Complexité et facteurs spéciaux

Julien Cassaigne

Résumé

Dans l'ensemble des facteurs d'une suite infinie à valeurs dans un ensemble fini, certains éléments jouent un rôle particulier : les facteurs *spéciaux* et *bispéciaux*. Nous montrons comment ils peuvent servir à calculer la complexité de suites, c'est-à-dire le nombre de facteurs de longueur donnée, et à prouver que certaines fonctions peuvent être obtenues comme complexités de suites alors que d'autres ne le peuvent pas.

Abstract

Among the factors of an infinite sequence on a finite alphabet, some elements have a particular importance: *special* and *bispecial* factors. We show how they can be used to compute the complexity of sequences, i.e. the number of factors with a given length, and to prove that certain functions are obtainable as sequence complexity whereas other functions are not.

1 Introduction

Pour apprécier la structure d'une suite $\mathbf{u} = (u_n)_{n\in\mathbb{N}}$ de symboles d'un alphabet fini Σ, et en particulier pour mesurer la diversité des motifs qui apparaissent dans cette suite, on peut utiliser la *fonction de complexité* de la suite, qui est une fonction de $\mathbb{N}$ dans $\mathbb{N}$, habituellement notée $p_\mathbf{u}$ ou simplement p, qui à tout entier n associe le nombre de facteurs de longueur n de $\mathbf{u}$, c'est-à-dire le nombre de mots $w \in \Sigma^n$ tels que $w = u_k u_{k+1} \dots u_{k+n-1}$ pour un certain entier k.

Ainsi, la suite la plus simple possible, la suite constante a^ω avec $a \in \Sigma$, a pour complexité la fonction constante $p(n) = 1$, alors qu'une suite aléatoire a presque

Received by the editors May 95.
Communicated by M. Boffa.
1991 *Mathematics Subject Classification :* 68R15, 11B85.
Key words and phrases : Subword complexity, special factors, substitutive sequences.

sûrement pour complexité $p(n) = (\#\Sigma)^n$, où $\#\Sigma$ désigne le cardinal de l'alphabet Σ. Entre ces deux extrêmes, on trouve par exemple les suites ultimement périodiques, dont la complexité est bornée, les suites sturmiennes, de complexité $p(n) = n+1$, les suites automatiques, pour lesquelles $p(n) = O(n)$, et les suites substitutives, dont la complexité peut atteindre $O(n^2)$. Dans son survol [2], Allouche cite de nombreux autres exemples de suites ou de familles de suites dont on connaît la complexité.

Dans cet article, nous présentons un outil qui permet d'obtenir un certain nombre de résultats intéressants sur la complexité des suites : les *facteurs spéciaux*, qui sont les facteurs de la suite qui peuvent être prolongés de plusieurs manières, et les *facteurs bispéciaux* qui sont spéciaux dans les deux directions à la fois. Il se trouve que les nombres de facteurs spéciaux et bispéciaux sont reliés aux différences première et seconde de la fonction de complexité. Dans le cas, qui est celui d'un grand nombre de suites classiques, où la complexité est suffisamment basse, le nombre de facteurs bispéciaux est très faible et il est souvent aisé de l'évaluer, et d'obtenir ainsi la complexité beaucoup plus facilement qu'avec un calcul direct. Nous montrons notamment comment cette technique s'applique aux suites substitutives définies par un morphisme circulaire.

Nous envisageons également un autre point de vue sur la complexité des suites : trouver quelles fonctions peuvent être des fonctions de complexité. On se rend rapidement compte en effet que ce n'est pas le cas de n'importe quelle fonction, et que les quelques conditions nécessaires naturelles (p est une fonction croissante et $1 \leqslant p(n) \leqslant (\#\Sigma)^n$, par exemple) sont loin d'être suffisantes. En particulier, un résultat classique [10] dit que les fonctions de complexité non bornées sont supérieures à $n+1$: une fonction qui croît comme $\sqrt{n}$ par exemple n'est donc certainement pas une fonction de complexité. Une autre zone interdite, moins connue, se trouve à l'autre extrémité du spectre : si $p(n)$ n'est pas $(\#\Sigma)^n$ pour tout n, alors $p(n) = O(\alpha^n)$ avec $\alpha < \#\Sigma$, à cause de la relation $p(m+n) \leqslant p(m)p(n)$, valable quels que soient les entiers m et n. Une fonction telle que $(\#\Sigma)^n/\log n$ n'est donc pas une fonction de complexité. L'utilisation des facteurs spéciaux nous permet d'approcher des deux côtés la frontière des fonctions de complexité : d'une part nous construisons des familles de suites qui permettent de prouver que certaines fonctions sont réalisées, et d'autre part nous donnons une condition nécessaire plus fine qui limite les variations de $p(n+1) - p(n)$ quand $p(n)$ croît lentement.

