text{and} \quad s = 5/2$$

or

$$v = f_n \quad \text{and} \quad s = 2 + (F_{n-1} - 2)/F_n$$

for some $n \geq 4$.

Proof. In this case there exist a proper non-empty left factor v' of v and a proper non-empty right factor v'' of v such that $v = v'v''$ and $v^s = v'v''v'v''v'$. In analogy with the previous cases we first prove that $v'v''v' = h_m$ for some $m \geq 3$.

Clearly the case $m = 3$ is impossible.

If $m = 4$ then $v'v''v' = aba$, $v'v'' = ab$ and $v' = a$. Hence, $v = ab$ and $s = 5/2$.

Now, let $m \geq 5$. By Lemma 8, we have $v'v'' = f_{m-2}$ or $v'v'' = f_{m-1}$.

(ia) If $v'v'' = f_{m-2}$ then, again by Lemma 8, we have $F_{m-1} - 2 = |v'| < |v'v''| = F_{m-2}$ and so $m = 4$. Contradiction.

(ib) If $v'v'' = f_{m-1}$ then, again by Lemma 8, $v' = h_{m-2}$. Hence, for some $n = m - 1 \geq 4$, $v = f_n$ and $s = 2 + (F_{n-1} - 2)/F_n$. \square

Proposition 5. Let $v \in F(f)$. If v^3 has a maximal occurrence in f then $v = aba$.

Proof. In analogy with the proof of Proposition 3, the unique possibility is $v = h_4 = aba = f_3$. \square

Proposition 6. Let $v \in F(f)$ and s be a rational such that $3 < s$. If v^s has a maximal occurrence in f then

$$v = f_n \quad \text{and} \quad s = 3 + (F_{n-1} - 2)/F_n$$

for some $n \geq 4$.

Proof. First, as in the proofs of Propositions 3 and 5, we have that the rational s is not an integer. So we can suppose that there exist a non-empty left factor v' of v and a non-empty right factor v'' of v such that $v = v'v''$ and $v^s = (v'v'')^k v'$ for some integer $k \geq 3$. In analogy with the previous cases we prove that $(v'v'')^{k-1} v' = h_m$ for some $m \geq 3$.

Clearly the cases $m = 3$ and $m = 4$ are impossible.

Let $m \geq 5$. By Lemma 8, we have $v'v'' = f_{m-1}$ or $v'v'' = f_{m-2}$.

(ia) The case $v'v'' = f_{m-1}$ is impossible. In fact, again by Lemma 8, we would have $F_{m-1} = |v'v''| < |(v'v'')^{k-2}v'| = F_{m-2} - 2$, which is clearly a contradiction.

(ib) If $v'v'' = f_{m-2}$ then, again by Lemma 8, we have that $(v'v'')^{k-2}v' = h_{m-1}$ and so $k = 3$ and $v' = h_{m-3}$. As $0 < |v'| = F_{m-3} - 2$ we must have $m \geq 6$. Hence, for some $n = m - 2 \geq 4$, $v = f_n$ and $s = 3 + (F_{n-1} - 2)/F_n$. \square

Remark. The word $abaabaa = (f_3)^2 a$ is not a maximal power, being always a factor of a (maximal) occurrence of $abaabaaba = (f_3)^3$ in f . The word $abaabaabaa = (f_3)^3 a$ does not belong to $F(f)$.

Remark. Consider the sequences $2 + (F_{n-1} - 2)/F_n$ and $3 + (F_{n-1} - 2)/F_n$; their elements are exponents of powers having a maximal occurrence in f and the numbers $1 + \Phi$ and $2 + \Phi$ are their respective limits for n going to infinity. By Propositions 2–6, no other value greater than $(2 + \Phi)/2$ is the limit of such a sequence. On the other hand, as one can easily see, in the interval $[1, (2 + \Phi)/2]$ infinitely many values have such a property.

Proposition 7. Let s be a rational number greater than $(2 + \Phi)/2$ and $v \in F(f)$. If v^s has a maximal occurrence in f then there exists $n \geq 1$ such that

$$v = f_n.$$

Proof. It follows by Propositions 2–6. \square

Proposition 8. Let $n \geq 1$ and s be a rational number greater than $(2 + \Phi)/2$. If $(f_n)^s$ has a maximal occurrence in f then:

if $n = 1$ then $s = 2$;

if $n = 2$ then $s = 5/2$;

if $n = 3$ then $s = 2$ or $s = 3$;

if $n \geq 4$ then $s = 2 + (F_{n-1} - 2)/F_n$ or $s = 3 + (F_{n-1} - 2)/F_n$.

Proof. Again by Propositions 2–6. \square

Remark. For $n \geq 4$ the two values are effectively realized.

Proposition 9. Let r be a rational number greater than $(2 + \Phi)/2$ and $u \in F(f)$. If $u^r \in F(f)$, then there exists $n \geq 1$ such that u is a conjugate of f_n and, moreover, each occurrence of u^r is contained in a maximal one of $(f_n)^s$ for some $s \in [2, 2 + \Phi)$.

Proof. By Propositions 1, 7 and 8 and by Lemma 11. \square

Proposition 10. For each $\varepsilon > 0$ there exist a rational $t \in [(2 + \Phi)/2 - \varepsilon, (2 + \Phi)/2)$ and a word w , such that w^t has a maximal occurrence in f and $|w| \neq F_n$ for each $n \geq 3$.

Proof. Consider, for $n \geq 3$, the factorization

$$f_{n+4} = f_{n+1} f_n f_n f_n g_{n-1} f_{n-1} f_n.$$

Clearly, for $n \geq 3$, $(f_n f_n) f_n h_{n-1}$ has a maximal occurrence. As, for $n \geq 3$, $2F_n$ is not a Fibonacci number and $(2 + \Phi)/2$ is the limit of $(2F_n + F_{n+1} - 2)/2F_n$ for n going to infinity, the statement is proved. \square

The following Propositions 11–13 are known results on the Fibonacci infinite word and are easy consequences of Proposition 9.

Proposition 11 (Séébold [9]). Let $u \in F(f)$. If $u^2 \in F(f)$ then u is a conjugate of f_n for some $n \geq 1$.

Proof. This follows immediately from Proposition 9 and from $2 > (2 + \Phi)/2$. \square

Proposition 12 (Karhumäki [3]). Let $u \in F(f)$. Then $u^4 \notin F(f)$.

Proof. As already remarked by Séébold, Proposition 12 is a consequence of Proposition 11. \square

Proposition 13 (Mignosi and Pirillo [8]). Let $u \in F(f)$ and r rational such that $u^r \in F(f)$. Then $r < 2 + \Phi$.

Proof. By way of contradiction, suppose that for some $u \in F(f)$ and for some rational $r > 2 + \Phi$, $u^r \in F(f)$. By Proposition 9, each occurrence of u^r in f is contained in a maximal one of $(f_n)^s$ for some $s \in [2, 2 + \Phi)$ and some $n \geq 1$ such that $|u| = F_n$. So $r \leq s < 2 + \Phi$. Contradiction. \square

Remark. A previous version of this paper is in the *Actes du Séminaire Lotharingien de Combinatoire* (30^e Session, 1993)

References

- [1] J. Berstel, Mots de Fibonacci, Séminaire d'informatique théorique, LITP, Paris 1980–81, 57–78.
- [2] A. de Luca, A combinatorial property of the Fibonacci words, Inform. Process. Lett. 4 (1981) 193–195.
- [3] J. Karhumäki, On cube-free ω -words generated by binary morphism, Discr. Appl. Math. 5 (1985) 279–297.
- [4] D.E. Knuth, The Art of Computer Programming (Addison-Welsey, Reading, MA, 1968).

- [5] D.E. Knuth, Sequences with precisely $k + 1$ k -blocks, Solution to problem E2307, *Amer. Math. Monthly* 79 (1972) 773–774.
- [6] D.E. Knuth, J.H. Morris and V.R. Pratt, Fast pattern matching in strings, *SIAM J. Comput.* 6 (1977) 323–350.
- [7] Lothaire, *Combinatorics on words* (Addison-Wesley, Reading, MA, 1983).
- [8] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word, *Inform. théorique Appl.* 26 (1992) 199–204.
- [9] P. Séébold, *Propriétés combinatoires des mots infinis engendrés par certains morphismes*, Thèse de doctorat, L.I.T.P., Paris, 85-15, 1985.

REPETITIONS IN THE FIBONACCI INFINITE WORD (*)

by F. MIGNOSI ⁽¹⁾ and G. PIRILLO ⁽²⁾

Communicated by J. BERSTEL

Abstract. – *Let φ be the golden number; we prove that the Fibonacci infinite word contains no fractional power with exponent greater than $2 + \varphi$ and we prove that for any real number $\varepsilon > 0$ the Fibonacci infinite word contains a fractional power with exponent greater than $2 + \varphi - \varepsilon$.*

Résumé. – *Soit φ le nombre d'or; nous prouvons que le mot infini de Fibonacci ne contient pas la puissance fractionnaire d'exposant supérieur à $2 + \varphi$, et nous prouvons qu'il contient des puissances d'exposant supérieur à $2 + \varphi - \varepsilon$, quel que soit le nombre réel $\varepsilon > 0$.*

INTRODUCTION

Many papers are concerned with the existence of integer powers in “long enough” words or in infinite words; a classical combinatorial property is whether a given infinite word is k power-free or not, with k natural number.

No word on a two letters alphabet can avoid a square but it is well known that the Thue infinite word \mathbf{t} on a two letter alphabet does not contain cubes and that the Thue infinite word \mathbf{m} on a three letter alphabet does not contain squares (see [9], [10]).

The notion of overlap-free word and more generally the notion of fractional power are considered in many papers (see for instance [4], [7], [9], [10]).

