

TABLE DES MATIÈRES

Introduction générale	7
1 Préliminaires	10
1.1 Alphabet, mots.	10
1.2 Facteurs.	11
1.3 Morphismes.	12
1.4 Complexité abélienne	14
1.5 Fréquences	17
1.6 Mots de retour	18
1.6.1 Utilisation des facteurs bispéciaux	19
1.6.2 Utilisation des ancêtres dans les mots morphiques	19
1.7 Mots privilégiés	20
2 Combinatoire du mot de Thue-Morse binaire	22
2.1 Introduction	22
2.2 Définitions	23
2.2.1 Relation de récurrence 1	23
2.2.2 Relation de récurrence 2	23
2.2.3 Définition directe	23

INTRODUCTION GÉNÉRALE

La combinatoire des mots est une branche des mathématiques et de l'informatique théorique qui applique l'analyse combinatoire aux mots finis ou infinis. Cette branche s'est développée à partir de plusieurs branches des mathématiques, telles que la théorie des nombres, la théorie des groupes, les probabilités et bien sûr la combinatoire. Elle a des liens avec divers thèmes informatiques, comme l'algorithmique du texte, la recherche de motifs, la compression de textes...

La combinatoire des mots a pour objet d'étudier les propriétés des mots finis ou infinis. Elle remonte aux travaux du mathématicien norvégien Thue sur les suites sans répétitions de symboles, au début du 20^e siècle.

C'est Schützenberger qui est le fondateur de la combinatoire des mots moderne. Notamment, ses travaux avec Lyndon et Leutic ont donné naissance au livre *Combinatorics on words*, ouvrage collectif signé du nom de plume *M. Lothaire* et paru en 1983 [29]. La combinatoire de mots se développa rapidement à partir de cette date [3, 10, 9, 17, 28, 27].

Le mot t_2 de Thue-Morse sur l'alphabet binaire $\{0, 1\}$ est le mot infini engendré par le morphisme μ_2 défini par $\mu_2(0) = 01$, $\mu_2(1) = 10$. Ce mot a été utilisé pour la première fois de façon implicite par le mathématicien français Prouhet en 1851, pour donner une solution à un problème de théorie des nombres appelé depuis le problème de Prouhet-

Tarry-Escott que nous rappellerons dans la suite du travail. Thue l'a découvert et utilisé dans un article publié en 1912 [44] qui, avec un autre article datant de 1906 [43], est l'article fondateur de la combinatoire des mots. En 1921 Morse va réutiliser ce mot pour donner un exemple d'une suite récurrente non périodique, résolvant ainsi un problème de géométrie différentielle [32]. Une généralisation naturelle du mot de Thue-Morse sur l'alphabet à q lettres $\mathcal{A}_q = \{0, 1, \dots, q-1\}$ est le mot infini $\mathbf{t}_{b,q}$ engendré par le morphisme $\mu_{b,q}$ défini par : $\mu_{b,q}(k) = k(k+1)\dots(k+b-1)$, où $b \geq 2$ et les lettres sont exprimées modulo q . Une étude sur ce mot a été faite dans [41].

La complexité abélienne est une notion combinatoire utilisée dans l'étude des mots infinis. Cette idée est apparue quand Coven et Hedlund réalisèrent que les mots ultimement périodiques et les mots sturmiens peuvent être caractérisés en utilisant les vecteurs de Parikh [18]. Cette étude a été reprise et généralisée récemment par Richome, Saari et Zamboni dans [37], où la terminologie *Complexité abélienne* a été introduite. Depuis lors, l'étude des propriétés abéliennes des mots connaît un essor [15, 19, 38, 45, 6].

Récemment, la notion de mots de retour a joué un rôle important dans l'étude des systèmes dynamiques (symboliques). Un facteur séparant deux occurrences successives d'un mot w dans un mot infini est appelé mot de retour de w . Cette définition est donnée par Durand dans [1]. Cette notion a été utilisée pour la caractérisation de certaines classes de mots [21, 31, 49]. Les mots de retour d'autres mots classiques ont été étudiés [7, 33].

Dans ce mémoire nous nous intéressons à l'étude du mot de Thue-Morse sur l'alphabet $\mathcal{A}_3 = \{0, 1, 2\}$. Il s'agit du mot \mathbf{t}_3 engendré par le morphisme μ_3 défini par $\mu_3(0) = 012, \mu_3(1) = 120$ et $\mu_3(2) = 201$. Nous établissons des propriétés combinatoires de ce mot, axées sur l'équilibre, la complexité abélienne et les mots de retour. Le mot \mathbf{t}_3 se généralise naturellement sur un alphabet \mathcal{A}_q de taille $q \geq 3$. Il s'agit, sur l'alphabet $\mathcal{A}_q = \{0, 1, \dots, q-1\}$, du mot infini $\mathbf{t}_q = \mathbf{t}_{q,q}$ engendré par le morphisme μ_q défini par : $\mu_q(k) = k(k+1)\dots(k+q-1)$, où les lettres sont exprimées modulo q .

Le mémoire est organisée de la façon suivante :

Le chapitre 1 est consacré au rappel des définitions et propriétés utiles pour les thèmes abordés dans le reste du travail.

Dans le chapitre 2, nous présentons une étude combinatoire du mot de Thue-Morse binaire, \mathbf{t}_2 . Nous rappelons d'abord diverses définitions de \mathbf{t}_2 , puis nous présentons quelques problèmes mathématiques dont la résolution a nécessité l'utilisation du mot \mathbf{t}_2 .

Dans le chapitre 3, nous établissons des propriétés du mot de Thue-Morse ternaire portant sur les facteurs triprolongeables puis nous montrons que \mathbf{t}_3 est 2-équilibré. Enfin, nous déterminons la fonction de complexité abélienne de \mathbf{t}_3 .

Dans le chapitre 4, nous nous intéressons aux carrés de lettres de \mathbf{t}_3 et aux facteurs qui les séparent. Nous déterminons précisément les longueurs et la structure des facteurs séparateurs des carrés de lettres. Nous donnons également une estimation du nombre de carrés de lettres dans les facteurs de \mathbf{t}_3 . Nous terminons par l'étude des mots de retour et les facteurs privilégiés. Nous obtenons l'ensemble donnant le nombre de mots de retour des facteurs de \mathbf{t}_3 .

1.1 Alphabet, mots.

On appelle alphabet \mathcal{A} un ensemble fini de symboles. Les éléments de \mathcal{A} sont appelés des lettres. Par exemple $\mathcal{A} = \{1, a, x, \}$ est un alphabet à quatre lettres.

Soit \mathcal{A} un alphabet. Un mot fini est une suite finie de lettres de \mathcal{A} . La longueur d'un mot w sur \mathcal{A} est le nombre de lettres qui le constituent. Elle est notée $|w|$. Pour toute lettre a de \mathcal{A} , on note $|w|_a$ le nombre d'occurrences de a dans w . Par exemple, $ababa$ est un mot fini sur l'alphabet $\mathcal{A} = \{a, b\}$ et nous avons $|w| = 5$ et $|w|_a = 3$. Le mot vide, noté ϵ , est la suite ne contenant pas de lettre. L'ensemble des mots finis sur \mathcal{A} est noté \mathcal{A}^* . L'ensemble des mots finis non vides sur \mathcal{A} est noté \mathcal{A}^+ .

Un mot infini sur l'alphabet \mathcal{A} est une suite de lettres de \mathcal{A} indexée par \mathbb{N} . Si w est un mot infini, on peut l'écrire sous la forme $w = w_1w_2w_3...w_n...$ où les w_i sont des lettres de \mathcal{A} . L'ensemble des mots infinis sur \mathcal{A} est noté \mathcal{A}^ω . L'ensemble des mots finis ou infinis sur \mathcal{A} est noté \mathcal{A}^∞ .

Un mot infini u sur \mathcal{A} est dit périodique s'il existe un mot w dans \mathcal{A}^+ tel que $u = w^p$. L'entier $p = |w|$ est une période de u . Le mot u est dit ultimement périodique s'il existe deux mots $v \in \mathcal{A}^*$ et $w \in \mathcal{A}^+$ tels que $u = vw^p$. Un mot non ultimement périodique est

facteur v est dit spécial à droite (resp. à gauche) s'il admet au moins deux prolongements à droite (resp. à gauche). Il est dit bispécial s'il est à la fois spécial à droite et spécial à gauche.

Soient u et v deux mots finis non vides. Si v est un préfixe (resp. un suffixe) de u , alors $v^{-1}u$ (resp. uv^{-1}) est le mot obtenu de u en enlevant le préfixe (resp. le suffixe) v .

Exemple 1.2.1 *Considérons le mot $w = ababb$. Alors :*

$F(w) = \{ \varepsilon, a, b, ab, ba, bb, aba, abb, bab, abab, babb, ababb \}$;

les facteurs spéciaux à droite (resp. à gauche) de w sont ε, b et ab (resp. ε et b) ;

les facteurs bispéciaux de w sont ε et b ;

pour le préfixe ab , on a $(ab)^{-1}w = abb$. Pour le suffixe bb , on a $w(bb)^{-1} = aba$.

Un mot infini u est dit récurrent si tout facteur de u apparaît une infinité de fois dans u . Il est dit uniformément récurrent si pour tout entier n , il existe un entier n_0 tel que tout facteur de longueur n_0 contienne tous les facteurs de longueur n . Soient u un mot infini, uniformément récurrent et non ultimement périodique et v un facteur de u . Il existe un facteur bispécial de u contenant v . Le plus court facteur bispécial de u contenant v est appelé facteur bispécial étendu de v .

Exemple 1.2.2 *Soit $\mathcal{A} = \{a, b, c\}$. Considérons le mot infini w défini sur \mathcal{A} par : $w = (abc)^\infty$. On remarque que tout facteur commençant par a (resp. b ou c) apparaît à toutes les positions $n \equiv 1[3]$ (resp. $n \equiv 2[3]$ ou $n \equiv 0[3]$). Par suite, tout facteur de w de longueur k est contenu dans tout facteur de w de longueur $k + 2$. Par conséquent w est uniformément récurrent.*

1.3 Morphismes.

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

primitif si, il existe un entier positif non nul n tel que, pour toute lettre a de \mathcal{A} , $\varphi^n(a)$ contient toutes les lettres de \mathcal{A} ;

k -uniforme, si $| \varphi(a) | = k$ pour tout a dans \mathcal{A} ;

marqué à gauche (resp. à droite) sur l'alphabet $\mathcal{A} = \{a_1, a_2, \dots, a_d\}$, si les premières (resp. dernières) lettres de $\varphi(a_i)$ et $\varphi(a_j)$ diffèrent, pour tout $i \neq j$. Il est dit marqué s'il est à la fois marqué à droite et marqué à gauche.

Un mot infini u est engendré par un morphisme φ s'il existe une lettre a de \mathcal{A} telle que les mots $a, \varphi(a), \dots, \varphi^n(a), \dots$ sont des préfixes de plus en plus longs de u . On note $u = \varphi^\omega(a)$ et on l'appelle mot purement morphique.

Soient u un mot infini engendré par un morphisme f et w un facteur de u vérifiant $|w| \geq \max\{ | \varphi(a) | : a \in \mathcal{A} \}$. Alors w peut être décomposé sous la forme

$$p_0 \varphi(a_1) \varphi(a_2) \dots \varphi(a_n) s_{n+1},$$

où

$$n \geq 0;$$

$$a_0, a_1, \dots, a_{n+1} \in \mathcal{A};$$

$$p_0 \text{ est un suffixe de } \varphi(a_0);$$

$$s_{n+1} \text{ est un préfixe de } \varphi(a_{n+1}).$$

Le mot $v = a_1 a_2 \dots a_n$ est appelé un ancêtre de u . Cette décomposition est appelée *synchronisation* si elle est unique [12].

Soit φ un morphisme défini sur \mathcal{A} . Un mot $w \in \mathcal{A}^\omega$ est appelé point fixe de φ si $\varphi(w) = w$. Dans ce mémoire nous nous intéressons aux points fixes infinis. Un morphisme φ est dit prolongeable en une lettre $a \in \mathcal{A}$ si $\varphi(a) = av$, où $v \in \mathcal{A}^*$ et $\varphi^n(v) = \varphi^{n+1}(a)$ pour tout entier n . Si φ est prolongeable en a , alors $\varphi^n(a)$ est un préfixe propre de $\varphi^{n+1}(a)$ pour tout entier n .

Le théorème suivant est dû à Gottschalk.

Théorème 1.3.1 [22] *Soit φ un morphisme primitif, prolongeable en une lettre a . Alors, le mot infini $\varphi^\omega(a)$, engendré par φ en a , est uniformément récurrent.*

Exemple 1.3.1 : Le mot de Thue-Morse \mathbf{t}_2 . Il s'agit du mot engendré par le morphisme de Thue-Morse μ_2 défini par : $\mu_2(0) = 01$, $\mu_2(1) = 10$: $\mathbf{t}_2 = \mu_2(0)$. On a :

$$\mathbf{t}_2 = \mu_2(0) = 01101001100101101001011001101001100101\dots$$

Le morphisme de Thue-Morse est 2-uniforme, primitif et prolongeable en 0. Ainsi, \mathbf{t}_2 est uniformément récurrent.

Exemple 1.3.2 *Le mot de Fibonacci \mathbf{f} . Il s'agit du mot infini \mathbf{f} engendré par le morphisme de Fibonacci défini par : $(0) = 01, (1) = 0$. On a*

$$\mathbf{f} = (0) = 010010100100101\dots$$

Le nom *mot de Fibonacci* s'explique par la relation que \mathbf{f} a avec la suite d'entiers de Fibonacci, définie par récurrence par : $f_0 = 0, f_1 = 1$ et $f_n = f_{n-1} + f_{n-2}$, pour tout $n \geq 2$. En particulier, le mot de Fibonacci peut être engendré par une récurrence analogue. Soit $x_0 = 0, x_1 = 1$ et $x_n = x_{n-1}x_{n-2}$ pour tout $n \geq 3$. Alors, $\mathbf{f} = \lim_{n \rightarrow \infty} x_n$.

1.4 Complexité abélienne

Soient u un mot infini sur un alphabet $\mathcal{A}_q = \{a_0, a_1, \dots, a_{q-1}\}$ et v un facteur de u . Le vecteur de Parikh de v est le q -uplet $\psi(v) = (|v|_{a_0}, |v|_{a_1}, \dots, |v|_{a_{q-1}})$. On désigne par $\Psi_n(u)$, l'ensemble des vecteurs de Parikh des facteurs de longueur n de u :

$$\Psi_n(u) = \{\psi(v) : v \in F_n(u)\}.$$

Notons que $|v|_{a_0} + |v|_{a_1} + \dots + |v|_{a_{q-1}} = |v|$. La fonction de complexité abélienne de u est l'application définie de \mathbb{N} dans \mathbb{N} par :

$$\rho_u^{ab}(n) = \text{card}(\Psi_n(u)).$$

Ainsi, la complexité abélienne ρ_u^{ab} d'un mot u est la fonction qui compte le nombre de vecteurs de Parikh de u de longueur donnée.

Soient a et b deux lettres d'un alphabet $\mathcal{A} = \{0, \dots, q-1\}$ et v un mot fini sur \mathcal{A} . Lorsque $a = b$, alors $\psi(av) = \psi(vb)$. Lorsque $a \neq b$, alors $\psi(av) - \psi(vb)$ est le vecteur dont toutes les composantes sont 0 sauf sa $(a+1)^e$ composante dont la valeur est 1 et sa $(b+1)^e$ composante qui vaut -1 . Ceci montre comment évoluent les vecteurs de Parikh lorsque nous considérons deux facteurs successifs de même longueur d'un mot w . Par conséquent, nous avons le résultat suivant, établi par Richomme, Saari et Zamboni dans [37].

Lemme 1.4.1 *Soient v et w deux facteurs d'un mot infini u . Soient p et $p + c$ les i^e composantes respectives des vecteurs de Parikh de v et w , où p et c sont des naturels. Alors, pour tout $l \in \{0, 1, \dots, c\}$, il existe un facteur u_l de u dont la i^e composante du vecteur de Parikh est $p + l$.*

En conséquence, la complexité abélienne de u vérifie $\rho_u^{ab}(n) \geq c + 1$, pour tout entier naturel n , puisque pour $l_1, l_2 \in \{0, 1, \dots, c\}$, les facteurs u_{l_1} et u_{l_2} ont des vecteurs de Parikh différents lorsque $l_1 \neq l_2$.

Le terme suivant a été introduit par Turek dans [46] pour l'étude de la complexité abélienne du mot de Tribonacci.

Définition 1.4.1 *Soit u un mot infini. On désigne par $u_{[n]}$ le préfixe de longueur n de u , $n \in \mathbb{N}$.*

Pour tout facteur w de longueur n de u , le vecteur de Parikh relatif de w est défini par :

$$\psi^{rel}(w) = \psi(w) - \psi(u_{[n]}).$$

L'ensemble des vecteurs de Parikh relatifs des facteurs de longueur n de u est donné par :

$$\Psi_n^{rel}(u) = \{\psi^{rel}(w) : w \in F_n(u)\}.$$

Remarque 1.4.1 [46] *L'idée de vecteur de Parikh relatif ψ^{rel} s'inspire de la technique utilisée par Adamczewski dans [1]. En effet, il utilise les fréquences des lettres pour simplifier l'étude des propriétés d'équilibre des substitutions primitives.*

Le symbole $u_{[n]}$ est utilisé dans la suite pour l'étude de la complexité abélienne. Observons que $\Psi_n(u)$ a le même cardinal que $\Psi_n^{rel}(u)$. On obtient alors que :

$$\rho_n^{ab}(u) = \text{card}(\Psi_n^{rel}(u))$$

Un mot infini u est dit α -équilibré, si pour tout $a \in \mathcal{A}$ et pour tous facteurs v et w de u de même longueur, on a :

$$||v|_a - |w|_a| \leq \alpha$$

Le Lemme suivant a été établi dans [37].

Lemme 1.4.2 *Pour tout mot u in ni , la fonction de complexité abélienne ρ_u^{ab} est bornée si et seulement si u est \leq -équilibré pour un entier naturel n .*

Remarque 1.4.2 *Pour tout entier naturel n , l'ensemble des vecteurs de Parikh relatifs a les propriétés suivantes :*

$$0 \in \Psi_n^{rel}(u);$$

si $(\psi'_0, \psi'_1, \dots, \psi'_{q-1}) \in \Psi_n^{rel}(u)$ et si u est \leq -équilibré, alors $|\psi'_k| \leq n$, pour tout $k \in \mathcal{A}$.

Il suffit de remarquer que 0 est le vecteur de Parikh relatif de $u_{[n]}$. Considérons à présent un facteur w de u tel que $(\psi'_0, \psi'_1, \dots, \psi'_{q-1}) = \psi^{rel}(w)$, $|w| = n$. Alors $|\psi'_a| = ||w|_a - |u_{[n]}|_a|$ pour toute lettre a . Comme u est \leq -équilibré, alors pour tous facteurs w et v de u de même longueur et pour tout $a \in \mathcal{A}$, $||w|_a - |v|_a| \leq n$. En particulier, pour $v = u_{[n]}$ on a le deuxième point.

Remarque 1.4.3 [46] *(Comparaisons entre ψ et ψ^{rel}).*

Soit u un mot \leq -équilibré. Les avantages de l'utilisation des vecteurs de Parikh relatifs au lieu des vecteurs de Parikh standards sont :

Les composantes de $\psi(w)$ augmentent avec la longueur de w . En effet, $|w|_1 + |w|_2 + \dots + |w|_{q-1} = |w|$. Par contre les composantes de $\psi^{rel}(w)$ sont bornées.

L'ensemble des vecteurs de Parikh $\{\psi(w) : w \in F(u)\}$ est in ni . Par contre, l'ensemble des vecteurs de Parikh relatifs est ni .

Exemple 1.4.1 *Considérons le mot $u = 01212001120201012$. Alors*

$$F_2(u) = \{00, 01, 02, 10, 11, 12, 20, 21\}.$$

Par suite,

$$\Psi_2(u) = \{(2, 0, 0), (1, 1, 0), (1, 0, 1), (0, 2, 0), (0, 1, 1), \dots\}.$$

Le préfixe $u_{[2]}$ est 01 et on a $\psi(u_{[2]}) = (1, 1, 0)$. Ainsi,

$$\Psi_2^{rel}(w) = \{(-1, 1, 0), (0, 0, 0), (0, -1, 1), (1, -1, 0), (-1, 0, 1)\}.$$

La synchronisation des facteurs dans les mots morphiques simplifie l'étude des vecteurs de Parikh relatifs.

1.5 Fréquences

Dé nition 1.5.1 [11] *Soit u un mot infini sur un alphabet \mathcal{A} . On dit que u admet :*

des fréquences des lettres si pour toute lettre a et pour toute suite (u_n) de préfixes de u tel que $\lim_{n \rightarrow +\infty} |u_n| = +\infty$, alors $\lim_{n \rightarrow +\infty} \frac{|u_n|_a}{|u_n|}$ existe.

des fréquences uniformes de lettres si pour toute lettre a et pour toute suite (v_n) de facteurs de u tel que $\lim_{n \rightarrow +\infty} |v_n| = +\infty$, alors $\lim_{n \rightarrow +\infty} \frac{|v_n|_a}{|v_n|}$ existe.

Dé nition 1.5.2 *Soit u un mot infini sur un alphabet \mathcal{A} . u admet :*

des fréquences de facteurs si pour tout facteur v de u et pour toute suite (u_n) de préfixes de u tel que $\lim_{n \rightarrow +\infty} |u_n| = +\infty$, alors $\lim_{n \rightarrow +\infty} \frac{|u_n|_v}{|u_n|}$ existe.

des fréquences uniformes de facteurs si pour tout facteur v et pour toute suite (u_n) de facteurs de u tel que $\lim_{n \rightarrow +\infty} |u_n| = +\infty$, alors $\lim_{n \rightarrow +\infty} \frac{|u_n|_v}{|v_n|}$ existe.

Remarque 1.5.1 : *Si un mot u admet des fréquences uniformes de lettres (resp. de facteurs), alors il admet des fréquences de lettres (resp. de facteurs).*

Notation : Si un mot infini u admet des fréquences de lettres (resp. de facteurs), alors, pour toute lettre a (resp. pour tout facteur v), la fréquence de a (resp. de v) est notée $f_a(u)$ (resp. $f_v(u)$) ou plus simplement f_a (resp. f_v).

