Université de la Méditerranée
Faculté des sciences de Luminy

THÈSE

pour obtenir le grade de

Docteur de l'Université de la Méditerranée
Aix-Marseille Université

Spécialité : **Mathématiques**

préparée à l' **Institut de Mathématiques de Luminy**

dans le cadre de l'École Doctorale **en Mathématiques et Informatique de Marseille**

présentée et soutenue publiquement
par

Elise Vaslet

le 23 juin 2011

Titre:
RÉPÉTITIONS DANS LES MOTS ET SEUILS D'ÉVITABILITÉ

Directeur de thèse: **Sébastien Ferenczi**
Co-directeur de thèse: **Julien Cassaigne**

Rapporteurs

Anna Frid
Michel Rigo

Jury

Anna Frid,	Rapporteur
Michel Rigo,	Rapporteur
Pascal Hubert,	Examinateur
Idrissa Kaboré,	Examinateur
Michaël Rao,	Examinateur
Serge Troubetzkoy,	Examinateur
Sébastien Ferenczi,	Directeur de thèse
Julien Cassaigne,	Co-directeur de thèse

Est-ce que la théorie est une place forte ? Ou est-ce qu'elle est une chasse gardée ? Ou bien plutôt qu'est-ce que la théorie ? Les minoritaires, dans quelque société que ce soit, sont dans une position singulière en ce qui concerne les productions intellectuelles : le plus souvent ils haïssent la théorie, la connaissant pour ce qu'elle est, le verbiage sacerdotal de ceux qui les dominent, ce qui sort de la tête et de la bouche de ceux qui disposent de la force (outils, armes concrètes, police, armée) et de la nourriture (salaires, terres, biens . . .). Dans la relation majoritaire/minoritaire la force, les biens et la liberté individuelle qui en découlent étant des caractéristiques du dominant, l'expression institutionnalisée de sa conscience et de sa vue de la situation est la seule a être publiée, diffusée, et glosée. Cela alors se nomme « théorie ». De plein droit. [...] La pensée qui s'élabore, là, jamais n'est appelée théorie.

Colette Guillaumin, *Femmes et théories de la société : remarques sur les effets théoriques de la colère des opprimées*, dans *Sexe, Race et Pratique du pouvoir*

La Bibliothèque est illimitée et périodique. S'il y avait un voyageur éternel pour la traverser dans un sens quelconque, les siècles finiraient par lui apprendre que les mêmes volumes se répètent toujours dans le même désordre - qui, répété, deviendrait un ordre : l'Ordre. Ma solitude se console à cet élégant espoir.

Jorge Luis Borges, *La bibliothèque de Babel*, dans *Fictions*

Résumé

Nous étudions dans cette thèse différents problèmes d'évitabilité des répétitions dans les mots infinis. Soulevée par Thue et motivée par ses travaux sur les mots sans carrés, la problématique s'est développée au cours du XXe siècle, et est aujourd'hui devenue un des grands domaines de recherche en combinatoire des mots. En 1972, Dejean proposa une importante conjecture, dont la validation étape par étape s'est terminée récemment (2009). La conjecture concerne le seuil des répétitions d'un alphabet, i.e., la borne inférieure des exposants évitables sur cet alphabet. La notion de seuil, comme frontière entre évitabilité et non-évitabilité d'un ensemble donné de mots, est le fil directeur de nos travaux. Nous nous intéressons d'abord à une généralisation du seuil des répétitions (nous donnons des encadrements de sa valeur). Cette notion permet d'ajouter, pour décrire l'ensemble des répétitions à éviter, au paramètre de l'exposant, celui de la longueur des répétitions. Puis, nous étudions des problèmes d'existence de mots dans lesquels, simultanément, certaines répétitions sont interdites et d'autres sont forcées. Nous répondons, pour l'alphabet ternaire, à la question : quels réels sont l'exposant critique d'un mot infini sur un alphabet fixé ? Nous introduisons ensuite une notion de haute répétitivité, et établissons une description partielle des couples d'exposants paramètrant une double contrainte de haute répétitivité et d'évitabilité. Pour finir, nous utilisons des résultats et techniques issus de ces problématiques pour résoudre une question de coloration de graphes : nous introduisons un seuil des répétitions, calqué sur celui connu pour les mots, et donnons sa valeur pour deux classes de graphes, les arbres et les graphes de subdivisions.

Abstract

In this thesis we study various problems on repetition avoidance in infinite words. Raised by Thue and motivated by his work on squarefree words, the topic developed during the 20th century, and has nowadays become a principal area of research in combinatorics on words. In 1972, Dejean proposed an important conjecture whose verification in steps was completed recently (2009). The conjecture concerns the repetition threshold for an alphabet, i.e., the infimum of the avoidable exponents for that alphabet. The notion of threshold as a borderline between avoidability and unavoidability for a given set of words is the guiding line of our work. First, we focus on a generalization of the repetition threshold. This concept allows us to include, in addition to the exponent, the length of the repetitions as a parameter in the description of the set of repetitions to avoid. We obtain various bounds in that respect. We then study existence problems for words in which simultaneously some repetitions are forbidden, and others are forced. For the ternary alphabet, we answer the question : what real numbers are the critical exponent of some infinite word over a given alphabet ? Also, we introduce a notion of highly repetitive words and give a partial description of the pairs of exponents which parameterize the existence of words both highly repetitive and repetition-free. Finally, we use results and techniques stemming from those problems to solve a question on graph colouring : we introduce a repetition threshold adapted from the thresholds we know for words, and give its value for two classes of graphs, namely, trees and subdivision graphs.

Remerciements

Je remercie chaleureusement, vivement et sincèrement, tou-te-s celleux sans qui rien n'aurait été pareil. Pour l'écoute, les discussions et les débats, les aiguillages, la bienveillance, la disponibilité, l'attention, la patience, la présence, les coups de mains, les coups de pouce, l'hébergement, les corrections, les rencontres, l'accueil, les invitations, la bouffe, les échanges, le quotidien, les encouragements, le réconfort, les thunes, la réconciliation informatique, le transfo de freebox, les sauvetages administratifs, les conseils, les relectures, les explications, les luttes, les voyages, les virées en virages, la compréhension, l'amitié, les veillées de la Timone, les feuilles de TD, les remplacements de surveillance, le bon voisinage, les cafés de luxe et ceux de la Plaine, et toutes les petites choses...

Un merci tout particulier à Julien Cassaigne et Sébastien Ferenczi, qui ont été les chefs qu'il me fallait, exactement ; à Anna Frid et Michel Rigo, qui ont accepté la charge de rapporteurs ; à Pascal Hubert, Idrissa Kabore, Michaël Rao, et Serge Troubetzkoy, qui ont bien voulu faire partie du jury ; à Francesca Fiorenzi, Pascal Ochem, et Narad Rampersad, mes co-auteurs préférés ; à Pierre, Tomasz, Pierre, Thomas et tous les membres de l'équipe DAC, Jeff, Vincent, Marie, Tarek, Étienne, Laurent, Aurélia, Jean-Bruno, Sophie, Sonia, Pauline et l'équipe poétique, Marc-Hubert, James, Kalle, les gens de Turku, de Montréal, de Liège, de Bordeaux, d'Amiens, de St Jérôme et d'ailleurs, Florence, Dédé, Arnaud Pinguet, Pythéas Fogg, Valérie, Thierry, l'Ecool, Martin, Rachel, Jean, l'AG de Luminy, la CNT, Bibo, Caro, Ju Omega, Christophe, Sibylle, la Béscherelle intelligentia, Julien, Lolo, Clive, Juan, Victoria, Julie, Seb, Zé, Tchoupi, Brice, Alice, mes espagnols, Patricia, Nicole, Annaëlle, Soledad, la smala pipo, Monsieur Fab, les animaux, les gens, et les bêtes à quatre pattes.

Table des matières

Introduction

L'histoire des mots comme sujet d'étude systématique et à part entière commence au début du XXe siècle, quand un mathématicien norvégien, Axel Thue, publie deux articles, [60] et [61], dans lesquels il étudie la combinatoire des suites de symboles et, plus particulièrement, les répétitions dans ces suites. Il y donne la construction d'un mot infini sans carrés (i.e., sans facteurs répétés deux fois consécutivement) sur un alphabet à trois lettres et sur un alphabet à quatre lettres. Il introduit ensuite le mot qu'on appelle aujourd'hui mot de Thue-Morse, construit par substitution itérée, et prouve ainsi qu'il existe un mot binaire sans chevauchements (i.e., sans facteurs de la forme $xyxyx$). Ces travaux restent mal connus pendant longtemps, et ces résultats sont redécouverts indépendemment plusieurs fois, notamment par Morse et Hedlund [40] dans les années 1940. Puis, en 1972, Françoise Dejean publie un article [23] dans lequel elle construit un mot ternaire sans répétitions d'exposant strictement supérieur à $7/4$. Elle définit alors la notion de seuil des répétitions pour un alphabet à k lettres, comme la borne inférieure des exposants évitables, et elle propose la conjecture suivante : le seuil des répétitions pour un alphabet à k lettres est $k/(k-1)$, sauf quand $k = 3$ ou $k = 4$, pour lesquels les seuils valent respectivement $7/4$ et $7/5$. La validation de cette conjecture s'est faite étape par étape, avec les contributions de nombreux auteurs, au cours des quarante dernières années. Elle n'a été terminée que récemment, en 2009 [52, 20]. Elle a joué un rôle phare dans le développement de la problématique, de par les techniques mises en œuvre pour sa résolution, et les nouvelles questions, riches et variées, soulevées par les recherches qu'elle a suscitées. Les travaux présentés dans cette thèse se situent dans cette lignée.

L'étude des répétitions dans les mots à la manière de Thue et Dejean peut être formulée de manière plus générale de la façon suivante. Considérant un alphabet et un ensemble E de mots finis, existe-t-il un mot infini dont aucun facteur n'est dans E ? Et, connaissant une description de cet ensemble E de mots “à éviter” par des paramètres raisonnables, quels sont les paramètres rendant E évitable ? Les répétitions sont de bons candidats pour un ensemble intéressant de mots “à éviter”, et les exposants un paramètre simple et naturel pour décrire cet ensemble. Le seuil des répétitions n'est alors autre que la “frontière” entre l'évitabilité et l'inévitabilité d'un ensemble de mots décrit par des exposants de répétitions.

Signalons que, parallèlement à l'étude des répétitions autour des travaux de Thue et de la conjecture de Dejean, s'est développée celle des motifs évitables, les motifs étant une manière de décrire les mots d'ensembles E “à éviter” comme des blocs de différents types disposés dans un certain ordre. Les répétitions peuvent être vues comme des cas particuliers de motifs. Nous n'adopterons pas ce point de vue dans cette thèse, mais nous signalons sur le sujet le chapitre *unavoidable patterns* de l'ouvrage [38] du collectif M. Lothaire.

Nous aborderons dans cette thèse différents aspects de cette problématique de “frontière” entre évitabilité et inévitabilité, ceci au travers de quatre axes d'étude correspondant chacun à un chapitre annexe : le seuil des répétitions généralisé, la réalisation des réels comme exposants critiques, les mots hautement répétitifs, et le seuil des répétitions des graphes colorés. Chacune des annexes A, B, C et D correspond à un article, publié ou à paraître, co-écrit ou non. Ces articles ont été reproduits tels qu'ils ont été soumis ou publiés. C'est pourquoi on pourra relever certaines redondances et parfois certains conflits de notation et de définition entre les différentes annexes. Nous avons estimé qu'à condition de pointer et expliquer ces contradictions et ces redondances, ce que nous ferons de manière détaillée dans le corps du document, la compréhension n'était pas altérée. Par ailleurs, notre choix de garder une terminologie non unifiée fait écho à l'aspect largement non homogène de la terminologie dans la problématique.

Nous consacrons le premier chapitre à présenter les notions et les résultats classiques dont nous ferons usage dans la suite. En particulier, nous introduisons deux objets, fondamentaux en combinatoire des mots en général et dans nos travaux en particulier, les substitutions et les répétitions. Ces dernières sont abordées, dans un premier temps, du point de vue de la périodicité, puis, du point de vue des α-puissances de Dejean et Brandenburg.

Dans le deuxième chapitre, nous établissons un panorama de problématiques et de résultats concernant l'évitabilité des répétitions. Nous commençons par un rapide état de l'art de la conjecture de Dejean et de sa validation par étapes, en donnant un aperçu des idées et des techniques mises en œuvre. Puis, nous présentons quelques résultats de structure et de dénombrement des mots sans répétitions, parce qu'ils illustrent un autre aspect de la notion de seuil, et qu'ils révèlent le statut particulier des répétitions sur l'alphabet binaire. Enfin, nous traitons de l'évitabilité des répétitions pour lesquelles on tient compte de la longueur de la période, et non plus seulement de l'exposant. Nous définissons le seuil des répétitions généralisé $RT(k, \ell)$, et nous établissons des bornes inférieures et supérieures de sa valeur :

* $RT(2, \ell) \geq 1 + \frac{2}{\ell+2}$,

* $\forall k \geq 2, RT(k, 2) \geq 1 + \frac{1}{1+\lfloor \frac{3}{2}(k-1) \rfloor}$,

* $\forall k \geq 3, \forall \ell \geq 3, RT(k, \ell) \geq 1 + \frac{1}{(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1})k}$,

* $RT(k, \ell) \leq 1 + \frac{2 \ln \ell}{\ell \ln \lambda} + O(\frac{1}{\ell})$, pour k fixé, ℓ tendant vers l'infini, et $\lambda = \frac{(k-1)+\sqrt{(k-1)(k+3)}}{2}$.

Dans le troisième chapitre, nous abordons deux problèmes d'existence de mots soumis à la fois à une contrainte d'évitabilité, et à un forçage de certaines autres répétitions. Le premier de ces problèmes est celui de savoir quels sont les nombres réels réalisables comme exposant critique. Nous obtenons le résultat suivant :

* Tout nombre réel supérieur à $RT(3) = 7/4$ est l'exposant critique d'un mot ternaire infini.

Le second problème consiste à renforcer encore la contrainte de forçage des répétitions, en imposant non plus l'occurrence simple d'une répétition donnée, mais son occurrence une infinité de fois à chaque position dans le mot. Nous proposons pour cela la notion de mot hautement répétitif, et nous montrons plusieurs résultats concernant les mots à la fois hautement répétitifs et d'exposant critique fini, notamment :

* Pour tout réel $\beta < 2$, il existe un mot infini binaire qui est 2^+-free et hautement β-répétitif.

* Pour tout réel $\beta < 7/4$, il existe un mot infini ternaire qui est $7/4^+$-free et hautement β-répétitif, alors qu'il n'existe pas de mot ternaire $7/4^+$-free et hautement $7/4$-répétitif.

* Il existe un mot infini ternaire qui est 2-free et hautement $7/4$-répétitif.

Enfin, dans le dernier chapitre, nous transférons, de manière très naturelle, la problématique du seuil des répétitions aux colorations de graphes. Nous établissons alors la valeur de ce seuil des répétitions pour deux classes de graphes, celle des arbres notée $\mathfrak{T}$, et celle des graphes de grandes subdivisions, notée $\mathfrak{S}$. Plus précisément, nous montrons que :

* $RT(2,\mathfrak{T}) = 7/2$, $RT(3,\mathfrak{T}) = 3$, et $\forall k \geq 4, RT(k,\mathfrak{T}) = 3/2$.

* $RT(2,\mathfrak{S}) = 7/3$, $RT(3,\mathfrak{S}) = 7/4$, et $\forall k \geq 4, RT(k,\mathfrak{S}) = 3/2$.

Chapitre I

Préliminaires

1 Mots finis et infinis

On appelle *lettre* un symbole, par exemple 1, a, ou x, et *alphabet* un ensemble fini et non vide de symboles. Dans la suite, l'alphabet binaire sera noté $B = \{0, 1\}$, et l'alphabet à k lettres, pour $k \geq 3$, sera noté $A_k = \{1, 2, ..., k\}$. Un *mot fini* est la concaténation d'un nombre fini de lettres. Par exemple, 123456, *mot*, et *ababa* sont des mots finis. En particulier, le *mot vide* noté ϵ est le résultat de la concaténation de zéro lettre. Si $w = w_1w_2...w_n$ est un mot fini sur un alphabet Σ, on note $|w|$ sa longueur, soit $|w| = n$. Si a est une lettre dans Σ, on note $|w|_a$ le nombre d'occurences de a dans w. Enfin, on note $\overline{w}$ le mot miroir $w_nw_{n-1}...w_2w_1$ de w. On dit que w est un *palindrome* si $w = \overline{w}$. Par exemple, *rotor* et faf sont des palindromes.

L'ensemble des mots finis sur l'alphabet Σ est noté Σ^*, et l'ensemble des mots de longueur $n \in \mathbb{N}$ est noté Σ^n. Muni de la concaténation, l'ensemble Σ^* est un monoïde libre (de base Σ).

Deux mots u et v dans Σ^* sont dits conjugués s'il existe deux mots x et y tels que $u = xy$ et $v = yx$. La conjugaison induit une relation d'équivalence. La classe d'équivalence d'un mot $u = u_1u_2...u_n$ est l'ensemble des mots $\sigma^k(u)$, pour $0 \leq k \leq n-1$, où σ est l'application

$$\sigma : \begin{cases} \Sigma^* & \to \Sigma^* \\ w_1w_2...w_{|w|} & \mapsto w_2...w_{|w|}w_1. \end{cases}$$

Un *mot infini* sur l'alphabet Σ est une application de $\mathbb{N}^*$ dans Σ. 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 Σ. L'ensemble des mots infinis sur Σ est noté Σ^ω.

Soit $w \in \Sigma^\omega \cup \Sigma^*$ un mot fini ou infini. Un *facteur* de w est un mot fini $w' \in \Sigma^*$ tel que $w = uw'v$, avec $u \in \Sigma^*$ et $v \in \Sigma^\omega \cup \Sigma^*$. Si $u = \epsilon$ (resp. $v = \epsilon$) on dit que w' est un *préfixe* (resp. *suffixe*) de w. L'ensemble des facteurs de w est noté $\mathrm{Fact}(w)$. Une *occurence* du facteur w' dans w est un indice n tel que $w' = w_nw_{n+1}...w_{n+|w'|-1}$.

On peut munir l'ensemble Σ^ω d'une *distance* d définie comme suit : si $x = x_1x_2...$ et $y = y_1y_2...$ sont deux mots infinis,

$$d(x, y) = \begin{cases} 2^{-n}, \text{ où } n = min\{k \geq 1 \mid x_k \neq y_k\} & \text{si } x \neq y, \\ 0 & \text{sinon.} \end{cases}$$

L'ensemble Σ^ω est alors un *espace métrique compact.* La compacité de Σ^ω est un corollaire d'un résultat fondamental connu sous le nom de lemme de König. Ce lemme sera souvent implicitement utilisé. Pour la formulation originale, on pourra se reporter à [35, 36].

Théorème 1 (Lemme de König, [38] chapitre 1)**.** *Soient Σ un alphabet et X un sous-ensemble infini de Σ^*. Il existe un mot $w \in \Sigma^\omega$ tel que tout préfixe de w est le préfixe d'un mot de X.*

Notons $\lim_{n\to\infty} w_n = w$ la convergence de la suite de mots $(w_n)_{n\in\mathbb{N}}$ vers le mot w dans l'espace métrique Σ^ω. Supposons maintenant qu'une suite de mots finis $(v_n)_{n\in\mathbb{N}}$ est telle que chaque v_n est un préfixe propre (i.e., distinct) de v_{n+1}. Par abus de notation (v_n peut être vu comme le mot infini $v_n00...0...$), on dira que $(v_n)_{n\in\mathbb{N}}$ converge vers l'unique mot infini v tel que chaque v_n est préfixe de v, et on notera $\lim_{n\to\infty} v_n = v$.

Par exemple, donnons-nous une suite $(c_i)_{i\geq 0}$ de lettres de l'alphabet binaire : $\forall i \geq 0,\ c_i \in B$. Puis, définissons la suite de mots finis suivante :

$$\begin{array}{rcl} F_0 & = & c_0 \\ F_1 & = & F_0\ c_1\ \rho(\overline{F_0}) \\ & ... & \\ F_n & = & F_{n-1}\ c_n\ \rho(\overline{F_{n-1}}) \\ & ... & \end{array}$$

où ρ est la fonction $0 \mapsto 1$, $1 \mapsto 0$ qui échange les lettres 0 et 1. Alors, $F = \lim_{n\to+\infty} F_n$ est un mot infini, appelé *mot de pliage de papier.*

2 Substitutions

La notion de substitution est une notion fondamentale en combinatoire des mots. C'est en effet un outil simple (des règles de remplacement d'une lettre par un mot fini), et c'est essentiellement le seul, qui permet de construire, à la main, par itération, des mots infinis non triviaux. De plus, construire un mot par l'itération de règles simples a cela d'intéressant qu'elle rend possible un certain contrôle des facteurs, et notamment des répétitions, en "remontant" l'itération, opération qu'on appelle parfois *désubstitution.*

Soient Σ_1 et Σ_2 deux alphabets. On appelle *substitution* un morphisme de monoïde non effaçant, i.e., une application $h : \Sigma_1^* \mapsto \Sigma_2^*$ telle que $h(\epsilon) = \epsilon$, telle que $h(uv) = h(u)h(v)$ pour tous mots u, v dans Σ_1^*, et telle que $h(x) \neq \epsilon$, pour tout mot $x \neq \epsilon$ dans Σ_1^*. Une substitution est entièrement déterminée par les images $h(a)$ des lettres a de Σ_1. On peut alors facilement étendre la définition aux mots infinis, de la façon suivante : si $w = w_1w_2...w_n...$ est dans Σ_1^ω, on définit son image $h(w) \in \Sigma_2^\omega$ par $h(w) = h(w_1)h(w_2)...h(w_n)....$

Une substitution $h : \Sigma^* \to \Sigma^*$ est dite *prolongeable* en la lettre a si $h(a) = ax$, où $x \in \Sigma^* \setminus \{\epsilon\}$. Une telle substitution peut alors être itérée de manière à construire un mot infini appelé *point fixe* de h, noté $h^\omega(a)$, défini par :

$$h^\omega(a) = \lim_{n\to\infty} h^n(a) = \lim_{n\to+\infty} axh(x)h^2(x)...h^{n-1}(x).$$

Un tel mot est aussi appelé *mot purement substitutif.*

Par exemple, la substitution $\mu_F : \{0,1\}^* \mapsto \{0,1\}^*$ définie par

$$\mu_F : \begin{cases} 0 & \mapsto 01 \\ 1 & \mapsto 0 \end{cases}$$

est appelée *substitution de Fibonacci*. Elle est prolongeable en 0 et son point fixe

$$w_F = \mu_F^{\infty}(0) = 0100101001001010010100100101001001...$$

est appelé *mot de Fibonacci*. Cette substitution et son point fixe jouent un rôle fondamental en combinatoire des mots. En particulier, le mot de Fibonacci est *sturmien* (un mot sturmien est un mot qui, pour tout entier n, possède exactement $n+1$ facteurs de longueur n).

Une substitution $h : \Sigma_1^* \mapsto \Sigma_2^*$ est dite *k-uniforme*, avec $k \in \mathbb{N}$, si $|h(a)| = k$ pour toute lettre a de Σ_1. Elle est dite *uniforme* si elle est k-uniforme pour un certain k.

Par exemple, la substitution $\mu_{TM} : \{0,1\}^* \mapsto \{0,1\}^*$ définie par

$$\mu_{TM} : \begin{cases} 0 & \mapsto 01 \\ 1 & \mapsto 10 \end{cases}$$

est appelée *substitution de Thue-Morse*. Elle est 2-uniforme. Elle est prolongeable en 0 et son point fixe

$$w_{TM} = \mu_{TM}^{\infty}(0) = 0110100110010110100101100110 1001...$$

est appelé *mot de Thue-Morse*. Elle joue un rôle majeur dans l'étude des répétitions et en particulier dans cette thèse, en raison de ses bonnes propriétés vis-à-vis des répétitions (voir le chapitre II pour plus de détails). Une autre substitution uniforme qui aura une grande importance dans la suite, est la substitution qu'on appelle *substitution de Dejean*, donnée par :

$$\mu_D : \begin{cases} a & \mapsto abcacbcabcbacbcacba \\ b & \mapsto bcabacabcacbacabacb \\ c & \mapsto cabcbabcabacbabcbac \end{cases}$$

Dans les problèmes d'évitabilité des répétitions, ou plus généralement quand on cherche à construire des mots infinis ayant certaines propriétés sur leurs facteurs, on cherche souvent à itérer de "bonnes" substitutions, c'est-à-dire des substitutions qui permettent de gérer l'évolution des facteurs lors de l'itération.

Une substitution $h : \Sigma_1^* \to \Sigma_2^*$ est dite *comma-free* si $\forall a, b, c \in \Sigma_1$, et $\forall s, r \in \Sigma_2^*$, si $h(ab) = sh(c)r$, alors soit $r = \epsilon$ et $b = c$, soit $s = \epsilon$ et $a = c$. Par exemple, pour un entier $k \geq 3$, la substitution

$$\sigma : \begin{cases} \Sigma_k^* & \to B^* \\ i & \mapsto 0^{k-i+1}1^i, \end{cases}$$

est comma-free. En effet, le mot 10 n'a pas d'occurrence dans $\sigma(\Sigma_k)$, alors qu'il est toujours facteur des mots de la forme $\sigma(i)\sigma(j)$, et par ailleurs, σ est $(k+1)$-uniforme.

Une substitution prolongeable $g : \Sigma^* \to \Sigma^*$ est dite *synchonisante* s'il existe un entier $L > 0$ tel que pour tout facteur u du point fixe w de g, si $|u| \geq L$, alors il existe deux uniques lettres x, y de Σ, et trois uniques mots s, p et v de Σ^*, tels que $u = sg(v)p$, où $s \neq \epsilon$ est un suffixe de $g(x)$ et $p \neq \epsilon$ est un préfixe de $g(y)$. Par exemple, la substitution μ_{TM} de Thue-Morse est synchronisante avec $L = 4$ (implicite par exemple dans le lemme 8 de [52]). Par ailleurs, on montre dans l'annexe B que la substitution de Dejean μ_D est synchronisante, avec $L = 18$ (en fait, on montre même une propriété plus forte). Ces deux exemples illustrent bien l'importance des propriétés de synchronisation dans de nombreux résultats concernant les répétitions.

Pour d'autres définitions et résultats de désubstitution (par exemple, découpage, reconnaissabilité, détermination, etc, des substitutions), on pourra se reporter à [41], [42], et [10].

3 Périodicité et répétitions

Un mot infini w est dit *périodique* s'il existe un mot fini non vide x tel que

$$w = x^\omega = xx...x...,$$

il est dit *ultimement périodique* s'il existe deux mots finis x et y tels que $w = xy^\omega$. Un mot infini non ultimement périodique est dit *apériodique*.

Un mot infini w est dit *récurrent* si tout facteur de w possède une infinité d'occurences dans w. Si de plus, pour tout facteur u de w, il existe un entier n tel que tout facteur de w de longueur n contient une occurrence de u, alors w est dit *uniformément récurrent*.

Si l'on veut parler de périodicité pour un mot fini, on peut définir assez naturellement un mot fini périodique comme un mot qui s'écrit sous la forme x^n, pour un entier n. Cependant, cette définition est très restrictive : tous les facteurs d'un mot infini périodique ne sont pas eux-mêmes périodiques, seuls le sont ceux dont la longueur est multiple de la longueur de la période. Afin de disposer d'une notion de périodicité stable par passage aux facteurs, on préfère adapter la notion de période aux mots finis en "coupant" les mots périodiques infinis, comme suit. Soit $w = x^\omega$ un mot infini, périodique, de période le mot fini x. Tout préfixe de w peut s'écrire sous la forme $x^n y$, où y est un préfixe de x. Ceci nous amène à la définition suivante :

Définition 1. *Un mot fini v est une répétition de période p s'il est préfixe du mot infini p^ω. Le rationnel $\frac{|v|}{|p|}$ est appelé exposant de v, et le préfixe e de p tel que $v = p^{\lfloor \frac{|v|}{|p|} \rfloor} e$ est appelé excès de v.*

Par exemple, le mot français *entente* est préfixe du mot $(ent)^\omega$. C'est une répétition de période *ent* et d'exposant $7/3$. Il est aussi préfixe du mot $(entent)^\omega$, et c'est donc, aussi, une répétition de période *entent* et d'exposant $7/6$. En fait, on remarque qu'un même mot fini peut posséder différentes périodes et différents exposants. De plus, il est clair qu'un mot fini v est toujours une répétition de période lui-même et d'exposant 1, car il est préfixe de v^ω.