In this paper we prove that the Fibonacci infinite word contains no fractional power with exponent greater than $2 + ((\sqrt{5} + 1)/2)$ and that for any real number $\varepsilon > 0$ the Fibonacci infinite word contains a fractional power with exponent greater than $2 + ((\sqrt{5} + 1)/2) - \varepsilon$.

(*) Received 1989, revised November 1990.

⁽¹⁾ Dept. Mat. et Appl., via Archirafi 34, 90123 Palermo (Italia), L.I.T.P.-Université Paris-VII, 2, place Jussieu, 75252 Paris Cedex 05.

⁽²⁾ I.A.G.A.-I.A.M.I. C.N.R., Viale Morgagni 67/A, 50134 Firenze (Italia).

To our knowledge this is the first time that this property for a non rational value is looked for in a given infinite word.

DEFINITIONS AND PRELIMINARY RESULTS

We refer to [6] for the terminology.

Let A be an alphabet. We denote by A^* the *free monoid* on A . The elements of A^* are called *words* and the elements of A are called *letters*. We denote by 1 the empty word which is the identity of A^* ; we also denote by $|v|$ the length of a word v .

A word v is a *factor* of a word w if there exist $u, u' \in A^*$ such that

$$w = uvu'$$

and we say that v is a *left factor* of w if u is the empty word.

If a word w is of the form

$$w = v \dots v = v^k$$

with $u \neq 1$, we say that w is a *k-power* of v ; k is called the *exponent* of the power and v is the *base* of the power.

If a word w is of the form

$$w = v \dots vu = v^k u$$

with $u \neq 1$, $k \geq 1$ and u left factor of v , we say that w is a *fractional power* of u of exponent $e = |w|/|v|$ and v is the base of the power.

An infinite word s on an alphabet A is a map from the set of positive integers into A ; we denote by A^ω the set of all infinite words on the alphabet A .

A word $v \in A^*$ is a factor of the infinite word s if there exist $u \in A^*$, $s' \in A^\omega$ such that $s = uvs'$. If u is the empty word then v is a left factor of s .

The Fibonacci infinite word \mathbf{f} on the alphabet $A = \{a, b\}$ is obtained by iterating the morphism $\psi : \{a, b\} \rightarrow \{a, b\}$ given by

$$\psi(a) = ab, \quad \psi(b) = a$$

starting with the letter a (see [1]). Therefore

$$\mathbf{f} = abaababaabaabab\dots$$

We define the sequence of the finite Fibonacci words by the rule:

$$\begin{aligned} \mathbf{f}_0 &= b, \\ \mathbf{f}_{n+1} &= \psi(\mathbf{f}_n). \end{aligned}$$

It is easy to see that $\mathbf{f}_{n+2} = \mathbf{f}_{n+1} \mathbf{f}_n$ and, consequently, the sequence $|\mathbf{f}_n|$, $n \in \mathbb{N}$ is the sequence of Fibonacci numbers; moreover for any $n \geq 1$, \mathbf{f}_n is a left factor of \mathbf{f}_{n+1} and of \mathbf{f} .

For $n \geq 2$ we denote by \mathbf{g}_n the word $\mathbf{f}_{n-2} \mathbf{f}_{n-1}$. It is easy to see that for each $n \geq 2$ there exists a word \mathbf{v}_n such that $\mathbf{f}_n = \mathbf{v}_n xy$ and $\mathbf{g}_n = \mathbf{v}_n yx$ with $x, y \in \{a, b\}$ and $x \neq y$ and also that $\mathbf{f}_{n+2} = \mathbf{f}_n \mathbf{f}_n \mathbf{g}_{n-1}$.

The following fact is straightforward

Fact. – If u is a left factor of \mathbf{f}_n and also of \mathbf{g}_{n-1} then u is a left factor of \mathbf{v}_{n-1} and, consequently

$$|u| \leq |\mathbf{v}_{n-1}| = |\mathbf{g}_{n-1}| - 2 = |\mathbf{f}_{n-1}| - 2.$$

In the sequel we will use the following results.

PROPOSITION 1 (Karhumäki [4]): *The Fibonacci infinite word \mathbf{f} contains no 4-power.*

PROPOSITION 2 (Séebold [8]): *Let $v \neq 1$; if v^2 is a factor of the Fibonacci infinite word \mathbf{f} then there exists n such that $|v| = |\mathbf{f}_n|$; more precisely $v = wz$ with $zw = \mathbf{f}_n$ for some words z and w , $|w| > 0$, i. e. v is a conjugate of \mathbf{f}_n .*

Now let $u \neq 1$, $u \in A^*$ and let $u = x_1 \dots x_n$, $x_i \in A$; we denote by \hat{u} the mirror image of u , that is $x_n \dots x_1$.

We say that a factor u of \mathbf{f} is special if ua and ub are both factors of \mathbf{f} .

PROPOSITION 3 (Berstel [1]): *If u is a special factor of the Fibonacci infinite word \mathbf{f} then \hat{u} is a left factor of \mathbf{f} .*

Since the sequence $|\mathbf{f}_n|$, $n \in \mathbb{N}$, is the sequence of Fibonacci numbers, we have the following proposition.

PROPOSITION 4 (Hardy and Wright [5]): *For any $n > 1$*

$$\frac{|\mathbf{f}_{n+1}| - 2}{|\mathbf{f}_n|} = \frac{|\mathbf{f}_n| + |\mathbf{f}_{n-1}| - 2}{|\mathbf{f}_n|} < \frac{\sqrt{5} + 1}{2}$$

and

$$\lim_{n \rightarrow \infty} \frac{|\mathbf{f}_n| + |\mathbf{f}_{n-1}| - 2}{|\mathbf{f}_n|} = \frac{\sqrt{5} + 1}{2}.$$

PROPOSITION 5 (de Luca [2]): *For each i the word \mathbf{f}_i is primitive; therefore for each i the conjugates of \mathbf{f}_i are distinct.*

RESULTS AND PROOFS

Let us prove the following lemma.

LEMMA: *No fractional power with exponent greater than $1 + (\sqrt{5} + 1)/2$ can be a left factor of the Fibonacci infinite word \mathbf{f} . More precisely, if vuv is a fractional power which is a left factor of \mathbf{f} then $v = \mathbf{f}_n$ for some n and $|vuv| \leq |\mathbf{f}_n| + |\mathbf{f}_n| + |\mathbf{f}_{n-1}| - 2$.*

Proof: Let vuv be a fractional power which is a left factor of \mathbf{f} .

By using Proposition 2 we have that $|v| = |\mathbf{f}_n|$ for some n , and, consequently vv is a left factor of \mathbf{f} with length $2|\mathbf{f}_n|$. By inspection one can easily see that n is greater than or equal to 3.

As \mathbf{f}_n is a left factor of \mathbf{f} we have that $v = \mathbf{f}_n$ for some $n \geq 3$. Thus $vuv = \mathbf{f}_n \mathbf{f}_n u$ and either u is a left factor of \mathbf{f}_n or \mathbf{f}_n is a left factor of u .

But for $n \geq 3$ $\mathbf{f}_{n+2} = \mathbf{f}_n \mathbf{f}_n \mathbf{g}_{n-1}$ is a left factor of \mathbf{f} .

Hence, since \mathbf{g}_{n-1} is not a left factor of \mathbf{f}_n , we have that u is necessarily a left factor of \mathbf{g}_{n-1} ; by the fact

$$|u| \leq |\mathbf{f}_{n-1}| - 2.$$

Thus $|vuv| \leq |\mathbf{f}_n| + |\mathbf{f}_n| + |\mathbf{f}_{n-1}| - 2$ and, by Proposition 4,

$$\frac{|vuv|}{|v|} \leq \frac{|\mathbf{f}_n| + |\mathbf{f}_n| + |\mathbf{f}_{n-1}| - 2}{|\mathbf{f}_n|} < 1 + \frac{\sqrt{5} + 1}{2}, \quad \square$$

We are now ready to prove our main result.

PROPOSITION 6: *The Fibonacci infinite word \mathbf{f} contains no fractional power with exponent greater than $2 + ((\sqrt{5} + 1)/2)$ and, for any real number $\varepsilon > 0$, it contains a fractional power with exponent greater than $2 + ((\sqrt{5} + 1)/2) - \varepsilon$.*

Proof: Let $vvvu$ be a fractional power factor of \mathbf{f} . As in \mathbf{f} there are no 4 powers (Proposition 1) one can find in \mathbf{f} a factor

$$u'xu''u'xu''u'xu''u'y$$

where $u'xu''=v$, u is a left factor of u' , $u'' \in \{a, b\}^*$ and $x, y \in \{a, b\}$ with $x \neq y$.

It follows that $u'xu''u'xu''u'$ is a special factor of \mathbf{f} . By Proposition 3, $\hat{u}'\hat{u}''x\hat{u}'\hat{u}''x\hat{u}'$ is a left factor of \mathbf{f} . From the Lemma

$$\frac{|\hat{u}'\hat{u}''x\hat{u}'\hat{u}''x\hat{u}'|}{|\hat{u}'\hat{u}''x|} = \frac{|vvu'|}{|v|} < 1 + \frac{\sqrt{5}+1}{2},$$

and, consequently,

$$\frac{|vvvu|}{|v|} \leq \frac{|vvvu'|}{|v|} < 2 + \frac{\sqrt{5}+1}{2}.$$

At last, for $n \geq 3$, $\mathbf{f}_{n+4} = \mathbf{f}_{n+1}\mathbf{f}_n\mathbf{f}_n\mathbf{g}_{n-1}\mathbf{f}_{n-1}\mathbf{f}_n$.

Hence, for $n \geq 3$, $\mathbf{f}_n\mathbf{f}_n\mathbf{f}_n\mathbf{v}_{n-1}$ is always a factor of \mathbf{f} .