La notion de fréquences dans les mots infinis ressort abondamment dans la littérature. Dans [3], Allouche et Shallit ont établi des résultats sur les fréquences des lettres dans les mots automatiques. Plus tard, dans [37], Saari a donné une condition nécessaire et suffisante pour qu'un mot morphique admette des fréquences de lettres. De même, dans [22] et [12], les auteurs ont établi que tout mot β -équilibré admet des fréquences uniformes de lettres. Dans [14], Cassaigne et Kaboré ont établi les deux résultats suivants :

Théorème 1.5.1 *Soit u un mot binaire infini. Alors, $\rho_u^{ab}(n) = o(n)$ si et seulement si u admet des fréquences uniformes de lettres.*

Les auteurs ont ensuite généralisé ce résultat à tout mot infini sur un alphabet à k lettres, $k \geq 3$.

Théorème 1.5.2 *Soit u un mot infini sur un alphabet à k lettres, $k \geq 3$. Nous avons les propriétés suivantes.*

Si $\rho_u^{ab}(n) = o(n)$ alors, u admet des fréquences uniformes de lettres.

Si u admet des fréquences uniformes de lettres alors, $\rho_u^{ab}(n) = o(n^{k-1})$.

Dans [36], Queelec a montré que tout point fixe d'un morphisme primitif admet des fréquences de facteurs.

Dans [5], Balková a entrepris des investigations sur les fréquences de certains mots morphiques. Elle parvient à mettre en relation les fréquences des facteurs de ces mots et leurs ancêtres.

Proposition 1.5.1 *Soient u un mot infini engendré par un morphisme marqué, primitif, k -uniforme et v un facteur de u . Si w est l'unique ancêtre de v alors $f_v = \frac{f_w}{k}$.*

Corollaire 1.5.1 *Soit w un facteur d'un mot infini u dont la fréquence existe.*

Si w a un unique prolongement b à droite, alors $f_w = f_{wb}$

Si w a un unique prolongement a à gauche, alors $f_w = f_{aw}$.

Ce corollaire implique que si un facteur w admet un unique prolongement à gauche aw et un unique prolongement à droite wb , alors $f_w = f_{aw} = f_{wb} = f_{awb}$.

Corollaire 1.5.2 *Soient u un mot infini, récurrent et non ultimement périodique et w un facteur de u . On suppose que la fréquence de w dans u existe. Soit v le facteur bispécial étendu de w . Alors, $f_w(u) = f_v(u)$.*

1.6 Mots de retour

Soient u un mot infini récurrent, sans chevauchement et v, w deux facteurs de u tels que v sépare deux occurrences successives de w dans u . Alors le facteur wv est appelé mot de retour de w dans u [7]. L'ensemble des mots de retour de w est noté $Ret(w)$.

1.6.1 Utilisation des facteurs bispéciaux

Soient u un mot infini et w un facteur de u non spécial à droite. Alors, le nombre d'occurrences de w et wa dans u coïncide, pour tout $a \in \mathcal{A}$ et on a :

$$Ret(w) = Ret(wa).$$

Si w admet un unique prolongement $b \in \mathcal{A}$ à gauche, alors

$$Ret(bw) = bRet(w)b^{-1} = \{bvb^{-1} \mid v \in Ret(w)\}$$

où bvb^{-1} est le facteur de u obtenu de v en le prolongeant à gauche par la lettre b et en supprimant sa dernière lettre qui est b . En effet, si v est un mot de retour de w alors, v se termine par la lettre b puisque w est uniquement prolongeable à gauche par b .

Soit u un mot infini, uniformément récurrent et non ultimement périodique. Alors, tout facteur w de u possède un facteur bispécial étendu. Ainsi, pour établir le cardinal et la structure des mots de retour d'un mot récurrent et non ultimement périodique u , il suffit de considérer ses facteurs bispéciaux [7].

1.6.2 Utilisation des ancêtres dans les mots morphiques

Soit u un mot morphique. Alors, tout facteur (w) a au moins un ancêtre, à savoir le facteur w . Pour tout facteur w d'un mot morphique u , il existe une étroite relation entre les mots de retour de w et ceux de son image (w) à travers le lemme suivant.

Lemme 1.6.1 [7] *Soient u un mot engendré par un morphisme φ et w un facteur de u . Si w est le seul ancêtre de (w) , alors :*

$$Ret((w)) = \varphi(Ret(w)).$$

Preuve : Soit v un mot de retour de w . Alors, le facteur v s'écrit sous la forme $v = wx$, où x sépare deux occurrences successives de w . Ainsi, vw est un facteur de $u = \varphi(u)$, w est un préfixe de vw et w apparaît exactement deux fois dans vw . Comme w est le seul

ancêtre de (w) , alors (vw) admet (w) uniquement comme préfixe et suffixe. Ainsi, (v) est un mot de retour de (w) et donc $(Ret(w)) = Ret((w))$.

Supposons à présent que v' est un mot de retour de (w) . Alors, il existe un facteur x de u tel que $v' = (w)x$. Par suite, $v' (w) = (w)x (w)$. Ainsi, tout ancêtre de $v' (w)$ est de la forme vw , où v est un mot de retour de w puisqu'il admet exactement deux occurrences de w , avec une comme préfixe et l'autre comme suffixe. Comme $(w)x (w)$ est dans u , il existe un facteur x' de u tel que $x = (x')$. Par suite, $v' = (wx')$. Donc $Ret((w)) = Ret(w)$. ■

1.7 Mots privilégiés

Les mots privilégiés ont été définis pour la première fois dans [25] et étudiés par la suite dans [34].

Soit u un mot fini. Un premier retour complet de u est un mot qui commence et se termine par u et qui contient exactement deux occurrences de u . Un premier retour complet d'un mot est appelé mot de retour complet de ce mot. L'ensemble des mots privilégiés sur un alphabet \mathcal{A} noté $Pri(\mathcal{A})$, est défini par récurrence comme suit :

$u \in Pri(\mathcal{A})$ si u est un premier retour complet d'un mot privilégié plus court $v \in Pri(\mathcal{A})$,

Les mots privilégiés les plus courts sont les lettres de \mathcal{A} et le mot vide ϵ .

L'ensemble des facteurs privilégiés d'un mot w est noté $Pri(w)$. On note $Pri_w(n)$ l'ensemble des facteurs privilégiés de longueur n de w : $Pri_w(n) = Pri(w) \cap F_n(w)$. On appelle fonction de complexité privilégiée de w la fonction A_w définie par $A_w(n) = \#Pri_w(n)$, pour tout naturel n .

Exemple 1.7.1 *Considérons le mot $w = abbabaabbaababbabaa$. Alors*

$$Pri_w(1) = \{a, b\};$$

$$Pri_w(2) = \{aa, bb\};$$

$$Pri_w(3) = \{aba, bab\};$$

$$Pri_w(4) = \{abba, baab\}.$$

Rappelons quelques propriétés sur les mots privilégiés établies dans [34].

Lemme 1.7.1 *Soient u un mot privilégié et v un pré xe (resp. su xe) privilégié de u . Alors, v est un su xe (resp. pré xe) de u .*

Preuve : Soit v un préfixe privilégié de u . Si $|u| \leq 1$, c'est évident, puisque v est soit le mot vide soit une lettre. De plus, si $u = v$, alors v est à la fois un préfixe et un su xe de u . À présent, supposons que $|u| \geq 2$ et $|u| > |v|$. Par définition, u est un premier retour complet d'un mot privilégié plus court w . Si $|v| \geq |w|$, alors w est un su xe de v puisque v et w sont su xes de u . Ainsi, u contient au moins trois occurrences de w et ceci est impossible. Si $|w| = |v|$, alors v est su xe de u . Pour terminer, supposons que $|w| > |v|$. Alors, v est un préfixe propre de w . Comme w est un facteur privilégié, alors v est aussi un su xe de w et donc de u . On procède de même lorsque v est un su xe de u . ■

Lemme 1.7.2 *Soient u un mot privilégié et v son plus long pré xe (resp. su xe) privilégié. Alors, u est un premier retour complet de v .*

Preuve : Soient u un mot privilégié et v son plus long préfixe privilégié. Si $|u| \leq 1$, c'est évident. Supposons que $|u| \geq 2$ et que u est un premier retour complet du mot privilégié w . Si $|v| > |w|$, alors w est un préfixe de v et donc su xe de v d'après le Lemme 1.7.1. Par suite, u contient au moins trois occurrences de w et ceci contredit le fait que u soit un premier retour complet de w . Ainsi, $|v| \leq |w|$ et comme v est le plus long préfixe privilégié de u , alors $|v| = |w|$. De la même façon, on traite le cas où u est le plus long su xe de u . ■

Lemme 1.7.3 *Chaque position dans un mot infini uniformément récurrent introduit un facteur privilégié.*

Preuve : Soit u un mot infini et uniformément récurrent sur un alphabet \mathcal{A} . Nous savons que toute position dans u correspond à une lettre a de \mathcal{A} . Comme u est uniformément récurrent, la lettre a apparaît une infinité de fois dans u . Ainsi, la position a correspond au début d'un facteur privilégié commençant par a . ■

CHAPITRE 2

COMBINATOIRE DU MOT DE THUE-MORSE BINAIRE

2.1 Introduction

Dans ce chapitre nous rappelons quelques propriétés combinatoires du mot de Thue-Morse binaire. C'est le mot infini dont les premiers termes sont :

$$\mathbf{t}_2 = 0110100110010110\dots$$

Le mot \mathbf{t}_2 a été largement étudié et figure parmi les mots infinis classiques les plus étudiés [4, 33, 34].

Nous évoquerons d'abord un problème de théorie des nombres connu sous le nom de *Problème de Prouhet-Tarry-Escott*. Ensuite, nous rappelons un résultat dans lequel Robbins utilise le mot de Thue-Morse pour calculer la limite d'une suite particulière. Puis nous présentons une construction de Allouche de \mathbf{t}_2 à partir d'une partition des entiers. Nous rappelons également l'usage du mot de Thue-Morse dans la construction des carrés magiques. Enfin, nous présentons diverses propriétés combinatoires comme les chevauchements, les facteurs bispéciaux, la complexité abélienne, les fréquences, les mots de retour et les facteurs privilégiés.

Actuellement, le mot \mathbf{t}_2 est connu sous diverses définitions.

2.2 Définitions

2.2.1 Relation de récurrence 1

La suite de Thue-Morse est la suite $\mathbf{t} = (\mathbf{t}_n)$ définie par : $\mathbf{t}_0 = 0$ et pour tout $n \geq 1$,

$$\begin{cases} \mathbf{t}_{2n} = \mathbf{t}_n \\ \mathbf{t}_{2n+1} = 1 - \mathbf{t}_n. \end{cases}$$

Cette définition peut s'interpréter comme suit : si l'on ne conserve, dans la suite \mathbf{t} , que les termes d'indices pairs, on retrouve la suite \mathbf{t} . Si en revanche on ne garde que les termes d'indices impairs, on obtient la suite opposée, où les 0 et 1 sont échangés.

2.2.2 Relation de récurrence 2

Soient $(u_n)_n$ et $(v_n)_n$ les suites de mots définies par : $u_0 = 0$, $v_0 = 1$ et pour tout entier n non nul, $u_{n+1} = u_n v_n$ et $v_{n+1} = v_n u_n$. Alors,

$$\mathbf{t} = \lim_{n \rightarrow +\infty} u_n$$

2.2.3 Définition directe

Le n^{e} terme de la suite \mathbf{t} est égal à la somme, modulo 2, des chiffres dans le développement binaire de n .

Par exemple :

$$(18)_2 = 10010, \text{ donc } \mathbf{t}_{18} = 0;$$

$$(19)_2 = 10011, \text{ donc } \mathbf{t}_{19} = 1;$$

$$(20)_2 = 10100, \text{ donc } \mathbf{t}_{20} = 0.$$

2.2.4 Définition par morphisme

Le mot \mathbf{t} est également obtenu en itérant le morphisme μ appelé morphisme de Thue-Morse défini par $\mu(0) = 01$ et $\mu(1) = 10$.

Les premières itérations à partir de 0 donnent :

$$\mu^0(0) = 0$$

$$\mu^1(0) = 01$$

$$\mu^2(0) = 0110$$

$$\mu^3(0) = 01101001$$

$$\mu^4(0) = 0110100110010110.$$

Si l'on itère à partir de 1, on obtient la suite complémentaire, c'est à dire que les 0 et 1 sont échangés.

Le mot de Thue-Morse a joué et continue de jouer un rôle important dans divers domaines des mathématiques grâce à ses nombreuses propriétés. En e et, en plus des problèmes cités dans l'introduction, il intervient aussi dans la partition des entiers.

2.3 Problème de Prouhet-Tarry-Escott

La suite **t** apparaît implicitement dans un papier de Prouhet en 1851 [35]. Prouhet s'est intéressé a un problème de théorie des nombres qui sera étudié 50 ans plus tard par Tarry et Escott. Ce problème est de nos jours connu sous le nom de "Prouhet-Tarry-Escott" ou "Multigrades".

Dé nition 2.3.1 *Le problème de Prouhet-Tarry-Escott consiste à trouver une partition de l'ensemble $\{0, 1, \dots, 2^N - 1\}$ en deux ensembles disjoints I et J tels que*

$$\sum_{i \in I} i^k = \sum_{j \in J} j^k$$

où $k \in \{0, 1, 2, \dots, t\}$ avec t un entier naturel.

En prenant $0^0 = 1$, le cas particulier $k = 0$ montre que les ensembles I et J doivent avoir les mêmes nombres d'éléments.

Le problème peut être reformulé simplement comme suit :

Dé nition 2.3.2 *Pour tout entier naturel m , trouver un entier naturel r et deux suites d'entiers différentes (a_1, a_2, \dots, a_r) et (b_1, b_2, \dots, b_r) telles que :*

$$\begin{aligned}
a_1 + a_2 + a_3 + \dots + a_r &= b_1 + b_2 + b_3 + \dots + b_r \\
a_1^2 + a_2^2 + a_3^2 + \dots + a_r^2 &= b_1^2 + b_2^2 + b_3^2 + \dots + b_r^2 \\
a_1^3 + a_2^3 + a_3^3 + \dots + a_r^3 &= b_1^3 + b_2^3 + b_3^3 + \dots + b_r^3 \\
a_1^4 + a_2^4 + a_3^4 + \dots + a_r^4 &= b_1^4 + b_2^4 + b_3^4 + \dots + b_r^4 \\
&\vdots \\
&\vdots \\
&\vdots \\
a_1^m + a_2^m + a_3^m + \dots + a_r^m &= b_1^m + b_2^m + b_3^m + \dots + b_r^m.
\end{aligned}$$

Si (a_1, a_2, \dots, a_r) et (b_1, b_2, \dots, b_r) forment une solution du problème pour $m \in \mathbb{N}$, alors, l'entier r est appelé **taille** de la solution et l'entier m est son **degré**. On note $(a_1, a_2, \dots, a_r) \stackrel{m}{=} (b_1, b_2, \dots, b_r)$

Exemple 2.3.1 Les suites $(0, 3, 5, 6)$, $(1, 2, 4, 7)$, $(0, 3, 5, 6, 8, 11)$ et $(1, 2, 4, 7, 9, 10)$

$$0^1 + 3^1 + 5^1 + 6^1 = 1^1 + 2^1 + 4^1 + 7^1 = 14;$$

$$0^2 + 3^2 + 5^2 + 6^2 = 1^2 + 2^2 + 4^2 + 7^2 = 70.$$

Par conséquent, $(0, 3, 5, 6) \stackrel{2}{=} (1, 2, 4, 7)$. Cette solution est de taille 4 et de degré 2.

Par contre les suite $(0, 3, 5, 6, 8, 11)$ et $(1, 2, 4, 7, 9, 10)$ forment une solution de taille 6 et de degré 1

Prouhet fut le premier à donner une solution générale au problème Prouhet-Tarry-Escott. En effet, il a donné une solution de taille 2^m et de degré m , pour tout $m \geq 1$ en partitionnant l'ensemble des entiers $[0, 2^{m+1} - 1]$ en deux ensembles en utilisant le mot de Thue-Morse. Il établit ainsi le théorème suivant :

Théorème 2.3.1 Pour tout $m \geq 1$, il existe une solution de taille 2^m au problème de Prouhet-Tarry-Escott.

Pour prouver ce théorème, il considère d'abord le mot \mathbf{t}_2 de Thue-Morse et suppose que $m \geq 1$. Ensuite, pour tout $1 \leq i \leq 2^{m+1}$, il note a_i l'indice du i^e 0 et b_i celui du i^e 1 du mot de Thue-Morse. Il montre enfin que les suites $(a_1, a_2, \dots, a_{2^m})$ et $(b_1, b_2, \dots, b_{2^m})$ forment une solution du problème de Prouhet-Tarry-Escott. Pour plus de détails sur cette preuve, nous référons le lecteur à [50].

Ce résultat peut être énoncé le plus simplement comme suit :

Théorème 2.3.2 *La suite $\mathbf{t}_2 = (t_n)_{n \geq 0}$ vérifie les propriétés suivantes : si*

$$I = \{i \in \{0, 1, \dots, 2^N - 1\} : t_i = 0\}$$

$$J = \{j \in \{0, 1, \dots, 2^N - 1\} : t_j = 1\}$$

Alors, pour tout $0 \leq k \leq N - 1$, on a

$$\sum_{i \in I} i^k = \sum_{j \in J} j^k.$$

Exemple 2.3.2 *Pour $k = 0, 1, 2, 3$, on montre aisément que*

$$0^k + 3^k + 5^k + 6^k + 9^k + 10^k + 12^k + 15^k = 1^k + 2^k + 4^k + 7^k + 8^k + 11^k + 13^k + 14^k.$$

En effet, les 16 premiers termes de la suite \mathbf{t}_2 sont 0110100110010110. Les indices de la lettre 0 en ordre croissant sont 0, 3, 5, 6, 9, 10, 12 et 15. Ceux de la lettre 1 sont 1, 2, 4, 7, 8, 11, 13 et 14.

Prouhet a ensuite regardé le problème de façon générale. Plus précisément, son étude a consisté à trouver, pour tout naturel N , une partition de $[0, 2^N - 1]$ en q ensembles I_1, I_2, \dots, I_q tels que les sommes $\sum_{i \in I_j} i^k$, (avec $j = 1, 2, \dots, q$ et $k = 0, 1, 2, \dots, N - 1$) ne dépendent pas de j . Il résout ainsi ce problème en construisant la suite $\mathbf{t}_q = (t_q(n))_{n \geq 0}$ définie par $t_q(n) = s_q(n)$, où $s_q(n)$ est la somme des chiffres de la décomposition de l'entier n en base q . Il pose ensuite $I_j = \{0 \leq i \leq 2^N - 1 : t_q(i) = j\}$. Pour plus de détails sur ce travail, nous référons le lecteur à l'article [26].

2.4 Le curieux produit in ni

Dans [48] Woods pose la question suivante : Quelle est la limite de la suite

$$\frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \dots? \quad (1)$$

Robbins a répondu à cette question dans [39]. En effet, il prouve que cette suite a pour limite $\frac{\sqrt{2}}{2}$ en utilisant curieusement le mot de Thue-Morse à travers la proposition suivante.

Proposition 2.4.1 *Soit $v_n = (-1)^{t_n}$, où $(t_n)_{n \geq 0}$ est la suite de Thue-Morse. Alors,*

$$\left(\frac{1}{2}\right)^{v_0} \left(\frac{3}{4}\right)^{v_1} \left(\frac{5}{6}\right)^{v_2} \dots = \prod_{n=0}^{\infty} \left(\frac{2n+1}{2n+2}\right)^{v_n} = \frac{\sqrt{2}}{2}. \quad (|)$$

Preuve : Remarquons d'abord que tout terme de (1) correspond à $\frac{1}{n}$ à un certain rang. Par exemple, les 10 premiers termes de la suite de Thue-Morse sont 0, 1, 1, 0, 1, 0, 0, 1, 0, 1. Par suite, les 10 premiers termes de la suite $(v_n)_{n \geq 0}$ sont 1, -1, -1, 1, -1, 1, 1, -1, 1, -1. Par conséquent, le produit $\prod_{n=0}^9 \frac{1}{n+1}$ au rang 10 est donné par

$$\frac{1}{2} \cdot \frac{4}{3} \cdot \frac{6}{5} \cdot \frac{7}{8} \cdot \frac{10}{9} \cdot \frac{11}{12} \cdot \frac{13}{14} \cdot \frac{16}{15} \cdot \frac{17}{18} \cdot \frac{20}{19}.$$

Ce produit correspond au 6^e terme de la suite (1).