On appelle *carré* une répétition d'exposant 2, *cube* une répétition d'exposant 3, et *chevauchement* une répétition d'exposant strictement supérieur à 2.

Une autre vision, qui mène à des définitions équivalentes, a été proposée par Brandenburg et Dejean. Elle est basée sur l'observation que tout mot fini peut s'écrire sous la forme $x^n x'$, où x est un mot fini, x' un préfixe de x, n un entier non nul.

Définition 2. *Soit α un rationnel. Un mot fini v est une α-puissance s'il existe un mot fini x, un préfixe x' de x, et un entier $n \neq 0$ tels que $v = x^n x'$ et $\alpha = n + \frac{|x'|}{|x|}$. Les mots x et x' sont appelés respectivement période et excès de v.*

Par exemple, *ionisation* est une $10/7$-puissance de période *ionisat* et d'excès *ion*.

Les deux définitions 1 et 2 sont clairement équivalentes, et il y a redondance de la terminologie. Une répétition v d'exposant α, de période p et d'excès e selon la définition 1, est une α-puissance de période p et d'excès e selon la définition 2.

Enfin, dans de nombreux contextes, et notamment dans les annexes B et C, est utilisée une notion plus forte de répétition. Il peut en effet être pertinent de disposer d'un triplet de paramètres (exposant, période et excès) unique pour une répétition donnée. Pour cela, on décide de privilégier une période parmi les différentes possibilités. De façon plutôt intuitive par rapport aux problèmes qui nous occuperont par la suite, qui consisteront à éviter les grands exposants, on choisit la période la plus courte (et donc l'exposant le plus grand).

Définition 3. *Soit v un mot fini. Il peut se factoriser de manière unique sous la forme suivante :*

$$v = p^k e,$$

où $k \geq 1$ est un entier, e est un préfixe de p, et où la longueur $|p|$ de p est minimale. On dit alors que v est la répétition de période (unique) p, d'excès (unique) e et d'exposant (unique) $E(v) = \frac{|v|}{|p|}$.

Ainsi, selon cette dernière définition, *entente* est la répétition de période *ent* et d'exposant 7/3. En fait, étant donné un mot fini v, sa période (resp. son exposant) au sens de la définition 3 est le plus courte de ses périodes (resp. le maximum de ses exposants) au sens des définitions 1 ou 2.

Continuons maintenant nos allers retours entre mots infinis et mots finis. La périodicité est une propriété des mots infinis. Adaptée aux mots finis, elle mène à la définition d'un nouvel objet, les répétitions, dont une des caractéristiques principales est l'exposant. Qu'en est-il si on essaie d'adapter cette notion d'exposant aux mots infinis ? On en arrive à la définition suivante.

Définition 4. *Soit w un mot infini. L'exposant critique de w, noté $E_c(w)$, est le réel :*

$$E_c(w) = \sup\{r \in \mathbb{Q} \mid \exists u \in \mathrm{Fact}(w),\ u \text{ est une } r-\text{puissance}\}.$$

Clairement, tout mot périodique a pour exposant critique $+\infty$. Il en est de même pour le mot $01011013...01^n...$, qui contient la n-puissance 1^n pour tout entier n. L'exposant critique du mot de Fibonacci, calculé par Mignosi et Pirillo [39], est $2 + \Phi$, où $\Phi = \frac{1+\sqrt{5}}{2}$ est le nombre d'or. L'exposant critique du mot de Thue-Morse est 2 d'après un résultat de Thue [61] (voir le chapitre II).

4 Bibliographie généraliste

Terminons ce chapitre préliminaire en signalant quelques ouvrages et articles plus ou moins généraux sur la combinatoire des mots et les notions que nous venons d'aborder :

* M. Lothaire, “Combinatorics on Words”, *Encyclopedia of Mathematics* Vol. 17, Addison-Wesley, 1983.

* M. Lothaire, “Algebraic Combinatorics on Words”, Cambridge University Press, 2002.

* M. Lothaire, “Applied Combinatorics on Words”, Cambridge University Press, 2005.

* N. Pytheas Fogg, “Substitutions in Dynamics, Arithmetics and Combinatorics”, *Lecture Notes in Mathematics* 1794, Springer-Verlag, 2002.

* C. Choffrut and J. Karhumäki, “Combinatorics on words”, *Handbook of Formal Languages*, Springer-Verlag, 1997, 329–438.

* J. Berstel and J. Karhumäki, “Combinatorics on words, a tutorial”, *Bull. Eur. Assoc. Theor. Comput. Sci. (EATCS)* 79 :178–228, 2003.

* J. Berstel, D. Perrin, “The origin of combinatorics on words”, *European J. Comb.* 28 (2007) 996–1022.

* J. Berstel, A. Lauve, C. Reutenauer, F. V. Saliola, “Combinatorics on Words : Christoffel Words and Repetitions in Words”, *American Mathematical Society*, 2009.

Chapitre II

Éviter les répétitions : de la recherche des mots sans carrés aux seuils des répétitions

L'évitabilité (de répétitions, de motifs) compte parmi les problématiques à la fois les plus anciennes et les plus actives, en combinatoire des mots et plus généralement en théorie des languages formels. Typiquement, la question est la suivante. Étant donnés un alphabet Σ et un ensemble E de mots finis, est-il possible de construire un mot infini sur l'alphabet Σ qui ne contienne comme facteur aucun des mots de E ? Ou bien l'ensemble E est-t-il *inévitable*, c'est-à-dire que tout mot suffisamment long sur Σ possède un facteur dans E ? Par exemple, sur l'alphabet binaire $\{0,1\}$, l'ensemble $\{0,11\}$ est clairement inévitable.

Ce cadre large d'étude des ensembles de mots évitables ou inévitables est riche en résultats et en problèmes remarquables. Un aperçu en est donné dans le premier chapitre du livre [38] du collectif M. Lothaire. Dans ce qui suit, nous nous restreindrons au cas où l'ensemble E est infini, et même plus, au cas où E est un ensemble de répétitions. Signalons que dans le cas plus général où E est infini mais pas forcément un ensemble de répétitions, le concept de motif est une façon d'en avoir une description finie, et ainsi de travailler avec des ensembles décidables. Pour plus de détails sur l'étude des motifs évitables, qui est une autre branche fructueuse du domaine, on pourra se reporter au quatrième chapitre de [38].

1 La conjecture de Dejean

Dans deux articles publiés en 1906 [60] et 1912 [61], le mathématicien norvégien Axel Thue pose le problème de l'existence d'un mot infini sans carrés. Il donne la construction de tels mots sur l'alphabet à 3 lettres, et sur l'alphabet à 4 lettres. Il constate de plus que sur l'alphabet binaire, un tel mot n'existe pas, et il construit alors un mot sans chevauchements : c'est le mot de Thue-Morse. L'étude et la résolution de ces problèmes marquèrent les débuts d'une problématique plus générale et intensivement étudiée, celle des mots sans répétitions, avec comme fil directeur la conjecture de Dejean, proposée en 1972 [23] et résolue, étape par étape, jusqu'en 2009.

Definition 1. *Soit α un réel. Un mot, fini ou infini, est dit α-free (resp. α^+-free) si aucun de ses facteurs n'est une β-puissance, pour un rationnel $\beta \geq \alpha$ (resp. $\beta > \alpha$).*

Théorème 2 (Thue, [61])**.** *Le mot de Thue-Morse*

$$w_{TM} = 0110100110010110100101100110100110010 1...$$

est 2^+*-free.*

Ce mot est optimal dans le sens où, clairement, tout mot binaire de longueur ≥ 4 contient un carré.

Proposition 1 (Thue, [60])**.** *Il existe un mot ternaire* 2*-free.*

Cette fois, ce résultat n'est pas optimal. En effet,

Théorème 3 (Dejean, [23])**.** *Il existe un mot ternaire* $7/4^+$*-free.*

Dejean pose alors le problème suivant : pour un entier donné k, quelle est la borne inférieure des réels α tels qu'il existe un mot infini α-free sur l'alphabet à k lettres A_k ?

Définition 5. *Soit k un entier. On appelle seuil des répétitions de l'alphabet à k lettres, et on note $RT(k)$, la borne inférieure des réels α tels qu'il existe un mot infini α-free sur l'alphabet A_k.*

D'après les travaux de Thue, il est clair que $RT(2) = 2$. Dejean, elle, montre que $RT(3) = 7/4$, d'après le théorème 3, et en vérifiant informatiquement que par ailleurs, tout mot ternaire suffisament long contient des répétitions d'exposant $7/4$. Puis, toujours informatiquement, elle observe que sur l'alphabet A_4, tout mot suffisament long contient une répétition d'exposant supérieur à $7/5$. Elle conjecture alors le résultat suivant, connu sous le nom de *conjecture de Dejean* :

$$\forall k \geq 2, RT(k) = r_k,$$

où $(r_k)_{k \geq 2}$ est la suite de rationnels définie par :

$$r_k = \begin{cases} 2 & \text{si } k = 2 \\ \frac{7}{4} & \text{si } k = 3 \\ \frac{7}{5} & \text{si } k = 4 \\ \frac{k}{k-1} & \text{sinon.} \end{cases}$$

Le théorème suivant, complété par les preuves informatiques de Dejean, établit de façon efficace et élégante que pour tout entier $k \geq 2$, $RT(k) \geq r_k$.

Théorème 4 (Moulin Ollagnier, [43])**.** *Pour tout entier $k \geq 5$, tout mot sur l'alphabet A_k de longueur $k+2$ possède une répétition d'exposant au moins $\frac{k}{k-1}$.*

Moulin Ollagnier ayant ainsi réglé un sens de la conjecture, c'est sur la preuve du fait que $RT(k) \leq r_k$ que ce sont concentrées les recherches après 1984. Or, montrer que $RT(k) \leq r_k$ signifie prouver, de façon constructive ou non, l'existence d'un certain mot infini, en l'occurence un mot infini sans répétitions. Un outil fondamental et naturel dans les preuves de ces résultats a été l'utilisation de substitutions particulières, qui se comportent bien vis-à-vis des répétitions.

Définition 6. *Soit α un réel. Une substitution $h : \Sigma_1^* \to \Sigma_2^*$ est dite α-free (resp. α^+-free) si elle vérifie : pour tout mot $w \in \Sigma_1^*$,*

$$h(w) \text{ est } \alpha\text{-free (resp. } \alpha^+\text{-free)} \Leftrightarrow w \text{ est } \alpha\text{-free (resp. } \alpha^+\text{-free).}$$

Le point fixe d'une substitution α-free est clairement un mot α-free. D'où l'importance de telles substitutions quand on cherche à construire des mots sans répétitions.

Proposition 2 (Brandenburg, [7]). *La substitution de Thue-Morse* $\mu_{TM} : 0 \mapsto 01,\ 1 \mapsto 10$*, est* 2^+*-free.*

De plus, la substitution de Thue-Morse est “l’unique” substitution 2^+-free, dans le sens où toute substitution binaire 2^+-free est de la forme $\mu_{TM}^n \circ \sigma$, avec n un entier, et σ la permutation $0 \mapsto 1, 1 \mapsto 0$ (voir [61]).

Dans sa preuve du fait que $RT(3) = 7/4$, Dejean construit un mot ternaire $7/4^+$-free, elle aussi, comme point fixe d’une substitution, la substitution μ_D décrite dans le chapitre 1 et définie par :

$$\mu_D : \begin{cases} a & \mapsto abcacbcabcbacbcacba \\ b & \mapsto bcabacabcacbacabacb \\ c & \mapsto cabcbabcabacbabcbac \end{cases}.$$

Proposition 3 (Dejean, [23]). *La substitution de Dejean* μ_D *est* $7/4^+$*-free.*

Enfin, Brandenburg établit le résultat suivant, fondamental dans la compréhension du problème des seuils des répétitions.

Proposition 4 (Brandenburg, [7]). *Il n’existe pas de substitution* $7/5^+$*-free sur l’alphabet à* 4 *lettres* A_4*. De plus, pour tout entier* $k \geq 5$*, il n’existe pas de substitution* $(\frac{k}{k-1})^+$*-free sur l’alphabet à* k *lettres* A_k*.*

Ainsi, pour prouver la conjecture pour $k \geq 4$, il n’est plus possible de construire des mots r_k^+-free par itération d’un morphisme r_k^+-free. À partir de ce constat, nous donnerons dans la suite de cette partie un aperçu des travaux des différents auteurs qui ont mené à la preuve complète de la conjecture. Nous avons choisi de parcourir ces résultats et ces méthodes dans l’ordre chronologique de leur publication, afin de rendre compte de l’évolution des idées et ainsi de mieux souligner la cohésion entre elles, malgré leur diversité et une apparente indépendance.

Pour résoudre le cas de l’alphabet à 4 lettres, Pansiot [47] se ramène à un raisonnement sur l’alphabet binaire grâce à un codage, le *codage de Pansiot*. En effet, il remarque qu’un mot w qui est $7/4^+$-free sur l’alphabet A_3, $7/5^+$-free sur l’alphabet A_4, ou $\frac{k}{k-1}^+$-free sur l’alphabet $A_k, k \geq 5$, est tel que chacun de ses facteurs de longueur $k-1$ ont toutes leurs lettres distinctes, et donc que tout facteur $w_i w_{i+1} ... w_{i+k-2}$ de w se prolonge à droite de seulement deux façons, soit par w_i, soit par la seule lettre de l’alphabet A_k qui n’apparaît pas dans $w_i w_{i+1} ... w_{i+k-2}$. On peut donc coder ce choix binaire. Le *code de Pansiot* du mot w est le mot binaire $P_k(w)$ tel que pour tout indice $1 \leq j \leq |w| - k + 1$ ($j \geq 1$ si w est infini), la j-ième lettre de $P_k(w)$ est

$$P_k(w)[j] = \begin{cases} 0 \text{ si } w_{k-1+j} = w_j \\ 1 \text{ sinon.} \end{cases}$$

Ainsi, par exemple, le code de Pansiot de $w_1 = 1234516243 \in A_6^*$ est $P_6(w_1) = 01101$, et celui de $w_2 = 312123212132 \in A_3^*$ est $P_3(w_2) = 1001010011$. On voit facilement que w est défini de façon unique par son préfixe $w_1 w_2 ... w_{k-1}$ de longueur $k-1$ et par son code de Pansiot $P_k(w)$. On peut aussi définir l’opération inverse du codage :

$$M_k(v)[j] = \begin{cases} j \text{ si } j < k \\ M_k(v)[j-k+1] \text{ si } j \geq k \text{ et si } v[j-k+1] = 0 \\ x, \text{ où } A_k = \{x, M_k(v)[j-k+1], ..., M_k(v)[j-1]\}, \text{ sinon.} \end{cases}$$

Par exemple, le mot binaire $v = 01100101$ est le code de Pansiot sur A_4 du mot $M_4(v) = 12314214312$, et sur A_5 du mot $M_5(v) = 123415241325$. Pansiot montre alors que le mot $M_4(\mu^\omega(1))$, où μ est la

substitution binaire
$$\mu : \begin{cases} 0 & \mapsto 101101 \\ 1 & \mapsto 10, \end{cases}$$
est $7/5^+$-free, ce qui résoud la conjecture de Dejean pour $k = 4$. Moulin Ollagnier [43] améliore alors la méthode de Pansiot, en faisant l'observation essentielle suivante : le codage de Pansiot P_k correspond à une action sur le groupe symétrique S_k. Ceci l'amène à introduire une distinction entre deux types de répétitions. Les répétitions dites *courtes* sont celles dont l'excès est de longueur strictement inférieure à $k - 1$, et les autres, dites *répétitions noyau*, correspondent dans le code binaire aux éléments du noyau de l'action sur le groupe symétrique. Moulin Ollagnier donne des conditions suffisantes et décidables pour qu'une substitution donnée génère un point fixe évitant les répétitions noyau, ce qui lui permet de se ramener à une recherche systématique des exposants des "petits" facteurs, correspondant aux répétitions courtes. Il résoud ainsi la conjecture pour $5 \leq k \leq 11$, en explicitant, pour chaque k, une substitution binaire h_k dont le point fixe est le codage de Pansiot d'un mot r_k^+-free :

$$h_5 : \begin{cases} 0 \mapsto 010101101101010110110 \\ 1 \mapsto 101010101101101101101 \end{cases}$$
$$h_6 : \begin{cases} 0 \mapsto 010101101101011010110 \\ 1 \mapsto 101011010110110101101 \end{cases}$$
$$h_7 : \begin{cases} 0 \mapsto 0110110110110101101101101010 \\ 1 \mapsto 1010110110110101101101101101 \end{cases}$$
$$h_8 : \begin{cases} 0 \mapsto 1011010101101011010110101010 \\ 1 \mapsto 1011010101011011011010101101 \end{cases}$$
$$h_9 : \begin{cases} 0 \mapsto 101011010110110101011010110110101010 \\ 1 \mapsto 101010110110110101011011010110101101 \end{cases}$$
$$h_{10} : \begin{cases} 0 \mapsto 1010101011011011011010101011011011010110 \\ 1 \mapsto 1010101011011011010110101010101011010101 \end{cases}$$
$$h_{11} : \begin{cases} 0 \mapsto 1010101010101101101011010110101010110110 \\ 1 \mapsto 1010101010101011011011011011011010101101 \end{cases}$$

Ces idées sont ensuite adaptées par Mohammad-Noori et Currie [16] aux mots sturmiens, qui étendent ainsi la validité de la conjecture aux cas $12 \leq k \leq 14$. Plus important, ils proposent de construire des mots de telle façon que la gestion des répétitions courtes et celle des répétitions noyau peuvent se faire de façons séparées et indépendantes : les mots r_k^+-free qu'ils donnent sont de la forme
$$M_k(g(h^\omega(0))),$$
où g est une substitution telle que $g(B^*)$ évite (après codage de Pansiot) les répétitions courtes d'exposant strictement supérieur à r_k, et h est choisie de manière à éliminer les répétitions noyau.

Puis, en 2007, Carpi [11] va plus loin dans cette direction, en cherchant des mots de la forme
$$M_k(f(w_m)),$$
où w_m est un mot sur A_m qui permet d'éliminer les répétitions noyaux (avec une construction de type pliage de papier), et où f est telle que $f(A_m^*)$ évite les répétitions courtes. Sa constuction d'un mot w_5 sur A_5 permet une avancée capitale dans la résolution de la conjecture, puisque celle-ci est dorénavant validée pour les grands alphabets, plus précisément pour tous les $k \geq 33$. C'est par ailleurs la première

méthode proposée qui n'est pas basée sur une recherche calculatoire par machine, toutes les méthodes précédentes consistant en fait à réduire le problème à un certain nombre de conditions décidables, et à en trouver des solutions par calculs (d'où bien sûr l'impossiblité, par ce type de méthode, de résoudre plus qu'un nombre fini de cas).

À partir de là, reste à combler le fossé entre $k = 14$ et $k = 33$. Il le sera, entre 2008 et 2009, essentiellement de deux types de façons, d'une part par des adaptations de la méthode de type Moulin Ollagnier, constructive et nécessitant des calculs sur machine, d'autre part par des améliorations de la construction de Carpi. Ainsi, tout d'abord, Currie et Rampersad étendent les résultats de Carpi, en trouvant un mot w_4 sur A_4 évitant les répétitions noyau, et en prouvant que $f(A_4^*)$ évite toujours les répétitions courtes. Ceci valide la conjecture pour $k \geq 30$ [17], et pour $k \geq 27$ [19]. Enfin, les mêmes auteurs utilisent et simplifient la méthode de recherche par machine de Moulin Ollagnier, pour trouver des substitutions binaires h_k, pour $15 \leq k \leq 26$, telles que $M_k(h_k^\omega(0))$ soit r_k^+-free, ce qui termine la preuve de la validité de la conjecture. Simultanément, Rao [52] établit une autre preuve des derniers cas de la conjecture, en revenant lui aussi à la méthode de Moulin Ollagnier, mais en l'améliorant d'une autre manière. Son idée est de réduire et diviser encore la gestion des répétitions, en catégorisant les répétitions noyau en deux nouveaux types de répétitions, les *répétitions noyau courtes* et les *répétitions noyau longues*. Il cherche alors des mots sous la forme

$$M_k(h_k(w_{TM})),$$

où w_{TM} est le mot de Thue-Morse, et h_k est une substitution bien choisie. L'utilisation du mot de Thue-Morse et de ses propriétés de synchronisation permettent de réduire les contraintes d'évitabilité à un ensemble de conditions décidables. Il donne alors des substitutions h_k pour $k = 4$ et $8 \leq k \leq 34$. D'où, finalement, le théorème :

Théorème 5 (Carpi, Currie, Dejean, Mohammad-Noori, Moulin-Ollagnier, Pansiot, Rampersad, Rao)**.** *Le seuil des répétitions de l'alphabet à k lettres vaut*

$$RT(k) = \begin{cases} 2 & \text{si } k = 2 \\ \frac{7}{4} & \text{si } k = 3 \\ \frac{7}{5} & \text{si } k = 4 \\ \frac{k}{k-1} & \text{sinon.} \end{cases}$$

Presque quarante années de recherches, une petite dizaine d'auteurs, une bonne dizaine d'articles, sans compter les travaux qui n'ont pas mené à publication, la résolution de la conjecture de Dejean a été le fruit de travaux riches et collectifs. Ils ont mené à une compréhension profonde des problématiques en jeu et de la structure des objets, à la mise en place d'idées et de techniques novatrices, élégantes et efficaces, et à l'ouverture à de nombreuses questions nouvelles.

Les méthodes de résolution de la conjecture, et plus spécifiquement les constructions de mots $RT(k)^+$-free sur A_k, restent morcelées, même si on y voit très clairement des liens et des idées communes. Unifier ces preuves par une construction valable pour tout alphabet serait une belle conclusion de l'histoire. Ainsi, Rao conjecture que sa construction fonctionne pour tout alphabet de taille supérieure à 8 (dans les cas où $k = 2, 3, 5, 6$ ou 7, il montre que le mot de Thue-Morse n'élimine pas les répétitions noyau longues) :

Conjecture 1. *Pour tout $k \geq 8$, il existe une substitution h_k telle que le mot $M_k(h_k(w_{TM})) \in A_k^\omega$ soit $RT(k)^+$-free.*

Une autre piste possible serait une construction à la Carpi, qui, comme nous l'avons vu, est la seule ne nécessitant pas de recherche calculatoire par machine.

Terminons cette section par un résumé des étapes de la preuve de la conjecture de Dejean, schématisé dans la figure II.1, et en donnant la liste, dans l'ordre chronologique, des articles correspondant à chacune de ces étapes.

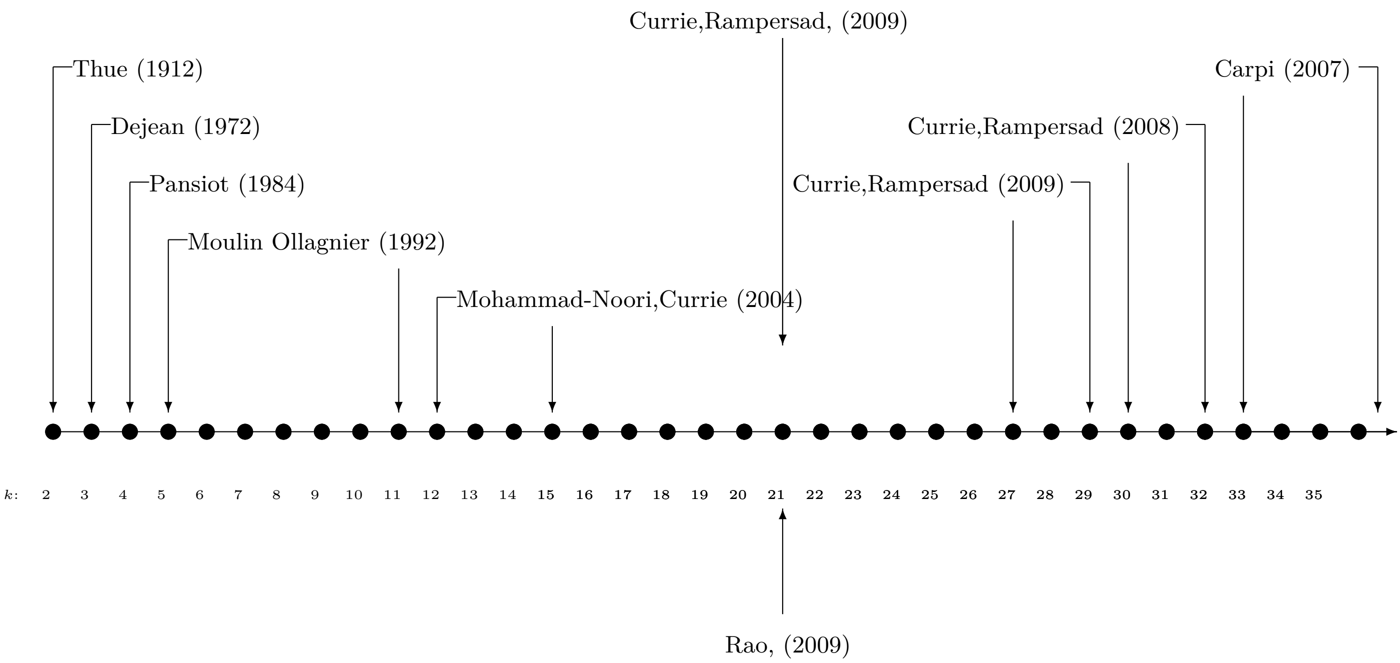


Figure II.1 – Historique de la conjecture de Dejean et de sa résolution

* A. Thue, “Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen”, *Norske Vid. Selsk. Skr. Mat. Nat. Kl.*, 10 (1912), 1–67.
  Pour une traduction, on se reportera à J. Berstel, “Axel Thue's papers on repetitions in words : a translation”, *Publication du LaCIM*, 20, 1995.

* F. Dejean, “Sur un théorème de Thue”, *J. Combin. Theory Ser. A* 13 (1972).

* J.-J. Pansiot, “A propos d'une conjecture de F. Dejean sur les répétitions dans les mots”, *Discrete Appl. Math.* 7 (1984).

* J. Moulin-Ollagnier, “Proof of Dejean's conjecture for alphabets with 5,6,7,8,9,10 and 11 letters”, *Theor. Comput. Sci.* 95 (1992).

* J. D. Currie et M. Mohammad-Noori, “Dejean's conjecture and Sturmian words”, *European J. Combin.* 28 (2007).

* A. Carpi, “On Dejean's conjecture over large alphabets”. *Theor. Comput. Sci.* 385 (2007), 137–151.

* J. D. Currie et N. Rampersad, “Dejean’s conjecture holds for $n \geq 30$”, *Theor. Comput. Sci.* 410 (2009), 2885–2888.

* J. D. Currie et N. Rampersad, “Dejean’s conjecture holds for $n \geq 27$”, *Theor. Inform. Appl.* 43 (2009), 775–778.

* J. D. Currie et N. Rampersad, “A proof of Dejean’s conjecture”, *Math. Comp.*, à paraître.

* M. Rao, “Last cases of Dejean’s conjecture”, *WORDS’09*.