Since

$$\frac{|\mathbf{f}_n\mathbf{f}_n\mathbf{f}_n\mathbf{v}_{n-1}|}{|\mathbf{f}_n|} = 2 + \frac{|\mathbf{f}_n| + |\mathbf{f}_{n-1}| - 2}{|\mathbf{f}_n|},$$

the second part of the proposition follows from Proposition 4. \square

In the proof of the above proposition we used the fact that for $n \geq 3$, $\mathbf{f}_n\mathbf{f}_n\mathbf{f}_n\mathbf{v}_{n-1}$ is a factor of \mathbf{f} . As a consequence all words of the form $wzwwz$ with $zw = \mathbf{f}_n$ and $|z| \leq |\mathbf{v}_{n-1}|$ are factors of \mathbf{f} ; by Proposition 5 all these words are distinct. Since $0 \leq |z| \leq |\mathbf{v}_{n-1}|$, the number of these words is $|\mathbf{v}_{n-1}| + 1$.

Let us suppose that vvv is a factor of \mathbf{f} and that $|v| = |\mathbf{f}_n|$ for some $n \geq 3$. By proposition 2, $v = wz$, $|w| > 0$, and $zw = \mathbf{f}_n$.

Suppose that $|z| > |\mathbf{v}_{n-1}|$; since $\mathbf{f}_n = \mathbf{f}_{n-1}\mathbf{f}_{n-2} = \mathbf{v}_{n-1}yx\mathbf{f}_{n-2}$ with $x, y \in \{a, b\}$ and $x \neq y$, we can write $\mathbf{f}_n = \mathbf{v}_{n-1}yuw$ with $z = \mathbf{v}_{n-1}yu$ and, consequently, $vvv = w\mathbf{v}_{n-1}yuw\mathbf{v}_{n-1}yuw\mathbf{v}_{n-1}yu$.

We know that $\mathbf{f}_n\mathbf{f}_n\mathbf{f}_n\mathbf{g}_{n-1} = \mathbf{v}_{n-1}yuw\mathbf{v}_{n-1}yuw\mathbf{v}_{n-1}yuw\mathbf{v}_{n-1}xy$ is a factor of \mathbf{f} ; thus $w\mathbf{v}_{n-1}yuw\mathbf{v}_{n-1}yuw\mathbf{v}_{n-1} = w\mathbf{v}_{n-1}(yuw\mathbf{v}_{n-1})^2$ is a special factor and by Proposition 3 its mirror image must be a prefix of \mathbf{f} . This is impossible by the Lemma because $|w| > 0$.

Hence we have proved the following proposition.

PROPOSITION 7: For $n \geq 3$ the number of distinct factors v of \mathbf{f} with length $|\mathbf{f}_n|$ such that vvv is also a factor of \mathbf{f} is exactly $|\mathbf{v}_{n-1}| + 1$. More precisely they are all the words of the form wz with $zw = \mathbf{f}_n$ and $|z| \leq |\mathbf{v}_{n-1}|$.

OBSERVATION: As $2 + ((\sqrt{5} + 1)/2)$ is an irrational number it cannot exist a fractional power with exponent equal to it.

In the Thue infinite word \mathbf{t} on a two letters alphabet A there are clearly squares but there are no overlaps (that is factors like $xvxy$, $x \in A$, $v \in A^*$). On the contrary it is easy to see that, for any $\varepsilon > 0$, in the Thue infinite word \mathbf{m} on a three letters alphabet there exists a fractional power with exponent greater than $2 - \varepsilon$ but it is a classical result that \mathbf{m} is square free.

Remark: Proposition 6 and 7 were firstly proved by using techniques of Sturmian words. Following the suggestion of P. Séébold we tried to find a simpler proof; actually our proof is simpler than the previous one and use only elementary properties of the Fibonacci infinite word.

ACKNOWLEDGEMENTS

We thank P. Séébold for his useful suggestions and Professor J. Justin for his friendly advice.

REFERENCES

1. J. BERSTEL, Mots de Fibonacci, *Séminaire d'informatique théorique*, L.I.T.P., Paris, Année 1980/1981, pp. 57-78.
2. A. DE LUCA, A combinatorial property of the Fibonacci words, *Inform. Process. Lett.*, 1981, 12, n. 4, pp. 193-195.
3. F. DEJEAN, Sur un théorème de Thue, *J. Comb. Theory*, Ser. A, 1972, 13, pp. 90-99.
4. J. KARHUMÄKI, On cube-free ω -words generated by binary morphism, *Discr. Appl. Math.*, 1983, 5, pp. 279-297.
5. G. H. HARDY and E. M. WRIGHT, An Introduction to the theory of Numbers, Oxford University Press, Fifth edition, 1983.
6. LOTHAIRES, Combinatorics on words, *Addison Wesley*, 1983.
7. A. RESTIVO and S. SALEMI, Overlap-free words on two symbols, *Lecture Notes in Comput. Sci.*, 1984, 192, pp. 198-206.
8. P. SÉÉBOLD, Propriétés combinatoires des mots infinis engendrés par certains morphismes, *Thèse de doctorat*, Rapp. Tec. L.I.T.P., 85-14, 1985.
9. A. THUE, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat.-Nat. Kl.*, Christiania 1906, Nr. 7. pp. 1-22.
10. A. THUE, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat.-Nat. Kl.*, Christiania 1912, Nr. 1, pp. 1-67.

Sur les produits directs de semi-groupes répétitifs

Giuseppe PIRILLO

Résumé — Nous prouvons que si S et T sont deux semi-groupes finiment engendrés dont le produit direct est répétitif, alors l'un au moins des deux semi-groupes de S et T est fini.

On direct products of repetitive semigroups

Abstract — We prove that if S and T are finitely generated semi-groups and their direct product is repetitive then S is finite or T is finite.

INTRODUCTION. — La notion de semi-groupe répétitif a été introduite par Justin qui l'a étudiée dans les articles [2] à [9] et, en collaboration avec l'auteur de cette Note, dans [10] et [11]. En particulier, il est prouvé dans [7] que les semi-groupes finis ainsi que le groupe additif des entiers sont répétitifs.

Donc, la finitude d'un semi-groupe finiment engendré est suffisante pour la répétitivité mais n'est pas nécessaire.

Plusieurs conditions, plus fortes que la répétitivité et qui assurent la finitude d'un semi-groupe finiment engendré, ont été étudiées, par exemple dans [11].

Le but de ce travail est de prouver que si le produit direct de deux semi-groupes finiment engendrés est répétitif, alors au moins l'un d'eux est fini. Donc, la finitude d'un semi-groupe finiment engendré S est une condition nécessaire et suffisante pour la répétitivité du produit direct $S \times S$.

DÉFINITIONS ET RÉSULTATS. — En utilisant la terminologie et les notations de Lothaire [13], notons A^* (resp. A^+) le *monoïde libre* (resp. *semi-groupe libre*) sur l'alphabet A . Un élément de A est une *lettre* et un élément de A^* est un *mot*. Nous notons $|u|$ la *longueur* du mot u .

Un mot u est *facteur* d'un mot v s'il existe deux mots u' et u'' tels que

$$v = u'uu''.$$

Un mot *infini* (à droite) s sur l'alphabet A est une application de l'ensemble des entiers naturels \mathbb{N} dans A . Nous écrivons

$$s = s(0)s(1) \dots s(i) \dots$$

où $s(i)$ est la lettre « de rang i » dans s .

Un mot u de A^+ est un *facteur* du mot infini s s'il existe deux entiers naturels $i \leq j$ tels que

$$u = s(i) \dots s(j).$$

Le lemme suivant est une conséquence bien connue du Lemme de Koenig.

LEMME 1. — Si A est un alphabet fini et si E est un sous-ensemble infini de A^+ , alors il existe un mot infini s tel que tout facteur de s est facteur d'au moins un mot de E .

Soient maintenant G un ensemble de générateurs d'un semi-groupe S et φ le morphisme de G^+ sur S défini par

$$\varphi(g) = g$$

Note présentée par Marcel-Paul SCHÜTZENBERGER.

pour tout $g \in G$.

Nous disons que $u \in G^+$ est *irréductible* si pour tout mot $w \in G^+$ tel que $\varphi(w) = \varphi(u)$ on a

$$|u| \leq |w|.$$

Le lemme suivant, déjà utilisé dans [2] par exemple, est une conséquence immédiate du lemme 1 et du fait qu'un facteur d'un mot irréductible est aussi irréductible.

LEMME 2. — Soient S un semi-groupe infini et G un ensemble fini de générateurs de S . Alors il existe un mot infini s sur G tel que tout facteur de s est irréductible.

DÉFINITION 1. — Soient S un semi-groupe, A un alphabet, $\varphi: A^+ \rightarrow S$ un morphisme et k un entier supérieur ou égal à 2. On dit qu'un mot $w \in A^+$ est une **puissance k -ième modulo φ** s'il existe w_1, w_2, \dots, w_k , mots de A^+ , tels que

$$w = w_1 w_2 \dots w_k$$

et

$$\varphi(w_1) = \varphi(w_2) = \dots = \varphi(w_k).$$

Si S est le semi-groupe abélien libre sur A et φ le morphisme induit par l'identité sur A , on dit que w est une **puissance k -ième abélienne**.

DÉFINITION 2. — Un semi-groupe S est **répétitif** si pour tout entier $k \geq 2$, pour tout alphabet fini A et pour tout morphisme $\varphi: A^+ \rightarrow S$, tout mot infini sur A contient un facteur qui est une puissance k -ième modulo φ .

Il est possible de prouver, en utilisant le théorème de Ramsey ou de façon plus directe, que tout semi-groupe fini est répétitif et, à l'aide du théorème de Van der Waerden, que le semi-groupe additif des entiers naturels et le groupe additif des entiers sont répétitifs [7].