Pour la preuve, il considère P et Q , les produits infinis définis par

$$P = \prod_{n=0}^{\infty} \left(\frac{2n+1}{2n+2}\right)^{v_n}, Q = \prod_{n=1}^{\infty} \left(\frac{2n}{2n+1}\right)^{v_n}$$

Alors,

$$PQ = \frac{1}{2} \prod_{n=1}^{\infty} \left(\frac{n}{n+1}\right)^{v_n} = \frac{1}{2} \prod_{n=1}^{\infty} \left(\frac{2n}{2n+1}\right)^{v_{2n}} \prod_{n=0}^{\infty} \left(\frac{2n+1}{2n+2}\right)^{v_{2n+1}}$$

Il utilise ensuite le Théorème d'Abel pour assurer la convergence du produit PQ . Nous savons d'après la définition de la suite de Thue-Morse (Récurrence 1) que $t_{2n} = t_n$ et que $t_{2n+1} = 1 - t_n$. Ainsi, $v_{2n} = v_n$ et $v_{2n+1} = -v_n$. Par conséquent,

$$PQ = \frac{1}{2} \prod_{n=1}^{\infty} \left(\frac{2n}{2n+1} \right)^{v_n} \left(\prod_{n=0}^{\infty} \left(\frac{2n+1}{2n+2} \right)^{v_n} \right)^{-1} = \frac{1}{2} \frac{Q}{P}.$$

Observons que Q est formé de produits de réels non nuls. Ainsi, $Q \neq 0$ et on a $P^2 = \frac{1}{2}$. Le résultat s'ensuit puisque P est positif. ■

2.5 Partition de l'ensemble des entiers

Soit \mathbf{A} l'ensemble (ordonné) des entiers contenant 0 et 1 tel que pour tout élément $n \geq 1$ de \mathbf{A} , $2n$ n'est pas un élément de \mathbf{A} . Les premiers éléments de \mathbf{A} sont

$$0, 1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 21, 23, \dots$$

Notons que les ensembles \mathbf{A} et $2\mathbf{A} = \{2x : x \in \mathbf{A}\}$ forment une partition de l'ensemble des entiers naturels [4, 42, 16]. La relation entre l'ensemble \mathbf{A} et le mot de Thue-Morse prouvée dans [4] est donnée par le théorème suivant :

Théorème 2.5.1 *Soit $(a_n)_{n \geq 0} = 0, 1, 3, 4, 5, 7, 9, 11, 12, \dots$ la suite d'éléments de \mathbf{A} . On définit la suite $\mathbf{Z} = (z_n)_{n \geq 0}$ par :*

$$\mathbf{Z} = 0^{a_1-a_0} 1^{a_2-a_1} 0^{a_3-a_2} 1^{a_4-a_3} \dots 0^{a_{2n+1}-a_{2n}} 1^{a_{2n+2}-a_{2n+1}} \dots$$

Alors, \mathbf{Z} est égal à la suite de Thue-Morse. ■

Exemple 2.5.1 *Considérons les 16 premiers termes de la suite $(a_n)_{n \geq 0}$. Ces termes sont $\{0, 1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 21\}$ dans cet ordre. Par suite, les premiers termes correspondants de la suite \mathbf{Z} sont :*

$$0^{1-0} 1^{3-1} 0^{4-3} 1^{5-4} 0^{7-5} 1^{9-7} 0^{11-9} 1^{12-11} 0^{13-12} 1^{15-13} 0^{16-15} 1^{17-16} 0^{19-17} 1^{20-19} 0^{21-20}.$$

Ceci donne

$$011010011001011010010.$$

Ce mot correspond aux 21 premiers termes de la suite de Thue-Morse.

2.6 Carrés magiques

Un carré magique d'ordre $m \in \mathbb{N}$ est une matrice d'ordre $m \times m$, à coefficients distincts et dans $\{1, 2, \dots, m^2\}$ tels que la somme des coefficients de chaque ligne, colonne et diagonale est la même. Cette somme est égale à $\frac{1}{2}m(m^2 + 1)$ [8]. Dans la suite, nous montrons comment utiliser le mot de Thue-Morse pour construire un carré magique d'ordre 2^m , pour tout $m \geq 2$. Cette méthode a été établie par Adler et Li dans [2].

Pour construire un carré magique M d'ordre 2^m , pour tout $m \geq 2$, on numérote d'abord les termes de la suite en commençant par 1. Ensuite, on remplit la matrice M en considérant n comme son n^e coefficient si $t_{n-1} = 1$. Enfin, à l'aide des numéros restants réorganisés par ordre décroissant, on remplit M de gauche à droite et du haut vers le bas.

Exemple 2.6.1 *Utilisons la méthode pour construire un carré magique d'ordre $2^2 = 4$. Considérons les $4^2 = 16$ premiers termes du mot de Thue-Morse.*

$$\begin{array}{l} n : 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12 \quad 13 \quad 14 \quad 15 \quad 16 \\ t_{n-1} : 0 \quad \mathbf{1} \quad \mathbf{1} \quad 0 \quad \mathbf{1} \quad 0 \quad 0 \quad \mathbf{1} \quad \mathbf{1} \quad 0 \quad 0 \quad \mathbf{1} \quad 0 \quad \mathbf{1} \quad \mathbf{1} \quad 0 \end{array}$$

Remarquons que $t_{n-1} = 1$ si $n \in \{2, 3, 5, 8, 9, 12, 14, 15\}$. Les numéros restants pris en ordre décroissant sont 16, 13, 11, 10, 7, 6, 4, 1. Ils sont utilisés pour remplir les cases vides de la matrice en préservant l'ordre. Le carré magique qui en découle est donné par la **Figure 2.1**.

Dans cette figure, le premier carré correspond à la matrice partiellement remplie à l'aide de l'ensemble $\{2, 3, 5, 8, 9, 12, 14, 15\}$. Le second est le carré magique obtenu en complétant le premier à l'aide des nombres 16, 13, 11, 10, 7, 6, 4, 1 pris dans cet ordre.

	2	3	
5			8
9			12
	14	15	

→

16	2	3	13
5	11	10	8
9	7	6	12
4	14	15	1

FIGURE 2.1 – Carré magique d'ordre 2^2 du mot de Thue-Morse \mathbf{t}_2

2.7 Chevauchements

Un chevauchement est un mot de la forme $auaua$, où a est une lettre et u un mot (éventuellement vide). Un mot w est dit sans chevauchement s'il ne contient aucun chevauchement comme facteur.

Par exemple, considérons le mot ternaire suivant :

$$ababbaabaaabbbabbabbacacbcacccbaa.$$

Dans ce mot, les facteurs suivants sont des chevauchements :

aaa , avec $u =$;

$abbabba$, avec $u = bb$;

$cacbcacccbc$, avec $u = accb$.

Le résultat suivant est établi par Thue dans [43].

Théorème 2.7.1 *Le mot de Thue-Morse \mathbf{t}_2 est sans chevauchement.*

La preuve de ce théorème utilise les deux lemmes suivants :

Lemme 2.7.1 *Considérons l'ensemble $C = \{01, 10\}$. Pour tout x dans C^* , ni $0x0$ ni $1x1$ n'est dans C^* .*

Preuve : Soit $w \in C^*$. Alors, $|w|_0 = |w|_1$. Pour $x \in C^*$, on a $|0x0|_0 = |x|_0 + 2 = |x|_1 + 2 > |x|_1 = |0x0|_1$. Ainsi, $0x0$ n'est pas dans C^* . On procède de même pour $1x1$. ■

Lemme 2.7.2 *Soit $w \in \{0, 1\}^*$ et μ_2 le morphisme de Thue-Morse. Alors, si w est sans chevauchement, $\mu_2(w)$ est sans chevauchement.*

Preuve : Procédons par l'absurde. Soit w un mot de longueur minimale tel que $\mu_2(w)$ contient un chevauchement. Par suite, il existe des mots x, y et u dans $\{0, 1\}^*$ et une lettre a dans $\{0, 1\}$ tels que $\mu_2(w) = xauauay$. Comme μ_2 est 2-uniforme, alors la minimalité de la longueur de w implique que $|x|, |y| \leq 1$.

Cas 1 : $|x| = 1$. Alors, $|y| = 0$ puisque $auaua$ est de longueur impaire. De plus, $x = a$ puisque aa ne peut être le début de l'image d'un facteur par μ_2 . Ainsi, $\mu_2(w) = \bar{a}auaua$, où $\bar{a} = 1 - a$. Si u est de longueur paire, alors u et aua sont dans C^* . Ceci contredit le Lemme 2.7.1. Par suite, u est de longueur impaire et donc ua est dans l'image de μ_2 . C'est à dire qu'il existe v dans $\{0, 1\}^*$ tel que $ua = \mu_2(v\bar{a})$. Ainsi $\mu_2(w) = \bar{a}a\mu_2(v\bar{a})\mu_2(v\bar{a}) = \mu_2(\bar{a}v\bar{a}v\bar{a})$. Comme le morphisme μ_2 est injectif, alors $w = \bar{a}v\bar{a}v\bar{a}$. Par conséquent, w est un chevauchement.

Cas 2 : $|x| = 0$. Alors, $|y| = 1$. En procédant de manière analogue au **Cas 1**, On montre que $\mu_2(w) = auaua\bar{a}$. Si u est de longueur paire, alors, les facteurs u et aua sont dans C^* et ceci est une contradiction d'après le Lemme 2.7.1. Par suite u est de longueur impaire et donc au s'écrit comme image d'un facteur av' . Ainsi, $\mu_3(w) = \mu_2(av'av'a)$ et par injectivité de μ_2 , $w = av'av'a$. Par conséquent, w est un chevauchement. ■

Preuve du Théorème : D'après le Lemme 2.7.2, $\mu_2^n(0)$ est sans chevauchement pour tout entier n . Par suite, les préfixes de $\mu_2^\infty(0)$ sont sans chevauchement. Par conséquent, t_2 est sans chevauchement. ■

2.8 Équilibre et complexité abélienne

La complexité abélienne du mot de Thue-Morse est explicitement déterminée par Richomme, Saari et Zamboni dans [37]. Dans cette section, nous rappelons quelques

propriétés de \mathbf{t}_2 avant d'étudier son équilibre et sa complexité abélienne. Pour commencer, rappelons le résultat suivant que nous utiliserons dans la suite du travail. Il est établi par Coven et Hedlund dans [18]

Lemme 2.8.1 *Soit u un mot in ni. Alors, u est périodique de période p si et seulement si $\rho_u^{ab}(p) = 1$.*

Le lemme suivant sera utilisé dans la suite du travail.

Lemme 2.8.2 [37] *Soit $u = u_0u_1u_2\dots$ un mot in ni non ultimement périodique sur \mathcal{A}_2 . Alors, pour tout $k \geq 2$, il existe des facteurs v et w de u de longueur k , avec $v = 0s1$ et $w = 1t0$, où s et t sont des mots (éventuellement vides) sur \mathcal{A}_2 .*

Preuve : Supposons que pour un certain $k \geq 2$, le mot u ne contienne aucun facteur de longueur k commençant par 0 et finissant par 1. Alors, pour tout $i \geq 1$, la suite $(u_{i+n(k-1)})_{n \geq 1}$ est soit de la forme 1^s0^∞ ou 1^∞ . Ainsi, pour tout i suffisamment grand, on a $u_{i+k-1} = u_i$. Ce qui fait de u un mot ultimement périodique. Ceci est impossible puisque u est non ultimement périodique. De même, si u ne contient pas de facteur w de longueur k commençant par 1 et se terminant par 0, alors, il est ultimement périodique.

■

Théorème 2.8.1 [37] *La complexité abélienne d'un mot binaire aperiodique u est*

$$\rho_u^{ab}(n) = \begin{cases} 2 & \text{si } n \text{ est impair} \\ 3 & \text{si } n \text{ est pair} \end{cases}$$

si et seulement si il existe un mot u' tel que $u = \mu_2(u')$, $u = 0\mu_2(u')$ ou $u = 1\mu_2(u')$

Preuve : Supposons que la complexité abélienne du mot u est 2 si n est impair et 3 si n est pair et non nul. Montrons d'abord par récurrence que pour tout entier naturel k , on a

$$\Psi_u(2k+1) = \{(k+1, k), (k, k+1)\},$$

$$\Psi_u(2k+2) = \{(k+1, k+1), (k, k+2), (k+2, k)\}.$$

Pour $k = 0$ on a $\Psi_u(1) = \{(0, 1), (1, 0)\}$ et $\Psi_u(2) = \{(2, 0), (1, 1), (0, 2)\}$. Supposons que

$$\Psi_u(2k+2) = \{(k+1, k+1), (k, k+2), (k+2, k)\}.$$

Alors, d'après le lemme 1.4.1, il existe des entiers naturels p et q tels que

$$\Psi_u(2k+3) = \{(p+1, q), (p, q+1)\}.$$

De plus

$$\Psi_u(2k+3) \cup \psi(2k+2) + \{(1, 0), (0, 1)\} = \{(k+2, k+1), (k+1, k+2), (k, k+3), (k+3, k)\}.$$

Ceci aboutit à l'une des égalités suivantes

$$\Psi_u(2k+3) = \{(k+2, k+1), (k+1, k+2)\};$$

$$\Psi_u(2k+3) = \{(k+2, k+1), (k+3, k)\};$$

$$\Psi_u(2k+3) = \{(k+1, k+2), (k, k+3)\}.$$

Observons que $\{(k+2, k+1), (k+3, k)\} \cup \{(1, 0), (0, 1)\}$ est un ensemble ne contenant pas $(k, k+2)$. Ainsi, la deuxième égalité n'est pas possible puisque $(k, k+2)$ est dans $\Psi_u(2k+2)$. De même, la troisième égalité n'est pas possible, car $(k+2, k)$ est dans $\Psi_u(2k+2)$. Par suite

$$\Psi_u(2k+4) = \Psi_u(2k+3) + \{(1, 0), (0, 1)\} = \{(k+3, k+1), (k+2, k+2), (k+1, k+3)\}.$$

Comme $\rho_u^{ab}(2k+4) = 3$, l'inclusion précédente est une égalité et la récurrence est prouvée.

Comme $\Psi_u(3) = \{(1, 2), (2, 1)\}$, alors les facteurs 000 et 111 ne sont pas dans u . Supposons à présent que u contient un facteur de la forme $11(01)^k 011$. Alors, ce mot admet $(k+4, k+1)$ comme vecteur de Parikh. Ce qui n'est pas possible d'après ce qui précède. Ainsi, u ne contient aucun facteur de la forme $11(01)^k 011$. De façon analogue, on montre que u ne contient aucun facteur de la forme $00(10)^k 100$. Par suite, entre deux occurrences consécutives de 00 il y a une occurrence de 11 et entre deux occurrences successives de 11, il y a une occurrence de 00. Comme u est apériodique, les facteurs 00

et 11 apparaissent une infinité de fois dans u . Ainsi, u peut se décomposer sous la forme $pu_1u_2...u_k...$, où chaque u_i est un mot dans $1(01)^*00(10)^*1$ et p un suffixe d'un mot dans $1(01)^*00(10)^*1$. Par conséquent, $u = \mu_2(u')$, $u = 0\mu_2(u')$ ou $u = 1\mu_2(u')$, où u' est un mot.

Réciproquement, soit u' un mot infini tel que $u = \mu_2(u')$, $u = 0\mu_2(u')$ ou $u = 1\mu_2(u')$. On suppose que u est apériodique.

Soit $n = 2k + 1$. Alors, tout facteur v de u de longueur n est de la forme $\mu_2(v')a$ ou $a\mu_2(v')$, avec v' un facteur de longueur k de u' et $a \in \{0, 1\}$. Par suite

$$\Psi_u(n) = \{(k+1, k), (k, k+1)\}.$$

D'après le Lemme 2.8.1, on a $\rho_u^{ab}(n) \geq 2$ et donc

$$\rho_u^{ab}(n) = 2$$

.

Supposons à présent que $n = 2k$. Alors, tout facteur v de longueur n de u est sous la forme $\mu_2(s)$ ou $a\mu_2(t)b$, où s et t sont des facteurs de longueur k et $k-1$ respectivement de u' et $a, b \in \mathcal{A}_2$. Par suite,

$$\Psi_u(n) = \{(k, k), (k-1, k+1), (k+1, k-1)\}.$$

D'après le lemme 2.8.2, u' contient un facteur s (resp. t) de longueur $k-1$ précédé par 0 et suivi par 1 (resp. précédé par 1 et suivi par 0). Ainsi, u contient les facteurs de longueur n de la forme $1\mu_2(s)1$ et $0\mu_2(t)0$ qui ont respectivement pour vecteurs de Parikh $(k-1, k+1)$ et $(k+1, k-1)$. Par suite,

$$\Psi_u(n) = \{(k, k), (k-1, k+1), (k+1, k-1)\}.$$

Donc

$$\rho_u^{ab}(n) = 3 \blacksquare$$

En conséquence on a le résultat suivant.

Théorème 2.8.2 *La complexité abélienne du mot de Thue-Morse est donnée par :*

$$\rho_{\mathbf{t}_2}^{ab}(n) = \begin{cases} 2 & \text{si } n \text{ est impair} \\ 3 & \text{si } n \text{ est pair} \end{cases}$$

Preuve : Il suffit d'utiliser le théorème précédent du fait que $\mathbf{t}_2 = \mu_2(\mathbf{t}_2)$. ■

De la preuve du Théorème 2.9.1, nous déduisons que le mot \mathbf{t}_2 est 2-équilibré.

2.9 Mots de retour et facteurs privilégiés

2.9.1 Mots de retour

Les mots de retour du mot de Thue-Morse ont été étudiés par Balková, Palentová et Steiner dans [7]. Ils retrouvent l'ensemble des mots de retour de tout facteur w de \mathbf{t}_2 à l'aide de celui des facteurs bispéciaux.

Proposition 2.9.1 *Soit w un facteur de \mathbf{t}_2 . Alors, w est spécial à droite (resp. à gauche) si et seulement si $\mu_2(w)$ est spécial à gauche (resp. à droite).*

Preuve : Soit w un facteur de \mathbf{t}_2 spécial à droite. Alors, wi est facteur de \mathbf{t}_2 , pour tout $i \in \mathcal{A}_2$. Par suite, $\mu_2(wi)$ est aussi dans \mathbf{t}_2 . Comme l'image de chaque lettre commence par elle-même, alors $\mu_2(w)$ est spécial à droite.

Réciproquement, supposons que $\mu_2(w)$ est spécial à droite. Alors, $\mu_2(w)0$ et $\mu_2(w)1$ sont dans \mathbf{t}_2 . Ainsi, les mots $\mu_2(w)01$ et $\mu_2(w)10$ sont aussi des facteurs de \mathbf{t}_2 puisque 0 (resp. 1) est le début de l'image de 0 (resp. 1) et \mathbf{t}_2 est récurrent. Ces facteurs s'écrivent respectivement $\mu_2(w0)$ et $\mu_2(w1)$. Par conséquent, les mots $w1$ et $w0$ sont dans \mathbf{t}_2 et donc w est spécial à droite.

On procède de même lorsque w est spécial à gauche. ■

En conséquence, un facteur w est bispécial si et seulement si $\mu_2(w)$ est bispécial.

Proposition 2.9.2 *Soit w un facteur bispécial de \mathbf{t}_2 de longueur supérieure à 4. Alors, il existe un facteur bispécial v de \mathbf{t}_2 tel que $w = \mu_2(v)$.*

Preuve : Soit w un facteur bispécial de \mathbf{t}_2 tel que $|w| \geq 4$. Alors, w se synchronise sous la forme $i\mu_2(v)j$ avec $v \in \mathbf{t}_2$ et i (resp. j) est le suffixe (resp. préfixe) de l'image d'une lettre. Comme w est à la fois spécial à gauche et spécial à droite, i et j sont nécessairement vides. Sinon, w admettrait un unique prolongement à gauche (resp. à droite). Par conséquent, $w = \mu_2(v)$ et v est bispécial d'après la Proposition 2.9.1. ■

Théorème 2.9.1 *Tout facteur de \mathbf{t}_2 possède 3 ou 4 mots de retour.*

Preuve : Soit v un facteur de \mathbf{t}_2 . Nous savons que pour avoir le cardinal de $Ret(v)$, il suffit de déterminer le cardinal de $Ret(w)$, où w est le facteur bispécial étendu de v (voir section 1.6). D'après le Lemme 1.6.1 et la Proposition 2.9.2, nous pouvons nous restreindre aux facteurs bispéciaux de longueur inférieure à 4. Comme les lettres jouent des rôles symétriques, nous pouvons considérer les facteurs bispéciaux commençant par 0. Il s'agit alors de 0, 01, 010. Ainsi, on a

$$Ret(0) = \{0, 01, 011\};$$

$$Ret(01) = \{01, 010, 011, 0110\},$$

$$Ret(010) = \{010, 0100110, 01011, 010110011\}.$$

Ce qui prouve le théorème. ■

2.9.2 Facteurs privilégiés

L'étude des facteurs privilégiés du mot de Thue-Morse a été initiée par Peltomäki. Il a notamment étudié dans [34] la fonction de complexité privilégiée de \mathbf{t}_2 . Il a obtenu une formule décrivant cette fonction. A l'aide de cette formule, il montre que la fonction d'une part est non bornée et d'autre part s'annule en une infinité de points. Pour la suite, on note pour tout facteur privilégié v de \mathbf{t}_2 , $Pri_n(v)$ l'ensemble des facteurs privilégiés de longueur n de \mathbf{t}_2 commençant par v et $A_n(v) = |Pri_n(v)|$.

Il établit la proposition suivante :

Proposition 2.9.3 *La fonction de complexité privilégiée du mot de Thue-Morse est donnée par*

$$\frac{1}{2}A_{\mathbf{t}_2}(n) = A_n(00) + A_n(010) + A_n(0110).$$

Pour la preuve de cette proposition, nous utilisons le morphisme échange E_2 défini de \mathcal{A}_2^* dans \mathcal{A}_2^* par $E_2(0) = 1$ et $E_2(1) = 0$. Pour tout facteur w de \mathbf{t}_2 , $E_2(w)$ est aussi dans \mathbf{t}_2 , puisque $E_2\mu_2 = \mu_2E_2$.

Preuve : Pour déterminer les facteurs privilégiés du mot \mathbf{t}_2 , nous considérons les facteurs commençant par 0. Les autres s'obtiennent par application de E_2 . Nous savons de plus que \mathbf{t}_2 ne contient pas de cube. Par suite, les mots de retour complet de 0 sont 00, 010 et 0110. Ainsi, les facteurs privilégiés de \mathbf{t}_2 commençant par 0 s'obtiennent à l'aide de ceux des trois premiers mots de retour complet de 0. Par conséquent,

$$Pri_n(0) = Pri_n(00) \quad Pri_n(010) \quad Pri_n(0110)$$

pour $n \geq 1$. Ainsi, en associant les facteurs privilégiés commençant par 1 on obtient

$$\frac{1}{2}A_{\mathbf{t}_2}(n) = A_n(00) + A_n(010) + A_n(0110). \blacksquare$$

Le mot \mathbf{t}_2 étant sans chevauchement, alors il ressort que $Pri_1(0) = \{0\}$, $Pri_2(0) = \{00\}$, $Pri_3(0) = \{010\}$, $Pri_4(0) = \{0110\}$. Ainsi, $A_{\mathbf{t}_2}(1) = A_{\mathbf{t}_2}(2) = A_{\mathbf{t}_2}(3) = A_{\mathbf{t}_2}(4) = 2$.

Dans la suite, nous donnons quelques résultats sur les facteurs privilégiés du mot de Thue-Morse établis dans [34].