2 Combinatoire et dénombrement des mots sans répétitions

En 1983, Brandenburg [7] montre que le nombre de mots binaires 3-free de longueur n est à croissance exponentielle. Puis, Restivo et Salemi [53, 54] prouvent que le nombre de mots binaires 2^+-free de longueur n est à croissance polynomiale. Pour des bornes de cette croissance, voir [5]. La question naturelle qui se pose alors est de savoir quel est l’exposant α pour lequel le nombre de mots binaires α-free passe d’une croissance polynomiale à une croissance exponentielle. Karhumäki et Shallit y répondent :

Théorème 6 (Karhumäki et Shallit, [29])**.** *Il y a un nombre polynomial de mots binaires $7/3$-free de longueur n, et un nombre exponentiel de mots binaires $7/3^+$-free de longueur n.*

Pour des alphabets plus gros, Ochem, puis Kolpakov et Rao ont prouvé que la croissance était toujours exponentielle :

Théorème 7 (Ochem, [45])**.** *Pour $k = 3$ et $k = 4$, il y a un nombre exponentiel de mots $RT(k)^+$-free de longueur n sur A_k.*

Théorème 8 (Kolpakov et Rao, [30])**.** *Soit $5 \leq k \leq 10$. Il y a un nombre exponentiel de mots $RT(k)^+$-free de longueur n sur A_k.*

L’alphabet binaire a donc un statut particulier quant au nombre de ses mots sans répétitions. Ceci s’explique au vu de la structure de ces mots. En effet, Restivo et Salemi [53, 54] prouvent que tout mot binaire 2^+-free est l’image, quitte à enlever un préfixe et un suffixe de longueur bornée (et petite), d’un autre mot 2^+-free par la substitution μ_{TM} de Thue-Morse. Karhumäki et Shallit [29] généralisent ce théorème de factorisation :

Théorème 9 (Karhumäki et Shallit, [29])**.** *Soit $x \in \{0,1\}^*$ un mot α-free, avec $2 < \alpha \leq 7/3$. Il existe $u, v \in \{\epsilon, 0, 1, 00, 11\}$ et un mot $y \in \{0,1\}^*$ α-free tels que $x = u\mu_{TM}(y)v$.*

Cette structure particulière des mots binaires α-free, pour $2 < \alpha \leq 7/3$, a d’autres conséquences que le résultat de dénombrement du théorème 6. Par exemple, dans la section 2 du chapitre III, on verra aussi apparaître un seuil à $7/3$ pour l’alphabet binaire, conséquence du théorème 9. On fera bien sûr le lien avec le fait que la substitution μ_{TM} de Thue-Morse est la “seule” substitution 2^+-free de l’alphabet binaire (voir la partie précédente), ainsi qu’avec la structure des carrés binaires sans chevauchements (voir la proposition 10) et des carrés binaires sans 7/3-puissances (voir la proposition 11).

Enfin, signalons une autre question relative à la structure des mots sans répétitions, et plus spécifiquement à la fréquence des lettres dans les mots sans répétitions. En effet, il est facile de voir, d’après le codage de Pansiot par exemple, que dans tout mot $(\frac{k}{k-1})^+$-free sur l’alphabet A_k, chaque lettre a une fréquence comprise entre $\frac{1}{k+1}$ et $\frac{1}{k-1}$. En se basant sur des observations numériques, Ochem [44] propose une version plus forte de la conjecture de Dejean :

Conjecture 2.

- *Pour tout* $k \geq 5$*, il existe un mot infini* $(\frac{k}{k-1})^+$*-free sur l'alphabet à* k *lettres, dont la fréquence de chaque lettre est* $\frac{1}{k+1}$.

- *Pour tout* $k \geq 6$*, il existe un mot infini* $(\frac{k}{k-1})^+$*-free sur l'alphabet à* k *lettres, dont la fréquence de chaque lettre est* $\frac{1}{k-1}$.

Chalopin et Ochem prouvent deux cas de cette conjecture :

Proposition 5 (Chalopin et Ochem, [13])**.** *Il existe un mot infini* $\frac{5}{4}^+$*-free sur* A_5 *dont chaque lettre a une fréquence* $1/6$.

Proposition 6 (Chalopin et Ochem, [13])**.** *Il existe un mot infini* $\frac{6}{5}^+$*-free sur* A_5 *dont chaque lettre a une fréquence* $1/5$.

Rao [52] observe qu'en adaptant la construction de mots $(\frac{k}{k-1})^+$-free de la forme $M_k(h_k(w_{TM}))$, qu'il propose pour résoudre la conjecture de Dejean, on peut générer des mots $(\frac{k}{k-1})^+$-free dont les fréquences des lettres sont forcées. Pour cela, il remarque qu'il suffit de chercher des substitutions h_k sous une certaine forme, ce qui lui permet de prouver la conjecture de Ochem pour $9 \leq k \leq 38$.

3 Les répétitions longues et le seuil des répétitions généralisé

Une répétition sur un alphabet donné est caractérisée entièrement (en tant que motif, pas en tant que mot réalisé sur un alphabet) par deux nombres : son exposant, et la longueur de sa période. Les problématiques liées au seuil des répétitions de Dejean consistent à construire, si possible, des mots infinis évitant un ensemble de facteurs décrits par leurs exposants. L'idée maintenant est de prendre en compte la longueur de la période dans la description des facteurs à éviter.

Sur l'alphabet binaire, tout mot suffisamment long contient un carré. Partant de ce fait, de nombreux auteurs se sont intéressés à ces carrés en terme de longueur : même s'ils sont inévitables dans leur ensemble, une partie d'entre eux peut-elle être évitable ? Entringer, Jackson et Schatz [25] construisent en 1974 un mot binaire sans carrés xx, avec $x \geq 3$, et prouvent que la borne 3 comme longueur de période est optimale :

Théorème 10 (Entringer, Jackson et Schatz, [25])**.** *Il existe un mot binaire infini sans carrés de période de longueur supérieure à* 3*. Par ailleurs, tout mot binaire de longueur* ≥ 18 *contient un carré dont la période est de longueur supérieure à* 2.

Allouche et Bousquet-Mélou [2] prouvent que les mots de pliage de papier (donc binaires) ont tous leurs carrés de période de longueur 1, 3 ou 5. Fraenkel et Simpson [26] prouvent un résultat plus fort, en construisant un mot infini binaire sans carrés exceptés 00, 11, et 0101. Rampersad, Shallit et Wang [51] donnent une construction plus simple, en utilisant un morphisme uniforme, d'un mot vérifiant ces mêmes propriétés. Cette constuction leur permet même d'engendrer en nombre exponentiels des mots de taille n avec ces propriétés.

Dans le cas de l'alphabet ternaire, Pansiot, dans la conclusion de [47], avait fait la remarque suivante concernant son mot $7/5^+$-free (qu'il note N) :

"Il est à noter que la notion de seuil de répétition est très locale puisque par exemple le mot N *ne contient pas de* $7/5$*-répétitions de longueur supérieure à* 14*. Des résultats analogues ont été obtenus*

pour les carrés dans les mots sur un alphabet à 2 lettres. Il pourrait donc être intéressant d'étudier les répétitions u^t en fonction non seulement de [leur exposant] t et du nombre de lettres, mais aussi de la longueur de [la période] u."

Poursuivant alors dans cette problématique, Ilie, Shallit et Ochem [28] introduisent en 2005 une généralisation du seuil des répétitions.

Definition 2. *Soit k un entier et A_k l'alphabet à k lettres. Soient α un réel et ℓ un entier. Un mot fini ou infini est dit (α, ℓ)-free (resp. (α^+, ℓ)-free) si aucun de ses facteurs n'est une répétition d'exposant $\beta \geq \alpha$ (resp. $\beta > \alpha$) et de période de longueur supérieure à ℓ.*

Definition 3. *Le seuil des répétitions généralisé $RT(k, \ell)$ est la borne inférieure des réels α tels qu'il existe un mot infini (α, ℓ)-free sur A_k.*

On remarque bien sûr que $RT(k, 1)$ est le seuil des répétitions défini dans la section 1. Les valeurs de $RT(k, 1) = RT(k)$ sont donc connues pour tous les k. Pour d'autres valeurs de ℓ, Ilie, Ochem et Shallit ont établi quelques résultats, et en ont conjecturés d'autres, en se basant sur des calculs informatiques. Ces valeurs sont réunies dans le tableau de la figure II.2, dans lequel les valeurs en gras sont celles qu'ils ont prouvées, et celles en italique sont celles qu'ils ont conjecturées.

k \ l	1	2	3	4	5	6	7	8
2	**2**	**2**	**8/5**	**3/2**	**7/5**	**4/3**	*31/24*	*24/19*
3	**7/4**	**3/2**	**4/3**	*5/4*	*6/5*	*7/6*	*8/7*	*9/8*
4	**7/5**	*5/4*	*6/5*	*7/6*	*8/7*	*9/8*	*10/9*	*11/10*
5	**5/4**	*6/5*	*8/7*	*9/8*	*10/9*			
6	**6/5**	*36/31*						
7	**7/6**	*8/7*						
8	**8/7**							
9	**9/8**							
10	**10/9**							
11	**11/10**							
12	**12/11**							
13	**13/12**							
14	**14/13**							
...	**...**							

FIGURE II.2 – Quelques valeurs, exactes ou conjecturées, de $RT(k, \ell)$

Mentionnons par ailleurs le problème de l'évitabilité des répétitions arbitrairement longues, avec en particulier le résultat suivant :

Théorème 11 (Cassaigne, [12])**.** *Il existe un mot infini binaire w tel que $\lim_{|xy| \to +\infty} \frac{|xyx|}{|xy|} = 1$ pour tout $xyx \in \mathrm{Fact}(w)$.*

Hormis les résultats exacts du tableau de la figure II.2 pour des petites valeurs de k et de ℓ, Ilie, Ochem et Shallit donnent des bornes plutôt grossières de $RT(k, \ell)$:

Proposition 7 (Ilie, Ochem et Shallit, [28])**.** *Pour tous entiers $k \geq 2$ et $\ell \geq 1$, le seuil des répétitions généralisé $RT(k, \ell)$ existe, et*

$$1 + \frac{\ell}{k^\ell} \leq RT(k, \ell) \leq 2.$$

Nous établissons de meilleures bornes dans l'annexe A. Plus précisément, nous prouvons les résultats suivants.

Théorème 12. *Pour tout $\ell \geq 2$,*

$$RT(2,\ell) \geq 1 + \frac{2}{\ell+2}.$$

Théorème 13. *Pour tout $k \geq 2$,*

$$RT(k,2) \geq 1 + \frac{1}{1 + \lfloor 3/2(k-1) \rfloor}.$$

Théorème 14. *Pour tout $k \geq 3$ et tout $\ell \geq 3$,*

$$RT(k,\ell) \geq 1 + \frac{1}{(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1})k}.$$

Théorème 15.

$$RT(k,\ell) \leq 1 + \frac{2\ln\ell}{\ell\ln\lambda} + O(\frac{1}{\ell}),$$

pour $k \geq 2$ fixé, ℓ tendant vers l'infini, et $\lambda = \frac{(k-1)+\sqrt{(k-1)(k+3)}}{2}$.

Certains de ces résultats ont été présentés à la conférences *WORDS 2009*, dans deux exposés séparés : Ochem et Fiorenzi, *More on generalized repetition thresholds*, et Vaslet, *Bounds for the generalized repetition threshold.* Cela a donné ensuite lieu à une collaboration entre les trois auteurs, et à la publication dans *Theoretical Computer Science* de l'article reproduit dans l'annexe A.

Depuis l'établissement de ces bornes, Rumyantsev [55] a établi, par une utilisation astucieuse de la complexité de Kolmogorov et du lemme de Lovász, une amélioration remarquable de la borne supérieure. Sa borne est la meilleure connue à ce jour :

Théorème 16 (Rumyantsev, [55])**.** *Il existe une constante c telle que pour tout $k \geq 2$ et tout $\ell \geq 1$,*

$$1 + \frac{1}{k\ell} \leq RT(k,\ell) \leq 1 + \frac{c}{k\ell}.$$

Chapitre III

Évitabilité et répétitivité

Nous venons de traiter de diverses questions qui concernent toutes des mots satisfaisant à des contraintes d'évitabilité, que ces contraintes soient décrites par les exposants et/ou par la longueur des répétitions. Dans ce qui suit, un nouveau type de contrainte est ajouté : des contraintes de répétitivité. C'est-à-dire que nous allons considérer des mots pour lesquels, d'une part, nous interdisons un certain ensemble de facteurs, et d'autre part, nous forçons un autre ensemble de facteurs - les facteurs dont il sera question étant toujours des répétitions, paramétrées par leur exposant.

1 Évitabilité des répétitions et exposants critiques

Les exposants critiques définis dans le chapitre préliminaire permettent de donner une autre définition, équivalente à celle utilisée dans le chapitre 1, du seuil des répétitions : pour un entier k donné,

$$RT(k) = \inf_{w \in A_k^\omega} \{E_c(w)\}.$$

Dans les différentes preuves de la conjecture de Dejean, les auteurs construisent des mots de A_k^ω qui sont $RT(k)^+$-free. Les répétitions d'exposant inférieur à $RT(k)$ sont par ailleurs inévitables. Ainsi, chaque rationnel $RT(k)$ est l'exposant critique d'un mot. Qu'en est-il pour les autres réels, supérieurs à $RT(k)$ (les exposants strictement inférieurs à $RT(k)$ étant inévitables, il est clair qu'aucun réel strictement inférieur à $RT(k)$ n'est un exposant critique) ? Construire un mot d'exposant critique un réel α donné signifie construire un mot qui évite les répétitions d'exposant plus grand que α, tout en imposant l'occurence de répétitions d'exposant plus petit. Krieger et Shallit établissent le résultat suivant :

Théorème 17 (Krieger et Shallit, [34])**.** *Pour tout réel $\alpha \geq 1$, il existe un alphabet A et un mot $w \in A^\omega$ tel que $E_c(w) = \alpha$.*

Mais la taille de l'alphabet A sur lequel ils construisent le mot w augmente vite avec α. Plus précisément, ils montrent, de façon constructive, l'existence d'un mot :

- à 4 lettres d'exposant critique α, si $\alpha > 2$,

- à 2 lettres (en fait le mot de Thue-Morse) d'exposant critique 2.

Mais dans le cas où $1 < \alpha < 2$, leur preuve utilise explicitement l'existence d'un mot v α^+-free. En notant k l'entier tel que $v \in A_k^\omega$, ils construisent alors, de manière effective, un mot infini sur l'alphabet

A_{3k} d'exposant critique α. Ainsi, si la taille de l'alphabet tend bien vers l'infini quand α tend vers 1, sa croissance est beaucoup trop rapide pour répondre au même problème en fixant cette fois la taille de l'alphabet. Ils laissent donc ouverte la question suivante : pour un entier k donné, tout réel $\alpha \geq RT(k)$ est-il l'exposant critique d'un mot de A_k^ω ? Currie et Rampersad [18] répondent par l'affirmative pour l'alphabet binaire :

Théorème 18 (Currie et Rampersad [18])**.** *Pour tout réel $\alpha \geq 2$, il existe un mot binaire infini d'exposant critique α.*

Ils donnent une construction effective de tels mots, en appliquant au mot de Thue-Morse un certain nombre d'opérations qu'on pourrait qualifier d'itération-effacement, afin de forcer l'apparition de répétitions d'exposants voulus, tout en gardant le contrôle sur les exposants à éviter. Leurs mots ont la forme suivante :

$$\lim_{n\to+\infty} \delta^{t_1}\mu_{TM}^{s_1}(\delta^{t_2}\mu_{TM}^{s_2}(\delta^{t_3}\mu_{TM}^{s_3}(...\delta^{t_n}\mu_{TM}^{s_n}(\epsilon)...))),$$

où δ est une opération d'effacement, où μ_{TM} est la substitution de Thue-Morse, et où les suites d'entiers $(t_n)_n$ et $(s_n)_n$ sont astucieusement choisies.

Nous répondrons à la même question dans le cas de l'alphabet ternaire :

Théorème 19. *Tout nombre réel supérieur à $RT(3) = 7/4$ est l'exposant critique d'un mot ternaire infini.*

Pour cela, dans la continuité des techniques utilisées dans la preuve du théorème 18, nous utilisons une méthode d'itération-effacement avec la substitution μ_D de Dejean, après avoir établi des propriétés de désubstitution de ce morphisme, qui permettent de contrôler le devenir des répétitions et de leur exposant lors de l'itération. Pour les détails de la preuve, on se reportera à l'annexe B, qui consiste en un article à paraître dans *Electronic Journal of Combinatorics*.

Par ailleurs, au vu de la structure des répétitions et des mots sans répétitions sur les alphabets de taille ≥ 3, par rapport au caractère particulier de ce qu'il se passe sur l'alphabet binaire (voir section 2), nous sommes portés à penser que des résultats similaires existent pour tous les alphabets, et, à la suite de Currie et Rampersad [18], on peut proposer la conjecture suivante :

Conjecture 3. *Pour tout entier $k \geq 2$, pour tout réel $\alpha \geq RT(k)$, il existe un mot $w \in A_k^\omega$ d'exposant critique α.*

Cependant, la proposition 2 de Brandenburg, et ses conséquences sur les mots $RT(k)^+$-free sur A_k qui valident la conjecture de Dejean (cf. section 1), nous assurent que la méthode d'itération-effacement avec une "bonne" substitution n'est plus directement applicable, mais qu'il faut l'adapter aux constructions utilisées dans les différentes étapes de la preuve de la conjecture. Ce qui, par ailleurs, fournit une nouvelle motivation à la recherche d'une construction unifiée de mots $RT(k)^+$-free.

Signalons pour finir qu'une étude remarquable des exposants critiques a été menée par Krieger [33], en particulier l'étude des exposants critiques des points fixes de substitutions [31], [32]. Elle établit le résultat suivant :

Théorème 20 (Krieger, [31])**.** *L'exposant critique d'un mot purement substitutif est soit infini, soit un nombre algébrique. De plus, si un mot purement substitutif est engendré par une substitution uniforme, alors, son exposant critique est un nombre rationnel.*

La preuve de ce résultat lui permet de décrire un algorithme de calcul de l'exposant critique des points fixes de certaines substitutions. Ainsi, elle donne dans sa thèse [33] des tables bien fournies d'exposants critiques de points fixes de substitutions. De telles tables peuvent s'avérer très utiles. Par exemple, dans la partie qui suit, la résolution des questions soulevées nécessitent de choisir des morphismes dont le point fixe a un exposant critique qui convient.

2 Évitabilité et haute répétitivité

Comme on l'a vu à la fin de la section II 3, le théorème 11 établit en substance que les répétitions arbitrairement longues sont toutes évitables, si l'on se situe sur l'ensemble, tout entier, des mots infinis. On peut alors chercher à savoir ce qu'il en est pour des ensembles de mots plus restreints, et par exemple, dans ce qui nous intéresse ici, pour des ensembles de mots sans répétitions. Plus précisément, considérons l'ensemble des mots binaires 2^+-free. Le seuil des répétitions $RT(2)$ vaut 2, donc en particulier, tous les mots binaires 2^+-free contiennent des carrés. Dekking établit qu'ils en contiennent même une infinité :

Proposition 8 (Dekking, [24])**.** *Tout mot binaire, infini et 2^+-free, contient des carrés arbitrairement longs.*

Shallit [57] étend ce résultat en montrant que :

Théorème 21. *Pour tout réel $2 < \alpha \leq 7/3$, tout mot infini binaire α-free contient des carrés arbitrairement longs. Par ailleurs, pour tout réel $\alpha > 7/3$, il existe un mot infini binaire qui est α-free et qui évite les carrés arbitrairement longs (i.e., dont les carrés sont de longueur bornée).*

Par exemple, Dekking [24] donne la construction d'un mot infini binaire 3-free et sans carrés de longueur supérieure à 4.

Voyons maintenant sous un autre angle l'étude des carrés dans les mots binaires. Clairement, pour tout mot binaire périodique, à chaque position donnée dans le mot, commence un carré arbitrairement long (autrement dit, commencent des carrés en nombre infini). On peut aussi observer qu'il existe des mots apériodiques dont les carrés vérifient cette même propriété. En fait, c'est le cas de tous les mots sturmiens [3]. Certains de ces sturmiens, par exemple le mot de Fibonacci, sont non seulement apériodiques, mais sont même d'exposant critique fini. Par ailleurs, cette propriété sur les carrés n'est clairement pas vérifiée dans tout mot d'exposant critique fini. Par exemple, le mot de Thue-Morse est 2^+-free, mais n'a aucun carré comme préfixe. Plus généralement,

Proposition 9 (Currie et Rampersad, [21])**.** *Si w est un mot binaire 2^+-free, alors il existe une position i dans w telle que w ne contient pas de carré commençant à cette position i.*

Ce résultat est une conséquence de la caractérisation, due à Shelton et Soni [58], des carrés sans chevauchement :

Proposition 10. *Soit l'ensemble*

$$A = \bigcup_{k \geq 0} \mu_{TM}^k(\{00, 11, 010010, 101101\}).$$

A est l'ensemble des carrés qui apparaissent dans le mot de Thue-Morse. De plus, si $u \in \{0,1\}^*$ est un carré 2^+-free, alors, c'est un conjugué d'un mot de A.

Signalons par ailleurs qu'une étude très précise des carrés dans le mot de Thue-Morse a été faite par Pansiot [46], par Brlek [8] et par Brown, Rampersad, Shallit et Vasiga [9]. Toutes ces observations motivent une question naturelle, proposée par Richomme en 2005 : quelle est la borne inférieure des réels α tels qu'il existe un mot binaire infini qui soit α^+-free et qui contienne des carrés arbitrairement longs commençant à chaque position ?

On a vu dans la section 1 que tout réel strictement supérieur à 2 est l'exposant critique d'un mot binaire infini, que tout réel $> 7/4$ est celui d'un mot ternaire infini, et qu'on peut conjecturer qu'il se passe la même chose pour toute taille d'alphabet. La question de Richomme est alors une façon d'imposer des contraintes de répétitivité supplémentaires, et peut donc être vue comme une généralisation des problématiques liées aux exposants critiques.

Avant d'étudier cette question, pour plus de commodité, introduisons une nouvelle définition.

Définition 7. *Soit α un réel. Un mot fini ou infini w est dit hautement α-répétitif (resp. hautement α^+-répétitif) s'il contient des β-puissances, avec $\beta \geq \alpha$ (resp. $\beta > \alpha$), arbitrairement longues, commençant à chaque position (i.e., si $w = w_1 w_2 ... w_n ...$, pour tout i et pour tout L, le mot $w_i w_{i+1} ...$ contient une β-puissance de longueur supérieure à L comme préfixe).*

En 2010, Currie et Rampersad répondent à la question de Richomme.

Théorème 22 (Currie et Rampersad, [21])**.** *Pour tout réel $\alpha > 7/3$, il existe un mot infini binaire qui est α-free et hautement 2-répétitif. Par ailleurs, pour tout réel $2 < \alpha \leq 7/3$, aucun mot infini binaire α-free n'est hautement 2-répétitif.*

La deuxième partie de ce théorème découle d'un théorème de structure des carrés binaires sans 7/3-puissances, qui généralise la caractérisation de la proposition 10 :

Proposition 11 (Currie et Rampersad, [21])**.** *Les carrés binaires sans 7/3-puissances sont les conjugués des mots de $A = \bigcup_{k \geq 0} \mu_{TM}^k(\{00, 11, 010010, 101101\})$.*

Dans l'article de l'annexe C (article co-écrit avec N. Rampersad, et à paraître dans *Lecture Notes in Computer Science*), nous étendons ce résultat, d'une part à l'alphabet ternaire, et d'autre part, à d'autres répétitions que les carrés. Voici les résultats qui y sont démontrés :

Théorème 23. *Pour tout réel $\beta < 2$, il existe un mot infini binaire qui est 2^+-free et hautement β-répétitif.*

Théorème 24. *Pour tout réel $\beta < 7/4$, il existe un mot infini ternaire qui est $7/4^+$-free et hautement β-répétitif.*

Proposition 12. *Il n'existe pas de mot ternaire $7/4^+$-free et hautement 7/4-répétitif.*

Théorème 25. *Il existe un mot infini ternaire qui est 2-free et hautement 7/4-répétitif.*

Remarquons que la preuve de la proposition 12 (voir Annexe C) est facilement adaptable à tout alphabet, et on en déduit le résultat suivant :

Théorème 26. *Pour tout entier $k \geq 2$, il n'existe pas de mot de A_k^ω qui soit $RT(k)^+$-free et hautement $RT(k)$-répétitif.*

Afin de résumer l'ensemble des résultats des théorèmes et propositions 22, 23, 24, 12, et 25, on peut, de façon très naturelle, introduire, pour tout entier k, l'ensemble des couples de réels

$$\mathfrak{S}_k = \{(\alpha, \beta) \in \mathbb{R}^2 \mid \exists w \in A_k^\omega \ \alpha\text{-free et hautement } \beta\text{-répétitif}\}.$$

Dans l'annexe C, on trouvera deux dessins qui récapitulent ce que l'on sait ou non de $\mathfrak{S}_k$ pour $k = 2$ et pour $k = 3$. La description complète des $\mathfrak{S}_k$ pour tout alphabet reste une question ouverte. Cependant, il semble raisonnable de conjecturer la description de $\mathfrak{S}_k$ donnée dans les figures III.1 et III.2. En effet, il semble possible d'adapter les preuves de l'annexe C à d'autres cas, en utilisant d'autres substitutions, bien choisies en fonction de l'exposant critique de leur point fixe, par exemple les substitutions données par Krieger [33].

Notons qu'on peut aussi affaiblir la contrainte de haute répétitivité en une contrainte, soit du type “contenir une infinité de β-puissances” (autrement dit, “contenir des β-puissances arbitrairement longues), soit du type “contenir au moins une β-puissance à chaque position dans le mot”. Ainsi, en ce qui concerne le premier type, Currie, Rampersad et Shallit montrent les deux résultats suivants :

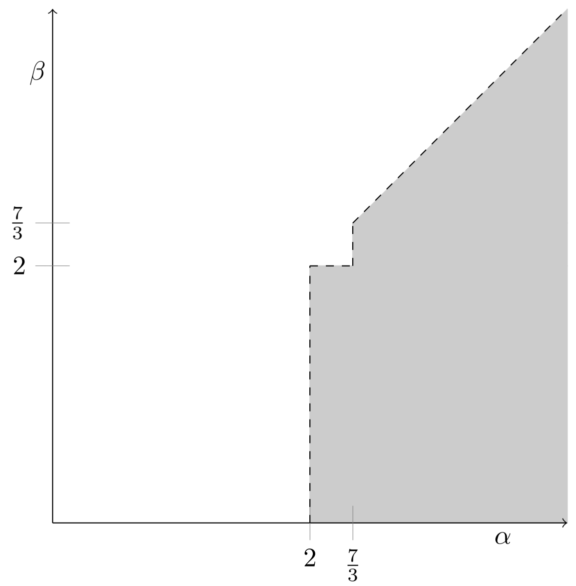


Figure III.1 – Conjecture pour $k = 2$ ($\mathfrak{S}_2$ en grisé.)