Donc, même pour un groupe finiment engendré, la condition d'être répétitif n'entraîne pas sa finitude. Par contre, certaines conditions plus fortes que la répétitivité (« forte répétitivité » [11], « uniforme répétitivité » [12]) assurent la finitude d'un semi-groupe finiment engendré.

Nous prouvons ici les propositions suivantes :

THÉORÈME. — Si S et T sont deux semi-groupes finiment engendrés dont le produit direct $S \times T$ est répétitif, alors l'un au moins des deux semi-groupes S et T est fini.

COROLLAIRE. — Soit S un semi-groupe finiment engendré. Les conditions suivantes sont équivalentes :

- 1) S est fini;
- 2) le produit direct $S \times S$ est répétitif.

PREUVES DES RÉSULTATS. — *Preuve du théorème.* — Soit G (resp. G') un ensemble fini de générateurs de S (resp. T). Supposons, par l'absurde, que S et T soient infinis.

Par le lemme 2, il existe deux mots infinis s et t sur les alphabets G et G' respectivement dont les facteurs sont irréductibles.

En vertu d'un résultat de [1] (voir aussi [8]), soit

$$d = d(0)d(1) \dots d(i) \dots$$

un mot infini sur $\{a, b\}$ sans puissances 4-ièmes abéliennes.

Factorisons s et t de la façon suivante

$$s = s_0 s_1 \dots s_i \dots$$

et

$$t = t_0 t_1 \dots t_i \dots$$

avec

$$|s_i| = \begin{cases} 1 & \text{si } d(i) = a \\ 2 & \text{si } d(i) = b \end{cases}$$

et

$$|t_i| = \begin{cases} 1 & \text{si } d(i) = b \\ 2 & \text{si } d(i) = a \end{cases}$$

L'ensemble $\{(s_i, t_i) \mid i \geq 0\}$ est fini. Comme $S \times T$ est répétitif il existe des entiers $\alpha < \beta < \gamma < \delta < \varepsilon$ tels que

$$s_\alpha \dots s_{\beta-1} = s_\beta \dots s_{\gamma-1} = s_\gamma \dots s_{\delta-1} = s_\delta \dots s_{\varepsilon-1}$$

et

$$t_\alpha \dots t_{\beta-1} = t_\beta \dots t_{\gamma-1} = t_\gamma \dots t_{\delta-1} = t_\delta \dots t_{\varepsilon-1}$$

Ces facteurs étant irréductibles, on obtient

$$|s_\alpha \dots s_{\beta-1}| = |s_\beta \dots s_{\gamma-1}| = |s_\gamma \dots s_{\delta-1}| = |s_\delta \dots s_{\varepsilon-1}|$$

et

$$|t_\alpha \dots t_{\beta-1}| = |t_\beta \dots t_{\gamma-1}| = |t_\gamma \dots t_{\delta-1}| = |t_\delta \dots t_{\varepsilon-1}|$$

En utilisant les définitions des factorisations de s et t on en déduit un système d'équations linéaires concernant les nombres d'occurrences de a et de b dans les mots

$$d(\alpha) \dots d(\beta-1)$$

$$d(\beta) \dots d(\gamma-1)$$

$$d(\gamma) \dots d(\delta-1)$$

$$d(\delta) \dots d(\varepsilon-1).$$

Il en résulte que ces mots sont abéliennement équivalents, c'est-à-dire, que d contient une puissance 4-ième abélienne. Contradiction.

Preuve du corollaire. — 1) \Rightarrow 2). — Si S est fini alors le produit direct $S \times S$ est fini et, donc, répétitif par un résultat déjà mentionné de [7].

2) \Rightarrow 1). — C'est une conséquence immédiate du théorème 1.

L'auteur remercie J. Justin dont les conseils ont été très utiles pour la rédaction de ce travail.

Note remise le 7 octobre 1991, acceptée le 22 octobre 1991.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] F. M. DEKKING, Strongly non repetitive sequences and progression-free sets, *J. Comb. Theory*, A, 27, 1979, p. 181-185.
- [2] J. JUSTIN, Propriétés combinatoires de certains semi-groupes, *C. R. Acad. Sci. Paris*, 269, série A, 1969, p. 1113-1115.
- [3] J. JUSTIN, Semi-groupes à générations bornées, dans *Problèmes Mathématiques de la Théorie des Automates, Séminaire Schützenberger, Lentin, Nivat 69/70*, Institut Henri-Poincaré, Paris, exposé n° 7, 1970, 10 p.
- [4] J. JUSTIN, Sur une Construction de Bruck et Reilly, *Semigroup Forum*, 3, 1971, p. 148-155.
- [5] J. JUSTIN, Groupes et semi-groupes à croissance linéaire, *C. R. Acad. Sci. Paris*, 273, série A, 1971, p. 212-214.

-
- [6] J. JUSTIN, Semi-groupes répétitif, dans *Logique et Automates, Séminaire I.R.I.A.*, Institut de Recherche d'Informatique et d'Automatique, Le Chesnay, France, 1971, p. 101-108.
- [7] J. JUSTIN, Généralisation du théorème de Van der Waerden sur les semi-groupes répétitifs, *J. Comb. Theory*, 12, 1972, p. 357-367.
- [8] J. JUSTIN, Characterization of repetitive commutative semigroups, *J. Algebra*, 21, 1972, p. 87-90.
- [9] J. JUSTIN, Groupes linéaires répétitifs, *C. R. Acad. Sci. Paris*, 292, série I, 1981, p. 349-350.
- [10] J. JUSTIN et G. PIRILLO, Two combinatorial properties of partitions of the free semigroup into finitely many parts, *Discrete Mathematics*, 52, 1984, p. 299-303.
- [11] J. JUSTIN et G. PIRILLO, On a Natural Extension of Jacob's Ranks, *J. Comb. Theory*, 43, 1986, p. 205-218.
- [12] J. JUSTIN, G. PIRILLO et S. VARRICCHIO, Unavoidable regularities and finiteness conditions for semi-groups, *Proceedings of the Third Italian Conference "Theoretical Computer Science"*, Mantova, BERTONI-BÖHM-MIGLIOLI éd., World Scientific, 2-4 nov. 1989.
- [13] M. LOTHAIRE, *Combinatorics on words*, Addison-Wesley, 1983.

I.A.G.A.-I.A.M.I. (C.N.R.), viale Morgagni 67/A, 50134 Firenze, Italia.

UNAVOIDABLE REGULARITIES AND FINITENESS CONDITIONS FOR SEMIGROUPS

Jacques JUSTIN, Giuseppe PIRILLO and Stefano VARRICCHIO

ABSTRACT. We prove that finitely generated and uniformly repetitive semigroups are finite.

1. INTRODUCTION

A classical problem in combinatorics on words consists in finding new unavoidable regularities, i.e. properties of the words of a free monoid necessarily verified by the "long enough" ones. Such regularities have applications in the Burnside problem and, more generally, in the study of finiteness conditions for finitely generated semigroups.

The interest of these problems in theoretical computer science is clear when one considers a well-known classical theorem: a language on a finite alphabet is recognizable by a finite automata if and only if its syntactic monoid is finite.

On the other hand, results in this field are due to Coudrain and Schützenberger [1], Simon [14], Hotzel [4], de Luca and Restivo [2], Restivo and Reutenauer [12], just to point out few examples. Very recent contributions can be found in [15] and [8] (see also [9]), hereby briefly summarized.

A theorem of Shirshov reads as follows: Any long enough word over a finite alphabet contains a factor which is n-divided or is a p-th power (see, for example chapter 7 in [10]); it should be stressed that using the properties of the Lyndon words Reutenauer has proved the theorem of Shirshov in a very elegant way, see [13].

Varricchio in [15], generalizing the techniques of Reutenauer, extended it to the one side infinite words, while Justin and Pirillo in [8] (see also [9]), using elementary techniques of symbolic dynamics, extended it to the two side infinite words.

The Restivo-Reutenauer theorem (a finitely generated and periodic semigroup is finite if and only if it is permutable, see [12]) was proved using the Shirshov theorem. In the frame of the above mentioned generalizations of this theorem, weaker conditions than permutability, but strong enough to ensure the finiteness of a finitely generated and periodic semigroup, are considered in [15], [8] and [9].

For definitions of n -divided word, permutable semigroup etc. and for proofs, see [12], [15], [8] and also [9].

This paper is devoted to a new finiteness condition for finitely generated semigroups, called uniform repetitivity.

2. UNIFORMLY REPETITIVE SEMIGROUPS

We denote by A^+ (resp A^*) the free semigroup (resp. free monoid) on the alphabet A . The elements of A^* are called words and the length of a word w of A^* is denoted by $|w|$. A word v is a factor of a word w if there exist two words u and u' of A^* such that $w = u v u'$. A sequence (or infinite word on the right) on an alphabet A is a mapping from $\mathbb{N} - \{0\}$ into A . If s is a sequence, we write

$$s = s(1) s(2) \dots s(i) \dots$$

We say that a word v over A is a factor of a sequence s over A if there exist positive integers i and j , $i \leq j$, such that $v = s(i) \dots s(j)$.

Now, let us recall the following definitions.

DEFINITION. A semigroup S is repetitive (resp. uniformly repetitive) if for each finite alphabet A , for each morphism $\varphi: A^+ \rightarrow S$ and for each positive integer k there exists an integer $R(\varphi, k)$ such that each word w of A^+ of length at least $R(\varphi, k)$ contains k consecutive factors of positive length, say w_1, w_2, \dots, w_k , such that $\varphi(w_1) = \dots = \varphi(w_k)$ (resp. $\varphi(w_1) = \dots = \varphi(w_k)$ and $|w_1| = \dots = |w_k|$).

Finite semigroups are repetitive, but also the infinite cyclic semigroup is repetitive (see [5], where this is proved using the van der Waerden theorem on arithmetic progressions); so repetitivity is not a finiteness condition for finitely generated

semigroups.