Posons $R = \{00, 010, 0110\}$. Alors on a

Lemme 2.9.1 *Soit $w \in Pri_n(0)$ tel que $|w| \geq 5$. Alors, w commence par un mot de retour complet d'un élément de R .*

Preuve : Comme $|w| \geq 5$, alors, il admet v comme préfixe propre, où v est l'un des mots 00, 010 ou 0110. Le facteur w étant privilégié, v est aussi son suffixe. Ainsi, w admet un mot de retour complet de v comme préfixe. \blacksquare

Proposition 2.9.4 *La fonctions. de complexité privilégiée du mot de Thue-Morse vérifie la condition suivante :*

$$A_{\mathbf{t}_2}(2n+1) = 0.$$

Preuve : Nous traitons le cas des facteurs privilégiés commençant par 0. Soit w un facteur privilégié de \mathbf{t}_2 commençant par 0 tel que $|w| \geq 5$. Alors, w commence par l'un des mots privilégiés 00, 010 ou 0110. En considérant les images des lettres par μ_2 , nous pouvons séparer ces mots comme suit : 0|0, 01|0, 0|10 et 01|10.

Supposons que w commence par 00 (resp. 0110). Alors, il se termine par 00 (resp. 0110).

Ainsi, w s'écrit sous la forme $0\mu_2(v)0$ (resp. $\mu_2(v)$) et donc a une longueur paire.

Supposons à présent que w commence par 010. Comme il est de longueur supérieure à 5, alors d'après le Lemme 2.9.1, w admet un des facteurs suivants comme préfixe : 01|01|10|10, 01|01|10|01|10|10, 0|10|01|0 ou 0|10|01|10|01|0. Comme w est privilégié, il se termine par l'un de ces facteurs. Par conséquent w est de longueur paire. ■

CHAPITRE 3

COMPLEXITÉ ABÉLIENNE DU MOT DE THUE-MORSE TERNAIRE

3.1 Introduction

Nous savons que le mot de Thue-Morse binaire \mathbf{t}_2 est 2-équilibré et que sa complexité abélienne vérifie $\rho_{\mathbf{t}_2}^{ab}(n) = 2$ si n est impair et $\rho_{\mathbf{t}_2}^{ab}(n) = 3$ si n est pair et non nul. Dans ce chapitre nous montrons que la complexité abélienne du mot de Thue-Morse ternaire \mathbf{t}_3 est donnée par la suite ultimement périodique $(\rho_{\mathbf{t}_3}^{ab}(n))_{n \geq 0} = 136(766)$.

Les résultats de ce chapitre ont été présentés à **WORDS 2017** et publiés dans **Lecture Notes in Computer Science** [24].

Nous débutons par quelques propriétés combinatoires, puis nous étudions les facteurs triprolongeables de \mathbf{t}_3 . Ensuite, nous montrons qu'il est 2-équilibré. Nous terminons le chapitre par l'étude de la complexité abélienne de \mathbf{t}_3 .

3.2 Définition et premières propriétés

En rappel, le mot de Thue-Morse ternaire est le mot \mathbf{t}_3 , engendré par le morphisme μ_3 défini par $\mu_3(0) = 012, \mu_3(1) = 120, \mu_3(2) = 201$ et dont les premières lettres sont :

$$\mathbf{t}_3 = 012120201120201012201012120120201012...$$

Pour tout entier naturel n , on montre que $\mu_3^n(a)$ est un préfixe de $\mu_3^{n+1}(a)$ pour toute lettre a de \mathcal{A}_3 . En effet, en considérant la lettre 0 on a :

$$\mu_3(0) = 012;$$

$$\mu_3^2(0) = \mu_3(012) = \mu_3(0)120201;$$

$$\mu_3^3(0) = \mu_3^2(012) = \mu_3^2(0)120201012201012120.$$

Supposons que la propriété est vérifiée jusqu'à l'ordre n . C'est à dire que $\mu_3^n(0)$ est un préfixe de $\mu_3^{n+1}(0)$. Montrons que $\mu_3^{n+1}(0)$ est un préfixe de $\mu_3^{n+2}(0)$. On a

$$\mu_3^{n+2}(0) = \mu_3(\mu_3^{n+1}(0)) = \mu_3^{n+1}(\mu_3(0)) = \mu_3^{n+1}(0)\mu_3^{n+1}(12).$$

Ainsi, $\mu_3^{n+1}(0)$ est un préfixe de $\mu_3^{n+2}(0)$. Par conséquent, $\mathbf{t}_3 = \lim_{n \rightarrow \infty} \mu_3^n(0)$.

Le morphisme μ_3 est primitif car l'image de toute lettre contient toutes les lettres de l'alphabet \mathcal{A}_3 . De plus μ_3 est 3-uniforme puisque pour toute lettre a de \mathcal{A}_3 , $|\mu_3(a)| = 3$. En plus, μ_3 est prolongeable en toute lettre de \mathcal{A}_3 . Par conséquent, le mot \mathbf{t}_3 est uniformément récurrent d'après le Théorème 1.3.1.

Soit $u = u_1u_2...u_n$ un mot sur un alphabet \mathcal{A} . Rappelons que le miroir de u est le mot $\bar{u} = u_n...u_2u_1$. Si u et v sont des mots de \mathcal{A}^* , alors $\overline{uv} = \bar{v}\bar{u}$. On appelle $(0, 2)$ -complémentation de $u \in \mathcal{A}_3^*$, le mot noté \tilde{u} et défini par $\tilde{u} = (2 - u_1)(2 - u_2)...(2 - u_n)$. Pour tout $u \in \mathcal{A}_3^*$, on a $\widetilde{\tilde{u}} = u$. Ainsi, nous posons $\hat{u} = \widetilde{\tilde{u}}$. L'opération $\hat{\cdot}$ est involutive et $\hat{\cdot}$ est l'identité : $\hat{\hat{u}} = u$ pour tout $u \in \mathcal{A}_3^*$.

Exemple 3.2.1 Considérons le facteur $u = \mu_3^2(001)$ de \mathbf{t}_3 .

Alors $u = 012120201012120201120201012$. On a :

$$\bar{u} = 210102021102021210102021210;$$

$$\tilde{u} = 210102021210102021102021210.$$

$$\hat{u} = 012120201120201012120201012$$

et

$$\hat{\hat{u}} = 012120201012120201120201012 = u.$$

Les propriétés suivantes sont inspirées de certaines propriétés vérifiées par le mot de Thue-Morse binaire \mathbf{t}_2 .

Proposition 3.2.1 *Considérons les suites (u_n) , (v_n) et (w_n) définies par $u_0 = 0$, $v_0 = 1$, $w_0 = 2$ et pour tout entier n non nul, $u_{n+1} = u_n v_n w_n$, $v_{n+1} = v_n w_n u_n$, et $w_{n+1} = w_n u_n v_n$. Alors :*

1. Pour tout $n \geq 0$, on a : $u_n = \mu_3^n(0)$, $v_n = \mu_3^n(1)$, $w_n = \mu_3^n(2)$.
2. Les suites (u_n) , (v_n) et (w_n) vérifient :

$$\hat{u}_{3n} = w_{3n} \text{ et } \hat{v}_{3n} = v_{3n};$$

$$\hat{u}_{3n+1} = u_{3n+1} \text{ et } \hat{v}_{3n+1} = w_{3n+1};$$

$$\hat{u}_{3n+2} = v_{3n+2} \text{ et } \hat{w}_{3n+2} = w_{3n+2}.$$

Preuve :

1. Procédons par récurrence sur n .

Les formules sont assurées pour $n = 0$ puisque nous avons par définition $u_0 = 0 = \mu_3^0(0)$, $v_0 = 1 = \mu_3^0(1)$ et $w_0 = 2 = \mu_3^0(2)$.

Supposons que pour tout $n \geq 1$, $u_n = \mu_3^n(0)$, $v_n = \mu_3^n(1)$ et $w_n = \mu_3^n(2)$. Par suite on a :

$$u_{n+1} = u_n v_n w_n = \mu_3^n(0) \mu_3^n(1) \mu_3^n(2) = \mu_3^n(012) = \mu_3^{n+1}(0);$$

$$v_{n+1} = v_n w_n u_n = \mu_3^n(1) \mu_3^n(2) \mu_3^n(0) = \mu_3^n(120) = \mu_3^{n+1}(1);$$

$$w_{n+1} = w_n u_n v_n = \mu_3^n(2) \mu_3^n(0) \mu_3^n(1) = \mu_3^n(201) = \mu_3^{n+1}(2).$$

2. Les relations sont vérifiées pour $n = 0$. Supposons que pour $n \geq 1$ on a :

$$\hat{u}_{3n} = w_{3n}, \text{ et } \hat{v}_{3n} = v_{3n};$$

$$\widehat{u}_{3n+1} = u_{3n+1} \text{ et } \widehat{v}_{3n+1} = w_{3n+1};$$

$$\widehat{u}_{3n+2} = v_{3n+2} \text{ et } \widehat{w}_{3n+2} = w_{3n+2}.$$

Alors, on a :

$$\begin{cases} \widehat{u}_{3(n+1)} = u_{3n+2} \widehat{v_{3n+2} w_{3n+2}} = \widehat{w_{3n+2} v_{3n+2}} \widehat{u_{3n+2}} = w_{3n+2} u_{3n+2} v_{3n+2} = w_{3(n+1)} \\ \widehat{v}_{3(n+1)} = v_{3n+2} \widehat{w_{3n+2} u_{3n+2}} = \widehat{u_{3n+2} w_{3n+2}} \widehat{v_{3n+2}} = v_{3n+2} w_{3n+2} u_{3n+2} = v_{3(n+1)} \end{cases}$$

On vérifie de même que

$$\widehat{u}_{3(n+1)+1} = u_{3(n+1)+1}, \widehat{v}_{3(n+1)+1} = w_{3(n+1)+1}$$

$$\widehat{u}_{3(n+1)+2} = v_{3(n+1)+2}, \widehat{w}_{3(n+1)+2} = w_{3(n+1)+2}. \blacksquare$$

Dé nition 3.2.1 On dé nit le morphisme E_3 sur \mathcal{A}_3 par :

$$E_3(0) = 1, E_3(1) = 2, E_3(2) = 0.$$

Nous avons les résultats suivants.

Lemme 3.2.1 Soit $u \in C^*$ où $C = \{012, 120, 201\}$. Alors, iui n'est pas dans C^* , pour tout $i \in \mathcal{A}_3$.

Preuve : Si $u \in C^*$, alors $|iui|$ n'est pas multiple de trois. Donc iui n'est pas dans C^* .

■

Proposition 3.2.2 1. $E_3(F_n(\mathbf{t}_3)) = F_n(\mathbf{t}_3)$.

2. Pour tout $a \in \mathcal{A}_3$, aaa n'est pas dans \mathbf{t}_3 ;

3. Pour tout $a \in \mathcal{A}_3$, $\mu_3(a)\mu_3(a)a$ n'est pas dans \mathbf{t}_3 ;

Preuve :

1. Soit $w \in F_n(\mathbf{t}_3)$. Alors $E_3(w)$ et w ont la même longueur puisque E_3 échange juste les lettres dans w . Par ailleurs, il existe un entier n_0 tel que w apparaît dans $u_{n_0} = \mu_3^{n_0}(0)$. Par suite, $E_3(w)$ apparaît dans $E_3(\mu_3^{n_0}(0))$. De plus, on a $E_3\mu_3 = \mu_3E_3$. Donc, $E_3(w)$ apparaît dans $\mu_3^{n_0}(1) = v_{n_0}$. Ainsi, $E_3(w) \in F_n(\mathbf{t}_3)$ puisque $v_{n_0} \in F_n(\mathbf{t}_3)$. Par conséquent $E_3(F_n(\mathbf{t}_3)) \subseteq F_n(\mathbf{t}_3)$. Comme E_3 est bijective, alors $E_3(F_n(\mathbf{t}_3)) = F_n(\mathbf{t}_3)$.

2. Il suffit de remarquer que tout carré de lettre commence (resp. finit) par la fin (resp. le début) de l'image d'une lettre. Comme aucune image de lettre ne commence ni se termine par un carré, alors \mathbf{t}_3 ne peut contenir de cube de lettre.
3. Procédons par l'absurde. Quitte à échanger les rôles des lettres, supposons que $a = 0$ et que $\mu_3(0)\mu_3(0)0 \in \mathbf{t}_3$. Observons que le dernier 0 est le préfixe de l'image de 0. Par ailleurs, $\mu_3(0)\mu_3(0)0 = 0\mu_3(1)\mu_3(1)$ et le 0 de $0\mu_3(1)\mu_3(1)$ est le suffixe de l'image de 1. En prolongeant $\mu_3(0)\mu_3(0)0$ (resp. $0\mu_3(1)\mu_3(1)$) à droite (resp. à gauche) par deux lettres, on obtient $\mu_3(0)\mu_3(0)\mu_3(0) \in \mathbf{t}_3$ (resp. $\mu_3(1)\mu_3(1)\mu_3(1) \in \mathbf{t}_3$). Ceci est impossible d'après 2. ■

Lemme 3.2.2 *Soit w un mot sur \mathcal{A}_3 . Si w est sans chevauchement, alors $\mu_3(w)$ est sans chevauchement.*

Preuve : Procédons par contraposition. Soit w un mot fini sur \mathcal{A}_3 de longueur minimale tel que $\mu_3(w)$ soit un chevauchement. Alors, il existe des facteurs x , y et u dans \mathcal{A}_3^* et une lettre a dans \mathcal{A}_3 tels que $\mu_3(w) = xauauay$. Comme le morphisme μ_3 est 3-uniforme et w est de longueur minimale, si le mot x (resp. y) est de longueur au moins 3, alors l'image de w privé de sa première (resp. dernière) lettre est un chevauchement. Ceci contredirait la minimalité de la longueur de w . Par conséquent, chacun des mots x et y est de longueur au plus 2. En considérant les longueurs de x et y , les cas suivants se dégagent.

Cas 1 : Le facteur x est de longueur 2. On peut alors écrire $\mu_3(w) = ijaauauay$, où i, j sont des lettres de \mathcal{A}_3 avec $i = j$ puisque ij est le préfixe de l'image d'une lettre. Par suite, ija représente l'image de la lettre i . Par suite, ua est l'image d'un facteur par μ_3 puisque u commence par l'image d'une lettre. Ainsi, y est le mot vide et on a $\mu_3(w) = ijauaia$. Il existe alors un mot v dans \mathcal{A}_3^* tel que $ua = \mu_3(vi)$. Par conséquent, $\mu_3(w) = \mu_3(ivivi)$ et $w = ivivi$ par injectivité.

Cas 2 : Le facteur x est de longueur 1. Alors $\mu_3(w) = iauauay$, où i est une lettre de \mathcal{A}_3 . Par suite, la première lettre de u correspond à la dernière lettre de l'image de i . Le mot u peut alors s'écrire $u = ju'$, où u' est un suffixe de u et j la dernière lettre

de l'image de i . Par suite, $\mu_3(w) = iaju'aju'ay$. Comme ija est le l'image de la lettre i , alors le mot u' commence par l'image d'une lettre. Ainsi, le mot u' se termine par la lettre i puisque aj est le su xte de l'image de i . Ainsi, $\mu_3(w) = iaju''iaju''ia y$, où u'' est un préfixe de u' . Comme y est de longueur au plus 2 et μ_3 est 3-uniforme, alors $y = j$ et donc u'' est l'image d'un mot v par μ_3 . Le mot $\mu_3(w)$ s'écrit alors $\mu_3(w) = \mu_3(ivivi)$ et w est un chevauchement par injectivité.

Cas 3 : Le facteur x est de longueur 0. Alors, en procédant comme au cas 1, on montre que le mot $\mu_3(w)$ s'écrit sous la forme $\mu_3(w) = auauaij$, où i et j sont des lettres di érentes. Par suite, u commence par ij et on a $\mu_3(w) = aiju'aiju'aij$, où u' est un su xte de u . Ainsi, u' est l'image d'un facteur v par μ_3 . Par conséquent, $\mu_3(w) = \mu_3(avava)$ et w est un chevauchement.

En définitive, $\mu_3(w)$ est sans chevauchement si w est sans chevauchement. ■

Théorème 3.2.1 *Le mot \mathbf{t}_3 est sans chevauchement.*

Preuve : D'après le Lemme 3.2.2, $\mu_3^n(0)$ est sans chevauchement, pour tout naturel n , puisque $\mu_3(0) = 012$ est sans chevauchement. Ainsi, tous les préfixes de $\mu_3^\infty(0)$ sont sans chevauchement. Par conséquent, \mathbf{t}_3 est sans chevauchement en tant que point fixe commençant par 0 du morphisme μ_3 . ■

3.3 Facteurs triprolongeables et équilibre

Dans cette section nous établissons quelques propriétés de \mathbf{t}_3 , puis nous montrons qu'il est 2-équilibré.

Lemme 3.3.1 *Soit u un facteur de \mathbf{t}_3 . Alors, il existe des facteurs $v, _1$ et $_2$ tels que $u = _1\mu_3(v)_2$ avec $|_1|, |_2| \leq 2$. La décomposition est unique si $|u| \geq 7$.*

Preuve : Pour l'unicité, voir [5]

Le mot \mathbf{t}_3 étant purement morphique, la décomposition $u = _1\mu_3(v)_2$ existe avec $|_1|, |_2| \leq 2$. Montrons que v est un facteur de \mathbf{t}_3 .

Si $|u| \leq 6$, alors v est de longueur au plus 2 et est dans \mathbf{t}_3 , puisque tout mot de longueur au plus 2 de \mathcal{A}_3 est dans \mathbf{t}_3 .

Supposons que v est non vide et de longueur au moins 7. Comme $\mu_3(v)$ est dans \mathbf{t}_3 et $\mathbf{t}_3 = \mu_3(\mathbf{t}_3)$, alors il existe w , un facteur de \mathbf{t}_3 tel que $\mu_3(v)$ est dans $\mu_3(w)$. Supposons que w est de longueur minimale. On a $|\mu_3(v)| \leq |\mu_3(w)|$. Par suite, il existe $v_1, v_2 \in F(\mathbf{t}_3)$ tels que $\mu_3(w) = v_1\mu_3(v)v_2$. Le morphisme μ_3 étant 3-uniforme et $|w|$ minimale, alors $|v_1| \leq 2$ (resp. $|v_2| \leq 2$). Par unicité de la décomposition, v_1 et v_2 sont vides. Par suite $\mu_3(w) = \mu_3(v)$ et $w = v$ par injectivité de μ_3 . Donc $v \in F(\mathbf{t}_3)$. ■

Ce lemme s'étend naturellement au mot \mathbf{t}_q de Thue-Morse sur un alphabet à q lettres, $q \geq 4$. Dans ce cas, la décomposition est unique si le facteur est de longueur supérieure à $2q$.

Proposition 3.3.1 *Soit u un facteur de \mathbf{t}_3 . Alors, u est triprolongeable à droite (resp. à gauche) si et seulement si $\mu_3(u)$ est triprolongeable à droite (resp. à gauche).*

Preuve : Soit u un facteur de \mathbf{t}_3 triprolongeable à droite. Alors, pour tout $i \in \mathcal{A}_3$, ui est dans \mathbf{t}_3 . Par suite, $\mu_3(u)i$ est dans \mathbf{t}_3 , puisque $\mu_3(i)$ commence par i .

Réciproquement, soit u un facteur de \mathbf{t}_3 tel que $\mu_3(u)$ est triprolongeable à droite avec $|u| \geq 2$ (le cas $|u| \leq 1$ est évident). Alors, $\mu_3(u)i$ est dans \mathbf{t}_3 , pour tout $i \in \mathcal{A}_3$. Ainsi, le facteur $\mu_3(u)i$ finit par la première lettre de l'image de $\mu_3(i)$, pour tout $i \in \mathcal{A}_3$; on utilise ici l'unicité du Lemme 3.3.1 car $|\mu_3(u)i| \geq 7$. Par suite, les facteurs $\mu_3(u)012$, $\mu_3(u)120$ et $\mu_3(u)201$ sont dans \mathbf{t}_3 . Ces trois facteurs s'écrivent respectivement $\mu_3(u)0$, $\mu_3(u)1$ et $\mu_3(u)2$. D'après le Lemme 3.3.1, $u0$, $u1$ et $u2$ sont dans \mathbf{t}_3 . Ce qui prouve que u est triprolongeable à droite dans \mathbf{t}_3 . On procède de même pour les facteurs triprolongeables à gauche. ■

Pour la suite, on désigne par $BST(\mathbf{t}_3)$, l'ensemble des facteurs de \mathbf{t}_3 à la fois triprolongeables à droite et triprolongeables à gauche.

En vertu de la Proposition 3.3.1, un facteur u est dans $BST(\mathbf{t}_3)$ si et seulement si $\mu_3(u)$ est dans $BST(\mathbf{t}_3)$.

Proposition 3.3.2 *Soit u un élément de $BST(\mathbf{t}_3)$. Si $|u| \geq 3$, alors il existe u' dans $BST(\mathbf{t}_3)$ tel que $u = \mu_3(u')$.*

Preuve : Soit $u \in BST(\mathbf{t}_3)$ tel que $3 \leq |u| \leq 6$ et commençant par 0. Alors, $u = 012$ ou $u = 012120$. Par suite, $u = \mu_3(0)$ ou $u = \mu_3(01)$. Les facteurs 0 et 01 sont à la fois triprolongeables à droite et triprolongeables à gauche dans \mathbf{t}_3 .

Supposons à présent que $|u| \geq 7$. Alors, le facteur u s'écrit de manière unique sous la forme $u = {}_1\mu_3(u'){}_2$, où u' , ${}_1$ et ${}_2$ sont des facteurs de \mathbf{t}_3 . Vérifions que les facteurs ${}_1$ et ${}_2$ sont vides. Comme u est triprolongeable à droite, les facteurs ${}_20$, ${}_21$ et ${}_22$ sont dans \mathbf{t}_3 . Ainsi, un des ${}_2i$, contient le carré d'une lettre. Ce qui est impossible car l'image d'aucune lettre ne contient de carré. De la même façon, on montre que ${}_1$ est vide. Ainsi $u = \mu_3(u')$. D'après la Proposition 3.3.1, u' est dans $BST(\mathbf{t}_3)$. ■

Les résultats précédents nous permettent de déterminer explicitement l'ensemble des facteurs bispéciaux triprolongeables de \mathbf{t}_3 .