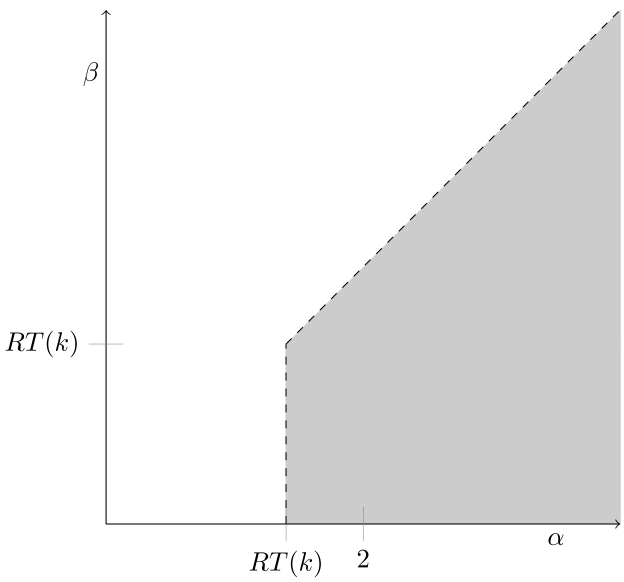


Figure III.2 – Conjecture pour $k \geq 3$ ($\mathfrak{S}_k$ en grisé)

Théorème 27 (Currie, Rampersad et Shallit, [22]). *Il existe un mot binaire infini* $7/3$*-free qui contient une infinité de* 2^+*-puissances.*

Théorème 28 (Currie, Rampersad et Shallit, [22]). *Pour tout réel* $\alpha > 2$*, il existe un réel* β *arbitrairement proche de* α*, tel qu'il existe un mot binaire infini* β^+*-free qui contient une infinité de* β*-puissances.*

Nous établissons dans l'annexe C le résultat suivant, analogue pour l'alphabet ternaire :

Théorème 29. *Pour tout réel* $\alpha \geq 7/4$*, il existe un réel* β*,* $7/4 \leq \beta < \alpha$*, arbitrairement proche de* α*, tel qu'il existe un mot ternaire* β^+*-free contenant une infinité de* β*-puissances.*

Pour l'autre type de contrainte faible de répétitivité, on demande aux mots de contenir une β-puissance à chaque position, mais pas forcément en nombre infini. On a tout d'abord le résultat suivant pour l'alphabet binaire :

Théorème 30 (Currie et Rampersad, [21]). *Pour tout réel* $\alpha > 2$*, il existe un mot binaire infini* α*-free* w *tel qu'à chaque position de* w*, commence un carré.*

Nous établirons dans l'annexe C le résultat analogue suivant, pour l'alphabet ternaire :

Théorème 31. *Pour tout réel* $\alpha > 7/4$*, il existe un mot ternaire* α*-free qui, à chaque position, contient une occurence d'une* $7/4$*-puissance.*

Chapitre IV

Coloration de graphes sans répétitions

Soit un entier k, une *k-coloration* d'un graphe G est une fonction qui assigne à chaque sommet de G une couleur parmi k couleurs données. On peut, de façon naturelle, généraliser la notion de répétitions dans les mots aux coloration de graphes, l'ensemble des couleurs jouant le rôle de l'alphabet, et la fonction de coloration celui de mot. On dit qu'une k-coloration c de G contient un carré s'il existe un chemin $v_1v_2...v_{2r}$, non intersectant (i.e., $\forall i,j, v_i \neq v_j$), dans G tel que $c(v_i) = c(v_{i+r})$ pour tout $i \in [1,r]$. Une *coloration non-répétitive* est alors une coloration qui ne contient pas de carrés. L'un des problèmes principaux concernant les colorations non-répétitives des graphes est celui du nombre chromatique. Si G est un graphe donné, son *nombre chromatique* $\pi(G)$ est le nombre minimal de couleurs tel qu'il existe une coloration non-répétitive de G. Le théorème suivant est une conséquence immédiate des résultats de Thue sur les mots sans carrés.

Théorème 32 (Grytczuk, [27])**.** *Si G est un chemin de plus de 4 sommets, $\pi(G) = 3$.*

Currie [15] s'intéresse au problème des répétitions dans les mots cycliques, et établit :

Théorème 33 (Currie, [15])**.** *Si G est un cycle de plus de 18 sommets, $\pi(G) = 3$.*

Le nombre chromatique vaut 4 pour les arbres [27], et 3 pour les grandes subdivisions de graphes [48], une subdivision d'un graphe étant le graphe obtenu par des subdivisions successives de ses arêtes, i.e. l'ajout d'un sommet et le remplacement de l'arête par deux arêtes.

Poursuivant l'étude des mots cycliques, Aberkane et Currie [1] considèrent le problème de la coloration non-répétitive dans l'autre sens. Au lieu de chercher le nombre de couleurs minimal pour éviter les carrés dans la coloration, on fixe le nombre de couleurs, et on cherche quelles répétitions il est possible d'éviter. Pour cela, de la même manière que pour les mots, on généralise la notion de carré pour définir les répétitions fractionnaires.

Soit k un entier, et c une k-coloration d'un graphe G. Un *facteur* est une suite de couleurs sur un chemin non-intersectant de G. Un tel facteur peut être vu comme un mot sur l'alphabet à k lettres. On peut donc définir son exposant, sa période et son excès, et lui appliquer toutes les problématiques vues précédemment. En particulier, Aberkane et Currie [1] montrent qu'il existe des mots binaires circulaires de toutes les longueurs évitant les répétitions d'exposant strictement supérieur à $5/2$, et que par ailleurs, tout cycle de longueur 5 contient une répétition d'exposant $5/2$. D'où la généralisation naturelle de la notion de seuil des répétitions aux graphes, avec cette petite subtilité qu'il est ici défini pour l'ensemble des graphes d'une classe donnée et pour un nombre de couleurs fixé, alors que le seuil de Dejean pour les mots est défini pour l'ensemble des mots sur un alphabet de taille fixée. Plus précisemment, si $\mathfrak{G}$ est une classe de graphes, et k un entier, on définit le seuil des répétitions $RT(k, \mathfrak{G})$ de la classe $\mathfrak{G}$ pour k couleurs par

$$RT(k, \mathfrak{G}) = \sup_{G \in \mathfrak{G}} RT(k, G),$$

où

$$RT(k, G) = \inf_{k\text{-coloration } c} \sup\{E(v), \mid v \text{ est un facteur dans la } k\text{-coloration } c \text{ de } G\}.$$

Ainsi, on a vu que $RT(2, \mathfrak{C}) = 5/2$ [1], où $\mathfrak{C}$ est la classe des cycles, et que $RT(4, \mathfrak{T}) \leq 2$ [27], où $\mathfrak{T}$ est la classe des arbres. Dans l'annexe D, nous établissons la valeur de $RT(k, \mathfrak{T})$ pour tout k :

Théorème 34.

$$RT(k, \mathfrak{T}) = \begin{cases} 7/2 & \text{si } k = 2, \\ 3 & \text{si } k = 3, \\ 3/2 & \text{si } k \geq 4. \end{cases}$$

L'ensemble des subdivisions de graphes n'est pas à proprement parler une classe de graphes, mais nous pouvons définir un seuil des répétitions pour la pseudo classe, notée $\mathfrak{S}$, des subdivisions suffisament grandes de graphes. En fait, considérons le nombre réel α tel que d'une part, pour tout graphe G, il existe une subdivision G_s de G telle que $RT(k, G_s) \leq \alpha$, et d'autre part, il existe un graphe G, tel que pour toute subdivision G_s de G, $RT(k, G_s) \geq \alpha$. Alors, par abus de notation, on note $RT(k, \mathfrak{S})$, et on appelle seuil des répétitions des graphes de subdivision, ce nombre α. On a vu que $RT(3, \mathfrak{S}) < 2$ [48]. Dans l'annexe D, nous établissons la valeur de $RT(k, \mathfrak{S})$ pour tout k :

Théorème 35.

$$RT(k, \mathfrak{S}) = \begin{cases} 7/3 & \text{si } k = 2, \\ 7/4 & \text{si } k = 3, \\ 3/2 & \text{si } k \geq 4. \end{cases}$$

L'article reproduit en annexe D, et auquel on se reportera pour les preuves des théorèmes 34 et 35, est co-écrit avec P. Ochem, et est à paraître dans *RAIRO Theoretical Informatics and Applications.*

Ainsi, le problème de trouver le seuil des répétitions pour toutes les tailles d'alphabets est résolu complètement dans le cas de deux classes de graphes, celle des arbres, et celle des graphes de subdivisions. Trouver $RT(k, \mathfrak{G})$ pour d'autres classes $\mathfrak{G}$, et en particulier compléter les résultats connus pour la classe des cycles, reste un problème ouvert.

Conclusion

En guise de conclusion, nous proposons quelques questions ouvertes qui émergent des problèmes abordés dans cette thèse.

Tout d'abord, une des principales questions concernant les seuils des répétitions, après la résolution de la conjecture de Dejean, est de savoir s'il est possible de trouver une façon de construire des mots $RT(k)^+$-free qui fonctionnerait pour tous les alphabets. Une piste dans ce sens est la conjecture 1 proposée par Rao.

Problème 1. *La conjecture suivante est-elle vraie : pour tout $k \geq 8$, il existe une substitution h_k telle que le mot $M_k(h_k(w_{TM})) \in A_k^\omega$ soit $RT(k)^+$-free ?*

En ce qui concerne les problèmes combinatoires et de dénombrement des mots $RT(k)^+$-free, nous signalons deux questions ouvertes :

Problème 2. *Pour $k \geq 3$, y a-t-il un nombre exponentiel de mots $RT(k)^+$-free de longueur n sur A_k ?*

Problème 3. *La conjecture 2 proposée par Ochem sur les fréquences des lettres dans les mots $RT(k)^+$-free est-t-elle vraie ?*

Dans la partie II 3, diverses bornes inférieures et supérieures de la valeur de $RT(k, \ell)$ ont été établies.

Problème 4. *Quelle est la constante c dans l'encadrement de Rumyantsev : $1 + \frac{1}{k\ell} \leq RT(k, \ell) \leq 1 + \frac{c}{k\ell}$?*

Plus généralement,

Problème 5. *Quelle est la valeur de $RT(k, \ell)$, pour tout $k \geq 2$, et tout $\ell \geq 1$?*

Dans le chapitre III, nous avons proposé la conjecture 3 :

Problème 6. *Pour tout $k \geq 2$, pour tout $\alpha \geq RT(k)$, existe-t-il un mot $w \in A_k^\omega$ d'exposant critique α ?*

Puis, concernant l'ensemble $\mathfrak{S}_k = \{(\alpha, \beta) \in \mathbb{R}^2 \mid \exists w \in A_k^\omega$ α-free et hautement β-répétitif$\}$, le problème général suivant reste ouvert (voir les figures III.1 et III.2 pour des conjectures) :

Problème 7. *Pout tout $k \geq 2$, que peut-on dire de l'ensemble de réels $\mathfrak{S}_k$?*

De même, on pourra étudier le problème analogue avec des conditions de répétitivité plus faibles. Par exemple :

Problème 8. *Pour tout $k \geq 2$, que peut-t-on dire de l'ensemble de réels $\mathfrak{I}_k = \{\beta \in \mathbb{R} \mid \exists w \in A_k^\omega$, β^+-free et contenant une infinité de β-puissances$\}$?*

Enfin, en ce qui concerne le seuil des répétitions pour les colorations de graphes, le problème qui suit est résolu pour les classes $\mathfrak{T}$ des arbres et $\mathfrak{S}$ des graphes de subdivisions, et partiellement pour la classe des cycles $\mathfrak{C}$:

Problème 9. *Pour une classe $\mathfrak{G}$ de graphes donnée, et pour tout $k \geq 2$, que vaut $RT(k, \mathfrak{G})$?*

Table des figures

Bibliographie

[1] A. Aberkane et J. D. Currie : There exist binary circular $5/2^+$-power-free words of every length. *Electron. J. Comb.*, 11, 2004.

[2] J.-P. Allouche et M. Bousquet-Mélou : Facteurs des suites de Rudin-Shapiro généralisées. *Bull. Belgian Math. Soc.*, 1:145–164, 1994.

[3] J.-P. Allouche, J. L. Davidson, M. Queffélec et L. Q. Zamboni : Transcendence of Sturmian or morphic continued fractions. *J. Number Theory*, 91:39–66, 2001.

[4] J. Berstel : Axel Thue's papers on repetitions in words : a translation. *Publications du LaCIM*, 20, 1995.

[5] J. Berstel : Growth of repetition-free words - a review. *Theoret. Comput. Sci.*, 340:280–290, 2005.

[6] J. Berstel et P. Séébold : A characterization of overlap-free morphisms. *Discr. Appl. Math.*, 46:275–281, 1993.

[7] F. J. Brandenburg : Uniformly growing k-th powerfree homomorphisms. *Theoret. Comput. Sci.*, 23:69–82, 1983.

[8] S. Brlek : Enumeration of factors in the Thue-Morse word. *Discrete Appl. Math.*, 24:93–96, 1989.

[9] S. Brown, N. Rampersad, J. shallit et T. Vasiga : Squares and overlaps in the Thue-Morse sequence and some variants. *Theor. Inform. Appl.*, 40:473–484, 2006.

[10] V. Canterini et A. Siegel : Automate des préfixes-suffixes associé à une substitution primitive. *Journal de théorie des nombres de Bordeaux*, 13(2):353–369, 2001.

[11] A. Carpi : On Dejean's conjecture over large alphabets. *Theoret. Comput. Sci.*, 385:137–151, 2007.

[12] J. Cassaigne : On extremal properties of Fibonacci word. *Theoret. Inform. Appl.*, 42(4):701–715, 2008.

[13] J. Chalopin et P. Ochem : Dejean's conjecture and letter frequency. *RAIRO - Theoret. Inf. Applic.*, 42(3):477–480, 2008.

[14] C. Choffrut et J. Karhumäki : Combinatorics of words. *In* A. Salomaa et G. Rozenberg, éditeurs : *Handbook of Formal Languages*, volume 1, pages 329–438, 1997.

[15] J. D. Currie : There are ternary circular square-free words of length n for $n \geq 18$. *Electron. J. of Combina.*, 9(1):1–7, 2002.

[16] J. D. Currie et M. Mohammad-Noori : Dejean's conjecture and Sturmian words. *Eur. J. Combin.*, 28, 2007.

[17] J. D. Currie et N. Rampersad : Dejean's conjecture holds for $n \geq 30$. *To appear in Theoret. Comput. Sci.*, http ://dx.doi.org/10.1016/j.tcs.2009.01.026, 2008.

[18] J. D. Currie et N. Rampersad : For each $\alpha > 2$ there is an infinite binary word with critical exponent α. *Electron. J. Combin.*, 15(34), 2008.

[19] J. D. Currie et N. Rampersad : Dejean's conjecture holds for $n \geq 27$. *Preprint*, http ://arxiv.org/abs/0901.3188, 2009.

[20] J. D. Currie et N. Rampersad : A proof of Dejean's conjecture. *Preprint*, http ://arxiv.org/abs/0905.1129, 2009.

[21] J. D. Currie et N. Rampersad : Infinite words containing squares at every position. *Theor. Inform. Appl.*, 44:113–124, 2010.

[22] J. D. Currie, N. Rampersad et J. Shallit : Binary words containing infinitely many overlaps. *Electron. J. Combin.*, 13, 2006.

[23] F. Dejean : Sur un théorème de Thue. *J. Comb. Theory A*, 13:90–99, 1972.

[24] F. M. Dekking : On repetitions of blocks in binary sequences. *J. Comb. Theory Ser. A*, 20:292–299, 1976.

[25] R. C. Entringer, D. E. Jackson et J. A. Schatz : On nonrepetitive sequences. *J. Comb. Theory A*, 16, 1974.

[26] A. S. Fraenkel et J. Simpson : How many squares must a binary sequence contain ? *Electronic J. Combina.*, 2, 1995.

[27] J. Grytczuk : Nonrepetitive colorings of graphs - a survey. *Int. J. Math. Sci.*, 2007.

[28] L. Ilie, P. Ochem et J. Shallit : A generalization of repetition threshold. *Theoret. Comput. Sci.*, 345, 2005.

[29] J. Karhumäki et J. Shallit : Polynomial versus exponential growth in repetition-free binary words. *J. Combin. Theory*, 105(2):335–347, 2004.

[30] R. Kolpakov et M. Rao : On the number of Dejean words over alphabets of 5, 6, 7, 8, 9, and 10 letters. *submis à Theoret. Comp. Sci.*

[31] D. Krieger : On critical exponents in fixed points of binary k-uniform morphisms. *Lecture Notes in Comput. Sci.*, 3884:104–114, 2006.

[32] D. Krieger : On critical exponents in fixed points of non-erasing morphisms. *Theoret. Comput. Sci.*, 376, 2007.

[33] D. Krieger : *Critical Exponents and Stabilizers of Infinite Words*. Thèse de doctorat, University of Waterloo, 2008.

[34] D. Krieger et J. Shallit : Every real number greater than 1 is a critical exponent. *Theoret. Comput. Sci.*, 381:177–182, 2007.

[35] D. König : Sur les correspondances multivoques des ensembles. *Fundamenta Math.*, 8:114–134, 1926.

[36] D. König : Über eine Schlussweise aus dem Endlichen ins Unensliche. *Acta Lit. Sci. (Szeged)*, 3:121–130, 1927.

[37] M. Lothaire : *Combinatorics on Words*, volume 17. Encyclopedia of Mathematics and its Applications, addison-wesley, reading édition, 1983.

[38] M. Lothaire : Algebraic combinatorics on words, Encyclopedia of mathematics and its applications 90. *Cambridge University Press, Cambridge*, 2002.

[39] F. Mignosi et G. Pirillo : Repetitions in the Fibonacci infinite word. *RAIRO Inform. Theor. Appl.*, 26:199–204, 1992.

[40] M. Morse et G. A. Hedlund : Unending chess, symbolic dynamics and a problem in semigroups. *Duke Math. J.*, 11:1–7, 1944.

[41] B. Mossé : Puissance des mots et reconnaissabilité des points fixes d'une substitution. *Theoret. Comp. Sci.*, 99:327–334, 1992.

[42] B. Mossé : Reconnaissabilité des substitutions et complexité des suites automatiques. *Bulletin de la S.M.F.*, 124:329–346, 1996.

[43] J. Moulin-Ollagnier : Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters. *Theoret. Comput. Sci.*, 95:187–205, 1992.

[44] P. Ochem : Letter frequency in infinite repetition-free words. *Theoret. Comput. Sci.*, 380:388–392, 2005.

[45] P. Ochem : A generator of morphisms for infinite words. *RAIRO - Theoret. Inf. Appl.*, 40:427–441, 2006.

[46] J. J. Pansiot : The Morse sequence and iterated morphisms. *Inform. Process. Lett.*, 12:68–70, 1981.

[47] J.-J. Pansiot : À propos de la conjecture de F. Dejean sur les répétitions dans les mots. *Discrete Appl. Math.*, 7:297–311, 1984.

[48] A. Pezarski et M. Zmarz : Non-repetitive 3-coloring of subdivided graphs. *Electron. J. Comb.*, 16(1), 2009.

[49] N. Pytheas Fogg : *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794. Lecture Notes in Mathematics, springer-verlag édition, 2001.

[50] N. Rampersad : *Infinite Sequences and Pattern Avoidance.* Thèse de doctorat, University of Waterloo, 2004.

[51] N. Rampersad, J. Shallit et M.-W. Wang : Avoiding large squares in infinite binary words. *In WORDS'03*, volume 27 de *TUCS General Publication*, pages 185–197, 2003.

[52] M. Rao : Last cases of Dejean's conjecture. *WORDS 2009*, http ://www.labri.fr/perso/rao/publi/dejean.ps, 2009.

[53] A. Restivo et S. Salemi : On weakly square free words. *Bull. European Assoc. Theoret. Comput. Sci.*, 21:49–56, 1983.

[54] A. Restivo et S. Salemi : Overlap free words on 2 symbols. *In Automata on Infinite Words*, volume 192, pages 198–206. Lecture Notes in Computer Science, springer, berlin édition, 1985.

[55] A. Rumyantsev : Upper bound for the generalized repetition threshold. http://arxiv.org/abs/1009.4454v2.

[56] K. Saari : Everywhere α-repetitive sequences and Sturmian words. *Europ. J. Combin.*, 31:177–192, 2010.

[57] J. Shallit : Simultaneous avoidance of large squares and fractional powers in infinite binary words. *Int. J. Found. Comput. Sci.*, 15:317–327, 2004.

[58] R. Shelton et R. Soni : Chains and fixing blocks in irreductible sequences. *Discrete Math.*, 54:93–99, 1985.

[59] A. M. Shur : The structure of the set of cube-free $\mathbb{Z}$-words in a two-letter alphabet (en russe). *Izv. Ross. Akad. Nauk Ser. Mat.*, 64:201–224, 2000. English translation in *Izv. Math.*, **64** (847–871), 2000.

[60] A. Thue : Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr., I. Mat. Nat. Kl., Christiana*, 7, 1906.

[61] A. Thue : Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. Mat. Nat. Kl.*, 10:1–67, 1912.

Annexe A

Sur le seuil des répétitions généralisé

F. Fiorenzi, P. Ochem, E. Vaslet, **Bounds for the generalized repetition threshold**,
à paraître dans *Theoretical Computer Science*.

BOUNDS FOR THE GENERALIZED REPETITION THRESHOLD

FRANCESCA FIORENZI, PASCAL OCHEM, AND ELISE VASLET

ABSTRACT. The notion of *repetition threshold*, which is the object of Dejean's conjecture (1972), was generalized by Ilie, Ochem, and Shallit in 2005, to include the lengths of the avoided words. We give a lower and an upper bound on this generalized repetition threshold.

1. INTRODUCTION

The study of repetitions in words has been one of the main topics in combinatorics on words. Thue [13] showed the existence of an infinite square-free word on three letters, that is without concatenated occurrences of the same non-empty factor. This fact was actually implied by the existence of an infinite binary overlap-free word (i.e., without factors of the form $uvuvu$ with u non empty).

A natural extension of this problem takes into account repetitions of a fractional exponent, where the exponent of a non-empty finite word is the ratio between its length and its period. This notion has been introduced by Dejean [7] and Brandenburg [1]. Dejean proved the existence of an infinite word over a three-letter alphabet without repetitions of exponent strictly greater than $7/4$. This bound is the best possible because every sufficiently long word over a three-letter alphabet contains a repetition of exponent at least $7/4$. The least real number $\alpha > 1$ such that there exists an infinite word on k letters avoiding repetitions of exponent strictly greater than α is called the repetition threshold on k letters. Thus Thue's result implies that the repetition threshold on two letters is 2, while Dejean's result means that the repetition threshold on three letters is $7/4$.

Dejean observed that for $k \geq 5$, the repetition threshold is not smaller than $\frac{k}{k-1}$, while for $k = 4$ it is not smaller than $7/5$. She conjectured that these are the actual values of the repetition thresholds. This conjecture has been proved true for $k = 4$ by Pansiot [11]. For $k \geq 5$, the conjecture has been solved thanks to the contribution of many authors: Moulin-Ollagnier [10], Currie and Mohammad-Noori [3], Rao [12], Currie and Rampersad [6, 4, 5], and Carpi [2].

In [8] the authors generalize the repetition threshold of Dejean to handle avoidance of repetitions with sufficiently long period. They define $R(k, \ell)$ as the least real number $\alpha > 1$ such that there exists an infinite word on k letters avoiding repetitions of exponent strictly greater than α and period not smaller than ℓ. This naturally extends the classical notion of repetition threshold. Moreover, in [8] its value has been calculated in some particular cases and general lower and upper bounds have been given. In this paper we improve these bounds by studying the asymptotics of the generalized repetition threshold. Part of this paper was presented at the conference *WORDS 2009*.

2. Definitions

Let $\alpha > 1$ be a rational number, and let $\ell \geq 1$ be an integer. A word w is an (α, ℓ)-*repetition* if $w = (uv)^n u$, where $|uv| = \ell$ and $\alpha = \frac{|w|}{\ell}$. In this case α is called *exponent* of the repetition and ℓ its *period*. Notice that in our definitions, exponent and period of a repetition are not univocally defined. For instance, the word *aabcaabcaa* on the alphabet $\{a, b, c\}$ is a $(\frac{5}{2}, 4)$-repetition, a $(\frac{5}{4}, 8)$-repetition and a $(\frac{10}{9}, 9)$-repetition. We say a word (α, ℓ)-*free* if it contains no factor that is an (α', ℓ')-repetition with $\alpha' \geq \alpha$ and $\ell' \geq \ell$. We say a word (α^+, ℓ)-*free* if it is (α', ℓ)-free for all $\alpha' > \alpha$. Finally, a word is α-*free* if it does not contain a repetition with exponent $\geq \alpha$ and is α^+-*free* if it is α'-free for all $\alpha' > \alpha$.

Let Σ_k denote the k-letter alphabet $\{0, 1, \ldots, k-1\}$. For integers $k \geq 2$ and $\ell \geq 1$, the *generalized repetition threshold* $R(k, \ell)$ is defined as the smallest real number α such that there exists an infinite (α^+, ℓ)-free word over Σ_k. Indeed, there always exists an infinite $(R(k, \ell)^+, \ell)$-free word over Σ_k. Nevertheless, as pointed out in [8], there is no known instance of a $(R(k, \ell), \ell)$-free infinite word. Finally, notice that by definition, $R(k+1, \ell) \leq R(k, \ell)$ and $R(k, \ell+1) \leq R(k, \ell)$.

The finiteness of $R(k, \ell)$ is due to the existence of an infinite binary overlap-free word. Ilie, Ochem and Shallit [8] also obtained a lower bound on $R(k, \ell)$, namely

$$1 + \frac{\ell}{k^\ell} \leq R(k, \ell) \leq 2.$$

The aim of the paper is to improve the above inequalities.

The case $\ell = 1$ corresponds to the classical repetition threshold and the values of $R(k, 1)$ are now all determined. Moreover, the proof of our upper bound explicitly uses the fact that $R(k, 1) = \frac{k}{k-1}$ for $k \geq 5$.

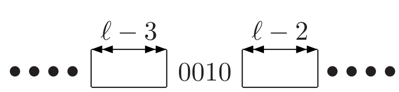
3. Lower Bound

A natural way of obtaining a bound of the form $R(k, \ell) \geq \alpha$ is to show that there is no infinite word over Σ_k^* which is (α, ℓ)-free. In this section we give lower bounds on $R(k, \ell)$. We treat separately the cases $k = 2, \ell \geq 2$ (Theorem 1), $k \geq 2, \ell = 2$ (Theorem 4) and $k, \ell \geq 3$ (Theorem 14).

Theorem 1. $R(2, \ell) \geq 1 + \frac{2}{\ell+2}$.

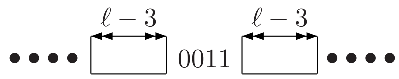
Proof. Suppose for the sake of contradiction that $R(2, \ell) < 1 + \frac{2}{\ell+2}$. That is, there exists an infinite binary word w with no repetition of period $\geq \ell$ and exponent $\geq 1 + \frac{2}{\ell+2}$. In particular, repetitions uvu such that $|u| = 2$ and $\ell \leq |uv| \leq \ell + 2$ cannot appear in w. Moreover, we can assume without loss of generality that w is double-infinite, that is, infinite in both directions. We say that a factor is *forbidden* if it cannot appear in w.

First, we check that the factor 0010 is forbidden. Indeed, if 0010 was a factor of w, then w would have a factor of the form $u_1 u_2 0010 v_1 v_2$, with $|u_1| = |v_2| = 4$, $|u_2| = \ell - 3$ and $|v_1| = \ell - 2$. The following picture illustrates the situation.



By previous considerations about the distances, we have that the blocks u_1 and v_2 coincide necessarily with 1111. This creates a repetition of period $2\ell + 3$ and exponent $1+\frac{4}{2\ell+3} > 1+\frac{2}{\ell+2}$. Hence the factor 0010 and, by symmetry, the factors 0100, 1101, and 1011, are forbidden.

Similarly, we check that the factor 0011 is forbidden. Otherwise we had a factor of the form $u_1u_2 0011 v_1v_2$, with $|u_1| = |v_2| = 4$ and $|u_2| = |v_1| = \ell - 3$.



The block u_1 on the left of the factor 0011 must be of the form $\bullet 10 \bullet$, then $110\bullet$, and finally 1100 since 1101 is forbidden. The block v_2 on the right of the factor 0011 must be of the form $\bullet 10 \bullet$, then $\bullet 100$, and finally 1100 since 0100 is forbidden. This creates a repetition of period $2\ell + 2$ and exponent $1 + \frac{2}{\ell+1} > 1 + \frac{2}{\ell+2}$.