In any case, many other repetition properties are of this type (for example, ramseyan or strongly repetitive or ω -repetitive finitely generated semigroups are finite, see [7]).

It is also known that uniformly repetitive semigroups are periodic, see [11]. The aim of this paper is to prove that these semigroups, when finitely generated, are finite.

To do this, we need the following definition and result.

DEFINITION. Two words u and v over an alphabet A are abelianly equivalent if and only if for each letter a of A the number of occurrences of a in u is exactly that of a in v (if and only if, in other words, u is a permutation of v). We say that a word w over A is an abelian n -power if there exist n words of positive length, say w_1, \dots, w_n , such that $w = w_1 \dots w_n$ and for each i and j , $1 \leq i, j \leq n$, the word w_i is abelianly equivalent to w_j .

THEOREM. Let $A = \{a, b\}$ a two letters alphabet. Then there exists a sequence s over A without factors which are abelian 4-power.

This theorem is due to Dekking [3] and improves an analogous result of Justin (see [6]) concerning a sequence over a two letters alphabet without abelian 5-power.

Now, we can state and prove our main theorem.

THEOREM. Let S be a finitely generated semigroup. Then S is uniformly repetitive if and only if it is finite.

Proof. The "if" part has been shown in chapter 4 of [10], using methods of [5].

For the proof of the "only if" part, let S be a semigroup, A be a finite set of generators of S and $\varphi: A^+ \rightarrow S$ be the canonical (epi)-morphism defined by $\varphi(a) = a$ for each $a \in A$.

We suppose by contradiction that S is uniformly repetitive and infinite. In these hypotheses the set L of the irreducible words is infinite (a word w is called irreducible if $\varphi(w) \neq \varphi(w')$ for any w' such that $|w'| < |w|$), otherwise S would be finite. By the König Lemma, as L is closed by factors, there exists an infinite word m on A such that all factors of m are irreducible.

By the recalled result of [3], there exists a sequence on the alphabet $\{1, 2\}$ that does not contain factors which are abelian 4-powers, let it be

$$l = l(1)l(2)\dots l(n)\dots$$

We can factorize m as

$$m = x_1 x_2 \dots x_n \dots$$

where $x_i \in A \cup A^2$ and $|x_i| = 1(i)$.

Let B be an alphabet such that $\text{card}(B) = \text{card}(A \cup A^2)$ and $\delta: B \rightarrow A \cup A^2$ be a bijection. Let $\psi: B^* \rightarrow A^*$ be the morphism generated by the bijection δ . Let us consider the morphism

$$\gamma: B^* \rightarrow S$$

defined by $\gamma(w) = \varphi(\psi(w))$ for $w \in B^*$.

Let

$$t = t(1) t(2) \dots t(n) \dots$$

where $t(i) = \delta^{-1}(x_i)$.

Since γ is uniformly repetitive, there exist consecutive factors w_1, w_2, w_3 and w_4 of t such that

$$|w_1| = |w_2| = |w_3| = |w_4| > 0 \quad (1)$$

and

$$\gamma(w_1) = \gamma(w_2) = \gamma(w_3) = \gamma(w_4).$$

Now, if we pose $y_i = \psi(w_i)$ for $i = 1, \dots, 4$, we have

$$\varphi(y_1) = \varphi(y_2) = \varphi(y_3) = \varphi(y_4) \quad (2)$$

and $y_1 y_2 y_3 y_4$ is a factor of positive length of m .

So, as y_1, y_2, y_3 and y_4 are irreducible, we must have

$$|y_1| = |y_2| = |y_3| = |y_4|. \quad (3)$$

If we set for $i = 1, \dots, 4$ the integers j_i such that

$$w_i = x_{j_i+1} \dots x_{j_{i+1}}$$

and if we denote by a_i (resp. b_i) the number of 1 (resp. 2) that appear in $l(j_i+1) \dots l(j_{i+1})$, then from (1) and (3) we obtain, for $r, s \in \{1, \dots, 4\}$,

$$a_r + b_r = a_s + b_s,$$

$$a_r + 2b_r = a_s + 2b_s.$$

So, we get

$$a_r = a_s \text{ and } b_r = b_s,$$

therefore

$$(l(j_1+1)...l(j_2)) (l(j_2+1)...l(j_3))(l(j_3+1)...l(j_4))(l(j_4+1)...l(j_5))$$

is an abelian 4-power and that is a contradiction.

REFERENCES

1. M. COUDRAIN and M.P. SCHÜTZENBERGER, Une condition de finitude des monoides finiment engendrés, C. R. Acad. Sci. Paris, Série A, 262(1966) 1149-1151.
2. A. de LUCA and A. RESTIVO, A finiteness condition for finitely generated semigroups, Semigroup Forum, 28(1984) 123-134.
3. F.M. DEKKING, Strongly non repetitive sequences and progression-free sets, J. Comb. Theory, Ser. A, 27(1979)181-185.
4. E. HOTZEL, On finiteness conditions in semigroups, J. Algebra, 60(1979)352-370.
5. J. JUSTIN, Généralization du théorème de van der Waerden sur les semigroupes répétitif, J. Comb. Theory, 12(1972), 357-367.
6. J. JUSTIN, Characterization of the repetitive commutative semigroups, J. Algebra, 21(1972),87-90.
7. J. JUSTIN and G. PIRILLO, On a Natural Extension of Jacob's Ranks, J. Comb. Theory, 43(1986), 205-218.
8. J. JUSTIN and G. PIRILLO, Shirshov's Theorem and Ω -Permutability of Semigroups, Advances in Mathematics, to appear.
9. J. JUSTIN and G. PIRILLO, Théorème de Shirshov et Ω -permutabilité des semigroupes, Publication of LITP-Univ. Paris VII,89-22(mars 1989).
10. M. LOTHAIRE, Combinatorics on words, Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, London, 1983.
11. G. PIRILLO, The van der Waerden Theorem and the Burnside Problem for semigroups, Arch. Math., to appear.
12. A. RESTIVO and C. REUTENAUER, On the Burnside Problem for semigroups, J. Algebra, 89(1984)102-104.

13. C. REUTENAUER, Mots de Lyndon et théorème de Shirshov, Ann. Sc. Math. 10,2(1986)237-245.

14. I. SIMON, Conditions de finitude pour des semi-groupes, C.R. Acad. Sci. Paris, Série A, 290(1980) 1081-1082.

15. S. VARRICCHIO, Factorizations of free monoids and unavoidable regularities, Theoretical Computer Science, to appear.

J. JUSTIN
(LITP-CNRS-Univ. ParisVII)
19,rue de Bagnaux
92330 SCEAUX
FRANCE

G. PIRILLO
IAGA-IAMI-CNR
Viale Morgagni, 67/A
50134 FIRENZE
ITALIA

S. VARRICCHIO
Dipartimento di Matematica
Univ. di Roma "La Sapienza"
P.le A. Moro,2
OO185 ROMA
ITALIA

Références

- [1] P. Alessandri, V. Berthé. Three distance theorems and combinatorics on words. *Enseign. Math. (2)*, 44(1-2) :103–132, 1998.
- [2] J.-P. Allouche and J. Shallit. *Automatic Sequences : Theory, Applications, Generalizations*. Cambridge University Press, 2003.
- [3] P. Arnoux, G. Rauzy. Représentation géométrique de suites de complexité $2n + 1$. *Bull. Soc. Math. France*, 119(2) :199–215, 1991.
- [4] D. Arquès, C. Michel. A complementary circular code in the protein coding genes. *J. Theor. Biol.*, 182, 1996.
- [5] J. Berstel. Mots de Fibonacci. Séminaire d’informatique théorique Paris, 1980-1981. 57-78.
- [6] J. Berstel. Mots infinis. In *Théorie des langages et complexité des algorithmes*, pages 89–102, Lyon, France, 1984.
- [7] J. Berstel. *The Book of L*, chapter Fibonacci words - a survey, pages 13–27. Springer-Verlag, 1986.
- [8] J. Berstel. Properties of infinite words : recent results. In *International Symposium on Theoretical Aspects of Computer Science (STACS 1989)*, volume 349 of *LNCS*, pages 36–46. Springer-Verlag, 1989.
- [9] J. Berstel. Axel Thue’s work on repetitions in words. In *Séries Formelles et Combinatoire Algébrique*, volume 11 of *Monographies du LaCIM*, pages 65–80. 1992.
- [10] J. Berstel. Axel Thue’s papers on repetitions in words : a translation. volume 20 of *Monographies du LaCIM*. 1995. 85 pages.
- [11] J. Berstel. On the index of Sturmian words. In *Jewels are forever*, pages 287–294. Springer-Verlag, Berlin, 1999.
- [12] J. Berstel. An exercise on Fibonacci representations. *Informatique Theorique et Applications*, 35(6) :491–498, 2001. A tribute to Aldo de Luca.
- [13] J. Berstel. Recent results on extensions of Sturmian words. *Internat. J. Algebra Comput.*, 12(1-2) :371–385, 2002. International Conference on Geometric and Combinatorial Methods in Group Theory and Semi-group Theory (Lincoln, NE, 2000).
- [14] J. Berstel, A. de Luca. Sturmian words, Lyndon words and trees. *Theoretical Computer Science*, 178(1-2) :171–203, 1997.