Théorème 3.3.1 *L'ensemble $BST(\mathbf{t}_3)$ est donné par :*

$$BST(\mathbf{t}_3) = \bigcup_{n \geq 0} \{\mu_3^n(i), \mu_3^n(i(E_3(i))) : i \in \mathcal{A}_3\} \setminus \{\emptyset\}.$$

Preuve : Soit u un élément de $BST(\mathbf{t}_3)$ de longueur supérieure ou égale à 3. D'après la Proposition 3.3.2, il existe u' dans $BST(\mathbf{t}_3)$ tel que $u = \mu_3(u')$. Ainsi, pour obtenir l'ensemble $BST(\mathbf{t}_3)$, il suffit de trouver ses éléments de longueur au plus 2 puisque les autres s'obtiennent par applications successives de μ_3 . Ces facteurs sont i et $iE_3(i)$, $i \in \mathcal{A}_3$. ■

Corollaire 3.3.1 *Soit u un facteur de \mathbf{t}_3 triprolongeable à droite. Si $|u| = 3^k$ ou $|u| = 2 \cdot 3^k$, $k \geq 0$, alors u est triprolongeable à gauche.*

Preuve : Soit u un facteur de \mathbf{t}_3 triprolongeable à droite et vérifiant $|u| = 3^k$, $k \geq 1$. Alors u se décompose sous la forme ${}_1\mu_3(v) {}_2$ où $v, {}_1, {}_2 \in F(\mathbf{t}_3)$. Le facteur u étant triprolongeable à droite, ${}_2$ est le mot vide. Par suite $u = {}_1\mu_3(v)$. Nous savons que $|{}_1| \leq 2$ et $|\mu_3(v)|$ est multiple de 3. Le facteur u étant de longueur 3^k alors ${}_1$ est le mot vide. Ainsi $u = \mu_3(v)$ où v est un facteur triprolongeable à droite de longueur 3^{k-1} . Par le même procédé, le facteur v s'écrit $v = \mu_3(v')$, où v' est un facteur triprolongeable à droite de longueur 3^{k-2} . De façon successive, on aboutit à $u = \mu_3^k(i)$, $i \in \mathcal{A}_3$. A l'aide du Théorème 3.3.1, on conclut que u est triprolongeable à gauche. On procède de même pour ceux de longueur $2 \leq 3^k$. ■

Dans la proposition suivante, nous montrons que tout élément u de $BST(\mathbf{t}_3)$ admet un unique prolongement à gauche (resp. à droite) qui est triprolongeable à droite (resp. à gauche).

Proposition 3.3.3 *Soit $u \in BST(\mathbf{t}_3)$ non vide. Alors, il existe une unique lettre i dans \mathcal{A}_3 telle que iu (resp. ui) soit triprolongeable à droite (resp. à gauche).*

Preuve : Construisons les ensembles $F(\mathbf{t}_3) \setminus (\mathcal{A}_3 v \mathcal{A}_3)$ où $v \in BST(\mathbf{t}_3)$. Nous donnons ceux pour lesquels $|v| \leq 3$, et par récurrence nous montrons que pour ceux de longueur supérieure les prolongements respectent l'unicité. Nous avons, pour $i \in \mathcal{A}_3$:

$$F(\mathbf{t}_3) \setminus (\mathcal{A}_3 i \mathcal{A}_3) = \{ii(i+1), (i+1)i(i+1), (i+2)ii, (i+2)i(i+1), (i+2)i(i+2)\};$$

$$F(\mathbf{t}_3) \setminus (\mathcal{A}_3 i(i+1) \mathcal{A}_3) = \{0i(i+1)2, 1i(i+1)2, 2i(i+1)0, 2i(i+1)1, 2i(i+1)2\};$$

$$F(\mathbf{t}_3) \setminus (\mathcal{A}_3 \mu_3(i) \mathcal{A}_3) = \{0\mu_3(i)1, 1\mu_3(i)0, 1\mu_3(i)1, 1\mu_3(i)2, 2\mu_3(i)0, 2\mu_3(i)1\}.$$

où $i+1$ est pris modulo 3. De manière extensive on a :

$$F(\mathbf{t}_3) \setminus (\mathcal{A}_3 0 \mathcal{A}_3) = \{001, 101, 200, 201, 202\};$$

$$F(\mathbf{t}_3) \setminus (\mathcal{A}_3 1 \mathcal{A}_3) = \{010, 011, 012, 112, 212\};$$

$$F(\mathbf{t}_3) \setminus (\mathcal{A}_3 2 \mathcal{A}_3) = \{020, 120, 121, 122, 220\};$$

à droite.

En résumé, un facteur de \mathbf{t}_3 triprolongeable d'un coté est de l'autre coté soit non spécial soit triprolongeable. ■

Proposition 3.3.5 *Pour tout entier naturel non nul n , \mathbf{t}_3 admet exactement trois facteurs triprolongeables à droite (resp. à gauche) de longueur n .*

Preuve : Nous savons que 0, 1 et 2 sont les facteurs triprolongeables à droite de longueur 1. Montrons que tout facteur triprolongeable à droite de longueur n est su xé d'un unique facteur triprolongeable à droite de longueur $n + 1$.

Soit w un facteur triprolongeable à droite de longueur n . Si w admet un unique prolongement a à gauche, alors aw est un facteur triprolongeable à droite puisque \mathbf{t}_3 est récurrent. S'il admet au moins deux prolongements à gauche, alors w est dans $BST(\mathbf{t})$ par la Proposition 3.3.4 et un seul de ses prolongements à gauche est triprolongeable à droite par la Proposition 3.3.3. Ainsi, le nombre de facteurs triprolongeables à droite de longueur $n + 1$ est égal au nombre de facteurs triprolongeables à droite de longueur n de \mathbf{t}_3 . De même, on traite le cas des facteurs triprolongeables à gauche. ■

La remarque suivante est une conséquence de la proposition 3.3.5.

Remarque 3.3.1 *Soit u un facteur de \mathbf{t}_3 triprolongeable à droite (resp. à gauche). Si u est de longueur $3k$, $k \geq 1$, alors il existe un facteur v de \mathbf{t}_3 , triprolongeable à droite (resp. à gauche) tel que $u = \mu_3(v)$.*

Proposition 3.3.6 *Pour tout entier naturel non nul n , les facteurs triprolongeables à droite (resp. à gauche) de longueur n commencent (resp. se terminent) par des lettres distinctes.*

Preuve : Par la Proposition 3.3.5, on sait qu'il y'a trois facteurs de longueur n triprolongeables à droite. Or si u est triprolongeable à droite, $E_3(u)$ et $E_3^2(u)$ le sont aussi. Ces trois facteurs u , $E_3(u)$ et $E_3^2(u)$ commencent par des lettres distinctes. ■

Dans ce qui suit, nous montrons que le mot \mathbf{t}_3 est 2-équilibré à l'aide des différentes formes synchronisées des facteurs.

Théorème 3.3.2 *Le mot \mathbf{t}_3 est 2-équilibré.*

Preuve : Soient u_1 et u_2 deux facteurs de longueur $n \geq 7$ de \mathbf{t}_3 . Alors, u_1 et u_2 se synchronisent de façon unique sous les formes $u_1 = {}_1\mu_3(v_1) {}_2$ et $u_2 = {}'_1\mu_3(v_2) {}'_2$, $v_1, v_2, {}_1, {}_2, {}'_1, {}'_2 \in F(\mathbf{t}_3)$. Posons $|_b = |{}_1|_b + |{}_2|_b$, $|'_b = |{}'_1|_b + |{}'_2|_b$, pour tout $b \in \mathcal{A}_3$. Considérons les cas suivants.

Cas 1 : n est multiple de 3. Alors, u_1 (resp. u_2) s'écrit de façon unique sous la forme $\mu_3(v)$, $ij\mu_3(v)k$ ou $i\mu_3(v)jk$, avec $i, j, k \in \mathcal{A}_3$ et $v \in F(\mathbf{t}_3)$. Considérons les différentes formes prises par u_1 et u_2 .

Supposons que $u_1 = \mu_3(v_1)$ et $u_2 = \mu_3(v_2)$. Alors, nous avons $\psi(u_1) = \psi(u_2)$ et on a :

$$||u_1|_i - |u_2|_i| = 0,$$

pour toute lettre i .

Supposons que $u_1 = \mu_3(v_1)$ et $u_2 = i\mu_3(v_2)jk$. Ecrivons u_1 sous la forme $\mu_3(v'_1)\mu_3(a)$, $a \in \mathcal{A}_3$. Par suite, $|v'_1| = |v_2|$. Comme les lettres ont les mêmes nombres d'occurrences dans l'image de toute lettre, alors $|\mu_3(a)|_b = 1$, pour tout $b \in \mathcal{A}_3$. De plus, $|_b \leq 2$, pour tout $b \in \mathcal{A}_3$ puisque jk est le préfixe de l'image d'une lettre. Ainsi,

$$||u_1|_b - |u_2|_b| = ||\mu_3(a)|_b - |_b| \leq 1.$$

On procède de même lorsque $u_1 = \mu_3(v_1)$ et $u_2 = ij\mu_3(v_2)k$.

Supposons que $u_1 = ij\mu_3(v_1)k$ et $u_2 = l\mu_3(v_2)mn$, $i, j, k, l, m, n \in \mathcal{A}_3$. Comme précédemment, on vérifie que $|_b, |'_b \leq 2$, pour tout $b \in \mathcal{A}_3$. Ainsi,

$$||u_1|_b - |u_2|_b| = ||_b - |'_b| \leq 2.$$

En prenant $u_1 = 101212$ et $u_2 = 010120$, on observe que la borne 2 est atteinte.

Cas 2 : $n - 1$ est multiple de 3. Alors, u_1 (resp. u_2) est de la forme $i\mu_3(v)$, $\mu_3(v)k$ ou $ij\mu_3(v)kl$, $i, j, k, l \in \mathcal{A}_3$, $v \in F(\mathbf{t}_3)$.

Supposons que $u_1 = i\mu_3(v_1)$ et $u_2 = \mu_3(v_2)j$. Alors, nous avons $|v_1| = |v_2|$. Donc $|u_1|_b - |u_2|_b| \leq 1$, pour tout $b \in \mathcal{A}_3$.

Supposons que $u_1 = ij\mu_3(v_1)kl$ et $u_2 = i'j'\mu_3(v_2)k'l'$, où ij et $i'j'$ (resp. kl et $k'l'$) sont des suffixes (resp. des préfixes) d'images de lettres. Alors, notons que $(i, k) = (j, l)$ et $(i', k') = (j', l')$. Par suite, $|v_1| = |v_2|$. Par analogie au cas précédent on vérifie que $|u_1|_b - |u_2|_b| \leq 2$. Par conséquent,

$$||u_1|_b - |u_2|_b| = ||i\mu_3(v_1)|_b - |\mu_3(v_2)j|_b| \leq 2,$$

pour tout $i \in \mathcal{A}_3$. En prenant $u_1 = 01\mu_3(12)01$ et $u_2 = 20\mu_3(01)20$ on observe que la borne 2 est atteinte.

Supposons que $u_1 = i'\mu_3(v_1)$ et $u_2 = ij\mu_3(v_2)kl$. Alors, on écrit u_1 sous la forme $i'\mu_3(v'_1)\mu_3(a)$, $a \in \mathcal{A}_3$ et $v' \in F(\mathbf{t}_3)$. Par suite, $|v'_1| = |v_2|$ et $|i\mu_3(a)|_b - |u_2|_b| \leq 2$. Donc

$$||u_1|_b - |u_2|_b| = |i\mu_3(a)|_b - |u_2|_b| \leq 2.$$

Cas 3 : $n - 2$ est multiple de 3. Supposons que u_1 (resp. u_2) s'écrit sous la forme $ij\mu_3(v_1)$, $i\mu_3(v_1)k$ ou $\mu_3(v_1)kl$ (resp. $ij\mu_3(v_2)$, $i\mu_3(v_2)k$ ou $\mu_3(v_2)kl$). Alors, nous avons $|v_1| = |v_2|$. De façon analogue aux cas précédents, on vérifie que $|u_1|_b - |u_2|_b| \leq 2$, pour $b \in \mathcal{A}_3$. ■

Le mot \mathbf{t}_3 étant 2-équilibré, sa complexité abélienne est bornée. Nous montrons qu'elle prend 7 comme valeur maximale et 3 comme valeur minimale pour toute longueur non nulle, avant de la déterminer explicitement.

3.4 Complexité abélienne

Dans cette section nous déterminons explicitement la fonction de complexité abélienne ρ^{ab} de \mathbf{t}_3 . Nous montrons que la suite $(\rho_{\mathbf{t}_3}^{ab}(n))_{n \geq 2}$ de \mathbf{t}_3 est 3-périodique. Nous commençons par donner la valeur maximale de la fonction de complexité abélienne en utilisant des graphes. Cette approche a été utilisée par Richomme, Saari et Zamboni

dans [38] où ils déterminent les valeurs extrêmes de la complexité abélienne du mot de Tribonacci en utilisant le graphe des vecteurs de Parikh. Nous utiliserons une technique analogue, mais avec les vecteurs de Parikh relatifs.

Notation : Soient $\psi(u) = (|u|_1, |u|_2, |u|_3)$ et $\psi(v) = (|v|_1, |v|_2, |v|_3)$. Alors, on pose

$$||u - v|| = \max(|u|_1 - |v|_1, |u|_2 - |v|_2, |u|_3 - |v|_3).$$

Naturellement, un mot w est k -équilibré si pour tous facteurs u et v de même longueur de w , on a $||u - v|| \leq k$.

Théorème 3.4.1 *La fonction de complexité abélienne $\rho_{\mathbf{t}_3}^{ab}$ de \mathbf{t}_3 vérifie $\rho_{\mathbf{t}_3}^{ab}(n) \leq 7$, pour tout n .*

Preuve : Nous savons que tout facteur u de \mathbf{t}_3 se décompose sous la forme $u = \mu_1(v)\mu_2(w)$, avec $|\mu_1|, |\mu_2| \leq 2$. Par suite $\psi(u) = (|\mu_1|_1 + |\mu_2|_1, |\mu_1|_2 + |\mu_2|_2, |\mu_1|_3 + |\mu_2|_3)$, avec $|\mu_1|_i = |\mu_1|_i|_0 + |\mu_1|_i|_1$, $|\mu_2|_i = |\mu_2|_i|_1 + |\mu_2|_i|_2$. Écrivons $\mathbf{t}_{3[n]} = \mu'_1(v')\mu'_2(w')$. Alors, $\psi(\mathbf{t}_{3[n]}) = (|\mu'_1|_1 + |\mu'_2|_1, |\mu'_1|_2 + |\mu'_2|_2, |\mu'_1|_3 + |\mu'_2|_3)$, avec $|\mu'_1|_i = |\mu'_1|_i|_0 + |\mu'_1|_i|_1$, $|\mu'_2|_i = |\mu'_2|_i|_1 + |\mu'_2|_i|_2$. On peut toujours avoir $|v| = |v'|$ quitte à décomposer u ou $\mathbf{t}_{3[n]}$ comme dans la preuve du Théorème 3.3.2. Par suite $\psi^{rel}(u) = \psi(u) - \psi(\mathbf{t}_{3[n]}) = (|\mu_1|_1 - |\mu'_1|_1, |\mu_1|_2 - |\mu'_1|_2, |\mu_1|_3 - |\mu'_1|_3)$ avec $|\mu_1|_1 + |\mu_1|_2 + |\mu_1|_3 = 0$, $|\mu_1|_i = |\mu_1|_i|_i$ et $|\mu_1|_1, |\mu_1|_2, |\mu_1|_3 \leq 2$, puisque \mathbf{t}_3 est 2-équilibré. Les valeurs possibles de $\psi^{rel}(u)$ sont données dans l'ensemble formé par les 19 vecteurs suivants :

$$\{(-2, 0, 2), (-2, 1, 1), (-2, 2, 0), (-1, -1, 2), (-1, 0, 1), (-1, 1, 0), (-1, 2, -1), (0, -2, 2), (0, -1, 1), (0, 0, 0), (0, 1, -1), (0, 2, -2), (1, -2, 1), (1, -1, 0), (1, 0, -1), (1, 1, -2), (2, -2, 0), (2, -1, -1), (2, 0, -2)\}.$$

La **Figure 3.1** représente le graphe formé par ces vecteurs. Les sommets du graphe sont constitués des 19 vecteurs de Parikh relatifs. Il existe une arrête entre deux vecteurs v_1 et v_2 si et seulement si $||v_1 - v_2|| = 1$. Nous noterons comme longueur d'un côté du graphe le nombre de vecteurs qui forment ce côté. Ainsi, le graphe de la **Figure 3.1** est circonscrit par un hexagone régulier dont les côtés sont de longueur 3. Nous dirons qu'un chemin est admissible sur le graphe si pour tous sommets u_1 et u_2 pris sur ce chemin, on a $||u_1 - u_2|| \leq 2$.

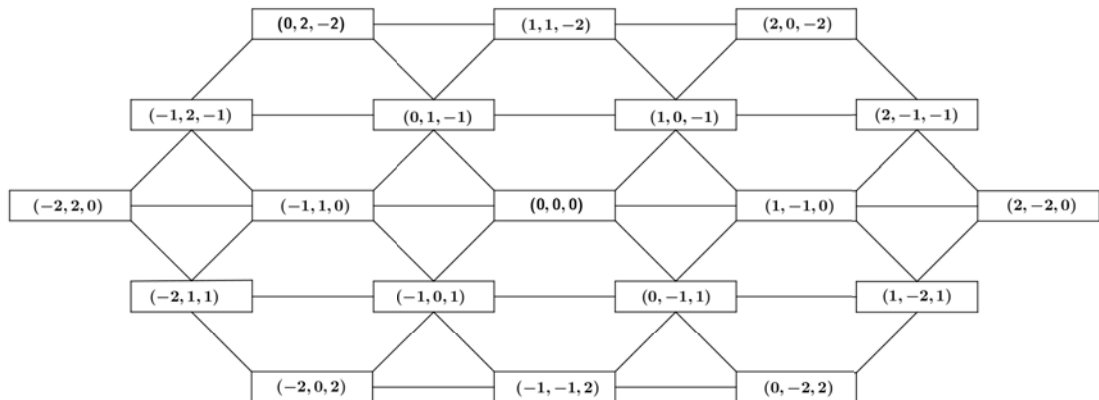


FIGURE 3.1 – graphe des vecteurs de parikh

Le mot \mathbf{t}_3 étant 2-équilibré, $\Psi^{rel}(\mathbf{t}_3)$ doit être un sous ensemble de l'ensemble formé par un hexagone régulier de côté 2 (la **Figure 3.2** par exemple) ou l'ensemble formé par un triangle équilatéral de côté 3 (la **Figure 3.3** par exemple), puisque deux vecteurs n'appartenant ni à un même triangle de côté 3 ni à un même hexagone de côté 2 n'appartiennent pas à un même chemin admissible. Ces ensembles (triangle équilatéral de côté 3 et hexagone régulier de côté 2) ont pour cardinaux respectifs 6 et 7. Par conséquent $\rho_{\mathbf{t}_3}^{ab}(n) \leq 7$.

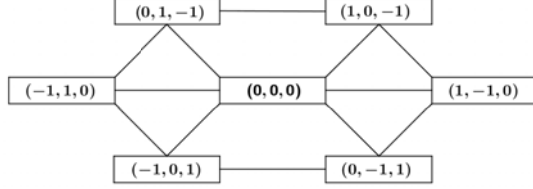


FIGURE 3.2 – Hexagone régulier de côté 2

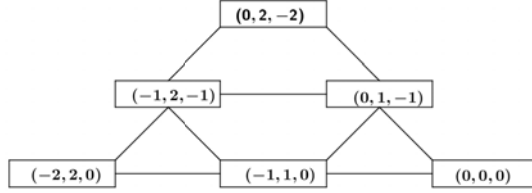


FIGURE 3.3 – Triangle équilatéral de côté 3 ■

Proposition 3.4.1 *Pour tout $n \geq 1$, $\rho^{ab}(3n) = 7$.*

Preuve : Soit u un facteur de \mathbf{t}_3 de longueur $3n$, $n \geq 1$. Alors, u se synchronise sous la forme $\mu_3(v)$, $i\mu_3(v)jk$ ou $ij\mu_3(v)k$ avec $i, j, k \in \mathcal{A}_3$, $ij, jk \in \{01, 12, 20\}$ et $v \in F(\mathbf{t}_3)$. Comme u est choisi quelconque on vérifie que ces trois formes sont prises par u . Comme le préfixe $\mathbf{t}_{3[3n]}$ commence par l'image d'une lettre, il est sous la forme $\mu_3(v)$. Pour la suite, notons $\mathbf{t}_{3[3n]} = \mu_3(v_1)$. Trois cas se présentent.

Cas 1 : Le facteur u est sous la forme $\mu_3(v_2)$. Alors, $|v_1| = |v_2|$. Par suite, u et $\mathbf{t}_{3[3n]}$ ont les mêmes nombres d'occurrences de chaque lettre. Donc $\psi^{rel}(u) = (0, 0, 0)$.