The factor 001 is forbidden because 0010 and 0011 are forbidden. By symmetry, 100, 110, and 011 are also forbidden. Thus, the only remaining possibilities for w are the words ${}^{\omega}0^{\omega}$, ${}^{\omega}1^{\omega}$, and ${}^{\omega}(01)^{\omega}$, which obviously contain repetitions of arbitrarily great exponent and period. □

Theorem 1 is certainly not optimal, since numerical evidences suggest that $R(2,\ell) = 1 + \frac{1}{\ell/3+1}$ for $\ell \geq 6$, $\ell = 0 \pmod 3$. We also mention that Kolpakov and Rao [9] have proved that $R(3,\ell) \geq 1 + \frac{1}{\ell}$, which would be a tight lower bound for the conjecture in [8] that $R(3,\ell) = 1 + \frac{1}{\ell}$ for $\ell \geq 2$.

We now consider the problem for a general k. The (α,ℓ)-freeness of a word imposes conditions on the length of some of its factors. Such conditions are described in the following, and will be used in the proof of Theorem 4 and Theorem 14.

Definition 2. Let k,ℓ be integers, let $m \geq \ell$ be a real number, and let a be a letter in Σ_k. We denote by $\mathfrak{L}(k,\ell,m,a)$ the language of the words w over Σ_k such that the following conditions $(C_i^{\ell,m})$ hold for $1 \leq i \leq \ell$

$(C_i^{\ell,m})$: if $a^i x a^i$ is a subword of w, then $|a^i x| > im$ or $|a^i x| < \ell$.

We denote by $\mathfrak{L}(k,\ell,m)$ the language $\bigcap_{a\in\Sigma_k} \mathfrak{L}(k,\ell,m,a)$.

Proposition 3. *Let k,ℓ be integers, $m \geq \ell$ be a real number, and $w \in \Sigma_k^*$. If w is $(1+\frac{1}{m},\ell)$-free, then, $w \in \mathfrak{L}(k,\ell,m)$.*

Proof. Let a be a letter in Σ_k. Let $i \in \{1,\ldots,\ell\}$ be an integer. Suppose that w does not satisfy condition $(C_i^{\ell,m})$, i.e., there exists a subword $a^i x a^i$ of w such that $\ell \leq |a^i x| \leq im$. Then, $a^i x a^i$ is a repetition with period $|a^i x| \geq \ell$, and exponent

$$\frac{|a^i x a^i|}{|a^i x|} = 1 + \frac{i}{|a^i x|} \geq 1 + \frac{1}{m}.$$

Hence w is not $(1+\frac{1}{m},\ell)$-free. □

First we consider the case $\ell = 2$, and look for a lower bound of $R(k,2)$.

Theorem 4. *Let $k \geq 2$. We have*

$$R(k,2) \geq 1 + \frac{1}{m},$$

where $m = 1 + \left\lfloor \frac{3}{2}(k-1) \right\rfloor$.

Proof. As $m = 1 + \left\lfloor \frac{3}{2}(k-1) \right\rfloor$, we observe that $2m+3 \geq 3k+1$. Hence we have $(2m+3)k > 3k^2 + (k-1)$. Consider $r = 3k+1$ and a word $w \in \Sigma_k^*$ such that $|w| = rk+1$. Then, there exists a letter $a \in \Sigma_k$ such that

$$|w|_a \geq r+1 = 3k+2.$$

Suppose that w is $(1+\frac{1}{m}, 2)$-free. Then, as $m = 1+\left\lfloor \frac{3}{2}(k-1) \right\rfloor \geq 2$, by Proposition 3, the word $w \in \mathfrak{L}(k, 2, m, a)$.
Conditions $(C_1^{2,m})$ and $(C_2^{2,m})$ imply that in w there are at least m letters between two non-consecutive occurrences of the letter a and at least $2m-1$ letters between two occurrences of the factor aa.

We prove by induction on $s \in \mathbb{N}$ that any word in $\mathfrak{L}(k, 2, m, a)$ containing $3s+2$ occurrences of the letter a is not shorter than $(2m+3)s+2$.

This is evident for $s = 0$. If the statement is true up to s, consider a word $v \in \mathfrak{L}(k, 2, m, a)$ containing $3(s+1)+2$ occurrences of the letter a. Then v can be written as

$$v = v_1 v_2 \ldots v_{s+2}$$

where v_1 contains exactly two occurrences of a and for each $i \geq 2$, the factor v_i begins with a letter a and contains exactly three occurrences of this letter. Clearly we have that $|v_{s+2}| \geq 2m+2$.

If the last letter of v_{s+1} is different from a, we have that the last m letters of v_{s+1} are different from a and we apply the inductive hypothesis on the prefix of v of length $|v_1 v_2 \ldots v_{s+1}| - m$. Thus $|v| \geq (2m+3)s+2+m+2m+2 \geq (2m+3)(s+1)+2$.

Otherwise, suppose that the last letter of v_{s+1} is a. If for each $2 \leq i \leq s+1$, the last letter of the factor v_i is a, then this factor v_i has the form $ax_i a y_i a$ with $|x_i|_a = |y_i|_a = 0$. Thus

$$v = v_1 a \ x_2 \ a \ y_2 \ aa \ x_3 \ a \ y_3 \ a \ \ldots \ a \ x_s \ a \ y_s \ aa \ x_{s+1} \ a \ y_{s+1} \ a \ v_{s+2}$$

and for each $i \geq 3$ we have $|x_i| \geq m$ and $|y_i| \geq m$. Moreover $|v_1| \geq m+2$ and $|x_2|+|y_2| \geq 2m-1$. This implies $|v| \geq m+2+3+2m-1+(2m+3)(s-1)+2m+2 \geq (2m+3)(s+1)+2$.

Finally, if for some $2 \leq i \leq s$ the factor v_i does not end with a, we can apply the induction hypothesis on a proper prefix of v, and similar arguments on the lengths as before, to conclude that $|v| \geq (2m+3)(s+1)+2$.

Hence

$$\begin{aligned} |w| &\geq (2m+3)k+2 \\ &> 3k^2+k+1 = rk+1, \end{aligned}$$

which is a contradiction since $|w| = rk+1$. Thus, there is no infinite $(1+\frac{1}{m}, 2)$-free word over Σ_k, and

$$R(k,2) \geq 1 + \frac{1}{m}.$$

□

In Theorem 14, the result of Theorem 4 is generalized to any $\ell \geq 3$. The proof also uses considerations on letter frequencies in the words of $\mathfrak{L}(k, \ell, m)$. In the following, ℓ is a fixed integer greater than 2.

Definition 5. Let $a \in \Sigma_k$. A word $w = w_1 \ldots w_{|w|} \in \Sigma_k^*$ is *of type* S_a if $1 \leq |w| \leq \ell$, $w_1 = w_{|w|} = a$, and if $b_1 b_2 \notin \mathrm{Fact}(w)$ for each $b_1, b_2 \in \Sigma_k \setminus \{a\}$.

Notice that any word of type S_a is in $\mathfrak{L}(k, \ell, m, a)$, for any real number $m \geq \ell$.

Definition 6. Let $a \in \Sigma_k$. Let w be a word of type S_a. The *weight* $p_a(w)$ *of* w is defined by

$$p_a(w) = \max\{j \in \mathbb{N} \mid a^j \in \mathrm{Fact}(w)\}.$$

Notice that $1 \leq p_a(w) \leq \ell$ for each word w.

Proposition 7. *Let $a \in \Sigma_k$. Let w be a word of type S_a. We have*

$$|w|_a \leq \ell - \left\lfloor \frac{\ell}{p_a(w)+1} \right\rfloor = \left\lceil \frac{\ell \cdot p_a(w)}{p_a(w)+1} \right\rceil .$$

Proof. Let us denote $p = p_a(w)$, and write w as

$$w = z_1 z_2 \ldots z_n z,$$

where $n \in \mathbb{N}$, $|z_i| = p+1$, for $1 \leq i \leq n$, and $|z| = r \leq p$. For each i, we have $|z_i|_a \leq p$ (because $a^{p+1} \notin \mathrm{Fact}(w)$), and $|z|_a \leq |z|$. Hence

$$|w|_a \leq np + r = |w| - \left\lfloor \frac{|w|}{p+1} \right\rfloor \leq \ell - \left\lfloor \frac{\ell}{p+1} \right\rfloor = \left\lceil \frac{\ell p}{p+1} \right\rceil .$$

□

Definition 8. Let $a \in \Sigma_k$. A word $w = w_1 \ldots w_{|w|} \in \Sigma_k^*$ is *of type* S'_a if $|w| \leq \ell$, $w_1 = w_{|w|} = a$.

Definition 9. Let $a \in \Sigma_k$ and $m \geq \ell$ be a real number. A word $w \in \Sigma_k^*$ is *of type* L_a if $|w| > m-1$ and $|w|_a = 0$.

Notice that any word of type L_a is in $\mathfrak{L}(k, \ell, m, a)$.

Lemma 10. *Let k, ℓ be integers, let $m \geq 2\ell - 2$ be a real number, $a \in \Sigma_k$, and $w \in \mathfrak{L}(k, \ell, m, a)$. Then w can be written as*

$$w = x s'_1 l_1 s'_2 l_2 \ldots l_{r-1} s'_r y,$$

where $r \in \mathbb{N}$ (if $r = 0$ then $w = x$), each word s'_i is of type S'_a, each word l_i is of type L_a and $|x|_a = |y|_a = 0$.

Proof. As $w \in \mathfrak{L}(k, \ell, m, a)$, w satisfies condition $(C_1^{\ell,m})$: if axa is a subword of w, then, $|axa| > m+1$ or $|axa| < \ell + 1$.
If $|w|_a = 0$, we have the result with $x = w$ and $r = 0$.

Now, if $|w_a| > 0$, let us write $w = xw'y$, where w' begins and ends with a, while x and y do not contain a. The maximal factors without a's in w' either have length $< \ell - 1$, or length $> m - 1$. Take the latter factors as $l_1, \ldots, l_{r-1}$ to get the factorization $w = xs'_1 l_1 s'_2 \ldots l_{r-1} s'_r y$. Each s'_i begins and ends with a, all other blocks do not contain a, and by definition, $|l_i| > m-1$ for all i. Thus, to prove the lemma, it suffices to show that $|s'_i| \leq \ell$ for any i. If it is not the case, then for some $s'_i = aua$, we have $|s'_i| > m+1$. Then $|u| > m-1$, and u contains the letter a by construction. So we have $s'_i = au_1au_2a$. At least one of the words au_1a and au_2a has length $> m+1$, because if both have length $> \ell - 1$, then $|s'_i| < 2\ell - 3 \leq m+1$. Thus, u_1 also contains a, and we repeat the argument to get a contradiction after a finite number of steps.

□

Proposition 11. *Let k, ℓ be integers, let $m \geq 2\ell - 2$ be a real number, $a \in \Sigma_k$ and $w \in \mathfrak{L}(k, \ell, m, a)$. There exists a word $\tilde{w}$ in $\mathfrak{L}(k, \ell, m, a)$, such that $|\tilde{w}| = |w|$, $|\tilde{w}|_a = |w|_a$ and $\tilde{w}$ can be written as*

$$\tilde{w} = xs_1l_1s_2l_2 \dots l_{r-1}s_ry,$$

where $r \in \mathbb{N}$ (if $r = 0$ then $\tilde{w} = x$), each word s_i is of type S_a, each word l_i is of type L_a, and $|x|_a = |y|_a = 0$.

Proof. By Lemma 10, w is of the form:

$$w = x's'_1l'_1s'_2 \dots l'_{r-1}s'_ry',$$

where $r \in \mathbb{N}$, $x', y' \in (\Sigma_k \setminus \{a\})^*$, s'_i are words of type S'_a, and l'_i are words of type L_a. For each $1 \leq i \leq r$, let us consider the factor s'_i, and iterate the rewriting rules $b_1b_2a \mapsto b_1ab_2, \forall b_1, b_2 \in \Sigma_k \setminus \{a\}$, on it. After each iteration, with regard to the lexicographic order (where $a \prec b$, $\forall b \in \Sigma_k \setminus \{a\}$), we get a smaller word, because $b_1ab_2 \prec b_1b_2a$. So the iteration finishes, and we get a word of the form s_ix_i, where s_i is a word of type S_a, and where $x_i \in (\Sigma_k \setminus \{a\})^*$. We remark that $|s_i| \leq |s'_i|$, $|s_ix_i| = |s'_i|$, $|s_i|_a = |s'_i|_a$, and $|s_ix_i|_a = |s'_i|_a$, since the iteration conserves the number of occurrences of each letter in s'_i.

Finally, let $\tilde{w} = xs_1l_1s_2l_2 \dots l_{r-1}s_ry$, where $x = x'$, $y = x_ry'$, and $l_i = x_il'_i$. It is clear that $|\tilde{w}| = |w|$ and $|\tilde{w}|_a = |w|_a$. It is then sufficient to prove that $\tilde{w} \in \mathfrak{L}(k, \ell, m, a)$. Condition $(C_1^{\ell,m})$ is clearly satisfied by definition of the types S_a and L_a. Let now $j \in [2, \dots, \ell]$ and $z \in \Sigma_k^*$ be such that $a^jza^j \in \mathrm{Fact}(\tilde{w})$.

Then, as $w \in \mathfrak{L}(k, \ell, m, a)$, we have $|a^jz| = |a^jz'| > jm+1$, and condition $(C_j^{\ell,m})$ holds.

The letter a moved by the iteration of the rewriting rules always has some letter b on its immediate left. Hence the iteration cannot result in a new factor a^j. So, if $\tilde{w}$ contains a^jza^j, then w contains $a^jz'a^j$, where $|z'| = |z|$. Then, as $w \in \mathfrak{L}(k, \ell, m, a)$, we have $|a^jz| = |a^jz'| > jm$, and condition $(C_j^{\ell,m})$ holds. $\square$

Lemma 12. *Let $p \geq 1$ be an integer. Let v be a word in $\mathfrak{L}(k, \ell, m, a)$, of the form described in Proposition 11:*

$$v = xs_1l_1s_2l_2 \dots l_{r-1}s_ry,$$

and where the words s_i of type S_a are such that $p_a(s_i) = p$ for each $i = 1, \dots, r$. Then,

$$r < \begin{cases} 1 + \frac{|v|-1}{m} \text{ if } p = 1 \\ 1 + \frac{|v|-p}{p(m+1)-\ell} \text{ if } p > 1. \end{cases}$$

Proof. Consider two consecutive blocks s_i and s_{i+1}, with $1 \leq i \leq r-1$. As they have weight p, they both contain a^p as a factor. So, $s_il_is_{i+1}$ has a factor a^pua^p, and by condition $(C_p^{\ell,m})$, we have $|a^pua^p| > p(m+1)$. So $|s_il_is_{i+1}| > p(m+1)$. But $|s_{i+1}| \leq \ell$, by definition of the type S_a. Thus, $\forall 1 \leq i \leq r-1$,

$$|s_il_i| > p(m+1) - \ell.$$

Moreover, $|s-r| \geq p$. Then,

$$|v| \geq \sum_{i=1}^{r-1} |s_il_i| + |s-r| > (r-1)\left(p(m+1) - \ell\right) + p,$$

and we deduce that

$$r < 1 + \frac{|v| - p}{p(m+1) - \ell}.$$

In the case $p = 1$, we can have a better bound: as $|l_i| > m - 1$ and $|s_i| \geq 1$, we have

$$|s_i l_i| > m,$$

hence

$$r < 1 + \frac{|v| - 1}{m}.$$

□

Proposition 13. *Let k, ℓ be integers, let $\xi \geq 2$ be a real number, let $m \geq \xi(\ell - 1)$ be a real number, $a \in \Sigma_k$, and $w \in \mathfrak{L}(k, \ell, m, a)$. We have*

$$\frac{|w|_a}{|w|} < \begin{cases} \frac{2}{|w|} + \frac{1}{m} \cdot \frac{3\xi - 1}{2\xi - 1} \text{ if } \ell = 2 \\ \\ \frac{\ell}{|w|} + \frac{1}{m}\left(\ell - \frac{(\xi-1)(\ell-1)}{2(2\xi-1)} - \frac{\xi^2(\ell-2)}{(\xi\ell-1)(2\xi-1)}\right) \text{ if } \ell \geq 3. \end{cases}$$

Proof. By Proposition 11, we can assume without loss of generality that

$$w = x s_1 l_1 s_2 l_2 \ldots l_{r-1} s_r y,$$

where $r \in \mathbb{N}$, each word s_i is of type S_a, each word l_i is of type L_a, and $|x|_a = |y|_a = 0$. Words of type L_a contain no occurrences of a thus, by Proposition 7,

$$|w|_a = \sum_{i=1}^{r} |s_i|_a \leq \sum_{p=1}^{\ell} n_p(w) \left\lceil \frac{\ell p}{p+1} \right\rceil,$$

where

$$n_p(w) = \#\{i \in [1, \ldots, r] \mid p_a(s_i) = p\}.$$

For any $1 \leq j \leq \ell$, we denote

$$q_j(w) = \sum_{p \geq j} n_p(w) = \#\{i \in [1, \ldots, r] \mid p_a(s_i) \geq j\}.$$

For any $1 \leq p \leq \ell - 1$, we have $n_p(w) = q_p(w) - q_{p+1}(w)$, and $n_\ell(w) = q_\ell(w)$. Hence

$$\begin{aligned} |w|_a &\leq \sum_{p=1}^{\ell-1} (q_p(w) - q_{p+1}(w)) \left\lceil \frac{\ell p}{p+1} \right\rceil + q_\ell(w) \left\lceil \frac{\ell^2}{\ell+1} \right\rceil \\ &= \sum_{p=1}^{\ell} q_p(w) \left(\left\lceil \frac{\ell p}{p+1} \right\rceil - \left\lceil \frac{\ell(p-1)}{p} \right\rceil \right). \end{aligned}$$

Let $n = |w|$. Because of Proposition 11, it makes sense to speak of $q_p(v)$ for any word $v \in \mathfrak{L}(k, \ell, m, a)$. Thus for any $1 \leq p \leq \ell$, we have

$$q_p(w) \quad \leq \quad \max\{q_p(v) \mid v \in \mathfrak{L}(k, \ell, m, a), \text{ and } |v| = n\}.$$

If v reaches this maximum, by replacing some occurrences of a by any other letter, we construct a word v' such that $|v'| = |v|$ and $n_i(v') = 0$ for each $i \neq p$. So,

$$q_p(w) \quad \leq \quad \max\{q_p(v) \mid v \in \mathfrak{L}(k, \ell, m, a), \ |v| = n, \text{ and } n_i(v) = 0, \forall i \neq p\}.$$

Finally, if $n_i(v) = 0$ for each $i \neq p$, then $q_p(v) = n_p(v)$, and we obtain

$$\begin{aligned} q_p(w) &\leq \max\{n_p(v) \mid v \in \mathfrak{L}(k,\ell,m,a),\ |v| = n,\ \text{and } n_i(v) = 0, \forall i \neq p\} \\ &< \begin{cases} 1 + \frac{n-1}{m} \text{ if } p = 1 \\ 1 + \frac{n-p}{p(m+1)-\ell} \text{ if } p > 1, \end{cases} \end{aligned}$$

by Lemma 12. Hence

$$\begin{aligned} |w|_a &< \sum_{p=1}^{\ell} \left(\left\lceil \frac{\ell p}{p+1} \right\rceil - \left\lceil \frac{\ell(p-1)}{p} \right\rceil \right) + \frac{(n-1)\left\lceil \frac{\ell}{2} \right\rceil}{m} + \sum_{p=2}^{\ell} \frac{(n-p)\left(\left\lceil \frac{\ell p}{p+1} \right\rceil - \left\lceil \frac{\ell(p-1)}{p} \right\rceil \right)}{p(m+1)-\ell} \\ &\leq \ell + \frac{n\left(\ell - \left\lfloor \frac{\ell}{2} \right\rfloor\right)}{m} + \sum_{p=2}^{\ell} \frac{n\left(\left\lfloor \frac{\ell}{p} \right\rfloor - \left\lfloor \frac{\ell}{p+1} \right\rfloor \right)}{p(m+1)-\ell} \\ &\leq \ell + \frac{n}{m}\left(\ell - \left\lfloor \frac{\ell}{2} \right\rfloor + \sum_{p=2}^{\ell} \frac{\left\lfloor \frac{\ell}{p} \right\rfloor - \left\lfloor \frac{\ell}{p+1} \right\rfloor}{p - \frac{1}{\xi}} \right). \end{aligned}$$

Indeed, from $m \geq \xi(\ell-1)$ follows $p(m+1) - \ell \geq (p - \frac{1}{\xi})m$. Then, if $\ell = 2$,

$$\frac{|w|_a}{|w|} < \frac{2}{n} + \frac{1}{m} \cdot \frac{3\xi - 1}{2\xi - 1},$$

and if $\ell \geq 3$ we have

$$|w|_a < \ell + \frac{n}{m}\left(\ell - \left\lfloor \frac{\ell}{2} \right\rfloor \cdot \frac{\xi - 1}{2\xi - 1} - \sum_{p=3}^{\ell} \left\lfloor \frac{\ell}{p} \right\rfloor \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)} \right).$$

Since

$$\begin{aligned} \sum_{p=3}^{\ell} \left\lfloor \frac{\ell}{p} \right\rfloor \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)} &\geq \sum_{p=3}^{\ell} \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)} \\ &= \frac{\xi^2(\ell-2)}{(\xi\ell - 1)(2\xi - 1)}, \end{aligned}$$

we obtain

$$\frac{|w|_a}{|w|} < \frac{\ell}{n} + \frac{1}{m}\left(\ell - \frac{(\xi-1)(\ell-1)}{2(2\xi-1)} - \frac{\xi^2(\ell-2)}{(\xi\ell-1)(2\xi-1)} \right).$$

□

Theorem 14. *For any $k \geq 3$, $\ell \geq 3$, we have $R(k,\ell) \geq 1 + \frac{1}{m}$, where*

$$m = \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell - 1} \right) k.$$

Proof. Let $N > 0$ be an integer and $m_N = \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1} + \frac{1}{N} \right) k$. Let $v \in \Sigma_k^*$. Suppose v is $(1 + \frac{1}{m_N}, \ell)$-free, and $|v| \geq N \cdot \ell \cdot m_N$.
As $k \geq 3$, we have $m_N \geq \frac{5}{2}\ell - \frac{3}{2} \geq \ell$. Thus, by Proposition 3, $v \in \mathfrak{L}(k,\ell,m_N)$. Moreover, as Σ_k is the k-letter alphabet, there exists a letter a in Σ_k such that

$\frac{|v|_a}{|v|} \geq \frac{1}{k}$. Then, by Proposition 13 with $\xi = 2$ (notice that $m_N \geq \frac{5}{2}\ell - \frac{3}{2} \geq 2\ell - 2$), and since $|v| \geq N \cdot \ell \cdot m_N$, we have

$$\frac{1}{k} \leq \frac{|v|_a}{|v|} < \frac{\ell}{|v|} + \frac{1}{m_N}\left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell - 1}\right) \leq \frac{1}{m_N}\left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell - 1} + \frac{1}{N}\right).$$

Then, we have

$$m_N < \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell - 1} + \frac{1}{N}\right) k.$$

Hence we have a contradiction because $m_N = \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1} + \frac{1}{N}\right) k$, and

$$R(k,\ell) \geq 1 + \frac{1}{m_N}.$$

Finally, as $m = \lim_{N \to +\infty} m_N$, we have

$$R(k,\ell) \geq 1 + \frac{1}{m}.$$

□

Remark 15. For $\ell = 2$, the result of Proposition 13 and the same arguments as in the proof of Theorem 14 give $R(k,2) \geq 1 + \frac{1}{m}$ with $m = \frac{5k}{3}$, which is not as good as the bound of Theorem 4.

Remark 16. Theorem 14 gives a general result and its proof is based on estimations used in Proposition 13. In order to have better results in some particular cases, one could improve these estimations as follows.

- For small values of ℓ, we can explicitly compute the value of the sum

  $$\sum_{p=3}^{\ell} \left\lfloor \frac{\ell}{p} \right\rfloor \frac{1}{(p - \frac{1}{\xi})(p - 1 - \frac{1}{\xi})}$$

  in Proposition 13, instead of taking a lower bound for it. For example, for $\ell = 6$, we get $R(k,6) \geq 1 + \frac{1}{m}$, with $m = \frac{1457k}{330}$.
- In the proof of Theorem 14, we choose $\xi = 2$ in applying Proposition 13. As it appears clear from the statement of this latter proposition, a greater ξ would give a better upper bound on the frequency $\frac{|w|_a}{|w|}$. Though, we need $\xi \leq \frac{m}{\ell - 1}$ and this implies a constraint on k. For example, if $k \geq 5$ one can choose $\xi = 3$ and get $m = (\frac{4}{5}\ell - \frac{2}{5} + \frac{3}{3\ell-1})k$.

4. Upper bound

In this section we explicitly use the fact that Dejean's conjecture is proved for $K \geq 5$. We describe a morphism from a K-letter alphabet to a k-letter one that transforms an infinite $\left(\frac{K}{K-1}\right)^+$-free word into a word in which the sufficiently large repetitions are of exponent not much greater than $\frac{K}{K-1}$.

Let $S_{k,t}$ be the set of words of length t over Σ_k of the form $0^e w$ where $e \geq 2$, $|w| \geq 1$, the first and the last letter of w are different from 0, and w does not contain 00 as a factor. For example, we have $S_{2,5} = \{00001, 00011, 00101, 00111\}$. Let $K = |S_{k,t}|$ and let h be a t-uniform morphism $h : \Sigma_K^* \to \Sigma_k^*$ such that the set of

h-images of letters in Σ_K is $S_{k,t}$. Now, if $K \geq 5$, we consider the h-image of some infinite $\left(\frac{K}{K-1}\right)^+$-free word over Σ_K.

A uniform morphism $h : \mathcal{A}^* \to \mathcal{B}^*$ is said to be *comma-free* if for any $a, b, c \in \mathcal{A}$ and $s, r \in \mathcal{B}^*$, $h(ab) = rh(c)s$ implies that either $r = \varepsilon$ and $a = c$ or $s = \varepsilon$ and $b = c$.

Remark 17. *A comma-free morphism h is always injective (actually it is injective on the set $\mathcal{A}$ of monoid generators). Moreover, if it is t-uniform, then for each factor u of a word in $h(\mathcal{A}^*)$ such that $|u| \geq 2t-1$, there exists a unique factorization $u = xh(u')y$ where $u' \in \mathcal{A}^*$ and $0 \leq |x|, |y| < t$.*

Lemma 18. *The t-uniform morphism $h : \Sigma_K^* \to \Sigma_k^*$ defined above is comma-free.*

Proof. Suppose that the h is not comma-free. Then there exist $a, b, c \in \Sigma_K$ and $s, r \in \Sigma_k^*$ such that $h(ab) = rh(c)s = w[1, \ldots, 2t]$ with $0 < |r| < t$. We obtain a contradiction for every possible value of $|r|$:

- if $|r| = 1$ or $|r| = 2$, then the letter $w[t + |r|]$ is 0 in $h(ab)$ and is not 0 in $rh(c)s$,
- if $|r| = t-1$ or $|r| = t-2$, then the letter $w[t]$ is not 0 in $h(ab)$ and is 0 in $rh(c)s$,
- if $2 < |r| < t-2$, then $h(c)$ contains the factor $w[t, \ldots, t+2]$. In $h(ab)$, this factor is of the form $x00$ with $x \neq 0$, whereas factors of this form do not exist in $h(c)$.

□

In order to get the mentioned repetition-freeness property in $h(\Sigma_K^*)$, we use the following lemma. For any real number $\ell \geq 1$, we write (α^+, ℓ)-free to mean $(\alpha^+, \lceil \ell \rceil)$-free and hence $R(k, \ell)$ to mean $R(k, \lceil \ell \rceil)$.

Lemma 19. *Let $\alpha, \beta \in \mathbb{R}$, $1 < \alpha < \beta < 2$. Let $h : \mathcal{A}^* \to \mathcal{B}^*$ be a comma-free t-uniform morphism. If $w \in \mathcal{A}^*$ is α^+-free, then $h(w)$ is $(\beta^+, \frac{2t-2}{\beta-\alpha})$-free.*

Proof. Let uvu be a β'-repetition in $h(w)$ with $\beta' > \beta$. Suppose $|u| \geq 2t-1$. Hence u contains an h-image and it can be uniquely written as $xh(u')y$ with $0 \leq |x|, |y| < t$. Thus this h-image $\bar{u} = h(u')$ appears at the same position of u in uvu.