- [15] J. Berstel, J. Karhumäki. Combinatorics on words - a tutorial. *Bulletin of the European Association of Theoretical Computer Science*, (79) :178–228, 2003.
- [16] J. Berstel, D. Perrin, J. F. Perrot, A. Restivo. Sur le théorème du défaut. *Journal of Algebra*, 60(1) :169–180, 1979.
- [17] J. Berstel, D. Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1985.
- [18] J. Berstel, M. Pocchiola. A geometric proof of the enumeration formula for Sturmian words. *International Journal of Algebra and Computation*, 3(3) :349–355, 1993.
- [19] J. Berstel, P. Séébold. A characterization of Sturmian morphisms. In *International Symposium on Mathematical Foundations of Computer Science (MFCS 1993)*, volume 711 of *LNCS*, pages 281–290. Springer-Verlag, 1993.
- [20] J. Berstel, P. Séébold. A remark on morphic Sturmian words. *RAIRO Inform. Théor. Appl.*, 28(3-4) :255–263, 1994.
- [21] J. Berstel, P. Séébold. *Algebraic combinatorics on words*, chapter Sturmian words, pages 40–97. Cambridge University Press, 2002.
- [22] J. Berstel, L. Boasson. M.-P. Schützenberger et la théorie des langages algébriques, Jean Berstel, page web, 2005.
- [23] V. Berthé. Autour du système de numération d’Ostrowski. *Bull. Belg. Math. Soc. Simon Stevin*, 8(2) :209–239, 2001. Journées Montoises d’Informatique Théorique (Marne-la-Vallée, 2000).
- [24] R. D. Blyth. Rewriting products of group elements. I. *J. Algebra*, 116(2) :506–521, 1988.
- [25] R. D. Blyth. Rewriting products of group elements. II. *J. Algebra*, 119(1) :246–259, 1988.
- [26] J.-P. Borel, F. Laubie. Quelques mots sur la droite projective réelle. *J. Théor. Nombres Bordeaux*, 5(1) :23–51, 1993.
- [27] L. Carlitz. Fibonacci representations. *Fibonacci Quart*, 6(4) :193–220, 1968.
- [28] C. Choffrut, J. Karhumäki. Combinatorics of words. In *Handbook of formal languages, Vol. 1*, pages 329–438. Springer, Berlin, 1997.

- [29] M. Coudrain, M.-P Schützenberger. Une condition de finitude des monoïdes finiment engendrés. *C. R. Acad. Sci. Paris Sér. A-B*, 262 :A1149–A1151, 1966.
- [30] D. Crisp, W. Moran, A. Pollington, P. Shiue. Substitution invariant cutting sequences. *J. Théor. Nombres Bordeaux*, 5(1) :123–137, 1993.
- [31] M. Crochemore, J. Désarménien, D. Perrin. A note on the Burrows-Wheeler transformation. *Theoret. Comput. Sci.*, 332(1-3) :567–572, 2005.
- [32] M. Crochemore and W. Rytter. *Text algorithms*. Oxford University Press, 1994.
- [33] M. Crochemore and W. Rytter. *Jewels in Stringology*. World Scientific, 2002.
- [34] A. de Luca. A combinatorial property of the Fibonacci words. *Inform. Process. Lett.*, 12(4) :193–195, 1981.
- [35] A. de Luca. Standard Sturmian morphisms. *Theoret. Comput. Sci.*, 178(1-2) :205–224, 1997.
- [36] A. de Luca. Sturmian words : structure, combinatorics, and their arithmetics. *Theoret. Comput. Sci.*, 183(1) :45–82, 1997.
- [37] A. de Luca, A. Restivo. A finiteness condition for finitely generated semigroups. *Semigroup Forum*, 28(1-3) :123–134, 1984.
- [38] A. de Luca, S. Varricchio. Combinatorial properties of uniformly recurrent words and an application to semigroups. *Internat. J. Algebra Comput.*, 1(2) :227–245, 1991.
- [39] A. de Luca, S. Varricchio. Finiteness and iteration conditions for semigroups. *Theoret. Comput. Sci.*, 87(2) :315–327, 1991.
- [40] A. de Luca, S. Varricchio. A finiteness condition for semigroups generalizing a theorem of Hotzel. *J. Algebra*, 136(1) :60–72, 1991.
- [41] A. de Luca, S. Varricchio. A finiteness condition for semigroups generalizing a theorem of Coudrain and Schützenberger. *Adv. Math.*, 108(1) :91–103, 1994.
- [42] A. de Luca, S. Varricchio. Regularity and finiteness conditions. In *Handbook of formal languages, Vol. 1*, pages 747–810. Springer, Berlin, 1997.

- [43] A. de Luca, S. Varricchio. *Finiteness and regularity in semigroups and formal languages*. Monographs in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 1999.
- [44] P. Degano, G. Pirillo, U. Vaccaro, editors. *Theoretical computer science*, River Edge, NJ, 1998. World Scientific Publishing Co. Inc.
- [45] F. Dejean. Sur un théorème de Thue. *J. Combinatorial Theory Ser. A*, 13 :90–99, 1972.
- [46] X. Droubay, J. Justin, G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1-2) :539–553, 2001.
- [47] X. Droubay, G. Pirillo. Palindromes and Sturmian words. *Theoret. Comput. Sci.*, 223(1-2) :73–85, 1999.
- [48] I. Fagnot. A little more about morphic sturmian words. preprint, 2002.
- [49] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16 :109–114, 1965.
- [50] H. Fürstenberg. Poincaré recurrence and number theory. *Bull. Amer. Math. Soc. (N.S.)*, 5(3) :211–234, 1981.
- [51] S. Gan. Sturmian sequences and the lexicographic world. *Proc. Amer. Math. Soc.*, 129(5) :1445–1451 (electronic), 2001.
- [52] A. Glen, J. Justin, G. Pirillo. Characterizations of finite and infinite episturmian words via lexicographic orderings. *European Journal of Combinatorics*. sous presse.
- [53] R. L. Graham. *Rudiments of Ramsey theory*, volume 45 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, Providence, R.I., 1981.
- [54] R. L. Graham, B. L. Rothschild, J. H. Spencer. *Ramsey theory*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, second edition, 1990. A Wiley-Interscience Publication.
- [55] G. H. Hardy, E. M. Wright. *An introduction to the theory of numbers*. The Clarendon Press Oxford University Press, New York, fifth edition, 1979.
- [56] K. Hashiguchi. Notes on finitely generated semigroups and pumping conditions for regular languages. *Theoret. Comput. Sci.*, 46(1) :53–66, 1986.

- [57] G. Inglese, G. Pirillo. A remark on a result by A. Thue. *Pure Math. Appl. Ser. A*, 1(1) :21–28, 1990.
- [58] J. Justin. Bornes des coefficients du polynôme cyclotomique et de certains autres polynômes. *C. R. Acad. Sci. Paris Sér. A-B*, 268 :A995–A997, 1969.
- [59] J. Justin. Propriétés combinatoires de certains semi-groupes. *C. R. Acad. Sci. Paris Sér. A-B*, 269 :A1113–A1115, 1969.
- [60] J. Justin. Semi-groupes à générations bornées. In *Séminaire M. P. Schützenberger, A. Lentin et M. Nivat, 1969/70 : Problèmes Mathématiques de la Théorie des Automates*, pages exp. 7, 10. Secrétariat mathématique, Paris, 1970.
- [61] J. Justin. Groupes et semi-groupes à croissance linéaire. *C. R. Acad. Sci. Paris Sér. A-B*, 273 :A212–A214, 1971.
- [62] J. Justin. Sur une construction de Bruck et Reilly. *Semigroup Forum*, 3(2) :148–155, 1971/72.
- [63] J. Justin. Characterization of the repetitive commutative semigroups. *J. Algebra*, 21 :87–90, 1972.
- [64] J. Justin. Généralisation du théorème de van der Waerden sur les semi-groupes répétitifs. *J. Combinatorial Theory Ser. A*, 12 :357–367, 1972.
- [65] J. Justin. Groupes linéaires répétitifs. *C. R. Acad. Sci. Paris Sér. I Math.*, 292(6) :349–350, 1981.
- [66] J. Justin. On a paper by M. Castelli, F. Mignosi, A. Restivo. Comment on : “Fine and Wilf’s theorem for three periods and a generalization of Sturmian words” [Theoret. Comput. Sci. **218** (1999), no. 1, 83–94; MR1687764 (2000c :68110)]. *Theor. Inform. Appl.*, 34(5) :373–377, 2000.
- [67] J. Justin. Episturmian words and morphisms (results and conjectures). In *Algebraic combinatorics and computer science*, pages 533–539. Springer Italia, Milan, 2001.
- [68] J. Justin. Episturmian morphisms and a Galois theorem on continued fractions. *Theor. Inform. Appl.*, 39(1) :207–215, 2005.
- [69] J. Justin, A. Kelarev. Factorization of infinite sequences in semigroups. *Ann. Mat. Pura Appl. (4)*, 174 :87–96, 1998.

- [70] J. Justin, B. Piochi, G. Pirillo. On fixed-point-free permutation properties in groups and semigroups. *Ann. Mat. Pura Appl. (4)*, 159 :45–64, 1991.
- [71] J. Justin, B. Piochi, G. Pirillo. Propriétés de permutabilité avec certaines restrictions. *Arch. Math. (Basel)*, 57(1) :14–20, 1991.
- [72] J. Justin, G. Pirillo. Two combinatorial properties of partitions of the free semigroup into finitely many parts. *Discrete Math.*, 52(2-3) :299–303, 1984.
- [73] J. Justin, G. Pirillo. On a natural extension of Jacob’s ranks. *J. Combin. Theory Ser. A*, 43(2) :205–218, 1986.
- [74] J. Justin, G. Pirillo. Une condition de finitude pour les semi-groupes finiment engendrés. *C. R. Acad. Sci. Paris Sér. I Math.*, 306(2) :63–65, 1988.
- [75] J. Justin, G. Pirillo. Comments on the permutation property for semigroups. *Semigroup Forum*, 39(1) :109–112, 1989.
- [76] J. Justin, G. Pirillo. Mots sans carré et permutabilité des semi-groupes. *C. R. Acad. Sci. Paris Sér. I Math.*, 309(7) :523–525, 1989.
- [77] J. Justin, G. Pirillo. A finiteness condition for semigroups generated by a finite set of elements of finite order. *Pure Math. Appl. Ser. A*, 1(1) :45–48, 1990.
- [78] J. Justin, G. Pirillo. Infinite words and permutation properties. *Semigroup Forum*, 40(1) :13–22, 1990.
- [79] J. Justin, G. Pirillo. Some remarks on the permutation property for semigroups. *European J. Combin.*, 11(2) :151–154, 1990.
- [80] J. Justin, G. Pirillo. Shirshov’s theorem and ω -permutability of semigroups. *Adv. Math.*, 87(2) :151–159, 1991.
- [81] J. Justin, G. Pirillo. Intracommutativity properties for groups and semigroups. *J. Algebra*, 153(2) :424–443, 1992.
- [82] J. Justin, G. Pirillo. Factorial languages and some combinatorial properties of semigroups. *Internat. J. Algebra Comput.*, 3(3) :295–316, 1993.
- [83] J. Justin, G. Pirillo. On some questions and conjectures in the combinatorial theory of semigroups. *Southeast Asian Bull. Math.*, 18(3) :91–104, 1994.