Cas 2 : Le facteur u est sous la forme $i\mu_3(v_2)jk$. Alors, on a :

$$\psi(u) = (|v_2| + |ijk|_0, |v_2| + |ijk|_1, |v_2| + |ijk|_2).$$

Montrons que l'ensemble des valeurs prises par ijk est

$$\{001, 012, 020, 101, 112, 120, 201, 212, 220\}.$$

D'après la Proposition 3.3.5, pour tout entier $n \geq 1$, \mathbf{t}_3 contient 3 facteurs triprolongeables à droite de longueur $3n$. Désignons par R_1 , R_2 et R_3 les facteurs triprolongeables à droite de longueur $3n - 3$. Comme ces facteurs commencent par des lettres différentes, on peut supposer, quitte à changer les indices, que $0R_1$, $1R_2$ et $2R_3$ sont les facteurs triprolongeables à droite de \mathbf{t}_3 de longueur $3n - 2$. Par suite, les mots $0R_101$, $0R_112$, $0R_120$, $1R_201$, $1R_212$, $1R_220$, $2R_301$, $2R_312$ et $2R_320$ sont des facteurs de longueur $3n$ de \mathbf{t}_3 . Ainsi, ijk parcourt l'ensemble annoncé. Donc, $\psi(ijk)$ prend toutes les valeurs de l'ensemble suivant

$$\{(2, 1, 0), (1, 1, 1), (2, 0, 1), (1, 2, 0), (0, 2, 1), (0, 1, 2), (1, 0, 2)\}.$$

Écrivons le préfixe $\mathbf{t}_{3[n]}$ sous la forme $\mu_3(v'_1)\mu_3(l)$, $l \in \mathcal{A}_3$. Alors, $|v'_1| = |v_2|$ et $\psi(\mu_3(l)) = (1, 1, 1)$. Ainsi, pour tous les facteurs u de longueur $3n$, $\psi^{rel}(u) = \psi(ijk) - \psi(\mu_3(l))$ prend toutes les valeurs de l'ensemble

$$\{(1, 0, -1), (0, 0, 0), (1, -1, 0), (0, 1, -1), (-1, 1, 0), (-1, 0, 1), (0, -1, 1)\}.$$

Cas 3 : Le facteur u est sous la forme $ij\mu_3(v_2)k$. Alors

$$\psi(u) = (|v_2| + |ijk|_0, |v_2| + |ijk|_1, |v_2| + |ijk|_2).$$

En procédant de façon analogue au cas 2 et en utilisant les facteurs triprolongeables à gauche, on vérifie que l'ensemble des valeurs prises par ijk est

$$\{010, 011, 012, 120, 121, 122, 200, 201, 202\}.$$

Par conséquent, pour tous les u remplissant ces conditions, $\psi^{rel}(u)$ prend toutes les valeurs de l'ensemble

$$\{(1, 0, -1), (0, 0, 0), (1, -1, 0), (0, 1, -1), (-1, 1, 0), (-1, 0, 1), (0, -1, 1)\}.$$

En définitive, on a :

$$\Psi_{3n}^{rel}(\mathbf{t}_3) = \{(1, 0, -1), (0, 0, 0), (1, -1, 0), (0, 1, -1), (-1, 1, 0), (-1, 0, 1), (0, -1, 1)\}. \blacksquare$$

Proposition 3.4.2 Pour tout $n \geq 1$, $\rho^{ab}(3n+1) = 6$.

Preuve : Soit u un facteur de \mathbf{t}_3 de longueur $3n+1$, $k \geq 1$. Alors, u se synchronise sous la forme $i\mu_3(v)$, $\mu_3(v)j$ ou $ij\mu_3(v)kl$, $i, j, k, l \in \mathcal{A}_3$, $ij, kl \in \{01, 12, 20\}$ et $v \in F(\mathbf{t}_3)$. Le préfixe $\mathbf{t}_{3[3n+1]}$ est sous la forme $\mu_3(v_1)b$, $b \in \mathcal{A}_3$. On a :

Cas 1 : $b = 0$. Alors, $\mathbf{t}_{3[3n+1]} = \mu_3(v_1)0$. Déterminons $\Psi_{3n+1}^{rel}(\mathbf{t}_3)$.

Soit v_2 un facteur de \mathbf{t}_3 tel que $u = i\mu_3(v_2)$. Alors, on a : $|v_1| = |v_2|$ et donc $\psi(\mu_3(v_1)) = \psi(\mu_3(v_2))$. En utilisant les facteurs triprolongeables à gauche de longueur $3n$, on vérifie que les valeurs prises par i sont 0, 1 et 2. Par conséquent, $\psi^{rel}(u)$ prend toutes les valeurs de $\{(0, 0, 0), (1, 1, 0), (2, 1, 0)\}$. De même, on vérifie que si $u = \mu_3(v_2)j$, $\psi^{rel}(u)$ parcourt tous les éléments de l'ensemble $\{(0, 0, 0), (1, 1, 0), (2, 1, 0)\}$.

Soit u un facteur de \mathbf{t}_3 de la forme $u = ij\mu_3(v_2)kl$. On écrit $\mathbf{t}_{3[3n+1]}$ sous la forme $\mu_3(v'_1)\mu_3(m)0$, $m \in \mathcal{A}_3$. Par suite, $|v'_1| = |v_2|$. Donc, $\psi^{rel}(u) = \psi(ijkl) \quad \psi(\mu_3(m)0)$ où $\psi(\mu_3(m)0) = (2, 1, 1)$. Nous savons que chaque facteur de la forme $ij\mu_3(v_2)kl$ est le prolongement à gauche d'un facteur de la forme $j\mu_3(v_2)kl$ dont l'ensemble des valeurs prises par jkl est

$$\{001, 012, 020, 101, 112, 120, 201, 212, 220\}.$$

Ainsi, l'ensemble des valeurs prises par $ijkl$ est

$$\{2001, 2012, 2020, 0101, 0112, 0120, 1201, 1212, 1220\}.$$

Donc, $\psi^{rel}(u)$ parcourt tous les éléments de l'ensemble

$$\{(0, 1, 1), (1, 1, 0), (0, 0, 0), (2, 1, 1), (1, 0, 1), (0, 1, 1)\}.$$

Finalement on obtient

$$\Psi_{3n+1}^{rel}(\mathbf{t}_3) = \{(0, 1, 1), (1, 1, 0), (0, 0, 0), (2, 1, 1), (1, 0, 1), (0, 1, 1)\}.$$

Cas 2 : $b = 1$. Alors, $\mathbf{t}_{3[3n+1]} = \mu_3(v_1)1$. Considérons les différentes formes de u .

Soit u un facteur de \mathbf{t}_3 de la forme $u = i\mu_3(v_2)$. Comme dans le cas 1, on vérifie que i prend les valeurs 0, 1 et 2. Par conséquent, $\psi^{rel}(u)$ parcourt tous les éléments de

$$\{(0, 0, 0), (1, -1, 0), (0, -1, 1)\}.$$

Soit u un facteur de \mathbf{t}_3 de la forme $u = ij\mu_3(v_2)kl$. Alors, on écrit $\mathbf{t}_{3[3n+1]}$ sous la forme $\mathbf{t}_{3[3n+1]} = \mu_3(v'_1)\mu_3(m)1$, $m \in \mathcal{A}_3$. En procédant comme au cas 1, on vérifie que l'ensemble des valeurs prises par $ijkl$ est

$$\{2001, 2012, 2020, 0101, 0112, 0120, 1201, 1212, 1220\}.$$

Ainsi, $\psi^{rel}(u)$ parcourt tous les éléments de l'ensemble

$$\{(1, 0, -1), (1, -1, 0), (0, 0, 0), (-1, 0, 1), (0, -1, 1), (1, -2, 1)\}.$$

En définitive on a :

$$\Psi_{3n+1}^{rel}(\mathbf{t}_3) = \{(1, 0, -1), (1, -1, 0), (0, 0, 0), (-1, 0, 1), (0, -1, 1), (1, -2, 1)\}.$$

Cas 3 : $i = 2$. Alors, $\mathbf{t}_{3[3n+1]} = \mu_3(v_1)2$. En procédant comme dans le cas précédent, on obtient :

$$\Psi_{3n+1}^{rel}(\mathbf{t}_3) = \{(1, 1, -2), (0, 1, -1), (1, 0, -1), (-1, 1, 0); (0, 0, 0), (1, -1, 0)\}. \blacksquare$$

Proposition 3.4.3 *Pour tout $n \geq 1$, $\rho^{ab}(3n+2) = 6$.*

Preuve : Soit u un facteur de longueur $3n+2$ de \mathbf{t}_3 ($n \geq 1$). Alors, u s'écrit sous la forme $i\mu_3(v_2)j$, $ij\mu_3(v_2)$ ou $\mu_3(v_2)kl$, $i, j, k, l \in \mathcal{A}_3$, $v_2 \in F(\mathbf{t}_3)$. Par ailleurs, le préfixe $\mathbf{t}_{3[3n+2]}$ est sous la forme $\mu_3(v_1)bc$, $bc \in \{01, 12, 20\}$.

Cas 1 : $bc = 01$. Alors, $\mathbf{t}_{3[3n+1]} = \mu_3(v_1)01$. Déterminons l'ensemble $\Psi_{3n+2}^{rel}(\mathbf{t}_3)$.

Soit u un facteur de \mathbf{t}_3 de la forme $i\mu_3(v_2)j$. Alors, v_1 et v_2 sont de même longueur. Donc, $\psi^{rel}(u) = \psi(ij) - \psi(01)$. A l'aide des facteurs triprolongeables à droite de longueur $n-1$, on vérifie que l'ensemble des valeurs prises par ij est $\{00, 01, 02, 10, 11, 12, 20, 21, 22\}$. Donc, $\psi^{rel}(u)$ prend toutes les valeurs de l'ensemble

$$\{(1, -1, 0), (0, 0, 0), (0, -1, 1), (-1, 1, 0), (-1, 0, 1), (-1, -1, 2)\}.$$

Soit u un facteur de \mathbf{t}_3 de la forme $ij\mu_3(v_2)$. Alors, v_1 et v_2 sont de même longueur. Donc, $\psi^{rel}(u) = \psi(ij) - \psi(01)$. Le facteur ij est le su de l'image d'une lettre. Il prend les valeurs 01, 12 et 20. Ainsi, $\psi^{rel}(u)$ prend toutes les valeurs de l'ensemble

$$\{(0, 0, 0), (0, -1, 1), (-1, 0, 1)\}.$$

De même, on vérifie que si u est de la forme $\mu_3(v_2)kl$, $\psi^{rel}(u)$ prend toutes les valeurs de l'ensemble $\{(0, 0, 0), (0, -1, 1), (-1, 0, 1)\}$.

On obtient finalement :

$$\Psi_{3n+2}^{rel}(\mathbf{t}_3) = \{(1, -1, 0), (0, 0, 0), (0, -1, 1), (-1, 1, 0), (-1, 0, 1), (-1, -1, 2)\}$$

Cas 2 : $bc = 12$. Alors $\mathbf{t}_{3[3n+2]} = \mu_3(v_1)12$. Déterminons l'ensemble $\Psi_{3n+2}^{rel}(\mathbf{t}_3)$.

Soit u un facteur de \mathbf{t}_3 de la forme $u = i\mu_3(v_2)j$. Ainsi, $\psi^{rel}(u) = \psi(ij) - \psi(12)$. Comme dans le cas précédent l'ensemble des valeurs prises par ij est

$$\{00, 01, 02, 10, 11, 12, 20, 21, 22\}.$$

Par conséquent, $\psi^{rel}(u)$ prend toutes les valeurs de l'ensemble

$$\{(2, -1, -1), (1, 0, -1), (0, 1, -1), (0, 0, 0), (0, -1, 1), (1, -1, 0)\}.$$

Soit u un facteur de \mathbf{t}_3 de la forme $u = ij\mu_3(v_2)$. Alors, les valeurs prises par ij sont 01, 12 et 20. De plus, v_1 et v_2 sont de même longueur. On montre ainsi, comme dans le cas 1 que $\psi^{rel}(u)$ prend toutes les valeurs de l'ensemble $\{(1, 0, -1), (1, -1, 0), (0, 0, 0)\}$.

En définitive on a

$$\Psi_{3n+2}^{rel}(\mathbf{t}_3) = \{(2, -1, -1), (1, 0, -1), (0, 1, -1), (0, 0, 0), (0, -1, 1), (1, -1, 0)\}.$$

Cas 3 : $bc = 20$. Alors $\mathbf{t}_{3[3n+2]} = \mu_3(v_1)20$. Comme dans les cas précédents on vérifie que :

$$\Psi_{3n+2}^{rel}(\mathbf{t}_3) = \{(1, 0, -1), (0, 1, -1), (0, 0, 0), (-1, 2, -1), (-1, 1, 0), (-1, 0, 1)\} \blacksquare$$

En somme, on a le théorème suivant.

Théorème 3.4.2 *La fonction de complexité abélienne de \mathbf{t}_3 est donnée par :*

$$\rho_{\mathbf{t}_3}^{ab}(n) = \begin{cases} 1 & \text{si } n = 0 \\ 3 & \text{si } n = 1 \\ 7 & \text{si } n \geq 2 \text{ et } n \text{ multiple de } 3 \\ 6 & \text{si } n \geq 2 \text{ et } n \text{ non multiple de } 3 \end{cases}$$

Preuve : Le résultat découle des Propositions 3.4.1, 3.4.2 et 3.4.3. ■

CHAPITRE 4

CARRÉS DE LETTRES, FACTEURS SÉPARATEURS ET MOTS DE RETOUR DANS LE MOT DE THUE-MORSE TERNAIRE

4.1 Introduction

Dans le chapitre 2, nous avons vu qu'un facteur donné du mot de Thue-Morse \mathbf{t}_2 admet soit 3 soit 4 mots de retour. Nous entreprenons ici de déterminer pour un facteur donné de \mathbf{t}_3 , le nombre de ses mots de retour. Nous commençons par l'étude des facteurs qui séparent les carrés de lettres dans \mathbf{t}_3 : leurs structures et leurs longueurs. Ensuite, nous déterminons les facteurs bispéciaux biprolongeables avant d'étudier les mots de retour dans \mathbf{t}_3 . Nous terminons le chapitre par les facteurs privilégiés de \mathbf{t}_3 .

Les résultats de ce chapitre ont fait l'objet d'un papier publié dans **International Journal of Applied Mathematics** [24] .

4.2 Fréquences

Dans [14], les auteurs ont établi que tout mot infini u sur un alphabet binaire $\mathcal{A}_2 = \{0, 1\}$ et ayant la même fonction de complexité abélienne que \mathbf{t}_2 possède des fréquences uniformes de lettres : $f_0(u) = f_1(u) = \frac{1}{2}$.

Proposition 4.2.1 *Le mot \mathbf{t}_3 admet des fréquences uniformes de lettres et on a :*

$$f_0(\mathbf{t}_3) = f_1(\mathbf{t}_3) = f_2(\mathbf{t}_3) = \frac{1}{3}.$$

Preuve : Soit w un facteur de \mathbf{t}_3 . Nous calculons la fréquence de la lettre 0 et nous obtenons celles des deux autres par symétrie de rôles. Les cas suivants se présentent :

Cas 1 : Le facteur $w \in F_{3n}(\mathbf{t}_3)$. Alors, w se synchronise sous la forme $\mu_3(u)$ ou $ij\mu_3(u)k$ ou encore $i\mu_3(u)jk$. Poursuivons avec les différentes formes synchronisées :

$w = \mu_3(u)$, alors $|u| = n$ et donc $|w|_0 = n$

$w = i\mu_3(u)jk$ ou $w = ij\mu_3(u)k$, alors $|u| = n - 1$. Comme jk (rep. ij) est le préfixe (resp. le suffixe) de l'image d'une lettre, il contient au plus une occurrence de 0. Donc ijk

contient au plus deux 0. Par suite $n - 1 \leq |w|_0 \leq n + 1$.

Cas 2 : Le facteur $w \in F_{3n+1}(\mathbf{t}_3)$. Alors, w est sous la forme $i\mu_3(u)$, $\mu_3(u)i$ ou $ij\mu_3(u)kl$. Par analogie au cas précédent, on vérifie que $n \leq |w|_0 \leq n + 1$ si w est sous la forme $i\mu_3(u)$ ou $\mu_3(u)i$, et $n - 1 \leq |w|_0 \leq n + 1$ si w est sous la forme $ij\mu_3(u)kl$.

Cas 3 : Le facteur $w \in F_{3n+2}(\mathbf{t}_3)$. Alors, w est sous la forme $i\mu_3(u)j$, $ij\mu_3(u)$ ou $\mu_3(u)kl$. Comme précédemment, on vérifie que $n \leq |w|_0 \leq n + 2$ si w est sous la forme $i\mu_3(u)j$, et $n \leq |w|_0 \leq n + 1$ si w est sous la forme $ij\mu_3(u)$ ou $\mu_3(u)kl$.

En définitive on a $n - 1 \leq |w|_0 \leq n + 2$. Par suite,

$$\frac{n - 1}{3n + 2} \leq \frac{|w|_0}{|w|} \leq \frac{n + 2}{3n + 2}.$$

On obtient le résultat en passant à la limite. ■

Dans [5], Balkova a donné un résultat général. Elle obtient les fréquences des facteurs de longueur quelconque du mot de Thue-Morse généralisé. Ce résultat s'appuie en grande partie sur les graphes de Rauzy.

Dé nition 4.2.1 Soit u un mot sur \mathcal{A}^∞ . On dé nît le graphe de Rauzy d'ordre n de u , noté Γ_n , le graphe orienté dont les sommets sont les facteurs de longueur n de u , deux sommets u_1 et u_2 sont liés par un arc (arête) si u_1 et u_2 apparaissent successivement dans u , c'est à dire $u_1 = bv$ et $u_2 = va$, où bva est un facteur de u , a et b des lettres, $v \in F_{n-1}(u)$. Cette arête est étiquetée par $bva \in F_{n+1}(u)$.

Remarquons que l'ensemble des arêtes de Γ_n s'identifie à $F_{n+1}(u)$, et celui des sommets à $F_n(u)$.

Le résultat principal établi par Balkova est une conséquence des résultats suivants, démontrés dans [5].

Lemme 4.2.1 Soit $\mathbf{t}_{b,q}$ le mot de Thue-Morse généralisé et posons $f = f_{01}$. Alors, $f = \frac{b^{m-1}}{q} \frac{b-1}{b^m-1}$, où m est le plus petit entier strictement positif vérifiant $m(b-1) \equiv 0 \pmod{q}$. Pour tout $1 \leq n \leq b$, les fréquences du sommet $w = 012\dots(n-1)$ ainsi que celles des arêtes se terminant en w vérifient

$$\begin{aligned} f_0 &= \frac{1}{q} \\ f_{012\dots(n-1)} &= (n-1)f + \frac{n-2}{q}, n \geq 2 \\ f_{(-1)012\dots(n-1)} &= nf + \frac{n-1}{q} \\ f_{(-1+k(b-1))01\dots(n-1)} &= \frac{f}{b^k}, k \in \{1, \dots, m-1\}, \end{aligned}$$

où les lettres sont exprimées modulo q .

Lemme 4.2.2 Soit $\mathbf{t}_{b,q}$ le mot de Thue-Morse généralisé et posons $f = f_{01}$. Alors, pour tout $b+1 \leq n \leq 2b-1$, les fréquences

1. du sommet bispécial $w = 01...(n-1)$ et des arêtes se terminant en w vérifiant

$$\begin{aligned} f_{01...(n-1)} &= \frac{1}{b^{m-1}}f - \frac{n-b-1}{b^m}f, \\ f_{(-1)01...(n-1)} &= \frac{1}{b^{m-1}}f - \frac{n-b}{b^m}, \\ f_{(b-2)01...(n-1)} &= \frac{1}{b^m}f, \end{aligned}$$

2. de l'arête $01...(b-1)1...(n+1-b)$ commençant au sommet spécial à gauche $v = 01...(b-1)1...(n-b)$ et des arêtes issues de v vérifiant

$$\begin{aligned} f_{01...(b-1)1...(n+1-b)} &= \frac{1}{b}f, \\ f_{(-1)01...(b-1)1...(n-b)} &= \frac{1}{b^m}f, \\ f_{((-1+(b-1))01...(b-1)1...(n-b)} &= \frac{1}{b}(2f - \frac{1}{q}), \\ f_{((-1+k(b-1))01...(b-1)1...(n-b)} &= \frac{1}{b^k}f, k \in \{2, \dots, m-1\}. \end{aligned}$$

Le résultat principal sur les fréquences des facteurs du mot de Thue-Morse généralisé établi par Balkova est le suivant

Théorème 4.2.1 Soit $b \geq 2$, $q \geq 1$, $b, q \in \mathbb{N}$ et $b = 1[q]$. Soit $\mathbf{t}_{b,q}$ le point fixe commençant par 0 du morphisme $\mu_{b,q}$ défini dans l'exemple 1.3.1 et posons $f = f_{01}$. Alors, les fréquences des facteurs de $\mathbf{t}_{b,q}$ prennent la forme suivante pour $N \in \mathbb{N}$.

Longueur N	Fréquences des facteurs de longueur $N+1$
0	$\frac{1}{q}$
1	$\frac{f}{b^k}, k \in \{0, \dots, m-1\}$
\vdots	\vdots

Pour le complément du tableau précédent nous référons le lecteur à [5].

Nous savons que tout point fixe d'un morphisme primitif admet des fréquences de facteurs [36]. Dans le même article, Queélec a établi que tous facteurs w et v d'un mot engendré par un morphisme marqué, primitif et k -uniforme vérifient $f_v = \frac{f_w}{k}$, où w est l'unique antécédent de v . Par conséquent, le mot \mathbf{t}_3 admet des fréquences de facteurs, puisque μ_3 vérifie toutes ces conditions. Nous avons la proposition suivante :

Proposition 4.2.2 *Les fréquences des facteurs de longueur 2 de \mathbf{t}_3 vérifient $f_{00} < f_{10} < f_{20}$.*

Preuve : Observons que toute occurrence de 00 apparaît uniquement dans 2001. De même, toute occurrence de 2001 apparaît uniquement dans 120012. Par conséquent $f_{00}(\mathbf{t}_3) = f_{120012}(\mathbf{t}_3)$ d'après le Corollaire 1.5.1. Le seul antécédent de 120012 étant 10, alors $f_{00}(\mathbf{t}_3) = \frac{1}{3}f_{10}(\mathbf{t}_3)$. En procédant de façon analogue avec 10, on vérifie que $f_{10}(\mathbf{t}_3) = \frac{1}{3}f_{20}(\mathbf{t}_3)$. En définitive, on a : $f_{00}(\mathbf{t}_3) = \frac{1}{3}f_{10}(\mathbf{t}_3) = \frac{1}{9}f_{20}(\mathbf{t}_3)$. ■

D'après le Lemme 4.2.1, $f_{01}(\mathbf{t}_3) = \frac{b^{m-1}}{q} \frac{b-1}{b^m-1}$, où m est le plus petit entier strictement positif vérifiant $m(b-1) = 0[q]$. Par définition de \mathbf{t}_3 , $b = 3$ et $q = 3$. Ainsi, $m = 3$ et par conséquent $f_{01}(\mathbf{t}_3) = \frac{3}{13}$. Donc $f_{ab}(\mathbf{t}_3) \in \{\frac{1}{39}, \frac{1}{13}, \frac{3}{13}\}$, pour tous $a, b \in \mathcal{A}_3$, d'après le Théorème 4.2.1. On en déduit que $f_{01}(\mathbf{t}_3) = \frac{3}{13}$, $f_{00}(\mathbf{t}_3) = \frac{1}{39}$ et $f_{10}(\mathbf{t}_3) = \frac{1}{13}$ d'après la Proposition 4.2.2.