The factor $\bar{v} = yvx$ is an h-image. We have that $\bar{u}\bar{v}\bar{u}$ is the h-image of a repetition in w and hence $\frac{|\bar{u}\bar{v}\bar{u}|}{|\bar{u}\bar{v}|} \leq \alpha$. Moreover, $\beta' = \frac{|\bar{u}\bar{v}\bar{u}|}{|\bar{u}\bar{v}|} + \frac{|x|+|y|}{|uv|}$ and $\beta' > \beta$ implies that $\frac{|x|+|y|}{|uv|} > \beta - \alpha$. Hence

$$|uv| < \frac{|xy|}{\beta - \alpha} \leq \frac{2t-2}{\beta - \alpha}.$$

Suppose now that $|u| \leq 2t-2$. Hence $\beta' > \beta$ implies that $\frac{|u|}{|uv|} > \beta - 1$. Thus $\frac{|uv|}{|u|} < \frac{1}{\beta - 1}$ and

$$|uv| < \frac{|u|}{\beta - 1} \leq \frac{2t-2}{\beta - 1} < \frac{2t-2}{\beta - \alpha}.$$

□

Recall that, by definition, if there exists a (β^+, ℓ)-free infinite word over Σ_k then $R(k, \ell) \leq \beta$.

Corollary 20. *Let $h : \Sigma_K^* \to \Sigma_k^*$ be a comma-free t-uniform morphism as above. If $\beta \in \mathbb{R}$, $\frac{K}{K-1} < \beta < 2$, and $K \geq 5$, then $R\left(k, \frac{2t-2}{\beta - \frac{K}{K-1}}\right) \leq \beta$.*

We now compute $K = K_{k,t}$. Consider the prefixes of length $(t-1)$ of the words in $S_{k,t}$:

- $K_{k,t-1}$ of them are such that the last letter is not 0,
- $K_{k,t-2}$ of them are such that the last letter is 0 and the penultimate letter is not 0,
- one them is the word 0^{t-1}.

Each prefix can be extended by one of the $(k-1)$ letters distinct from 0 to get a word in $S_{k,t}$, so $K_{k,t}$ satisfies the recurrence relation

$$K_{k,1} = K_{k,2} = 0, K_{k,t} = (k-1)(K_{k,t-1} + K_{k,t-2} + 1).$$

Solving this relation, we obtain that

$$K_{k,t} = \frac{k-1}{(2k-3)\sqrt{(k-1)(k+3)}}\left(\lambda^t - \mu^t - (k-2)(\lambda^{t-1} - \mu^{t-1})\right) - \frac{k-1}{2k-3},$$

where $\lambda = \frac{(k-1)+\sqrt{(k-1)(k+3)}}{2}$ and $\mu = \frac{(k-1)-\sqrt{(k-1)(k+3)}}{2}$.

We thus have

$$K_{k,t} = C_k \lambda^{t-1} - O(1), \text{ where } C_k = \frac{(k-1)(\sqrt{(k-1)(k+3)} - k + 3)}{2(2k-3)\sqrt{(k-1)(k+3)}}.$$

For each real number α, $\lfloor \alpha \rceil$ denotes the nearest integer to α.

Theorem 21. $R(k, \ell) \leq 1 + \frac{2\ln \ell}{\ell \ln \lambda} + O\left(\frac{1}{\ell}\right)$ *if k is fixed and ℓ tends to infinity.*

Proof. Let us fix $t = \left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil + 1$ and $\beta = 1 + \frac{2t-2}{\ell} + \frac{1}{K-1}$. For ℓ sufficiently large, we have $t \geq 6$ which ensures that $K \geq 5$, and we also have $\beta < 2$. We can thus use Corollary 20, which gives

$$R(k, \ell) = R\left(k, \frac{2t-2}{\beta - \frac{K}{K-1}}\right) \leq \beta = 1 + \frac{2t-2}{\ell} + \frac{1}{K-1}.$$

Since

$$\frac{2t-2}{\ell} = \frac{2\left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil}{\ell} = \frac{2\ln \ell}{\ell \ln \lambda} + O\left(\frac{1}{\ell}\right)$$

and

$$\frac{1}{K-1} = \frac{1}{C_k \lambda^{t-1} - O(1)} = \frac{1}{C_k \lambda^{\left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil} - O(1)} = O\left(\frac{1}{\ell}\right),$$

the result follows. $\square$

5. An example

Let us illustrate our results with a concrete example: $k = 8$ and $\ell = 100$. Theorem 14 gives $R(8, 100) \geq 1 + \left(\left(\frac{5}{6} \times 100 - \frac{1}{2} + \frac{2}{199}\right) \times 8\right)^{-1} = 1.001508871\ldots$. For the upper bound, we have to decide which morphism h will be used, or equivalently to choose the value of the parameter t. For a given t, we can compute $K = K_{k,t}$ and then the bound $\beta = 1 + \frac{2t-2}{\ell} + \frac{1}{K-1}$. So, we have to choose t so that β is minimized. The choice of $t = \left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil + 1$ in Theorem 21 is well-suited to get such an asymptotic result, but for a given pair (k, ℓ) like this example, it is better to make a specific case study.

- If $t = 3$, then $K = 7$ and $\beta = \frac{181}{150} = 1.20666666\ldots$.
- If $t = 4$, then $K = 56$ and $\beta = \frac{593}{550} = 1.07818181\ldots$.
- If $t = 5$, then $K = 448$ and $\beta = \frac{12094}{11175} = 1.08223713\ldots$.

Since β gets bigger if $t > 5$, the minimum is reached at $t = 4$, whereas $\left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil + 1 = 3$. We thus obtain $R(8, 100) \leq \frac{593}{550} \leq 1.078182$.

6. Conclusion

For k fixed and ℓ tending to infinity, we know now in particular that the asymptotics of the generalized repetition threshold $R(k, \ell)$ is between $1 + \Omega(1/\ell)$ and $1 + O(\ln \ell/\ell)$. New ideas are needed to settle this and other questions about $R(k, \ell)$, such as good estimates for $R(k, 2)$ or $R(k, k)$. The case $1.001848 < R(8, 100) < 1.078182$ suggests that there is still room for improvement.

References

[1] F.-J. Brandenburg. Uniformly growing k-th power-free homomorphisms, *Theor. Comput. Sci.* **23** (1983), 69–82.

[2] A. Carpi. On Dejean's conjecture over large alphabets, *Theoret. Comput. Sci.* **35** (2007), 137–151.

[3] J.D. Currie and M. Mohammad-Noori. Dejean's conjecture and Sturmian words, *Eur. J. Combin.* **28(3)** (2007), 876–890.

[4] J.D. Currie and N. Rampersad. Dejean's conjecture holds for $n \geq 27$, *ITA* **43(4)** (2009), 775-778.

[5] J.D. Currie and N. Rampersad. Dejean's conjecture holds for $n \geq 30$, *Theoret. Comput. Sci.* **345** (2005), 359–369.

[6] J.D. Currie and N. Rampersad. A proof of Dejean's conjecture, *Math. Comp.*, to appear.

[7] F. Dejean. Sur un théorème de Thue, *J. Combin. Theory. Ser. A* **13** (1972), 90–99.

[8] L. Ilie, P. Ochem, and J.O. Shallit. A generalization of repetition threshold, *Theoret. Comput. Sci.* **345** (2005), 359–369.

[9] R. Kolpakov and M. Rao. Personal communication.

[10] J. Moulin-Ollagnier. Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters, *Theoret. Comput. Sci.* **95** (1992), 187–205.

[11] J.-J. Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Disc. Appl. Math.* **7** (1984), 297–311.

[12] M. Rao. Last cases of Dejean's conjecture, *WORDS 2009*.

[13] A. Thue. Über unendliche Zeichenreihen, Christiania Vidensk.-Selsk. Skrifter. I. Mat. Nat. Kl, 7 (1906) 1–22. Reprinted in: T. Nagell (Ed.), Selected Mathematical Papers of Axel Thue, Universitetsforlaget, Oslo, 1977, pp. 139–158.

Francesca Fiorenzi, Laboratoire de Recherche en Informatique, UMR 8623, Université Paris-Sud, 91405 Orsay, France

E-mail address: fiorenzi@lri.fr

Pascal Ochem, Laboratoire de Recherche en Informatique, UMR 8623, Université Paris-Sud, 91405 Orsay, France & CNRS- Laboratoire J.-V.Poncelet, UMI 2615, Moscow

E-mail address: ochem@lri.fr

Elise Vaslet, Institut de Mathématiques de Luminy, UMR 6206, Université Aix-Marseille II, Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France

E-mail address: vaslet@iml.univ-mrs.fr

Annexe B

Sur l'ensemble des exposants critiques de l'alphabet ternaire

E. Vaslet, **Critical exponents of words over 3 letters**,
à paraître dans *Electronic Journal of Combinatorics.*

Critical exponents of words over 3 letters

Elise Vaslet
Institut de Mathématiques de Luminy, Université Aix-Marseille II
Marseille, France
vaslet@iml.univ-mrs.fr

Submitted: 2010; Accepted: 2011; Published: XX
Mathematics Subject Classification: 68R15

Abstract

For all $\alpha \geq RT(3)$ (where $RT(3) = 7/4$ is the repetition threshold for the 3-letter alphabet), there exists an infinite word over 3 letters whose critical exponent is α.

1 Introduction

Let A be a finite alphabet. Any finite word v over A, $v \neq \epsilon$, can be factorized as $v = p^k e$, where:

- $k \geq 1$
- e is a prefix of p
- $|p|$ is minimal

We then say that v has period p, excess e, and exponent $E(v) = |v|/|p|$. For example, the English word *church* has period *chur*, excess *ch*, and exponent 3/2, while the French word *entente* has period *ent*, excess *e*, and exponent 7/3. A (finite or infinite) word is said to be α-free (resp. α^+-free) if none of its subwords has exponent β, with $\beta \geq \alpha$ (resp. $\beta > \alpha$).

The critical exponent of an infinite word w over A is defined as

$$E_c(w) = \sup\{E(v) \in \mathbb{Q},\ v \text{ subword of } w\}.$$

For example, the binary word $abab^2 \cdots ab^n \cdots$ has critical exponent $+\infty$. The Thue-Morse word, fixed point of the morphism $0 \mapsto 01, 1 \mapsto 10$, has critical exponent 2 ([10] and [1]). The Fibonacci word, fixed point of the morphism $0 \mapsto 01, 1 \mapsto 0$, has critical exponent $2 + \phi$, where ϕ is the golden number [8].

The problem to determine if, for a given real number $\alpha > 1$, there is an infinite word w_α with critical exponent α, has been solved by Krieger and Shallit [7]. The number of letters they use to construct the word w_α grows very fast as α tends to 1, and they left open the construction of w_α over an alphabet with a fixed size.

Let k be a natural number, and let A_k be the k-letter alphabet. The repetition threshold on k letters is the real number (see [6] and [3] for more details)

$$RT(k) = \inf\{E_c(w), w \in A_k^\omega\}.$$

Dejean [6] conjectured that

$$RT(k) = \begin{cases} 2 & \text{if } k = 2 \\ 7/4 & \text{if } k = 3 \\ 7/5 & \text{if } k = 4 \\ k/(k-1) & \text{if } k > 4 \end{cases}$$

and this conjecture is now proved (see [9] and [5]).

It is clear that if $\alpha < RT(k)$, no word over A_k has critical exponent α. Currie and Rampersad [4] proved the following result for a binary alphabet:

> For each $\alpha \geq 2 = RT(2)$, there is an infinite binary word with critical exponent α.

And they conjectured:

> Let $k \geq 2$. For each $\alpha \geq RT(k)$, there is an infinite word over k letters with critical exponent α.

We will prove that this is true for $k = 3$. Let $A_3 = \{a, b, c\}$ be the 3-letter alphabet. Dejean [6] proved that $RT(3) = 7/4$. We immediately remark that if $\alpha \geq 2$, the result of Currie and Rampersad gives us a binary word, and then also a word over A_3, with critical exponent α. That's why we only have to consider the case $7/4 \leq \alpha < 2$. To demonstrate her result, Dejean considered the morphism μ (which we will call Dejean's morphism) defined as follows:

$$\mu : \begin{cases} a \mapsto abc\ acb\ cab\ c\ bac\ bca\ cba \\ b \mapsto \pi(\mu(a)) = bca\ bac\ abc\ a\ cba\ cab\ acb \\ c \mapsto \pi^2(\mu(a)) = cab\ cba\ bca\ b\ acb\ abc\ bac \end{cases}$$

where π is the permutation $(a\ b\ c)$, and proved that its fixed point $\mu^\infty(a)$ has critical exponent $7/4$.

2 Exponents and Dejean's morphism

Dejean [6] noticed the existence of specific subwords in $\mu(a)$, $\mu(b)$, and $\mu(c)$, which can be used to desubstitute μ. She called them characteristic factors:

Proposition 1. *The words $f_1 = abcacbc$, $f_2 = cabcbac$ and $f_3 = cbcacba$ only appear in $\mu(A_3^*)$ respectively as prefix, central factor, and suffix, of $\mu(a)$. Similarly, $\pi(f_1)$, $\pi(f_2)$, $\pi(f_3)$ only appear as prefix, central factor, and suffix of $\mu(b)$, and $\pi^2(f_1)$, $\pi^2(f_2)$, $\pi^2(f_3)$ as prefix, central factor, and suffix of $\mu(c)$.*

Definition 1. *The words f_1, f_2, and f_3 (resp. $\pi(f_1)$, $\pi(f_2)$, $\pi(f_3)$; resp. $\pi^2(f_1)$, $\pi^2(f_2)$, $\pi^2(f_3)$) are called characteristic factors of $\mu(a)$ (resp. $\mu(b)$; resp. $\mu(c)$).*

We use these characteristic factors to prove the following desubstitution results for μ.

Proposition 2. *Let w be a word over A_3, and u a subword of $\mu(w)$. If $|u| \geq 18$, then, there exists a unique $x \in A_3$, a unique $y \in A_3$, and there exist some unique $s, v, p \in A_3^*$, such that $u = s\mu(v)p$, where $s \neq \epsilon$ is a suffix of $\mu(x)$, and $p \neq \epsilon$ is a prefix of $\mu(y)$.*

Proof. As $|u| \geq 18$, u has a characteristic factor as a subword. The result is then clear since μ is a 19-uniform morphism, and since $\forall x \in A_3$, $\mu(x)$ begins and ends with x. $\square$

Theorem 1. *Let $w \in A_3^*$, and let u be a subword of $\mu(w)$. Assume u has period p, excess e and exponent $7/4 < |u|/|p| < 2$. Then, w has a subword v of length $|v| \geq \lceil |u|/19 \rceil$, with period q such that $|q| = \lceil |p|/19 \rceil$, and with exponent $E(v) \geq |u|/|p|$.*

Proof. There are two cases: either $|e| \geq 18$, or $|e| < 18$.

Suppose first that $|e| \geq 18$. Then, we also have $|u| \geq 18$. Then, by Proposition 2,

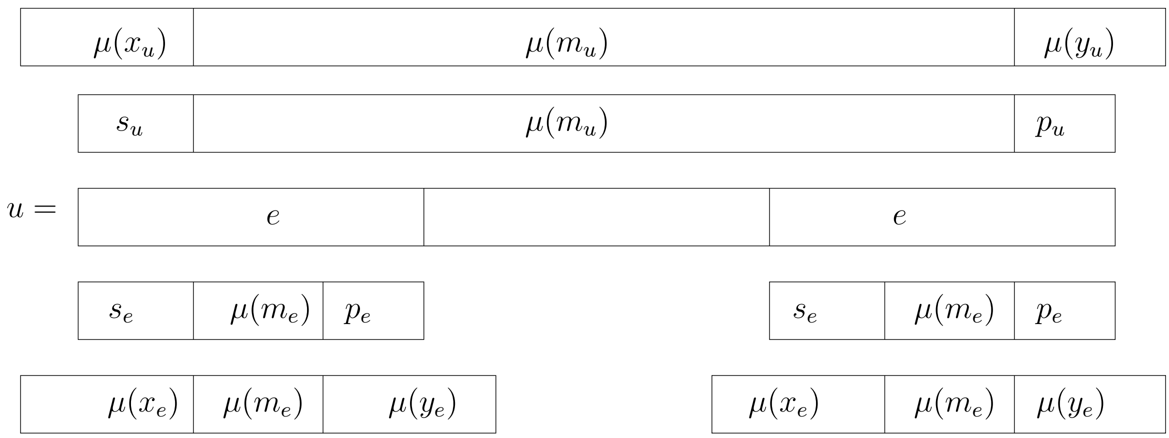
- There exist some unique $x_u, y_u \in A_3$, and some unique $m_u, s_u, p_u \in A_3^*$, $s_u \neq \epsilon$ suffix of $\mu(x_u)$, p_u prefix of $\mu(y_u)$, such that $u = s_u\mu(m_u)p_u$,
- There exist some unique $x_e, y_e \in A_3$, and some unique $m_e, s_e, p_e \in A_3^*$, $s_e \neq \epsilon$ suffix of $\mu(x_e)$, p_e prefix of $\mu(y_e)$, such that $e = s_e\mu(m_e)p_e$.

Let:

$$f = x_e m_e y_e$$

$$v = x_u m_u y_u$$

As e is a suffix of u, we have: $p_u = p_e$, $y_u = y_e$, and $\mu(x_e)\mu(m_e)$ is a suffix of $\mu(x_u)\mu(m_u)$, thus $x_e m_e$ is a suffix of $x_u m_u$. Moreover, e is a prefix of u, so $s_u = s_e$, $x_e = x_u$, $\mu(m_e)\mu(y_e)$ is a prefix of $\mu(m_u)\mu(y_u)$, and $m_e y_e$ is a prefix of $m_u y_u$ (see the following figure).



Therefore, f is a prefix and a suffix of v, and v has excess f ($|f|$ is maximal, otherwise $|e|$ would not be maximal in u). Let us denote its period by q, i.e. $v = qf$. It is clear that

$$|v| \geq \lceil \frac{|u|}{19} \rceil.$$

Moreover, q has length

$$\begin{aligned} |q| &= |v| - |f| = |m_u| - |m_e| \\ &= \frac{|\mu(m_u)| - |\mu(m_e)|}{19} \\ &= \frac{(|u| - |s_u| - |p_u|) - (|e| - |s_e| - |p_e|)}{19} \\ &= \frac{|u| - |e|}{19} \text{ (because } s_u = s_e \text{ and } p_u = p_e) \\ &= \frac{|p|}{19} \end{aligned}$$

Finally, v has exponent

$$E(v) = \frac{|v|}{|q|} \geq \lceil \frac{|u|}{19} \rceil / \frac{|p|}{19} \geq \frac{|u|}{19} . \frac{19}{|p|} = E(u)$$

Thus, v is the word we were looking for.

Suppose now that $|e| < 18$. Then, as $E(u) > 7/4$, $|u| \leq 24 + 18 = 42$. As μ is 19-uniform, u is a subword of a word $\mu(x)$, where $x \in A_3^*$, $|x| \leq 4$. Moreover, $E(u) > 7/4$, so x is not a subword of $\mu^\infty(a)$. Then, by looking at the $\mu(x)$ obtained if x is not a subword of $\mu^\infty(a)$, we can again reduce the set of possible x:

$$x \in \{aa, bb, cc, aaa, bbb, ccc, aaaa, bbbb, cccc, abab, acac, baba, bcbc, caca, cbcb\}$$

since in the other cases, $\mu(x)$ has no subword u such that $|u| \leq 42$ and $E(u) > 7/4$. Then, we have:

- if u is a subword of $\mu(x)$, with $x \in \{aa, bb, cc\}$, consider $v = x$. v has exponent 2, and period q of length $|q| = 1$. We can also remark that u necessarily has a period of length $19 = 19.|q|$, and has exponent $E(u) \leq 2 = E(v)$. Therefore, v is the word we were looking for.
- if u is a subword of $\mu(x)$, with $x \in \{aaa, bbb, ccc\}$, consider $v = x$. v has exponent 3, and period q of length $|q| = 1$. As u necessarily has a period of length $19 = 19.|q|$, and has exponent $E(u) \leq 3 = E(v)$, v is the word we were looking for.
- if u is a subword of $\mu(x)$, with $x \in \{aaaa, bbbb, cccc\}$, consider $v = x$. v has exponent 4, and period q of length $|q| = 1$. As u necessarily has a period of length $19 = 19.|q|$, and has exponent $E(u) \leq 4 = E(v)$, v is the word we were looking for.
- finally, if u is a subword of $\mu(x)$, with $x \in \{abab, acac, baba, bcbc, caca, cbcb\}$, consider $v = x$. v has exponent 2, and period q of length $q = 2$. As u necessarily has a period of length $2.19 = 19.|q|$, and has exponent $E(u) \leq 2 = E(v)$, v is the word we were looking for.

$\square$

3 Construction of an infinite α-free word over A_3

In the following, we will use the operator, denoted by δ, that removes the first letter of a word: for example, $\delta(0110) = 110$.

Lemma 1. *Let $\mathfrak{L}$ be the set $Fact(\mu(A_3^*))$ of all subwords of the words in $\mu(A_3^*)$. Let $\alpha \in]7/4, 2[$ and $v \in A_3^*$, such that:*

- *$abcbabcv \in \mathfrak{L}$*
- *$abcbabcv$ is α-free.*

Suppose that $babcbabcv = xuy$, where u has exponent $E(u) \geq \alpha$. Then, $x = \epsilon$, and $u = babcbabc$.

Proof. By hypothesis, $abcbabcv$ is α-free. Since $E(u) \geq \alpha$, u is necessarily a prefix of $babcbabcv$. Moreover, $babcbab$ has exponent $7/4 < \alpha \leq E(u)$. Therefore, $u = babcbabcv'$, where v' is a prefix of v. Suppose that $v' \neq \epsilon$.
By hypothesis, $abcbabcv' \in \mathfrak{L}$. Moreover, the subword $abcbabc$ only appears in $\mathfrak{L}$ as a subword of $\mu(c)$. So $v' = abacbabcbac....$
The excess of u is at most $babcbab$. Indeed, otherwise, the word $babcbabc$, whose exponent is 2, is a subword of $abcbabcv$, which is impossible since $abcbabcv$ is α-free. Then, u has excess e, with $|e| \leq 7$, and so has period p with $|p| \leq 9$. So u is a subword of $babcbabcabacbabc$. By looking at the factors of this word, we deduce that the only possibility is $u = babcbabc$.

$\square$

Lemma 2. *Let $\alpha \in]7/4, 2[$ be given. Let s, t be natural numbers such that $\mu^s(b) = xabcbabcy$, with $|x| = t$. Let $\beta = 2 - \frac{t}{4.19^s}$. Suppose that $7/4 < \beta < \alpha$, and that $abcbabcv \in \mathfrak{L}$ is α-free. Consider the word $w = \delta^t \mu^s(babcbabcv)$. Then, we have:*

1. *w has a prefix with exponent β.*
2. *If $abcbabcv$ has a subword with exponent γ and period p, then, w has a subword with exponent γ and a period of length $19^s|p|$.*
3. *w is α-free.*

Proof.
1. $\mu^s(babcbabc)$ has exponent 2 and period $\mu^s(babc)$. We have $|\mu^s(babc)| = 4.19^s$, and $|\mu^s(babcbabc)| = 8.19^s$, so the prefix $\delta^t\mu^s(babcbabc)$ of w has exponent
$$\frac{|w|}{|\mu^s(babc)|} = \frac{|\mu^s(babcbabc)| - t}{|\mu^s(babc)|} = \beta$$
2. Let u be a subword of $abcbabcv$, with exponent γ and period p. Then $\mu^s(u)$ is a subword of $\mu^s(abcbabcv)$, with exponent γ and period $19^s|p|$. Moreover, $\mu^s(abcbabcv)$ is a suffix of $\delta^t\mu^s(babcbabcv)$, since $t = |x| \leq |\mu^s(b)|$. So $\mu^s(u)$ is a subword of w, with the required properties.

3. Suppose that w has a subword u, with exponent $\kappa \geq \alpha$ and period p. Then, by iteration of Theorem 2, $babcbabcv$ has a subword u' with exponent $\kappa' \geq \kappa$ and q such that $|q| = \frac{|p|}{19^s}$. By Lemma 1, as $\kappa' \geq \alpha$, we deduce that $\kappa' = 2$ and that $u' = babcbabc$. Then $q = babc$, and $|p| = |q|.19^s = 4.19^s$.
   Moreover, u is not a subword of $\mu^s(abcbabcv)$, otherwise $abcbabcv$ has a subword with exponent $\geq \alpha$. u is not a subword of $\delta^t\mu^s(babcbabc)$ either, otherwise, we would have $|u| \leq |\delta^t\mu^s(babcbabc)| = 8.19^s - t$, and so:
   $$E(u) = \frac{|u|}{|p|} \leq \frac{8.19^s - t}{4.19^s} = \beta < \alpha,$$
   which is absurd since $E(u) \geq \alpha$.
   Therefore, u has a prefix z such that $z = z_1\mu^s(abcbabc)z_2$, where $z_1 \neq \epsilon$ is a suffix of $\mu^s(b)$, and $z_2 \neq \epsilon$ is a prefix of $\mu^s(a)$ (since we remarked that the first letter of v is a a). z being a prefix of u, z has a period of length 4.19^s. Then, z has period $z_1\mu^s(abc)z_1'$, with z_1' such that $\mu^s(b) = z_1'z_1$. So we have:
   $$z = z_1\mu^s(abc)z_1'z_1\mu^s(abc)z_2.$$
   We deduce that either z_2 is a prefix of z_1', or z_1' is a prefix of z_2. Yet neither is possible. Indeed, z_2 begins with the letter a, and z_1' begins with the letter b. Finally, w is α-free.

□

4 A word over A_3 with critical exponent $\alpha \geq RT(3)$

Definition 2. *A real number $\beta < \alpha$ is said to be obtainable if β can be written as $\beta = 2 - \frac{t}{4.19^s}$, where the natural numbers s and t verify:*

- $s \geq 3$
- *the word $\delta^t(\mu^s(b))$ begins with $abcbabc$.*

We note that for any given $s \geq 3$, it is possible to choose t such that

- $7/4 < \beta = 2 - \frac{t}{4.19^s} < \alpha$
- $|\alpha - \beta| \leq \frac{19^2}{4.19^s}$

Indeed, $\mu^2(a)$, $\mu^2(b)$, and $\mu^2(c)$ have length 19^2, and each have $abcbabc$ as a subword. Therefore, choosing a large enough s, we can always find some obtainable real numbers β, arbitrarily close to α.

Theorem 2. *Let $\alpha \geq RT(3) = 7/4$. Then, there is an infinite word over A_3 with critical exponent α.*

Proof. If $\alpha = 7/4$, we already know that $\mu^\infty(a)$ has critical exponent $7/4$. If $\alpha \geq 2$, by the theorem for $k = 2$, we can find a word over A_3^* with critical exponent α. Now, let $\alpha \in]7/4, 2[$.

Let $(\beta_i)_{i\in\mathbb{N}}$ be an increasing sequence of obtainable numbers, converging to α. For each i, we write β_i as:

$$\beta_i = 2 - \frac{t_i}{4.19^{s_i}}$$

where s_i and t_i are such that:

- $s_i \geq 3$
- $\delta^{t_i}\mu^{s_i}(b)$ begins with *abcbabc*.