- [84] J. Justin, G. Pirillo. On a question about factorization forests. *Theoret. Comput. Sci.*, 141(1-2) :351–355, 1995.
- [85] J. Justin, G. Pirillo. On a combinatorial property of Sturmian words. *Theoret. Comput. Sci.*, 154(2) :387–394, 1996.
- [86] J. Justin, G. Pirillo. Decimations and Sturmian words. *RAIRO Inform. Théor. Appl.*, 31(3) :271–290, 1997.
- [87] J. Justin, G. Pirillo. On some factorizations of infinite words by elements of codes. *Inform. Process. Lett.*, 62(6) :289–294, 1997.
- [88] J. Justin, G. Pirillo. Fractional powers in Sturmian words. *Theoret. Comput. Sci.*, 255(1-2) :363–376, 2001.
- [89] J. Justin, G. Pirillo. Episturmian words and episturmian morphisms. *Theoret. Comput. Sci.*, 276(1-2) :281–313, 2002.
- [90] J. Justin, G. Pirillo. On a characteristic property of Arnoux-Rauzy sequences. *Theor. Inform. Appl.*, 36(4) :385–388 (2003), 2002.
- [91] J. Justin, G. Pirillo. Episturmian words : shifts, morphisms and numeration systems. *Internat. J. Found. Comput. Sci.*, 15(2) :329–348, 2004.
- [92] J. Justin, G. Pirillo, S. Varricchio. Unavoidable regularities and finiteness conditions for semigroups. Proceedings of the Third Italian Conference THEORETICAL COMPUTER SCIENCE, 1989. Mantova, 2–4 November 1989.
- [93] J. Justin, G. Pirillo, S. Varricchio. Some conditions on the product of two elements of a semigroup. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 1(2) :85–87, 1990.
- [94] J. Justin, L. Vuillon. Return words in Sturmian and episturmian words. *Theor. Inform. Appl.*, 34(5) :343–356, 2000.
- [95] J. Karhumäki. On cube-free ω -words generated by binary morphisms. *Discrete Appl. Math.*, 5(3) :279–297, 1983.
- [96] A. Kelarev, G. Pirillo. A note on rewritable products in groups and semigroups. *Boll. Un. Mat. Ital. A (7)*, 11(3) :667–670, 1997.
- [97] D. E. Knuth. *The Art of Computer Programming*, volume 1 : Fundamental Algorithms. Addison-Wesley, 1968.
- [98] D.E. Knuth. Sequences with precisely $k+1$ k -blocks. solution of problem E2307. *Amer. Math. Monthly* 79 (1972), pp. 773–774, 1972.

- [99] D. E. Knuth, J. H. Morris, Jr., V. R. Pratt. Fast pattern matching in strings. *SIAM J. Comput.*, 6(2) :323–350, 1977.
- [100] G. Lallement. *Semigroups and combinatorial applications*. John Wiley & Sons, New York-Chichester-Brisbane, 1979. Pure and Applied Mathematics, A Wiley-Interscience Publication.
- [101] A. Lentin. *Equations dans le monoïdes libres*. Gauthier-Villars, Paris.
- [102] A. Lentin, M. Gross. *Notions sur les grammaire formelles*. Gauthier-Villars, Paris.
- [103] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1983.
- [104] M. Lothaire. *Mots*. Langue, Raisonnement, Calcul. [Language, Reasoning, Computation]. Editions Hermès, Paris, 1990. Mélanges offerts à M.-P. Schützenberger. [Miscellany offered to M.-P. Schützenberger].
- [105] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997. Corrected reprint of the 1983 original, with a new preface by Perrin.
- [106] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
- [107] M. Lothaire. *Applied combinatorics on words*, volume 105 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.
- [108] S. Mantaci, A. Restivo, M. Sciortino. Burrows-Wheeler transform and Sturmian words. *Inform. Process. Lett.*, 86(5) :241–246, 2003.
- [109] C. Michel, G. Pirillo, M. Pirillo. Varieties of comma free codes. *Comput. Math. Appl.* Special Issue, I FIMA Conference, sous presse.
- [110] F. Mignosi. Infinite words with linear subword complexity. *Theoret. Comput. Sci.*, 65(2) :221–242, 1989.
- [111] F. Mignosi, G. Pirillo. Repetitions in the Fibonacci infinite word. *RAIRO Inform. Théor. Appl.*, 26(3) :199–204, 1992.
- [112] F. Mignosi, P. Séébold. Morphismes sturmiens et règles de Rauzy. *J. Théor. Nombres Bordeaux*, 5(2) :221–233, 1993.
- [113] M. Morse, G. A. Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62 :1–42, 1940.

- [114] M. Morse, G. A. Hedlund. Unending chess, symbolic dynamics and a problem in semigroups. *Duke Math. J.*, 11 :1–7, 1944.
- [115] D. Perrin. *Combinatorics on words*, chapter Words, pages 1–17. Addison Wesley, 1983.
- [116] J.-E. Pin. *Combinatorics on words*, chapter Van der Waerden Theorem, pages 39–54. Addison Wesley, 1983.
- [117] B. Piochi, G. Pirillo. A proof of a theorem by G. Higman. In *Notes on algebraic systems, V*, volume 88, pages 7–14. Budapest, 1988.
- [118] B. Piochi, G. Pirillo. Sur une propriété de permutabilité des groupes finis. *C. R. Acad. Sci. Paris Sér. I Math.*, 307(3) :115–117, 1988.
- [119] B. Piochi, G. Pirillo. On a constant word in finite semigroups. *European J. Combin.*, 10(5) :457–467, 1989.
- [120] G. Pirillo. Morse and Hedlund’s skew sturmian words revisited. *Annals of Combinatorics*. sous presse.
- [121] G. Pirillo. Le sperimentazioni didattiche nell’ambito matematico, in relazione al dibattito in corso sulla riforma della scuola secondaria superiore ed alla revisione della scuola media dell’obbligo. Notiziario UMI, Giugno 1976, 1976. a cura di Piergiorgio Gherardini e Giuseppe Pirillo.
- [122] G. Pirillo. Indagine bibliografica sulle pubblicazioni matematiche periodiche italiane. Notiziario UMI, 1977.
- [123] G. Pirillo. Indagine sui libri di testo adottati nella scuola secondaria. Notiziario UMI, Agosto-Settembre 1978, 1978. Quarto convegno UMI sull’insegnamento della matematica, Ferrara 23–24 aprile 1978.
- [124] G. Pirillo. La scuola francese e l’insegnamento della matematica (da Napoleone ad Haby in una ventina di minuti ed una decina di lucidi). Notiziario UMI, 1979. Quinto convegno UMI sull’insegnamento della matematica, Ferrara 20–21 aprile 1979.
- [125] G. Pirillo. Généralisations du théorème de van der Waerden et semi-groupes à générations bornées. Thèse de 3ème cycle, L.I.T.P., C.N.R.S., Université Paris VI et Paris VII, Paris, 1981.
- [126] G. Pirillo. *Combinatorics on words*, chapter Repetitive mapping and morphisms, pages 55–62. Addison Wesley, 1983.
- [127] G. Pirillo. On permutation properties for semigroups. In *Group theory (Bressanone, 1986)*, volume 1281 of *Lecture Notes in Math.*, pages 118–119. Springer, Berlin, 1987.

- [128] G. Pirillo. On permutation properties for finitely generated semigroups. In *Combinatorics '86 (Trento, 1986)*, volume 37 of *Ann. Discrete Math.*, pages 375–376. North-Holland, Amsterdam, 1988.
- [129] G. Pirillo. A finiteness condition for finitely generated semigroups. *Rend. Circ. Mat. Palermo (2)*, 38(2) :212–216, 1989.
- [130] G. Pirillo. Two finiteness conditions for finitely generated and periodic semigroups. *Acta Sci. Math. (Szeged)*, 53(3-4) :233–236, 1989.
- [131] G. Pirillo. The van der Waerden theorem and the Burnside problem for semigroups. *Arch. Math. (Basel)*, 53(1) :1–3, 1989.
- [132] G. Pirillo. On a combinatorial property of Fibonacci infinite word. *Pure Math. Appl. Ser. A*, 1(2) :143–147, 1990.
- [133] G. Pirillo. On some properties of the Thue infinite word. In *Combinatorics '88, Vol. 2 (Ravello, 1988)*, Res. Lecture Notes Math., pages 325–329. Mediterranean, Rende, 1991.
- [134] G. Pirillo. Sur un théorème combinatoire de Shirshov. *C. R. Acad. Sci. Paris Sér. I Math.*, 313(9) :631–634, 1991.
- [135] G. Pirillo. Sur les produits directs de semi-groupes répétitifs. *C. R. Acad. Sci. Paris Sér. I Math.*, 314(8) :645–648, 1992.
- [136] G. Pirillo. On a combinatorial property of Fibonacci semigroup. *Discrete Math.*, 122(1-3) :263–267, 1993.
- [137] G. Pirillo. Properties of integers and finiteness conditions for semigroups. *Results Math.*, 24(1-2) :168–173, 1993.
- [138] G. Pirillo. Si può parlare delle parole di Fibonacci a studenti di liceo? *Notiziario UMI*, 1993. Atti del Quindicesimo Convegno sull'insegnamento della matematica, *Notiziario Unione Matematica Italiana*, Supplemento, n. 5, Maggio 1993.
- [139] G. Pirillo. Infinite words and biprefix codes. *Inform. Process. Lett.*, 50(6) :293–295, 1994.
- [140] G. Pirillo. On semigroups having the ω -iteration property. *Matematiche (Catania)*, 51(suppl.) :183–187 (1997), 1996.
- [141] G. Pirillo. A proof of Shirshov's theorem. *Adv. Math.*, 124(1) :94–99, 1996.
- [142] G. Pirillo. Fibonacci numbers and words. *Discrete Math.*, 173(1-3) :197–207, 1997.