4.3 Carrés de lettres et facteurs séparateurs

On désigne par facteur séparateur, tout facteur séparant deux carrés consécutifs de lettres dans \mathbf{t}_3 .

Proposition 4.3.1 *Soit w un facteur séparateur. Alors $|w| \in \{7, 16\}$.*

Preuve : A partir d'un carré quelconque de lettre nous construisons w à l'aide de μ_3 jusqu'au carré suivant. Nous savons que $E_3(F_n(\mathbf{t}_3)) = F_n(\mathbf{t}_3)$. Ainsi, il nous suffit de traiter le cas d'un seul carré. Considérons le carré 11.

Soient w_1, w_2 et v des facteurs de \mathbf{t}_3 tels que $v = w_1 11 w_2$ (où w_1 et w_2 sont de longueur au moins 25). Observons que 11 est issu de $\mu_3(21)$. On peut écrire $v = w'_1 \mu_3(21) w'_2$, où w'_1 (resp. w'_2) est un préfixe (resp. suffixe) de w_1 (resp. w_2). Comme 21 commence (resp. se termine) par la dernière (resp. la première) lettre de l'image d'une lettre, alors il admet un unique prolongement à gauche (resp. à droite). Donc $v = w''_1 \mu_3(012120) w''_2$ où w''_1 (resp. w''_2) est un préfixe (resp. suffixe) de w'_1 (resp. w'_2). Pour la suite, notons $v_1 = w''_1 \mu_3(012)$, $v_2 = \mu_3(120) w''_2$ et utilisons v_2 .

Par ailleurs 120 est triprolongeable à droite. Ainsi, 120012, 120120 et 120201 sont des facteurs de \mathbf{t}_3 . Par suite, v_2 prend l'une des formes suivantes : $\mu_3(120012)u_1$, $\mu_3(120120)u_2$, $\mu_3(120201)u_3$, où u_i est un suffixe de w'' , pour tout $i \in \{1, 2, 3\}$. Soit encore $v_2 = 120201012012120210u_1$, $v_2 = 120201012120201012u_2$ ou $v_2 = 120201012201012120u_3$

Pour $v_2 = 120201012201012120u_3$, il ressort que le carré suivant 11 est 22 et le facteur qui les sépare est de longueur 7.

Pour $v_2 = 120201012012120210u_1$, on a $v_2 = \mu_3(\mu_3(10))u_1$. Par ailleurs, 1 est l'unique prolongement à droite de 10. Ainsi, $v_2 = \mu_3(\mu_3(101))u'_1$, où u'_1 est un suffixe de u_1 . On obtient :

$$v_2 = 120201012012120201120201012u'_1$$

Pour $v_2 = 120201012012120210u_1$, on procède de façon analogue au cas précédent et on obtient $v_2 = \mu_3(\mu_3(112))u'_2$, où u'_2 est un suffixe de u_2 . Ainsi, on a :

$$v_2 = 120201012120201012201012120u'_2.$$

En définitive, les séparateurs sont soit de longueur 7 soit de longueur 16. ■

Observons qu'après le carré 11, les carrés de lettres qui suivent sont 11 ou 22. En conséquence, nous avons le résultat suivant.

Corollaire 4.3.1 *Soit w un facteur séparateur. Nous avons les propriétés suivantes.*

1. *Si w sépare ii et jj , alors $i = j$ ou $j = E_3(i)$.*
2. *L'ensemble des facteurs séparateurs de longueur 7 est*

$$\mathcal{S}(7) = \{i^{-1}\mu_3^2(i)j^{-1} : i, j \in \mathcal{A}_3, j = E_3(i)\}.$$

3. *Si $|w| = 16$, alors w sépare soit deux carrés identiques soit deux carrés différents. De plus on a :*

- (a) *Les facteurs séparateurs qui séparent des carrés identiques sont donnés par l'ensemble :*

$$\mathcal{S}_{i-i}(16) = \{i^{-1}\mu_3^2(ij)i^{-1} : i \in \mathcal{A}_3, j = E_3^2(i)\}.$$

Tout facteur séparateur dans cet ensemble admet l'image d'un carré de lettre comme facteur médian.

- (b) *Les facteurs séparateurs qui séparent des carrés différents sont donnés par l'ensemble :*

$$\mathcal{S}_{i-j}(16) = \{i^{-1}\mu_3^2(ii)j^{-1} : i \in \mathcal{A}_3, j = E_3(i)\}.$$

Preuve :

1. Rappelons que dans \mathbf{t}_3 , après le carré 11, le carré de lettres qui suit est 11 ou 22. Soit w_1 (resp. w_2) le facteur séparateur qui sépare le carré 11 du carré 11 (resp. 22). Par suite, posons $u = 11w_111$ et $v = 11w_222$. De la Proposition 3.2.2 on déduit que $E_3(u) = 22E_3(w_1)22$ (resp. $E_3(v) = 22E_3(w_2)00$) et $E_3^2(u) = 00E_3^2(w_1)00$ (resp. $E_3^2(v) = 00E_3^2(w_2)11$) sont dans \mathbf{t}_3 . Ainsi, on obtient les facteurs séparateurs précédés par 22 (resp. 00).
2. Le facteur séparateur de longueur 7 précédé par 11 est $w = 2020101 = 1^{-1}\mu_3(1)2^{-1}$ (voir la preuve de la Proposition 4.3.1.) En appliquant E_3 (resp. E_3^2) à w , on obtient celui précédé par 22 (resp. 00). Comme w sépare les carrés 11 et 22, alors $E_3(w)$ (resp. $E_3^2(w)$) séparent les carrés 22 et 00 (resp. 00 et 11).

3. Soit w un facteur séparateur de longueur 16. Considérons le cas où w est précédé par 11. Les deux autres cas sont donnés respectivement par $E_3(w)$ et $E_3^2(w)$. On a :

$w = 2020101201212020 = 1^{-1}\mu_3^2(10)1^{-1}$ si w est suivi par 11. Par ailleurs, l'image de 00 est un facteur médian de w .

$w = 2020101212020101 = 1^{-1}\mu_3^2(11)2^{-1}$ si w est suivi par 22. ■

Observons qu'aucun facteur séparateur de longueur 7 ne sépare deux carrés identiques.

Corollaire 4.3.2 *Tout facteur de \mathbf{t}_3 de longueur 19 contient au moins un carré de lettres et au plus deux. Cette longueur est minimale.*

Preuve : Soit w un facteur de \mathbf{t}_3 de longueur 19. Alors,

Si w ne contient pas de carré de lettre, alors on aurait un facteur séparateur de longueur 17 au moins. Ceci contredit la proposition 4.3.1.

Si w contient au moins trois carrés de lettre, on aurait un facteur séparateur de longueur au plus $\frac{19-6}{2} < 7$. Ceci contredit la proposition 4.3.1.

$\mu_3^2(10)$ est un facteur de longueur 18 sans carré de lettre. ■

Proposition 4.3.2 *Soit w un facteur séparateur de longueur 16. Alors w est précédé (resp. suivi) par un facteur séparateur de longueur 7.*

Preuve : Soit w un facteur séparateur de longueur 16 précédé par 11. Alors, w prend l'une des formes suivantes :

$$1^{-1}\mu_3^2(10)1^{-1}, 1^{-1}\mu_3^2(11)2^{-1}.$$

Nous savons que 10 (resp. 11) admet un unique prolongement à gauche et un unique prolongement à droite. Ainsi, toute occurrence de 10 (resp. 11) fait partie d'une occurrence de 0101 (resp. 0112). En appliquant μ_3^2 à ces deux facteurs, on vérifie que w est précédé (resp. suivi) par $0^{-1}\mu_3^2(0)1^{-1}$ (resp. $2^{-1}\mu_3^2(2)0^{-1}$ ou $1^{-1}\mu_3^1(0)2^{-1}$). Ces facteurs sont des

facteurs séparateurs de longueur 7 d'après le Corollaire 4.3.1.

On procède de même en précédant w par 00 ou 22. ■

Proposition 4.3.3 *Soient w_1 et w_2 deux facteurs séparateurs consécutifs de longueur 16. Alors, le facteur u séparant w_1 et w_2 vérifie $|u| \in \{11, 38\}$.*

Preuve : Soient w_1 et w_2 deux facteurs séparateurs consécutifs de longueur 16 et u un facteur séparant w_1 et w_2 . La longueur minimale de u est 11 d'après la Proposition 4.3.2 (le facteur séparateur de longueur 7 et les deux carrés de lettres qu'il sépare).

Montrons que u ne peut avoir une longueur intermédiaire entre 11 et 38.

Supposons que u contient deux facteurs séparateurs de longueur 7. Alors, u est de longueur 20 et prend la forme iiu_1jju_2kk , où $j = E_3(i)$, $k = E_3^2(i)$. Quitte à échanger les rôles des lettres, posons $i = 0$. Par suite, $u = 00u_111u_222 = 0\mu_3^2(01)2$. Ainsi, l'un des facteurs $00w_1uw_200$, $00w_1uw_222$, $22w_1uw_200$ et $22w_1uw_222$ est dans \mathbf{t}_3 d'après le Corollaire 4.3.2. Ces facteurs s'écrivent respectivement

$$0\mu_3^2(020122)0, 0\mu_3^2(020121)2, 2\mu_3^2(220122)0, 2\mu_3^2(220121).$$

Par contre, les mots 020122, 020121, 220122 et 2202121 ne sont pas dans \mathbf{t}_3 , puisque 02 et 22 (resp. 21 et 22) ne sont pas des suffixes (resp. des préfixes) d'images de lettres. Donc u ne peut pas prendre la forme iiu_1jju_2kk .

Supposons que u contient trois facteurs séparateurs de longueur 7. Alors, u est de longueur 29 et de la forme $iiu_1jju_2kku_3ii$, où $j = E_3(i)$ et $k = E_3^2(i)$. En procédant comme précédemment, on aboutit à une contradiction.

Montrons que u ne peut avoir une longueur supérieure à 38.

Supposons que u est de longueur 47 qui est la longueur possible suivant 38. Alors, il est de la forme $iiu_1jju_2kku_3iiu_4jju_5kk$. Quitte à échanger les rôles des lettres, posons $i = 0$. On a $u = 00u_111u_222u_300u_411u_522$ ou encore $u = 0\mu_3^2(01201)2$. Si w_1 est précédé par 00 (resp. 22), alors $0\mu_3^2(0201201)$ (resp. $2\mu_3^2(2201201)$) est dans \mathbf{t}_3 . Ce qui est impossible, car 0201201 (resp. 2201201) n'est pas dans \mathbf{t}_3 . On montre de même que u ne peut prendre la longueur 56 qui est la longueur possible qui suit.

Supposons que $|u| > 56$. Alors, on vérifie qu'il contient le facteur $\mu_3(012)\mu_3(012)\mu_3(012) = \mu_3(012012012)$. Ce qui est impossible d'après la Proposition 3.2.2.

Si u est de longueur 38 et commence par le carré 00, alors

$$u = 00121202011202010122010121200121202011 = 0\mu_3^2(0120)1.$$

Par ailleurs, 0120 est prolongeable à gauche (resp. à droite) par 2 (resp. par 1). Par conséquent, $0\mu_3^2(0120)1$ est dans \mathbf{t}_3 . On montre de même que lorsque u commence par le carré 11 (resp. 22), il est dans \mathbf{t}_3 .

En définitive $|u| \in \{11, 38\}$. ■

4.4 Estimation du nombre de carrés de lettres dans les facteurs de \mathbf{t}_3

On désigne par G l'ensemble défini par $G = \{i\mu_3^3(i)\mu_3^2(i)j : i, j \in \mathcal{A}_3, j = E(i)\}$. Observons que tout élément de G est dans \mathbf{t}_3 . D'après la Proposition 4.3.3, les éléments de G correspondent aux facteurs de longueur 38 séparant des facteurs séparateurs consécutifs de longueur 16. Par conséquent, les éléments de G ne s'entrecoupent pas.

Notation : Soient a et b deux lettres n'appartenant pas à \mathcal{A}_3 , $a \neq b$. On note \mathcal{R}_{ab} l'application définie de $F(\mathbf{t}_3)$ dans \mathcal{A}_3^* qui consiste à remplacer dans toute occurrence d'un élément de G le 3^e carré de lettre par ab .

Proposition 4.4.1 *Soit w un facteur de \mathbf{t}_3 de longueur n commençant par le carré d'une lettre. Soit ℓ le nombre d'occurrences des éléments de G dans w . Si n est multiple de 27, alors*

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2n}{27} + \ell.$$

Preuve : Soit w un facteur de longueur n de \mathbf{t}_3 , commençant par un carré de lettre. Nous savons que les éléments de G sont les plus longs facteurs de \mathbf{t}_3 , contenant uniquement des facteurs séparateurs de longueur 7. Par suite, si $|w|$ est multiple de 27, alors il existe un entier k tel que :

$\mathcal{R}_{ab}(w) = \prod_{n=1}^k i_n i_n X_n j_n j_n Y_n$, où X_n et Y_n sont des facteurs séparateurs. Ainsi,

$$|\mathcal{R}_{ab}(w)|_{00} + |\mathcal{R}_{ab}(w)|_{11} + |\mathcal{R}_{ab}(w)|_{22} = \frac{2n}{27}.$$

On obtient le résultat en ajoutant à l'égalité précédente. ■

Dans la suite nous évaluons le nombre de carrés de lettres dans un facteur donné de \mathbf{t}_3 .

Théorème 4.4.1 *Soit w un facteur de \mathbf{t}_3 de longueur n . Alors, il existe des entiers N et ℓ vérifiant $n - 43 \leq N \leq n$ et $\ell \in \{0, 1, 2, 3\}$ tels que*

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2N}{27} + \ell +$$

où ℓ est le nombre d'occurrences des éléments de G dans w .

Preuve : Pour n assez grand, décomposons w sous la forme $w = w_1 u w_2$, où w_1 est le préfixe de w précédant le premier carré de lettre, u le facteur de w commençant par le premier carré de lettre tel que $|u| \leq 27$ et w_2 le suffixe de w vérifiant $|w_2| < 27$. Posons $N = |u| = n - (|w_1| + |w_2|)$. Nous savons que $|u|_{00} + |u|_{11} + |u|_{22} = \frac{2N}{27} + \ell$ d'après la Proposition 4.4.1.

Déterminons le nombre possible de carrés de lettres dans w_2 . Deux cas se présentent : w_2 est un élément de G ou non.

Si w_2 est dans G , alors,

$$|w_2|_{00} + |w_2|_{11} + |w_2|_{22} = \begin{cases} 1 & \text{si } 2 \leq |w_2| \leq 10 \\ 2 & \text{si } 11 \leq |w_2| \leq 19 \\ 3 & \text{si } 20 \leq |w_2| \end{cases}$$

Si w_2 n'est pas dans G , alors w_2 contient au plus deux carrés de lettres. Posons le nombre de carrés de lettres dans w_2 . On obtient alors :

$$|w|_{00} + |w|_{11} + |w|_{22} = \frac{2N}{27} + \ell + \dots \quad \blacksquare$$

Proposition 4.4.2 *Chaque élément de G provient de l'image d'un carré de lettres par μ_3^3 .*

Preuve : Soit w un élément de G . Supposons que w commence par le carré 11. Alors, il est de la forme $1\mu_3^3(1)\mu_3^2(1)2 = 1\mu_3^2(1201)2$. Par ailleurs, le facteur 1201 est bispécial biprolongeable. Par suite, les facteurs 012012, 012010 et 212012 sont dans \mathbf{t}_3 . Observons que parmi ces facteurs, seule l'image de 012012 par μ_3^2 contient w et on a :

$$\mu_3^2(012012) = \mu_3^3(00) = 01212020w01012120.$$

On procède de même en commençant par le carré 00 (resp. 22) ■

En conséquence, pour tout $w \in G$, ni $\mu_3(w)$ ni $\mu_3^2(w)$ ne contient d'élément de G , puisque w ne contient pas d'image de carré de lettre.

Proposition 4.4.3 *Les plus longs facteurs de \mathbf{t}_3 contenant des carrés de lettres et non des images de carrés de lettres sont ceux de la forme u_1wu_2 , où $w \in G$, et $|u_1| = |u_2| = 10$.*

Preuve : Nous savons que tout facteur séparateur de carrés identiques est centré en l'image d'un carré de lettre d'après le Corollaire 4.3.1. Observons que parmi les facteurs séparateurs, seuls ceux de longueur 7 ne contiennent pas d'image de carré de lettres. De plus, les éléments de G sont les plus longs facteurs de \mathbf{t}_3 ne contenant pas de facteur séparateur de longueur 16. Ainsi, il s'agit de déterminer pour tout facteur w de G , le plus long prolongement à gauche (resp. à droite), ne contenant pas d'images de carrés de lettres. Supposons que w provient de l'image de 00 et posons $u = \mu_3^3(00)$. On a

$$u = \mu_3^3(00) = 01212020w01012120.$$

Par ailleurs, 2 (resp. 1) est l'unique prolongement à gauche (resp. à droite) de 00. En appliquant μ_3^3 à 2001 on constate que le premier carré 11 (resp. dernier carré 22) de w est précédé par 11 (resp. 22). Par suite, 01212020 (resp. 01012120) est prolongeable à gauche (resp. à droite) par 012 (resp. 120). Par conséquent, u_1 (resp. u_2) contiendrait l'image d'un carré de lettre si $|u_1| \geq 11$ (resp. $|u_2| \geq 11$).

On procède de la même façon lorsque w provient de l'image de 11 (resp. 22). ■

Proposition 4.4.4 *Soit w un facteur de \mathbf{t}_3 de longueur n . Soit α_n le nombre d'occurrences des éléments de G dans w . Alors $\frac{\alpha_n}{n}$ converge vers $\frac{1}{351}$ quand n tend vers ∞ .*

Preuve : Nous savons que $|w|_{00} + |w|_{11} + |w|_{22} = \frac{2N}{27} + \alpha_n$ d'après le Théorème 4.4.1. Par suite, $f_{00}(\mathbf{t}_3) + f_{11}(\mathbf{t}_3) + f_{22}(\mathbf{t}_3) = \frac{2}{27} + \frac{\alpha_n}{n}$, où $\frac{\alpha_n}{n} = \lim_{n \rightarrow \infty} \frac{\alpha_n}{n}$. L'existence de $\frac{1}{351}$ découle du fait que \mathbf{t}_3 admet des fréquences de facteurs. Par ailleurs, $f_{00}(\mathbf{t}_3) = f_{11}(\mathbf{t}_3) = f_{22}(\mathbf{t}_3) = \frac{1}{39}$. Ainsi, $\frac{1}{13} = \frac{2}{27} + \frac{1}{351}$. ■

4.5 Facteurs biprolongeables et mots de retour dans \mathbf{t}_3

Dans cette section, nous établissons quelques propriétés sur les facteurs bispéciaux biprolongeables de \mathbf{t}_3 . Ensuite, nous montrons que le nombre de mots de retour d'un facteur de \mathbf{t}_3 est soit 7, 8 ou 9. Enfin, nous étudions la fonction de complexité privilégiée de \mathbf{t}_3 .

4.5.1 Facteurs bispéciaux biprolongeables

Nous désignerons par $BSB(\mathbf{t}_3)$ l'ensemble des facteurs bispéciaux à la fois biprolongeables à droite et à gauche de \mathbf{t}_3 .

Proposition 4.5.1 *Soit w un facteur de \mathbf{t}_3 de longueur au moins 2. Alors, w est biprolongeable à droite (resp. à gauche) si et seulement si $\mu_3(w)$ est biprolongeable à droite (resp. à gauche).*

Preuve : Soit w un facteur de \mathbf{t}_3 biprolongeable à droite. On peut Supposer sans perte de généralité que w se prolonge par 0 et 1. Alors, $\mu_3(w)0$ et $\mu_3(w)1$ sont dans \mathbf{t}_3 , puisque $\mu_3(i)$ commence par i pour tout $i \in \mathcal{A}_3$.

Réciproquement, soit w un facteur de \mathbf{t}_3 tel que $\mu_3(w)$ est biprolongeable à droite. Quitte à échanger les rôles des lettres, supposons que $\mu_3(w)0$ et $\mu_3(w)1$ sont dans \mathbf{t}_3 . Comme que $|w| \geq 2$ et les facteurs $\mu_3(w)0$ et $\mu_3(1)$ admettent chacun une synchronisation unique.

Nous savons que 0 (resp. 1) sont les premières lettres de leurs images respectives. Par suite, les facteurs $\mu_3(w)012$ et $\mu_3(w)120$ sont dans \mathbf{t}_3 . Ces deux facteurs s'écrivent respectivement $\mu_3(w0)$, et $\mu_3(w1)$. Ce qui prouve que w est biprolongeable à droite dans \mathbf{t}_3 . On procède de même pour les facteurs biprolongeables à gauche. ■

En conséquence, tout facteur w de \mathbf{t}_3 est bispécial biprolongeable si $\mu_3(w)$ l'est et réciproquement.

Proposition 4.5.2 *Soit $w \in BSB(\mathbf{t}_3)$ tel que $|w| \geq 7$. Alors, il existe un unique v dans $BSB(\mathbf{t}_3)$ tel que $w = \mu_3(v)$.*

Preuve : Soit $w \in BSB(\mathbf{t}_3)$ vérifiant $|w| \geq 7$. Alors, w se synchronise de façon unique sous la forme ${}_1\mu_3(v) {}_2$. Comme w est bispécial biprolongeable, alors il commence (rep. se termine) par l'image d'une lettre. Par suite, ${}_1, {}_2 = \cdot$. Par conséquent, $w = \mu_3(v)$ et $v \in BSB(\mathbf{t}_3)$ d'après la Proposition 4.5.1. ■

Théorème 4.5.1 *L'ensemble $BSB(\mathbf{t}_3)$ est donné par :*

$$BSB(\mathbf{t}_3) = \bigcup_{n \geq 0} \{\mu_3^{n+1}(i)\mu_3^n(i), \mu_3^{n+1}(i)\mu_3^n(ij) : i, j \in \mathcal{A}_3, j = E_3(i)\}.$$

Preuve : En utilisant la Proposition 4.5.2, il suffit de trouver les éléments de $BSB(\mathbf{t}_3)$ de longueur au plus 7, puisque les autres s'obtiennent par applications successives de μ_3 . Ces facteurs sont $i(i+1)(i+2)i = \mu_3(i)i$ et $\mu_3(i)ij$, $i, j \in \mathcal{A}_3, j = E_3(i)$. ■

4.5.2 Mots de retour et facteurs privilégiés

Nous débutons la sous section par le principal résultat.