For all words $v \in \mathfrak{L}$, let $\Phi_i(v)$ denote the word $\delta^{t_i}\mu^{s_i}(bv)$, and consider the following sequence:

$$\begin{array}{rcl}
v_1 & = & \Phi_1(abcbabc) = \delta^{t_1}\mu^{s_1}(babcbabc) \\
v_2 & = & \Phi_1\Phi_2(abcbabc) = \delta^{t_1}\mu^{s_1}(b\delta^{t_2}\mu^{s_2}(babcbabc)) \\
v_3 & = & \Phi_1\Phi_2\Phi_3(abcbabc) \\
\vdots & & \\
v_n & = & \Phi_1\Phi_2\Phi_3...\Phi_n(abcbabc) \\
\vdots & &
\end{array}$$

By iteration of Lemma 2, as $abcbabc \in \mathfrak{L}$ is α-free, we deduce that each v_i is α-free.
Moreover, once again by Lemma 2, each v_i has a subword with exponent β_j, $j = 1, 2, ..., i$.
Finally, consider the word $w = \lim_{n\to\infty} \in A_3^\omega$ (it is possible to take this limit since each v_i is a prefix of v_{i+1}). w then has critical exponent α: it is α-free, yet has subwords with exponents β_i converging to α.

□

The conjecture proposed by Currie and Rampersad in [4] is then true for alphabets of size 2 and 3. It still have to be proved for alphabets of size ≥ 4. For that, another method must be found, because of Brandenburg's result in [2] : if $k \geq 4$, there is no $RT(k)$-free morphism, i.e., no morphism which maps, as Thue-Morse morphism for $k = 2$ or Dejean's morphism for $k = 3$, every $RT(k)$-free word to an $RT(k)$-free word.

References

[1] J. Berstel. Axel Thue's papers on repetitions in words: a translation. *Publications du LaCIM*, 20, 1995.

[2] F. J. Brandenburg. Uniformly growing k-th powerfree homomorphisms. *Theoret. Comput. Sci.*, 23:69–82, 1983.

[3] A. Carpi. On Dejean's conjecture over large alphabets. *Theoret. Comput. Sci.*, 385:137–151, 2007.

[4] J. D. Currie and N. Rampersad. For each $\alpha > 2$ there is an infinite binary word with critical exponent α. *Electron. J. Combin.*, 15(34), 2008.

[5] J. D. Currie and N. Rampersad. A proof of Dejean's conjecture. *Preprint*, http://arxiv.org/abs/0905.1129, 2009.

[6] F. Dejean. Sur un théorème de Thue. *J. Comb. Theory A*, 13:90–99, 1972.

[7] D. Krieger and J. Shallit. Every real number greater than 1 is a critical exponent. *Theoret. Comput. Sci.*, 381:177–182, 2007.

[8] F. Mignosi and G. Pirillo. Repetitions in the Fibonacci infinite word. *RAIRO Inform. Theor. Appl.*, 26:199–204, 1992.

[9] M. Rao. Last cases of Dejean's conjecture. *WORDS 2009*, http://www.labri.fr/perso/rao/publi/dejean.ps, 2009.

[10] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. Mat. Nat. Kl.*, 10:1–67, 1912.

Annexe C

Sur les mots hautement répétitifs d'exposant critique fini

N. Rampersad, E. Vaslet, **On highly repetitive and power free words**,
à paraître dans *Lecture Notes in Computer Science.*

On highly repetitive and power free words

Narad Rampersad[1] and Elise Vaslet[2]

[1] Department of Mathematics, University of Liège, Belgium
narad.rampersad@gmail.com
[2] Institut de Mathématiques de Luminy, Université Aix-Marseille II, France
vaslet@iml.univ-mrs.fr

Abstract. Answering a question of Richomme, Currie and Rampersad proved that 7/3 is the infimum of the real numbers $\alpha > 2$ such that there exists an infinite binary word that avoids α-powers but is highly 2-repetitive, i.e., contains arbitrarily large squares beginning at every position. In this paper, we prove similar statements about β-repetitive words, for some other β's, on the binary and the ternary alphabets.

1 Introduction

In this paper we study words that are non-repetitive, in the sense that they avoid α-powers for some α, but yet are highly repetitive, in the sense that for some β close to α, they contain infinitely many β-powers starting at every position. First we recall some basic definitions. A finite, non-empty word v can be written as $v = p^k e$, where $k \geq 1$, the word e is a prefix of p, and the length of p is minimal. We say that v has *period* p, *excess* e, and *exponent* $\text{expo}(v) = |v|/|p|$. A word with exponent α is called an α-power. An α^+-*power* is a word that is a β-power for some $\beta > \alpha$. A 2-power is called a *square*; a 2^+-power is called an *overlap*. A word is α-*free* if none of its factors is a β-power for any $\beta \geq \alpha$. A word is α^+-*free* if none of its factors is an α^+-power.

Thue [TH2] proved that there exist infinite overlap-free binary words. Dekking [DEK] later showed that any infinite overlap-free binary word must contain arbitrarily large squares. Currie, Rampersad, and Shallit [CRS] constructed 7/3-free words containing infinitely many overlaps. Their motivation came from the following result of Shur [SHU]: Any bi-infinite 7/3-free binary word is overlap-free. The analogue of this surprising result is not true for one-sided infinite words: There are infinite 7/3-free binary words that are not overlap-free. The result of Currie et al. shows that in fact the infinite 7/3-free binary words can be significantly different from the infinite overlap-free binary words.

Currie and Rampersad [CR] continued this line of study in order to respond to the following question of Richomme [RIC]: What is the infimum of the real numbers $\alpha > 2$ such that there exists an infinite word that avoids α-powers but contains arbitrarily large squares beginning at every position? They showed that over the binary alphabet the answer to Richomme's question is $\alpha = 7/3$. In this paper we show that if instead of requiring arbitrarily large squares at every position, we only ask for arbitrarily large β-powers at every position for some fixed $\beta < 2$, then the answer of 7/3 for Richomme's question can be replaced by 2. We also answer the analogous question for the ternary alphabet.

We also mention here the related work of Saari [SAA], who also studied infinite words containing squares (not necessarily arbitrarily large) beginning at every position. He called such words *squareful*, but he imposed the additional condition that the word contain only finitely many distinct minimal squares.

2 Preliminary definitions and results

In this paper, A_k will denote the k-letter alphabet, for an integer k.
Let α be a real number.

Definition 1. *A finite word is an α-repetition (resp. α^+-repetition) if it is a β-power for some $\beta \geq \alpha$ (resp. $\beta > \alpha$).*

Definition 2. *A word (finite or infinite) is highly α-repetitive (resp. highly α^+-repetitive) if it contains infinitely many α-repetitions (resp. α^+-repetitions) beginning at every position.*

Definition 3. *A word (finite or infinite) is α-free (resp. α^+-free) if it contains no α-repetition (resp. α^+-repetition).*

Definition 4. *A morphism $\mu : A^* \to B^*$ is α-free (resp. α^+-free) if for any word $w \in A^*$, the following equivalence holds:*

$$w \text{ is } \alpha\text{-free} \Leftrightarrow \mu(w) \text{ is } \alpha\text{-free.}$$

We now define three particular morphisms which we will use in the paper because of their α-freeness for certain α.
The Thue-Morse morphism is defined by

$$\mu_{TM} : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10. \end{cases}$$

Proposition 1 ([TH2]). *μ_{TM} is a 2^+-free morphism.*

Dejean's morphism is defined by

$$\mu_D : \begin{cases} a \mapsto abc\ acb\ cab\ c\ bac\ bca\ cba \\ b \mapsto \pi(\mu(a)) = bca\ bac\ abc\ a\ cba\ cab\ acb \\ c \mapsto \pi^2(\mu(a)) = cab\ cba\ bca\ b\ acb\ abc\ bac, \end{cases}$$

where π is the morphism that cyclically permutes the alphabet $\{a, b, c\}$.

Proposition 2 ([DEJ]). *μ_D is a $7/4^+$-free morphism.*

The following morphism was given by Brandenburg in [BRA]:

$$\mu_B : \begin{cases} a \mapsto abacbabcbac \\ b \mapsto abacbcacbac \\ c \mapsto abcbabcacbc. \end{cases}$$

Proposition 3 ([BRA]). *μ_B is a 2-free morphism.*

The results that follow will only be used in Section 5. They concern Dejean's morphism μ_D and its behaviour regarding repetitions and synchronizing properties, and were proved in [VAS]. If v is a word, we denote by δ the application that removes the first letter of v (and so, $\delta^t(v)$ removes the first t letters of v). Moreover, we denote by $\mathfrak{F}_D$ the set $\mathrm{Fact}(\mu_D(A_3^*))$ of factors of $\mu_D(A_3^*)$.

Lemma 1 ([VAS]). *Let $\alpha \in]7/4, 2[$ be given. Let s, t be natural numbers such that $\mu_D^s(b) = xabcbabcy$, with $|x| = t$. Let $\beta = 2 - \frac{t}{4.19^s}$. Suppose that $7/4 < \beta < \alpha$, and that $abcbabcv \in \mathfrak{F}_D$ is α-free. Consider the word $w = \delta^t \mu_D^s(babcbabcv)$. Then, we have:*

1. *w has a prefix with exponent β.*
2. *If $abcbabcv$ has a factor with exponent γ and period p, then, w has a factor with exponent γ and a period of length $19^s|p|$.*
3. *w is α-free.*

Definition 5 ([VAS]). *A real number $\beta < \alpha$ is said to be obtainable if β can be written as $\beta = 2 - \frac{t}{4.19^s}$, where the natural numbers s and t verify:*

* *$s \geq 3$*
* *the word $\delta^t(\mu_D^s(b))$ begins with abcbabc.*

We note that for any given $s \geq 3$, it is possible to choose t such that

* $7/4 < \beta = 2 - \frac{t}{4.19^s} < \alpha$
* $|\alpha - \beta| \leq \frac{19^2}{4.19^s}$.

Indeed, $\mu_D^2(a)$, $\mu_D^2(b)$, and $\mu_D^2(c)$ have length 19^2, and each has *abcbabc* as a factor. Therefore, choosing a large enough s, we can always find some obtainable real numbers β arbitrarily close to α.

3 Highly repetitive binary words

The results of Currie and Rampersad in [CR] Theorems 4 and 5 state that for any $\alpha > 7/3$, there exists an infinite binary word which is both α-free and highly 2-repetitive, and that such a word does not exist if $\alpha \leq 7/3$. The following proposition gives a similar result in the case of highly β-repetitive words, with $\beta < 2$: for any $\alpha > 2$ and any $\beta < 2$, there exists an infinite binary word which is both α-free and highly β-repetitive.

Theorem 1. *For any real number $\beta < 2$, there exists an infinite binary word which is both 2^+-free and highly β-repetitive.*

Proof. Let $\beta < 2$ be a real number. There exists an integer $m \geq 1$ such that $\beta \leq 2 - \frac{1}{2^m} < 2$. Let $r = 2 - \frac{1}{2^m}$, and let us construct a binary infinite word with infinitely many r-powers at every position, and which is 2^+-free. We recall that μ_{TM} is the Thue-Morse morphism.
We remark that $\mu_{TM}^m(11)$ is a factor of $\mu_{TM}^{m+3}(1)$. Indeed, $\mu_{TM}^3(1) = 10010110$, and so $\mu_{TM}^{m+3}(1) = \mu_{TM}^m(10010110) = u\mu_{TM}^m(11)\mu_{TM}^m(0)$, where $u = \mu_{TM}^m(10010)$. Let us define $q = m + 3$. We define the following sequence of finite words:

$$\begin{aligned} A_0 &= \mu_{TM}^m(11) \\ A_1 &= (u)^{-1}\mu_{TM}^q(A_0) \\ &\vdots \\ A_n &= (u)^{-1}\mu_{TM}^q(A_{n-1}), \end{aligned}$$

where we denote by $(x)^{-1}y$ the word obtained by removing the prefix x from the word y (here, the word $\mu_{TM}^q(A_n)$ always has u as a prefix, since A_n always has 1 as a prefix). As 11 and μ_{TM} are 2^+-free, then for any $n \geq 0$, A_n is 2^+-free.
We show that for any $n \geq 0$, A_n is a prefix of A_{n+1}. This is clearly true for $n = 0$. If A_{n-1} is a prefix of A_n, then we can write $A_n = A_{n-1}S$, where S is a binary word. Then, $A_{n+1} = (u)^{-1}\mu_{TM}^q(A_{n-1}S) = A_n\mu_{TM}^q(S)$. Thus, the sequence A_n converges to a infinite limit

word we denote by w, and this limit is 2^+-free. We now prove that w contains infinitely many r-powers beginning at every position. It can easily be seen that for any $n \geq 1$,

$$A_n = (u)^{-1}[\mu_{TM}^{q}(u)]^{-1} \cdots [\mu_{TM}^{q(n-1)}(u)]^{-1}\mu_{TM}^{qn}(A_0).$$

Let us denote by p_n the word $\mu_{TM}^{q(n-1)}(u) \cdots \mu_{TM}^{q}(u)u$. Then we have $A_n = (p_n)^{-1}\mu_{TM}^{qn}(A_0)$, since for any words u and v, $(uv)^{-1} = (v)^{-1}(u)^{-1}$. Moreover, we can see that p_n is a prefix of $\mu_{TM}^{qn}(1)$. Indeed,

$$\begin{aligned} |p_n| &= |u| \sum_{i=0}^{n-1} |\mu_{TM}(1)|^{qi} \\ &= |u| \sum_{i=0}^{n-1} 2^{qi} \\ &< 2^{qn} = |\mu_{TM}^{qn}(1)|, \end{aligned}$$

since $|u| \leq 2^q - 1$. So, we can write $\mu_{TM}^{qn}(1) = p_n s_n$, where s_n is a binary word. For any $n \geq 1$, $A_n = (p_n)^{-1}\mu_{TM}^{qn}(A_0) = (p_n)^{-1}\mu_{TM}^{qn}(\mu_{TM}^{m}(11))$. We know that $\mu_{TM}^{m}(1)$ begins with the letter 1, and let us denote its last letter by $a \in \{0,1\}$. We can write $\mu_{TM}^{m}(1) = 1xa$, where $x \in \{0,1\}^*$. We can also write $\mu^{qn}(a) = \tilde{p_n}\tilde{s_n}$, with $\tilde{p_n}$ and $\tilde{s_n}$ two binary words such that $|\tilde{p_n}| = |p_n|$ and $|\tilde{s_n}| = |s_n|$. Then, $A_n = s_n\mu_{TM}^{qn}(x)\tilde{p_n}\tilde{s_n}p_n s_n\mu_{TM}^{qn}(x)\tilde{p_n}\tilde{s_n}$. It follows that at all positions $0 \leq j \leq |s_n|$, there begins a repetition with period a conjugate of $s_n\mu_{TM}^{qn}(x)\tilde{p_n}\tilde{s_n}p_n$, and of length $|s_n\mu_{TM}^{qn}(x)\tilde{p_n}\tilde{s_n}p_n s_n\mu_{TM}^{qn}(x)\tilde{p_n}| = 2.|x|.2^{qn} + 3.2^{qn} = 2^{qn}(2^{m+1} - 1)$, since $|x| = |\mu_{TM}^{m}(1)| - 2 = 2^m - 2$. These repetitions have exponents $\frac{2^{qn}(2^{m+1}-1)}{2^{qn}|x|+2.2^{qn}} = \frac{2^{m+1}-1}{2^m} = r$. As n can be taken arbitrarily large, the result follows. □

Moreover, this result is optimal. Indeed, on the one hand, it is easy to see that for $\alpha \leq 2$, there is no infinite binary α-free word. On the other hand, Currie and Rampersad proved that

Proposition 4 ([CR]). *If w is an infinite overlap-free binary word, then there is a position i such that w does not contain a square beginning at position i.*

4 Highly repetitive ternary words

In this part, we will state some similar results in the case of the ternary alphabet. We recall that the repetition threshold of the ternary alphabet is $7/4$, so for any $\alpha \leq 7/4$, there is no infinite α-free ternary word.

Theorem 2. *For any real number $\beta < 7/4$, there exists an infinite ternary word that is both $7/4^+$-free and highly β-repetitive.*

Proof. For any real number $\beta < 7/4$, there exists an integer $m \geq 1$ such that $\beta \leq 7/4 - \frac{1}{19^{m+2}} < 7/4$. Let $r = 7/4 - \frac{1}{19^{m+2}}$, and let us construct a $7/4^+$-free word which contains infinitely many r-powers at every position. We recall that μ_D is Dejean's morphism.
We remark that $\mu_D^m(abcbabc)$ is a factor of $\mu_D^{m+2}(a)$. Indeed,

$$\mu_D^2(a) = \mu_D(a)\mu_D(b)cabcbabcabacbabcbac\mu_D(acbcabcbacbcacba),$$

and so

$$\mu_D^{m+2}(a) = u\mu_D^m(abcbabc)\mu_D^m(abacbabcbac)\mu_D^{m+1}(acbcabcbacbcacba),$$

where $u = \mu_D^{m+1}(ab)\mu_D^m(c)$. Let us define $q = m + 2$.
We define the following sequence of finite words:

$$\begin{aligned}
A_0 &= \mu_D^m(abcbabc)\\
A_1 &= (u)^{-1}\mu_D^q(A_0)\\
&\vdots\\
A_n &= (u)^{-1}\mu_D^q(A_{n-1}).
\end{aligned}$$

(Here, the word $\mu_D^q(A_n)$ always has u as a prefix, since A_n always has a as a prefix). As $abcbabc$ and μ_D are $7/4^+$-free, then for any $n \geq 0$, A_n is $7/4^+$-free.
Similar arguments as in the proof of Theorem 1 prove that the sequence A_n converges to an infinite limit word we denote by w, and this limit is $7/4^+$-free. We now prove that w contains infinitely many α-powers beginning at every position. Again with similar arguments as in the proof of Theorem 1, we can see that for any $n \geq 1$,

$$A_n = (p_n)^{-1}\mu_D^{qn}(A_0),$$

where p_n is the word $\mu_D^{q(n-1)}(u)\cdots\mu_D^q(u)u$. Moreover, we can see that p_n is a prefix of $\mu_D^{qn}(a)$. Indeed,

$$\begin{aligned}
|p_n| &= |u|\sum_{i=0}^{n-1}|\mu_D(a)|^{qi}\\
&= |u|\sum_{i=0}^{n-1}19^{qi}\\
&< 19^{qn} = |\mu_D^{qn}(a)|.
\end{aligned}$$

So, we can write $\mu_D^{qn}(a) = p_n s_n$, where s_n is a ternary word. For any $n \geq 1$, $A_n = (p_n)^{-1}\mu_D^{qn}(A_0) = (p_n)^{-1}\mu_D^{qn}(\mu_D^m(abcbabc))$. We know that $\mu_D^m(a)$ begins with the letter a, and let us denote its last letter by $x \in \{a,b,c\}$. We can write $\mu_D^m(a) = aXx$, where $X \in \{a,b,c\}^*$. Similarly, let us write $\mu_D^m(c) = cYy$, where $Y \in \{a,b,c\}^*$, and $y \in \{a,b,c\}$. We can also write $\mu_D^{qn}(y) = \tilde{p_n}\tilde{s_n}$, with $\tilde{p_n}$ and $\tilde{s_n}$ two ternary words such that $|\tilde{p_n}| = |p_n|$ and $|\tilde{s_n}| = |s_n|$. Then,

$$A_n = s_n\mu_D^{qn}(Xx\mu_D^m(b)cY)\tilde{p_n}\tilde{s_n}\mu_D^{qn+m}(b)p_n s_n\mu_D^{qn}(Xx\mu_D^m(b)cY)\tilde{p_n}\tilde{s_n}.$$

It follows that at all positions $0 \leq j \leq |s_n|$, there begins a repetition with period a conjugate of $s_n\mu_D^{qn}(Xx\mu_D^m(b)cY)\tilde{p_n}\tilde{s_n}\mu_D^{qn+m}(b)p_n$, and of length

$$\begin{aligned}
l &= |s_n\mu_D^{qn}(Xx\mu_D^m(b)cY)\tilde{p_n}\tilde{s_n}\mu_D^{qn+m}(b)p_n s_n\mu_D^{qn}(Xx\mu_D^m(b)cY)\tilde{p_n}|\\
&= 2.|Xx\mu_D^m(b)cY|.19^{qn} + 19^{qn+m} + 3.19^{qn}\\
&= 19^{qn}.(7.19^m - 1),
\end{aligned}$$

since $|X| = |Y| = |\mu_D^m(a)| - 2 = 19^m - 2$. These repetitions have exponent

$$\frac{19^{qn}(7.19^m - 1)}{19^{qn}|Xx\mu_D^m(b)cY| + 19^{qn+m} + 2.19^{qn}} = \frac{7.19^m - 1}{4.19^m} = \alpha.$$

As n can be taken arbitrarily large, the result follows. □

Now, we will consider the problem for highly 7/4-repetitive words. This would be Richomme's question for the ternary alphabet : what is the infimum of the real numbers α such that there exists an infinite ternary word which is both α-free and highly 7/4-repetitive?

Proposition 5. *There is no $7/4^+$-free ternary word which is highly $7/4$-repetitive.*

Proof. Let us suppose there exists such a word w. Then, any factor of w is left-special. Indeed, let u be a factor of w, and let i denote its position in w. w is recurrent because it is highly repetitive, so we can suppose $i \geq 1$. There is a long enough $7/4$-repetition v beginning at position i. We denote its period by p. Then, u is repeated in the excess of v : pu is a factor of v. Now, let $x_1 = w_{i-1}$ and $x_2 = p_{|p|}$. It is clear that $x_1 \neq x_2$, since otherwise, $x_1 v$ would be a $7/4^+$-repetition, which is impossible since w is $7/4^+$-free. So u has two left extensions. Moreover, as the factors aa, bb and cc are clearly forbidden, we can see that w contains the factor $abab$ (by extending b on the left). But $abab$ is a square, which contradicts the fact that w is $7/4^+$-free. □

Theorem 3. *There exists an infinite ternary word which is both 2-free and highly $7/4$-repetitive.*

Proof. We use the morphism of Brandenburg, which is a square-free morphism :

$$\mu_B : \begin{cases} a \mapsto abacbabcbac \\ b \mapsto abacbcacbac \\ c \mapsto abcbabcacbc. \end{cases}$$

We use it to give a similar construction as in the proofs of Theorems 1 and 2. We consider the sequence:

$$\begin{aligned} A_0 &= cbac\mu_B(cbc)abac = cbacabcbabcacbcabacbcacbacabcbabcacbcabac \\ A_1 &= (u)^{-1}\mu_B^3(A_0) \\ &\vdots \\ A_n &= (u)^{-1}\mu_B^3(A_{n-1}), \end{aligned}$$

where $u = \mu_B^2(a)\mu_B(ab)abacbab$. The sequence A_n converges to a limit we denote by w, which is a square-free word because μ_B is square-free. For any $n \geq 1$, if we denote by p_n the word $\mu_B^{3(n-1)}(u) \cdots \mu_B^3(u)u$, we can see that $A_n = (p_n)^{-1}\mu_B^{3n}(A_0)$. But p_n is a prefix of $\mu_B^{3n}(c)$, since

$$\begin{aligned} |p_n| &= |u| \sum_{i=0}^{n-1} |\mu_B(a)|^{3i} \\ &= |u| \sum_{i=0}^{n-1} 11^{3i} \\ &< 11^{3n} = |\mu_B^{3n}(c)|. \end{aligned}$$

So, we can write $\mu_B^{3n}(c) = p_n s_n$, where s_n is a ternary word. Then, for any $n \geq 1$,

$$\begin{aligned} A_n &= (p_n)^{-1}\mu_B^{3n}(A_0) \\ &= s_n\mu_B^{3n}(bacabcbabcacbcaba)p_n s_n \mu_B^{3n}(bca)p_n s_n \mu_B^{3n}(bacabcbabcacbcaba)p_n s_n. \end{aligned}$$

It follows that at all positions $0 \leq j \leq |s_n|$, there begins a repetition with period a conjugate of $s_n\mu_B^{3n}(bacabcbabcacbcaba)p_n s_n \mu_B^{3n}(bca)p_n$, and of length

$$\begin{aligned} l &= |s_n\mu_B^{3n}(bacabcbabcacbcaba)p_n s_n \mu_B^{3n}(bca)p_n s_n \mu_B^{3n}(bacabcbabcacbcaba)p_n| \\ &= 40.11^{3n}. \end{aligned}$$

These repetitions have exponents $40/22 = 20/11 \geq 7/4$. As n can be taken arbitrarily large, the result follows.

Remark 1. In fact, with this proof, we have the result: There exists an infinite 2-free ternary word which is highly 20/11-repetitive.

So the problem of finding the infimum of the real numbers α such that there exists an infinite ternary α^+-free word which is highly 7/4-repetitive is open. We proved that this infimum is between 7/4 and 2.

5 Weaker results on ternary words

In the following section, we will prove some weaker results for words over the ternary alphabet. First, instead of considering highly repetitive words, we consider words with powers at every positions. Then, we will discuss words containing infinitely many powers, but not necessary at every position. These questions were asked at the end of [CR] by Currie and Rampersad.

Theorem 4. *For every real number $\alpha > 7/4$, there exists an infinite α-free ternary word that contains 7/4-powers beginning at every position.*

Proof. Theorem 3.4 in [CR] established the result for any $\alpha > 2$. Let us consider $7/4 < \alpha \leq 2$. Let β be an obtainable number (see Definition 5), that is, it can be written as:

$$\beta = 2 - \frac{t}{4.19^s},$$

where $s \geq 3$, and $\delta^t \mu_D^s(b)$ begins with *abcbabc*.
For any word v in $\mathfrak{F}_D$, we denote by $\Phi(v)$ the word $\delta^t \mu_D^s(bv)$ (we recall that $\delta(u)$, for some word u, removes the first letter of u) and we consider the sequence:

$$\begin{aligned} v_1 &= \Phi(abcbabc) = \delta^t \mu_D^s(babcbabc) \\ v_2 &= \Phi(v_1) = \delta^t \mu_D^s(b\delta^t \mu_D^s(babcbabc)) \\ &\vdots \\ v_n &= \Phi(v_{n-1}) \\ &\vdots \end{aligned}$$

Let $w = \lim v_n$ (possible because each v_i is clearly a prefix of v_{i+1}). By iteration of Lemma 1, as *abcbabc* is α-free, each v_i is α-free, and then w is α-free.
For any $n \geq 0$, w begins with a prefix of the form

$$\delta^t \mu_D^s(b) \mu_D^s(\delta^t \mu_D^s(b)) \cdots \mu_D^{ns}(\delta^t \mu_D^s(b)) \mu_D^{(n+1)s}(abcbabc).$$

Thus, w contains the word $\mu_D^{ns}(\delta^t(\mu_D^s(b)) \mu_D^{(n+1)s}(abcbabc)$ at position

$$\begin{aligned} p_n &= |\delta^t \mu_D^s(b) \mu_D^s(\delta^t \mu_D^s(b)) \cdots \mu_D^{(n-1)s}(\delta^t \mu_D^s(b))| \\ &= \sum_{i=1}^{n} |\mu_D^{is}(b)| - \sum_{i=0}^{n-1} t.19^{is} \\ &= (19^s - t) \sum_{i=0}^{n-1} 19^{is} \\ &= (19^s - t) \frac{19^{ns} - 1}{19^s - 1} \end{aligned}$$

Moreover, the word $\mu_D^{ns}(\delta^t\mu_D^s(b))\mu_D^{(n+1)s}(abcbabc)$ contains 7/4-powers at every position between 1 and $q_n = |\mu_D^{ns}(\delta^t\mu_D^s(b))| = (19^s - t).19^{ns}$. So, in w, there is a 7/4-repetition starting at every position j, for every $j \in [p_n, p_n + q_n - 1]$. Since $p_{n+1} = p_n + q_n$, every position j in $\mathbb{N}$ is reached. □

Theorem 5. *For every real number $\alpha \geq 7/4$ there exists a real number β, with $7/4 \leq \beta < \alpha$, arbitrarily close to α, such that there is an infinite β^+-free ternary word containing infinitely many β-powers.*

Proof. Theorem 3.4 in [CR] established the result for any $\alpha > 2$. Let us consider $7/4 < \alpha \leq 2$. Let β be an obtainable number, that is, it can be written as:

$$\beta = 2 - \frac{t}{4.19^s},$$

where $s \geq 3$, and $\delta^t\mu_D^s(b)$ begins with $abcbabc$.
Similarly to the proof of Theorem 4, we consider the sequence:

$$\begin{aligned}
v_1 &= \Phi(abcbabc) = \delta^t\mu_D^s(babcbabc)\\
v_2 &= \Phi(v_1) = \delta^t\mu_D^s(b\delta^t\mu_D^s(babcbabc))\\
&\vdots\\
v_n &= \Phi(v_{n-1})\\
&\vdots
\end{aligned}$$

Let $w = \lim v_n$. w is α-free, and it is β^+-free, since $\beta < \alpha$. Moreover, for any $n \geq 1$, w contains the word $\mu_D^{ns}(\delta^t\mu_D^s(babcbabc))$, which has length $19^{ns}(8.19^s - t)$ and exponent β. □

6 Conclusion

We propose the following definition. For an integer k, consider the set

$$\mathfrak{S}_k = \{(\alpha,\beta) \in \mathbb{R}^2 : \text{there exists a word in } A_k^\omega \text{ which is } \alpha\text{-free and highly } \beta\text{-repetitive}\}.$$

The combined results of Currie and Rampersad [CR], and of Sections 3 and 4 of this paper, give a partial desciption of $\mathfrak{S}_k$ for $k = 2$ and $k = 3$. This is summarized in Figures 1 and 2.