- [143] G. Pirillo. From the Fibonacci word to Sturmian words. *Publ. Math. Debrecen*, 54(suppl.) :961–971, 1999. Automata and formal languages, VIII (Salgótarján, 1996).
- [144] G. Pirillo. Maximal circular codes and applications to theoretical biology. In *Mathematical and computational biology (Aizu-Wakamatsu City, 1997)*, volume 26 of *Lectures Math. Life Sci.*, pages 187–190. Amer. Math. Soc., Providence, RI, 1999.
- [145] G. Pirillo. A new characteristic property of the palindrome prefixes of a standard Sturmian word. *Sém. Lothar. Combin.*, 43 :Art. B43f, 3 pp. (electronic), 1999.
- [146] G. Pirillo. On a conjecture of Erdős. *Publ. Math. Debrecen*, 54(suppl.) :973–978, 1999. Automata and formal languages, VIII (Salgótarján, 1996).
- [147] G. Pirillo. Some factorizations of the Fibonacci word. *Algebra Colloq.*, 6(4) :361–368, 1999.
- [148] G. Pirillo. Remark on the Arquès-Michel code. pages 327–330, 2001. in : Proceedings of the Workshop “Theoretical Biology 2, History and Present Themes”, Arcidosso, September 1-3, 1999.
- [149] G. Pirillo. From the Fibonacci word to Sturmian words. II. *Pure Math. Appl.*, 11(3) :533–537, 2000.
- [150] G. Pirillo. A curious characteristic property of standard Sturmian words. In *Algebraic combinatorics and computer science*, pages 541–546. Springer Italia, Milan, 2001.
- [151] G. Pirillo. A characterization for a set of trinucleotides to be a circular code. pages 241–244, 2002. in : “Determinism, Holism, and Complexity” (Edited by C. Pellegrini, P. Cerrai, P. Freguglia, V. Benci and G. Israel).
- [152] G. Pirillo. Fibonacci word and Sturmian words. *Sci. Math. Jpn.*, 55(3) :641–647, 2002.
- [153] G. Pirillo. Inequalities characterizing standard Sturmian words. *Pure Math. Appl.*, 14(1-2) :141–144, 2003. Dedicated to the 65th birthday of Professor Franco Migliorini.
- [154] G. Pirillo. Inequalities characterizing standard Sturmian and episturmian words. *Theoret. Comput. Sci.*, 341(1-3) :276–292, 2005.

- [155] G. Pirillo. Numeri irrazionali e segmenti incommensurabili. *Nuova Secondaria* 7, pp. 87–91, 2005.
- [156] G. Pirillo. Some algebraic remarks on circular codes. *Pure Math. Appl.*, 16(1-2) :119–123, 2005.
- [157] G. Pirillo. Sulla frazione continua di $\sqrt{2}$. *Archimede*, 4 :197–198, 2005.
- [158] G. Pirillo. Sulla frazione continua di $\frac{\sqrt{5}+1}{2}$. *Bollettino dei docenti di matematica*, 52 :91–93, 2006.
- [159] G. Pirillo. Sulla frazione continua di $\sqrt{3}$. *Archimede*, 1 :23–25, 2006.
- [160] G. Pirillo. A hierarchy for circular codes. *Fibonacci*, Turku, 2006, sous presse
- [161] G. Pirillo, M. Pirillo. Growth function of self-complementary circular codes. *Biology Forum*, 98, 2005.
- [162] G. Pirillo, S. Varricchio. On uniformly repetitive semigroups. *Semigroup Forum*, 49(1) :125–129, 1994.
- [163] G. Pirillo, S. Varricchio. Some combinatorial properties of infinite words and applications to semigroup theory. In *Proceedings of the 5th Conference on Formal Power Series and Algebraic Combinatorics (Florence, 1993)*, volume 153, pages 239–251, 1996.
- [164] N. Pytheas-Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*. *Lectures Notes in Mathematics* 1794, Springer-Verlag, 2002.
- [165] F. P. Ramsey. On a problem of formal logic. *Proc. London Math. Soc.*, 30 :264–286, 1930.
- [166] G. Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. France*, 110(2) :147–178, 1982.
- [167] G. Rauzy. Suites à termes dans un alphabet fini. In *Seminar on number theory, 1982–1983 (Talence, 1982/1983)*, pages Exp. No. 25, 16. Univ. Bordeaux I, Talence, 1983.
- [168] G. Rauzy. Mots infinis en arithmétique. In *Automata on infinite words (Le Mont-Dore, 1984)*, volume 192 of *Lecture Notes in Comput. Sci.*, pages 165–171. Springer, Berlin, 1985.
- [169] A. Restivo, C. Reutenauer. Some applications of a theorem of Shirshov to language theory. *Inform. and Control*, 57(2-3) :205–213, 1983.
- [170] A. Restivo, C. Reutenauer. On the Burnside problem for semigroups. *J. Algebra*, 89(1) :102–104, 1984.

- [171] A. Restivo, C. Reutenauer. Rational languages and the Burnside problem. *Theoret. Comput. Sci.*, 40(1) :13–30, 1985. Special issue : Eleventh international colloquium on automata, languages and programming (Antwerp, 1984).
- [172] A. Restivo, S. Salemi. Overlap-free words on two symbols. In *Automata on infinite words (Le Mont-Dore, 1984)*, volume 192 of *Lecture Notes in Comput. Sci.*, pages 198–206. Springer, Berlin, 1985.
- [173] C. Reutenauer. Mots de Lyndon et un théorème de Shirshov. *Ann. Sci. Math. Québec*, 10(2) :237–245, 1986.
- [174] G. Richomme. Conjugacy and episturmian morphisms. *Theoret. Comput. Sci.*, 302(1-3) :1–34, 2003.
- [175] R. Risley, L. Zamboni. A generalization of Sturmian sequences : combinatorial structure and transcendence. *Acta Arith.*, 95(2) :167–184, 2000.
- [176] G.-C. Rota, editor. *Studies in combinatorics*. Mathematical Association of America, Washington, D.C., 1978. MAA Studies in Mathematics, 17.
- [177] J. Sakarovitch. *Éléments de théorie des automates*. Vuibert informatique, 2003.
- [178] M.-P. Schützenberger. Une théorie algébrique du codage. In *Séminaire Dubreil-Pisot, année 1955–1956*. Institut H. Poincaré, Paris, 1956. Exposé N° 15.
- [179] M.-P. Schützenberger. *poly rouge*. Paris, 1977.
- [180] P. Séébold. Propriétés combinatoires des mots infinis engendrés par certains morphismes. Thèse de doctorat, LITP, 85-15, Paris, 1985.
- [181] P. Séébold. Fibonacci morphisms and Sturmian words. *Theoret. Comput. Sci.*, 88(2) :365–384, 1991.
- [182] A. I. Shirshov. On rings with identity relations. *Mat. Sb. N.S.*, 43(85) :277–283, 1957.
- [183] I. Simon. Conditions de finitude pour des semi-groupes. *C. R. Acad. Sci. Paris Sér. A-B*, 290(23) :1081–1082, 1980.
- [184] S. Tabachnikov. Billiards. *Panor. Synth.*, (1) :vi+142, 1995.
- [185] A. Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I Math-Nat. Kl.*, 7 :1–22, 1906.

- [186] A. Thue. Über die gegenseitige Loge gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris.*, 1 :1–67, 1912.
- [187] B. L. van der Waerden. Beweis einer Baudet'schen Vermutung. *Nieuw Arch. Wisk.*, 15 :212–216, 1927.
- [188] B. L. van der Waerden. Wie der Beweis der Vermutung von Baudet gefunden wurde. In *Abhandlungen des Mathematischen Seminars der Hanseatischen Universität Hamburg*, pages 6–15, 1965. (also published as : [How the Baudet's conjecture was found, *Studies in Pure Mathematics*, Academic Press, 1971, pp. 251-260].
- [189] T. van Ravenstein. The three gap theorem (Steinhaus conjecture). *J. Austral. Math. Soc. Ser. A*, 45(3) :360–370, 1988.
- [190] D. Vandeth. Sturmian words and words with a critical exponent. *Theoret. Comput. Sci.*, 242(1-2) :283–300, 2000.
- [191] S. Varricchio. Bisections of free monoids and a new unavoidable regularity. In *Mathematical foundations of computer science, 1988 (Carlsbad, 1988)*, volume 324 of *Lecture Notes in Comput. Sci.*, pages 538–541. Springer, Berlin, 1988.
- [192] S. Varricchio. Factorizations of free monoids and unavoidable regularities. *Theoret. Comput. Sci.*, 73(1) :81–89, 1990.
- [193] S. Varricchio. A pumping condition for regular sets. *SIAM J. Comput.*, 26(3) :764–771, 1997.