Théorème 4.5.2 *Tout facteur de \mathbf{t}_3 admet 7, 8 ou 9 mots de retour.*

Preuve :

De la Section 2, nous savons que pour déterminer la structure et le cardinal de l'ensemble des mots de retour des facteurs de \mathbf{t}_3 , il suffit de considérer les facteurs bispéciaux, puisque \mathbf{t}_3 est non ultimement périodique et uniformément récurrent. De plus, d'après la Proposition 3.3.2, la Proposition 4.5.2 et le Lemme 1.6.1, il suffit de considérer les éléments initiaux $i, ij, \mu_3(i)$ de $BST(\mathbf{t}_3)$ et $\mu_3(i)i, \mu_3(i)ij$ de $BSB(\mathbf{t}_3)$, où $i, j \in \mathcal{A}_3$ et $j = E(i)$. En vertu de la Proposition 3.2.2, nous pouvons restreindre l'étude à 0, 01 de $BST(\mathbf{t}_3)$ et $\mu_3(0)0, \mu_3(0)01$ de $BSB(\mathbf{t}_3)$.

1. Déterminons $Ret(0)$.

Comme pour tout $i \in \mathcal{A}_3$, 0 apparaît dans $\mu_3(i)$, alors les mots de retour de 0 proviennent $\mu_3(ij)$, pour tout $i, j \in \mathcal{A}_3$. Ainsi, nous avons

$$Ret(0) = \{0, 01, 02, 012, 0112, 0122, 01212\}.$$

2. Déterminons $Ret(01)$.

Observons que 01 apparaît dans $\mu_3(0), \mu_3(2)$ et $\mu_3(11)$. Par suite, les mots de retour de 01 proviennent de $\mu_3(00), \mu_3(02), \mu_3(22), \mu_3(20), \mu_3(010), \mu_3(011), \mu_3(012), \mu_3(112)$ et $\mu_3(212)$. Par conséquent, l'ensemble des mots de retour de 01 est donné par

$$Ret(01) = \{01, 012, 0122, 01212, 01202, 011202, 012120, 0121202\}.$$

3. Déterminons $Ret(012)$.

Observons que $012 = \mu_3(0)$ et 012 apparaît dans $\mu_3(11)$ et $\mu_3(22)$. Par suite, les mots de retour de 012 proviennent de $\mu_3(00), \mu_3(010), \mu_3(020), \mu_3(0120), \mu_3(011), \mu_3(0122), \mu_3(012120), \mu_3(1120)$ et $\mu_3(220)$. Par conséquent, l'ensemble des mots de retour de 012 est donné par

$$Ret(012) = \{\mu_3(0), \mu_3(01), \mu_3(02), \mu_3(012), \mu_3(011)0^{-1}, \mu_3(0122)(01)^{-1}, \mu_3(01212), (12)^{-1}\mu_3(112), 2^{-1}\mu_3(22)\}.$$

4. Déterminons $Ret(0120)$.

Observons que 0120 apparaît dans $\mu_3(ii)$, pour tout $i \in \mathcal{A}_3$. Ainsi, les mots de

retour de 0120 sont issus de $\mu_3(iiwjj)$, où $i, j \in \mathcal{A}_3$ et w est un facteur séparateur. Par conséquent, en vertu du Corollaire 4.3.1, l'ensemble des mots de retour de 0120 est :

$$\begin{aligned} \text{Ret}(0120) = \{ & \mu_3(0\mu_3^2(0))0^{-1}, (12)^{-1}\mu_3(1\mu_3^2(1))(01)^{-1}, (2)^{-1}\mu_3(2\mu_3^2(2))(012)^{-1}, \\ & \mu_3(0\mu_3^2(02))(012)^{-1}, (12)^{-1}\mu_3(1\mu_3^2(10))(0)^{-1}, (2)^{-1}\mu_3(2\mu_3^2(21))(01)^{-1}, \\ & \mu_3(0\mu_3^2(00))(0)^{-1}, (12)^{-1}\mu_3(1\mu_3^2(11))(01)^{-1}, (2)^{-1}\mu_3(2\mu_3^2(22))(012)^{-1} \}. \end{aligned}$$

5. Déterminons $\text{Ret}(0120)$.

Observons que 01201 apparaît dans $\mu_3(00)$ et $\mu_3(22)$. Ainsi, les mots de retour de 01201 sont issus de $\mu_3(00w_100)$, $\mu_3(00w_222)$, $\mu_3(22w_300)$, $\mu_3(22w_422)$ où les w_i sont des facteurs de \mathbf{t}_3 . Posons u_{ij} (resp. v_{ij}) le facteur séparateur de longueur 16 (resp. longueur 7) qui sépare les carrés ii et jj , $i, j \in \mathcal{A}_3$. Par suite, les mots de retour de 01201 proviennent de $\mu_3(00u_{00}00)$, $\mu_3(00u_{01}11v_{12}22)$, $\mu_3(00v_{01}11v_{12}22)$, $\mu_3(00v_{10}11u_{11}11v_{12}22)$, $\mu_3(00v_{01}11u_{12}22)$, $\mu_3(22u_{22}22)$, $\mu_3(22u_{20}00)$ et $\mu_3(22v_{20}00)$. Par conséquent, l'ensemble des mots de retour de 01201 est donné par

$$\begin{aligned} \text{Ret}(01201) = \{ & \mu_3(00u_{00}00)2^{-1}, \mu_3(00u_{01}11v_{12}22), \mu_3(00v_{01}11v_{12}22), \\ & \mu_3(00v_{01}11u_{11}11v_{12}22), \mu_3(00v_{01}11u_{12}22), 2^{-1}\mu_3(22u_{22}22), 2^{-1}\mu_3(22u_{20}00)2^{-1}, \\ & 2^{-1}\mu_3(22v_{20}00)2^{-1} \} \blacksquare \end{aligned}$$

Proposition 4.5.3 *Pour tout entier naturel $n \geq 2$, la complexité privilégiée $A_{\mathbf{t}_3}(n)$ de \mathbf{t}_3 est donnée par*

$$\frac{1}{3}A_{\mathbf{t}_3}(n) = A_n(00) + A_n(010) + A_n(020) + A_n(0120) + A_n(01120) + A_n(01220) + A_n(012120).$$

Preuve : Comme $E_3(F_n(\mathbf{t}_3)) = F_n(\mathbf{t}_3)$, nous traitons le cas des facteurs commençant par 0. Les autres cas se traitent de façon analogue. De la preuve du Théorème 4.5.2, nous avons :

$$\text{Ret}(0) = \{0, 01, 02, 012, 0112, 0122, 01212\}.$$

Par suite, les mots de retour complet de 0 sont 00, 010, 020, 0120, 01120, 01220 et 012120. Ainsi, les facteurs privilégiés de \mathbf{t}_3 commençant par 0 et de longueur au moins 2 sont repartis en 7 groupes dépendants des 7 mots de retour complet de 0 et on a :

$$Pri_n(0) = \{0\} \quad Pri_n(00) \quad Pri_n(010) \quad Pri_n(020) \quad Pri_n(0120) \quad Pri_n(01120) \quad Pri_n(01220) \\ Pri_n(012120).$$

Remarquons qu'aucun mot de retour complet de 0 n'est préfixe ou suffixe de l'autre. Par suite, les ensembles $Pri_n(00)$, $Pri_n(010)$, $Pri_n(020)$, $Pri_n(0120)$, $Pri_n(01120)$, $Pri_n(01220)$ et $Pri_n(012120)$ sont deux à deux disjoints. Par conséquent, la fonction de complexité privilégiée de \mathbf{t}_3 est

$$\frac{1}{3}A_{\mathbf{t}_3}(n) = A_n(00) + A_n(010) + A_n(020) + A_n(0120) + A_n(01120) + A_n(01220) + A_n(012120). \blacksquare$$

Pour la suite, notons $R = \{00, 010, 020, 0120, 01120, 01220, 012120\}$.

Lemme 4.5.1 *Soit $w \in Pri_0(n)$ tel que $|w| \geq 7$. Alors, il existe $u \in R$ tel que w commence par un mot de retour complet de u .*

Preuve : Soit $w \in Pri_0(n)$. Alors, w admet un préfixe propre u dans R , puisque $|w| \geq 7$. Comme w est privilégié, w se termine par u . Ainsi, w contient au moins deux occurrences de u . Par conséquent, w commence par un mot de retour complet de u . \blacksquare

Théorème 4.5.3 *La fonction de complexité privilégiée $A_{\mathbf{t}_3}$ de \mathbf{t}_3 vérifie*

$$A_{\mathbf{t}_3}(n) = \begin{cases} 1 & \text{si } n = 0 \\ 3 & \text{si } n \in \{1, 2, 4, 6\} \\ 6 & \text{si } n \in \{3, 5\} \\ 3A_{010}(n) + 3A_{020}(n) + 3A_{012120}(n) & \text{si } n \equiv 0[3], n \geq 9 \\ 3A_{0120}(n) & \text{si } n \equiv 1[3], n \geq 7 \\ 3A_{00}(n) + 3A_{01120}(n) + 3A_{0120}(n) + 3A_{01220}(n) & \text{si } n \equiv 2[3], n \geq 8 \end{cases}$$

Preuve : Nous avons

$$Pri_0(1) = \{0\}, Pri_0(2) = \{00\}, Pri_0(3) = \{010, 020\}, Pri_0(4) = \{0120\}, Pri_0(5) = \{01120, 01220\}, Pri_0(6) = \{012120\}$$

Soit w un facteur privilégié de \mathbf{t}_3 commençant par 0 et de longueur n , $n \geq 7$. Alors w commence (resp. se termine) par un élément de R d'après le Lemme 4.5.1. Ainsi, en utilisant les éléments de R nous avons les cas suivants :

Cas 1 : Le facteur w commence (resp. se termine) par 00. Alors, w est sous la forme $00w'00$, avec $w' \in F(\mathbf{t}_3)$. Par ailleurs, dans le carré 00 le premier 0 est le suffixe de $\mu_3(1)$ tandis que le second 0 est le préfixe de $\mu_3(0)$. Par suite, w se synchronise sous la forme $w = 0\mu_3(w'')0$, $w'' \in F(\mathbf{t}_3)$. Ainsi, w est de longueur $3k + 2$ puisque μ_3 est 3-uniforme. De même, on montre que si w commence par 01120 (resp. 01220), alors il se synchronise sous la forme $01\mu_3(w'')$ (resp. $\mu_3(w'')20$), $w'' \in F(\mathbf{t}_3)$.

En définitive, lorsque w commence par 00, 01120 ou 01220, alors $|w| = 3k + 2$, $k \geq 1$.

Cas 2 : Le facteur w commence par 010. Alors, en procédant comme au cas 1, on montre que w se synchronise sous la forme $01\mu_3(w')0$ puisque 01 est le suffixe de $\mu_3(2)$. De même, si w commence par 012120, alors w se synchronise sous la forme $\mu_3(w')$ puisque 012120 est l'image de 01 par μ_3 .

Du cas 3 de la preuve du Théorème 4.5.2, on observe que tout mot de retour de 0120 est soit de longueur $3k$ soit de longueur $3k + 2$. Par conséquent, tout mot de retour complet de 0120 est soit de longueur $3(k + 1) + 1$ soit de longueur $3(k + 1) + 2$. ■

- [1] B. Adamczewski, *Balances for fixed points of primitive substitutions*, Theoret. Comput. Sci. 307 (2003) 47–75.
- [2] A. Adler, S.Y. R. Li, *Magic cubes and Prouhet sequences*, Amer. Math. Monthly 84 (8) :618–627, 1977.
- [3] J. P. Allouche, J. Shallit. *Automatique sequences*, Theory, Applications, Generalizations, Cambridge University Press, UK, 2003.
- [4] J. P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockush, S. Plouffe, B. E. Sagan, *A relative of the Thue-Morse sequence*, Discrete Math. 139 (1995) 455-461.
- [5] J. Balkova, *Factor frequencies in generalized Thue-Morse words*, Kybernetika, Vol. 48(2012), Num. 3, pp. 371–385.
- [6] J. Balkova, K. Benda, O. Turek, *Abelian complexity of infinite words associated with quadratic Parry numbers*, Theoret. Comput. Sci. 412 (2011) 6252-6260.
- [7] J. Balkova, E. Pelantová, W. Steiner, *Return words in fixed points of substitutions*, Monatsh. Math. 155(3-4) (2008), 251263.
- [8] J. Berstel, A. Lauve, C. Reutenauer, F. Saliola, *Combinatorics on Words : Christoffel Words and Repetitions in Words*, Amer. Math. Soc., 2008.

- [9] J. Berstel, D. Perrin, *The origins of combinatorics on words*, European J. Combinatorics 28 (2007), 996-1022.
- [10] J. Berstel, J. Karhumäki, *Combinatorics on words- a tutorial*, Bull. European Theor. Comput. Sci. 79 (2003), 178-228.
- [11] V. Berthé, M. Rigo, Eds, *Combinatorics, Automata and Number Theory*, Encyclopedia of Mathematics and its Applications 135, Cambridge University Press, 2010.
- [12] J. Cassaigne, *Facteurs spéciaux*, Bull. Belg. Math. 4(1997), 67-88.
- [13] J. Cassaigne, I. Kaboré, *A word without uniforme frequency*, preprint.
- [14] J. Cassaigne, I. Kaboré, *Abelian complexity and frequencies of letters in finite words*, Int. J. Found. Comput. Sci., Vol. 27(05), pp. 631- 649, 2016.
- [15] J. Cassaigne, G. Richomme, K. Saari, L. Q. Zamboni, *Avoiding Abelian powers in binary words with bounded Abelian complexity*, Int. J. Found. Comput. Sci. 22(2011), 905-920.
- [16] L. Carlitz, R. Scorville, V. Hoggatt, *Representation for a special sequence*, Fibonacci Quart. 10(1972), 499-518, 550
- [17] C. Chomsky, J. Karhumäki, *Combinatorics of words*, Handbook of Formal Languages, Vol. 1, chapter 6, pp. 329-438. Springer-Verlag, 1997.
- [18] E. M. Coven, G. A. Hedlund, *Sequences with minimal block growth*, Math. Syst. Theory 7(1973), 138-153.
- [19] J. Curie, N. Rampersad, *Recurrent words with constant abelian complexity*, Adv. Appl. Math. 47(2011), 116-124.
- [20] F. Durand, *A characterisation of substitutive sequences using return words*, Discrete Math. 179(1998), 89-101.
- [21] S. Ferenczi, C. Mauduit, A. Nogueira, *Substitution dynamical systems : algebraic characterization of eigenvalues*, Ann. Sci. Norm. Supér. 29(1996), 519-533.

- [22] W. H. Gottschalk, *Substitution on minimal sets*, Trans. Amer. Math. Soc. 109(1963).
- [23] I. Kaboré, B. Kientéga, *Some combinatorial properties of the ternary Thue-Morse word*, In. J. Appl. Math., Vol. 30, No 5(2018).
- [24] I. Kaboré, B. Kientéga, *Abelian complexity of the Thue-Morse word over a ternary alphabet*, In : Combinatorics on words (WORDS 2017), Brlek S. et al. (eds), LNCS 10432, pp 132-143, 2017.
- [25] J. Kellendonk, D. Lenz, and J. Savinien, *A Characterization of Subshifts with Bounded Powers*, Discrete Math. 313 (2013), 2881–2894.
- [26] D. H. Lehner, *The Tarry-Escott problem*, Scripta Math. 13(1947), 37-41.
- [27] M. Lothaire, *Applied Combinatorics on Words*, Vol. 105 of Encyclopedia of Mathematics. Cambridge University Press, 2005.
- [28] M. Lothaire, *Algebraic Combinatorics on Words*, Vol. 90 of Encyclopedia of Mathematics. Cambridge University Press, 2002.
- [29] M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, 1983(17).
- [30] F. Mignosi, P. Séebold, *If a DOL-language is k -power-free then it is circular*, Lect. Notes Comp. Sci. 700, 507-518, Springer.
- [31] M. Morse, G. A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. 62 (1940), 1–42.
- [32] M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. 44(1938), 632.
- [33] J. Peltomäki, *Privileged factors in the Thue-Morse Word- A comparison of privileged words and Palindromes*, Disc. Appl. Math., 187-199, 193(2015).
- [34] J. Peltomäki, *Introducing Privileged Words : Privileged Complexity of Sturmian Words*, Theoret. Comput. Sci. 500 (2013), 57–67.

- [35] E. Prouhet, *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sci. Paris 33(1851), 225.
- [36] M. Que'elec, *Substitution dynamical systems- Spectral analysis*, in Lecture Notes in Math. 1294, 1987.
- [37] G. Richomme, K. Saari, L. Q. Zamboni, *Abelian complexity of minimal subshifts*, J. London Math. Soc. 83(2011), 79-95.
- [38] G. Richomme, K. Saari, L. Q. Zamboni, *Balance and Abelian complexity of the Tribonacci word*, adv. Appl. Math. 45(2010), 212-231.
- [39] D. Robbins, *Solution to problem E 2692*, Amer. Math. Monthly 86(1979), 394-395
- [40] K. Saari, *On the frequency of letters in morphic sequences*, Grigoriev, D., Harrison, J. Hirsch, E.A (Eds) CSR 2006. LNCS, vol. 3967 : 334-345, 2006.
- [41] ' . Starosta, *Generalised Thue-Morse word and palindromic richness*, Kibernetika, Vol. 48(2012), No. 3, 361-370.
- [42] J. Tamura, *Partition of the set of positive integers, nonperiodic sequences and transcendence*, Anal. numb. Theory, Kyoto, 1995, Sūrikaiseikikenkyūsho Kōkyūroku, 961(1996), 161-182.
- [43] A. Thue, *Über unendliche Zeichenreihen*, Norske Vid. Selsk. Skr. Mat. Nat. Kl.7(1906), 1-22 .
- [44] A. Thue, *Über die gegenseilige lage gleicher Teile gewisser Zeichenreihen*, Norske vid. selsk. Mat. Nat. Kl. 1(1912), 1-67.
- [45] O. Turek, *Balance and Abelian complexity of a certain class of infinite ternary words*, RAIRO : Theoret. Informatics Appl. 44(2010), 313-337.
- [46] O. Turek, *Abelian complexity and abelian co-decomposition*, Theor. Comput. Sci. 469(2013), 77-91.
- [47] O. Turek, *Abelian complexity of the Tribonacci word*, J. Integer Sequences, Vol. 18(2015), pp. 15. 3. 4.

- [48] D. R. Woods, *Elementary problem proposal E 2692*, Amer. Math. Monthly 85(1978), 48.
- [49] L. Vuillon, *A characterization of sturmian words by return words*, Eur. J. Comb. 22(2001), 263-275.
- [50] E. M. Wright, *Prouhet's 1851 solution of the Tarry-Escott problem of 1910*, Amer. Math. Monthly 66 (3) :199–201, 1959. 86, 87

Résumé : Dans ce mémoire, nous étudions des propriétés combinatoires du mot de Thue-Morse ternaire. Il s'agit sur l'alphabet $\mathcal{A}_3 = \{0, 1, 2\}$ du mot \mathbf{t}_3 engendré par le morphisme μ_3 , défini par : $\mu_3(0) = 012$, $\mu_3(1) = 120$, $\mu_3(2) = 201$.

On appelle complexité abélienne d'un mot infini u , la fonction notée ρ_u^{ab} , qui compte le nombre de vecteurs de Parikh de longueur donnée dans u . Un mot de retour d'un facteur v de u est un facteur de u qui sépare deux occurrences successives de v dans u . Dans le chapitre 1 nous rappelons quelques concepts de base en combinatoire des mots. Le chapitre 2 est consacré au rappel de quelques propriétés du mot de Thue-Morse binaire \mathbf{t}_2 . Dans le chapitre 3 nous déterminons l'ensemble des facteurs bispéciaux biprolongeables de \mathbf{t}_3 puis nous montrons qu'il est 2 équilibré. Nous terminons en montrant que la complexité abélienne de \mathbf{t}_3 est donnée par la suite ultimement périodique 136(766) . Le dernier chapitre est consacré à l'étude des carrés de lettres de \mathbf{t}_3 ainsi que la structure des facteurs qui les séparent. Les mots de retour de \mathbf{t}_3 sont également étudiés. Nous montrons que tout facteur de \mathbf{t}_3 admet 7, 8 ou 9 mots de retour.

Abstract : In this thesis, we study some combinatorial properties of the Thue-Morse word over a ternary alphabet. It concerns the word \mathbf{t}_3 generated on the alphabet \mathcal{A}_3 by the morphism μ_3 defined by : $\mu_3(0) = 012$, $\mu_3(1) = 120$, $\mu_3(2) = 201$.

The abelian complexity of a word u is the function ρ_u^{ab} which enumerates the Parikh vectors number of a gived length in u . A return word of a factor v of u is a factor which separates two successive occurrences of v in u .

In the first chapter, we recall some basis concepts in combinatoric of words. The second chapter is devoted to recall of some properties of the binary Thue-Morse word \mathbf{t}_2 . In the third chapter, we deteminate the set of biprolongeable bispecial factors of \mathbf{t}_3 and we show that \mathbf{t}_3 is 2 balanced. Finaly, we show that the abelian complexity of \mathbf{t}_3 is gived by the ultimetely periodic sequence 136(766) . The last chapter is devoted to the study of squares of letters of \mathbf{t}_3 and factors which disconnect them. The return word of \mathbf{t}_3 are studied too. We show that every factor of \mathbf{t}_3 admits 7, 8 or 9 return words.