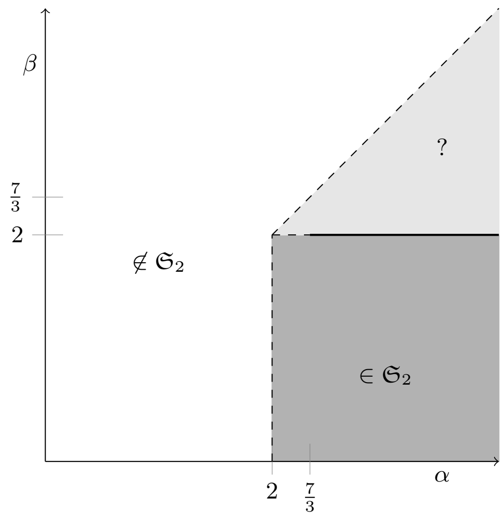


Fig. 1. $k = 2$

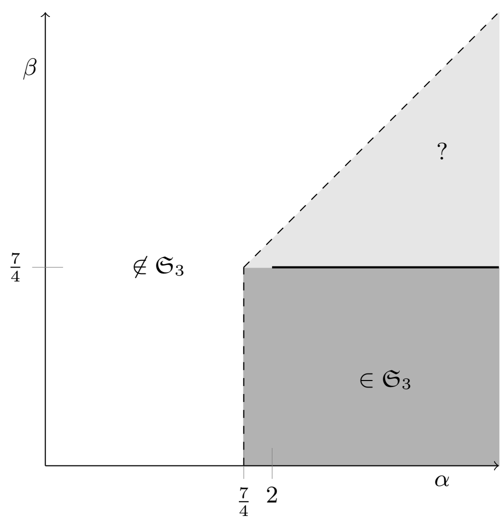


Fig. 2. $k = 3$

References

[BER] J. Berstel, "Axel Thue's papers on repetitions in words: a translation", *Publications du LACIM 20*, 1994.

[BRA] F. J. Brandenburg, "Uniformly growing k-th powerfree homomorphisms", *Theoret. Comput. Sci.* 23 (1983), 69–82.

[CAR] A. Carpi, "On Dejean's conjecture over large alphabets". *Theor. Comput. Sci.* 385 (2007), 137–151.

[CUR] J. Currie, N. Rampersad, "For each $\alpha > 2$ there is an infinite binary word with critical exponent α", *Electron. J. Combinatorics* 15 (2008), #N34.

[CURR] J. Currie, N. Rampersad, "Dejean's conjecture holds for $n \geq 30$", *Theoret. Comput. Sci.* 410 (2009), 2885–2888.

[CR] J. Currie, N. Rampersad, "Infinite words containing squares at every position", *Theor. Inform. Appl.* 44 (2010), 113–124.

[CRS] J. Currie, N. Rampersad, J. Shallit, "Binary words containing infinitely many overlaps", *Electron. J. Combinatorics* 13 (2006), #R82.

[DEJ] F. Dejean, "Sur un théorème de Thue", *J. Combin. Theory Ser. A* 13 (1972).

[DEK] F. M. Dekking, "On repetitions in binary sequences", *J. Comb. Theory Ser. A* 20 (1976), 292–299.

[KRI] D. Krieger, J. Shallit, "Every real number greater than 1 is a critical exponent", *Theoret. Comput. Sci.* 381 (2007).

[MIG] F. Mignosi, G. Pirillo, "Repetitions in the Fibonacci infinite word", *RAIRO Inform. Theor. Appl.* 26 (1992).

[PAN] J.-J. Pansiot, "A propos d'une conjecture de F. Dejean sur les répétitions dans les mots", *Discrete Appl. Math.* 7 (1984).

[RIC] G. Richomme. Personal communication, 2005.

[SAA] K. Saari, "Everywhere α-repetitive sequences and Sturmian words", *Europ. J. Combin.* 31 (2010), 177–192.

[SHU] A. M. Shur, "The structure of the set of cube-free $\mathbb{Z}$-words in a two-letter alphabet" (Russian), *Izv. Ross. Akad. Nauk Ser. Mat.* **64** (2000), 201–224. English translation in *Izv. Math.* **64** (2000), 847–871.

[TH1] A. Thue, "Über unendliche Zeichenreihen", *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana No.7* (1906).

[TH2] A. Thue, "Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen", *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana No.10* (1912).

[VAS] E. Vaslet, "Critical exponents of words over 3 letters", submitted.

Annexe D

Sur les colorations de graphes sans répétitions

P. Ochem, E. Vaslet, **Repetition threshold for subdivided graphs and trees**,
soumis à *Theoretical Informatics and Applications.*

Theoretical Informatics and Applications Will be set by the publisher
Informatique Théorique et Applications

REPETITION THRESHOLDS FOR SUBDIVIDED GRAPHS AND TREES

Pascal Ochem[1] and Elise Vaslet[2]

Abstract. The *repetition threshold* introduced by Dejean and Brandenburg is the smallest real number α such that there exists an infinite word over a k-letter alphabet that avoids β-powers for all $\beta > \alpha$. We extend this notion to colored graphs and obtain the value of the repetition thresholds of trees and "large enough" subdivisions of graphs for every alphabet size.

AMS Subject Classification. — Give AMS classification codes —.

1. Introduction

A non-repetitive coloring f of a graph is a vertex coloring containing no square, that is, there is no simple path $v_1, \ldots, v_{2r}$ such that $f(v_i) = f(v_{i+r})$ for all $i \in [1, r]$.

This notion can be extended by considering repetitions of fractional exponent, as it has been done in the framework of combinatorics on words.

Up to now, the most studied problem is the following: we fix the exponent (i.e., exponent 2, corresponding to squares) to be avoided and minimize the number of colors, i.e., the alphabet size. For large subdivisions of graphs, this so-called non-repetitive chromatic number is proved to be 3 [5]. The non-repetitive chromatic number of trees is 4 [3].

Aberkane and Currie [1] considered the problem the other way. They fix the alphabet size and study what exponents can be avoided. They show that there exist binary circular words of every length avoiding exponents $> 5/2$. The cycle of length 5 shows that this is the best possible bound.

[1] CNRS, LRI, Université Paris-Sud 11, 91405 Orsay, France
`ochem@lri.fr`

[2] IML, UMR 6206, Université Aix-Marseille II,
Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France
`vaslet@iml.univ-mrs.fr`

A notion of repetition threshold (similar to Dejean's repetition threshold [2]) can then be defined for graphs. Aberkane and Currie 's result sets it to 5/2 for cycles.

We settle the problem of finding the repetition thresholds of two graph classes, namely the subdivided graphs and the trees, for every alphabet size k. This improves earlier results on nonrepetitive coloring on these classes [3, 5].

2. Preliminaries

Let $k \in \mathbb{N}$ and let f be a k-coloring of a graph G. We call *factor* the sequence of colors on a non-intersecting path in a k-colored graph G. We recall the usual notions of period and exponent: a word w is a repetition with period p and excess e if $w = pe$, and e is a prefix of w. The exponent of the repetition w is the ratio $exp(w) = \frac{|pe|}{|p|}$.

Let G be a graph. We define the repetition threshold $RT(k, G)$ by

$$\inf_{k\text{-coloring } f} \sup \{exp(w) \mid w \text{ is a factor in the } k\text{-coloring } f \text{ of } G\}.$$

For a graph class $\mathcal{G}$, we define $RT(k, \mathcal{G}) = \sup_{G \in \mathcal{G}} RT(k, G)$. Let us express some known results with this definition: $RT(2, \mathcal{C}) = 5/2$ [1] and $RT(4, \mathcal{T}) \leq 2$ [3], where $\mathcal{C}$ the class of cycles and by $\mathcal{T}$ the class of trees. A subdivision of a graph G is a graph obtained from G by a sequence of edge subdivisions. A subdivision of an edge $\{u, v\}$ consists in the addition of a new vertex w and the replacement of the edge $\{u, v\}$ by the edges $\{u, w\}$ and $\{w, v\}$. It has been shown that for every graph G, there exists a subdivision G_s of G such that $RT(3, G_s) < 2$ [5].

The Thue-Morse word $w_{TM} = 0110100110010110100101100 11010\ldots$ is the fixed point of $0 \to 01,\ 1 \to 10$.

$w[i, \ldots, j]$ denotes the subword $w_i w_{i+1} \ldots w_j$ of the word w.

3. Repetition threshold for subdivided graphs

For a given k, we define the real number α such that:

- $\forall G, \exists G_s$ a subdivision of G s.t. $RT(k, G_s) \leq \alpha$
- $\exists G, \forall G_s$ a subdivision of G s.t. $RT(k, G_s) \geq \alpha$

By abuse of notation, we set $RT(k, \mathcal{S}) = \alpha$. We can see $\mathcal{S}$ as the pseudo-class of "large enough" subdivisions of graphs.

Theorem 1.

(1) $RT(2, \mathcal{S}) = \frac{7}{3}$
(2) $RT(3, \mathcal{S}) = \frac{7}{4}$
(3) $RT(k, \mathcal{S}) = \frac{3}{2}$, *for* $k \geq 4$.

Proof. We first prove the upper bounds. For any graph G, we construct a subdivision G_s of G and a suitable coloring of G_s. Without loss of generality, we can suppose that G is a (large) clomplete graph. For each k, we consider a set W

of words over the k-letter alphabet such that all words in W have a non-empty common prefix and a non-empty common suffix. An edge in G is replaced by a chain of vertices of degree 2 in G_s. A vertex of degree at least 3 in G_s thus corresponds to an original vertex in G and is said *big*. Such a chain consists of two half-chains. We color each half-chain with a distinct word $w \in W$ such that the first letter of w corresponds to the color of a big vertex. The last vertex of an half-chain is identified with the last vertex of the other half-chain of a same chain. This common vertex is the called the *center* of the chain and is colored with the common last letter of all words in W. A vertex is said to be *special* if it is a big vertex or a center. Two half chains sharing a special vertex are said to be *opposite* since they have opposite reading directions.

Case $k = 2$:
We consider a set W of (distinct) factors of the Thue-Morse word with prefix 0110010 and suffix 1011001. Suppose that there exists a repetition r in G_s with exponent $> \frac{7}{3}$. If r contains at most one big vertex (resp. at most one center), then either r is a factor of w_{TM} or r contains exactly one occurrence of at least one factor in $\{01001101, 10110010\}$. Those factors appear near the big vertex or the center because of the prefix 0110010 and the suffix 1011001 in the half chains, and since they are not factors of w_{TM}, we have a contradiction. If r contains at least two big vertices and at least two centers, then two half-chains should be matched in r. This is a contradiction because half-chains are distinct.

Case $k = 3$:
Let e denote the number of edges of G. Consider a set S of size $2e$ containing (distinct) $\left(\frac{7}{5}^+\right)$-free words over Σ_4 of length d with prefix 0123 and suffix 1230, such that the factors of length $d/5$ of words in S are distinct. Such a length d exists since there are exponentially many $\left(\frac{7}{5}^+\right)$-free words over Σ_4 [4]. Consider also the following 48-uniform morphism m:

$0 \mapsto 012010201202120121020102120121012010201210212021$
$1 \mapsto 012010201202120121012021020102120121012010212021$
$2 \mapsto 012010201202101210212012101202120102012021020121$
$3 \mapsto 012010201202101210201202120121012010212021020121$

The set W contains the words of the form $w = m(x)01201020120$, with $x \in S$. The prefixes and suffixes of length 11 and the factors of length 52 of words in W contain the factor 01201, whereas words in W do not contain the mirror factor 10210. Using Lemma 3.1 in [4], we can check that the words in W are $\left(\frac{7}{4}^+\right)$-free and $\left(\frac{3}{2}^+, 14\right)$-free.

Suppose that there exists a repetition $r = uvu$ in G_s with exponent $> \frac{7}{4}$, that is, $|u| > 3|v|$. Obviously, r cannot be contained in one half-chain. A computer check shows that the period $p = |uv|$ of r must be at least 68, by looking at the neighborhood of a special vertex. We first consider the case where r is contained in two (opposite) half-chains. If the common special vertex belongs to v, then $|uv| \geq 68$ and $|u| > 3|v|$ gives $|u| \geq 52$. This is a contradiction since the factor 01201 appearring in one occurence of u would induce a forbidden factor 10210 in the word in W corresponding to the opposite chain containing the other ocurrence of u. If the common special vertex belongs to u, then the restriction of r to one of the half-chain is a repetition of period $p \geq 68$ and exponent at most $\frac{3}{2}$. To complete it and obtain the repetition r, there must exist a factor u' of u with size $|u'| > p\left(\frac{7}{4} - \frac{3}{2}\right) \geq 68 \times \frac{1}{4} > 11$. So this occurence of u' contains the factor 01201 that is near the special vertex. Since the matching occurence of u' is contained in the opposite half-chain, we have again a contradiction with the factors 01201 and 10210. If r contains at least two special vertices, we have $|r| \geq d$, so that $|u| > \frac{3}{7}d$. Then, there exists a factor u' of u such that either $|u'| \geq 52$ and both occurrences of u' are in half-chains with opposite reading directions, or $|u'| \geq d/5$ and both occurrences of u' are in half-chains (possibly the same half-chain) with the same reading direction. In both cases, we have a contraction.

Case $k = 4$:
We consider the set S defined in the case $k = 3$ and the following 11-uniform morphism m:

$0 \mapsto 01320213032$
$1 \mapsto 01312023203$
$2 \mapsto 01232120323$
$3 \mapsto 01213231302$

The set W contains the words of the form $w = m(x)01$, with $x \in S$. The prefixes and suffixes of length 2 and the factors of length 12 of words in W contain the factor 01, whereas words in W do not contain the mirror factor 10. Using Lemma 3.1 in [4], we can check that the words in W are $\left(\frac{3}{2}^+\right)$-free.
Suppose that there exists a repetition $r = uvu$ in G_s with exponent $> \frac{3}{2}$, that is, $|u| > |v|$. Obviously, r cannot be contained in an half-chain We check that the period $p = |uv|$ of r must be at least 22, by looking at the neighborhood of a special vertex. We first consider the case where r is contained in two (opposite) half-chains. If the common special vertex belongs to v, then $|uv| \geq 22$ and $|u| > |v|$ gives $|u| \geq 12$. This is a contradiction since the factor 01 appearring in one occurence of u would induce a forbidden factor 10 in the word in W corresponding to the opposite chain containing the other ocurrence of u. If the common special vertex belongs to u, then u must contain a factor 01 or 10 that cannot be matched in the other occurrence of u. If r contains at least two special vertices, we have $|r| \geq d$, so that $|u| > \frac{1}{3}d$. Then, there exists a factor u' of u such that either $|u'| \geq 12$ and both occurrences of u' are in half-chains with opposite reading directions, or $|u'| \geq d/5$ and both occurrences of u' are in half-chains (possibly

the same half-chain) with the same reading direction. In both cases, we have a contraction.

Now, we prove the lower bounds.
Let us call *spider* a tree with at most one vertex of degree strictly greater than two. Notice that every subdivision G_s of a spider G contains G as a subgraph. For each k, we give a spider G such that any k-coloring of G contains a repetition of exponent at least the value of $RT(k, \mathcal{S})$ given in the theorem. Every subdivision G_s of this spider G contains G as a subgraph and thus a repetition of exponent at least $RT(k, \mathcal{S})$, which proves the lower bound.

Case $k = 2$:
Every $\left(\frac{7}{3}^+\right)$-free binary word of length 4 has a prefix in $F = \{00, 11, 010, 101, 0110, 1001\}$. Consider the spider G on 13 vertices containing a vertex v of degree 4 incident to 4 paths of length 3. Without loss of generality, v gets color 0. Since there are 4 paths to color in G and only 3 words in F starting with 0, two paths are colored with the same word in F. Now, for any factor $f[1, d]$ in F, the factor $f_d f_{d-1} \dots f_2 f_1 f_2 \dots f_{d-1} f_d$ is a repetition of exponent at least $\frac{7}{3}$.

Case $k = 3$:
Dejean [2] proved that the repetition threshold of a path on 39 vertices is $\frac{7}{4}$. Such a path is a spider.

Case $k \geq 4$:
We consider the spider $G = K_{1,k}$, that is, G contains a vertex v with degree k and its neighbors. We have to color $k + 1$ vertices, v and its neighbors, using k colors only, so at least two of them get the same color. If v has the same color as one of its neighbors then we get a square, and if two neighbors of v have the same color then we get a repetition of exponent $\frac{3}{2}$. □

4. Repetition threshold for trees

Theorem 2.

(1) $RT(2, \mathcal{T}) = \frac{7}{2}$
(2) $RT(3, \mathcal{T}) = 3$
(3) $RT(k, \mathcal{T}) = \frac{3}{2}$, *for* $k \geq 4$.

Proof. We define the family $U_n,\ n \geq 0$ of rooted trees as follows: U_0 is a vertex, U_{n+1} is obtained from $k^n + 1$ copies of U_n by adding a root vertex adjacent to the root of each copy. The family U_n is universal, that is, every (rooted) tree is the subgraph of U_n for some n. We thus have $RT(k, \mathcal{T}) = \lim_{n\to\infty} RT(k, U_n)$. A tree is *level-colored* if it is colored such that all vertices at the same depth have the same color.

We prove now that in any k-coloring of U_n, there exists a level-colored complete binary tree of height n. Suppose that the property holds for U_n and consider any coloring of U_{n+1}. We have $k^n + 1$ copies of U_n, each containing a level-colored binary tree, and there are at most k^n ways to color one of these binary trees. So two level-colored binary trees have the same coloring, and this creates the expected level-colored binary tree in U_{n+1}.

The repetition threshold of a k-coloring of U_n is at least the largest exponent of a repetition in a level-colored binary complete tree of height n that it contains. Moreover, the largest exponent of a repetition in a level-coloring is the same for U_n and for a binary complete tree of height n is the same, because the set of factors is the same in these two trees.

The study of $RT(k, \mathcal{T})$ is thus reduced to that of level-colorings of complete binary trees, so we can consider the word $w = w_1 w_2 \dots$ such that w_i is the color appearing at depth $i-1$. The factors of the level-colored complete binary tree are exactly those of the form $w_{m+l}w_{m+l-1}\dots w_{m+1}w_m w_{m+1}\dots w_{m+r-1}w_{m+r}$.

We first prove the upper bounds. For each k, we consider an infinite word w over the k-letter alphabet such that for all l, r, and m, the exponent of $w_{m+l}w_{m+l-1}\dots w_{m+1}w_m w_{m+1}\dots w_{m+r-1}w_{m+r}$ is less than the expected value of $RT(k, \mathcal{T})$. Without loss of generality, we suppose that $l \le r$.

Case $k = 2$:
We choose w to be the Thue-Morse word w_{TM}. Notice that the factor $w_{m+2}w_{m+1}w_m w_{m+1}w_{m+2}$ is in $\{10001, 01010, 11011, 00100, 10101, 01110\}$, so it is not a factor of w_{TM}. A repetition with period of length 1 has exponent at most 3. Suppose that $w_{m+l}w_{m+l-1}\dots w_{m+1}w_m w_{m+1}\dots w_{m+r-1}w_{m+r}$ is a repetition with a period of length at least 2 and exponent $> \frac{7}{2}$. The factor $w_{m+2}w_{m+1}w_m w_{m+1}w_{m+2}$ must then appear in $w[m, m+r]$, which is a factor of w_{TM}. A contradiction.

Case $k = 3$:
We choose w to be the image of any ternary square-free word by the morphism $0 \mapsto 00$, $1 \mapsto 11$, $2 \mapsto 22$ which doubles every letter. Clearly, the tree does not contain squares or cubes of period at least two.

Case $k = 4$:
We choose w to be any $\left(\frac{3}{2}\right)$-free word over Σ_4. Notice that w has no factors of the form aa nor aba where a and b are letters. So w contains no palindrome of size at least two. Suppose that uvu is a repetition of exponent $> \frac{3}{2}$, that is, $|u| > |v|$, and that 01 is a prefix of u. The first factor u cannot contain a vertex and two of its sons, since this creates a factor aba that cannot be matched in the other factor u. So the first occurrence of u is a prefix of the left branch $w_{m+l}w_{m+l-1}\dots w_{m+1}w_m$.

Without loss of generality, w_m is the last letter of u, since otherwise the factor $w_{m+l}w_{m+l-1}\dots w_{m+2}w_{m+1}w_{m+2}\dots w_{m+r-1}w_{m+r}$ rooted at level $m+1$ is the repetition $uv'u$ where v' is the suffix of v of length $|v|-2$, so that the exponent of $uv'u$ is greater than the exponent of uvu. Thus $|u| = l + 1$, and since $|v| < |u|$, we can consider the following cases on the right branch:

- $|v| = l$: the factor $w_{m+l}w_{m+l+1}$ should be 00, a contradiction.
- $|v| = l-1$: the factor $w_{m+l-1}w_{m+l}w_{m+l+1}$ should be 101, a contradiction.
- $|v| = l - t$, $t \geq 2$: the factor $w_{m+l-t+1} \dots w_{m+l}$ should be a palindrome of size at least two, a contradiction.

Now, we prove the lower bounds.
For each k, we prove that for any infinite word w over the k-letter alphabet, there exist l, r and m, such that the exponent of
$w_{m+l}w_{m+l-1} \dots w_{m+1}w_m w_{m+1} \dots w_{m+r-1}w_{m+r}$ is at least $RT(k, \mathcal{T})$.

Case $k = 2$:
It is easy to check that every long enough binary word either has a factor in $F = \{000, 111, 0101, 1010, 0110110, 1001001\}$ or is a factor of $(0011)^\omega$. Moreover, for any factor $f[1, d]$ in F, the factor $f_d f_{d-1} \dots f_2 f_1 f_2 \dots f_{d-1} f_d$ is a repetition of exponent at least $\frac{7}{2}$.

Case $k = 3$:
Let F be the set of words obtained by permutations of the letter alphabet of the words in $\{00, 0101, 0121012, 012012012\}$. A computer check shows that every ternary word w of length 58 contains a factor $w[b, e]$ such that the factor $w_e w_{e-1} \dots w_{b+1} w_b w_{b+1} \dots w_{e-1} w_e$ contains a cube.
The following word of length 57 does not contain such a factor:
010201210212012021021012012010210210120120210212012102010.

Case $k \geq 4$:
It is clear that any level-coloring of the complete binary tree of height 1 contains a repetition of exponent at least $\frac{3}{2}$.

$\square$

References

[1] A. Aberkane and J. Currie. There exist binary circular 5/2+ power free words of every length, *Electron. J. Comb.* **11** (2004), #R10
[2] F. Dejean. Sur un théorème de Thue, *J. Combin. Theory. Ser. A* **13** (1972), 90–99.
[3] J. Grytczuk. Nonrepetitive colorings of graphs - a survey, *Int. J. Math. Math. Sci.* (2007), Art. ID 74639, 10 pp.
[4] P. Ochem. A generator of morphisms for infinite words, *RAIRO: Theoret. Informatics Appl.* **40** (2006) 427–441.
[5] A. Pezarski and M. Zmarz. Non-repetitive 3-Coloring of subdivided graphs, *Electron. J. Comb.* **16(1)** (2009), #N15

Communicated by (The editor will be set by the publisher).
(The dates will be set by the publisher).

Titre

RÉPÉTITIONS DANS LES MOTS ET SEUILS D'ÉVITABILITÉ

Résumé

Nous étudions dans cette thèse différents problèmes d'évitabilité des répétitions dans les mots infinis. Soulevée par Thue et motivée par ses travaux sur les mots sans carrés, la problématique s'est développée au cours du XXe siècle, et est aujourd'hui devenue un des grands domaines de recherche en combinatoire des mots. En 1972, Dejean proposa une importante conjecture, dont la validation étape par étape s'est terminée récemment (2009). La conjecture concerne le seuil des répétitions d'un alphabet, i.e., la borne inférieure des exposants évitables sur cet alphabet. La notion de seuil, comme frontière entre évitabilité et non-évitabilité d'un ensemble donné de mots, est le fil directeur de nos travaux. Nous nous intéressons d'abord à une généralisation du seuil des répétitions (nous donnons des encadrements de sa valeur). Cette notion permet d'ajouter, pour décrire l'ensemble des répétitions à éviter, au paramètre de l'exposant, celui de la longueur des répétitions. Puis, nous étudions des problèmes d'existence de mots dans lesquels, simultanément, certaines répétitions sont interdites et d'autres sont forcées. Nous répondons, pour l'alphabet ternaire, à la question : quels réels sont l'exposant critique d'un mot infini sur un alphabet fixé ? Nous introduisons ensuite une notion de haute répétitivité, et établissons une description partielle des couples d'exposants paramètrant une double contrainte de haute répétitivité et d'évitabilité. Pour finir, nous utilisons des résultats et techniques issus de ces problématiques pour résoudre une question de coloration de graphes : nous introduisons un seuil des répétitions, calqué sur celui connu pour les mots, et donnons sa valeur pour deux classes de graphes, les arbres et les graphes de subdivisions.

Mot-clefs

Combinatoire des mots, évitabilité, répétitions, exposants critiques, conjecture de Dejean, seuil des répétitions, coloration de graphes

Abstract

In this thesis we study various problems on repetition avoidance in infinite words. Raised by Thue and motivated by his work on squarefree words, the topic developed during the 20th century, and has nowadays become a principal area of research in combinatorics on words. In 1972, Dejean proposed an important conjecture whose verification in steps was completed recently (2009). The conjecture concerns the repetition threshold for an alphabet, i.e., the infimum of the avoidable exponents for that alphabet. The notion of threshold as a borderline between avoidability and unavoidability for a given set of words is the guiding line of our work. First, we focus on a generalization of the repetition threshold. This concept allows us to include, in addition to the exponent, the length of the repetitions as a parameter in the description of the set of repetitions to avoid. We obtain various bounds in that respect. We then study existence problems for words in which simultaneously some repetitions are forbidden, and others are forced. For the ternary alphabet, we answer the question : what real numbers are the critical exponent of some infinite word over a given alphabet ? Also, we introduce a notion of highly repetitive words and give a partial description of the pairs of exponents which parameterize the existence of words both highly repetitive and repetition-free. Finally, we use results and techniques stemming from those problems to solve a question on graph colouring : we introduce a repetition threshold adapted from the thresholds we know for words, and give its value for two classes of graphs, namely, trees and subdivision graphs.

Keywords

Combinatorics on words, avoidability, repetitions, critical exponents, Dejean's conjecture, repetition threshold, graphs coloring

Adrr : Institut de Mathématiques de Luminy (UMR 6206) Campus de Luminy, Case 907 13288
MARSEILLE Cedex 9.