2 Suites et langages

2.1 Complexité d'un langage factoriel

La complexité d'une suite est en fait un cas particulier de la complexité d'un langage : étant donné un langage $L \subset \Sigma^*$, la fonction de complexité de L est la fonction p_L définie par $p_L(n) = \#(L \cap \Sigma^n)$. La complexité de la suite $\mathbf{u}$ n'est autre que la complexité du langage $F(\mathbf{u})$ formé de tous les facteurs de $\mathbf{u}$.

Il n'y a pas grand chose de général à dire sur les complexités de langages arbitraires : n'importe quelle fonction restant dans les bornes $0 \leqslant p(n) \leqslant (\#\Sigma)^n$ peut être obtenue. Il faut donc se restreindre à une classe donnée de langages, par exemple les familles classiques : langages rationnels, langages algébriques, etc. Nous

nous intéresserons ici exclusivement à la classe des *langages factoriels*, dont font partie tous les langages de facteurs de suites $F(\mathbf{u})$: un langage L est dit factoriel si tout facteur d'un élément de L est également élément de L.

Toutefois, la plupart des langages factoriels ne sont pas de la forme $F(\mathbf{u})$, et leur complexité n'est pas toujours la complexité d'une suite. En particulier, p_L n'est pas nécessairement une fonction croissante, alors que $p_{\mathbf{u}}$ l'est toujours : c'est par exemple le cas du langage des mots sans chevauchements sur l'alphabet $\{a, b\}$ [5, 6]. La raison de cette différence est assez simple : dans le langage des mots sans chevauchements, il existe des *mots non prolongeables à droite*, par exemple *abbabb* qui ne peut être prolongé ni par a ni par b sans faire apparaître un chevauchement (*abbabba* ou *bbb*). Cela ne se produit jamais avec les facteurs d'une suite, qui peuvent toujours être prolongés d'une manière au moins.

2.2 Le problème des suites non récurrentes

Un avantage des langages factoriels sur les suites est que les propriétés des premiers sont indépendantes du sens de lecture des mots. Les suites par contre sont dissymétriques (pour éviter cela il faudrait considérer des suites biinfinies, c'est-à-dire indexées par $\mathbb{Z}$), et en particulier la propriété de prolongeabilité à droite énoncée ci-dessus n'est pas vraie à gauche : il peut exister des préfixes de $\mathbf{u}$ dont aucun prolongement n'est facteur de $\mathbf{u}$. Les suites pour lesquelles cela ne se produit pas sont dites *récurrentes*, et la plupart des suites classiques en font partie.

En pratique, nous travaillerons soit avec des langages factoriels dont tous les éléments sont prolongeables à droite et à gauche (appelés langages factoriels *prolongeables*), soit avec des suites récurrentes, dont les langages de facteurs sont toujours factoriels prolongeables. Nous pourrons ainsi traiter de manière symétrique les deux extrémités des mots.

Pour calculer la complexité d'une suite non récurrente $\mathbf{u}$ sur un alphabet Σ, il suffira d'ajouter à l'alphabet Σ une nouvelle lettre z et d'étudier le langage

$$L = F(\mathbf{u}) \cup \{\, z^n w \mid n \in \mathbb{N} \text{ et } w \text{ préfixe de } \mathbf{u} \,\}$$

(qui est l'ensemble des facteurs de la suite biinfinie ${}^{\omega}z\mathbf{u}$, et donc factoriel et prolongeable). Il sera ensuite facile de passer à $\mathbf{u}$ au moyen de la relation

$$p_L(n) = p_{\mathbf{u}}(n) + n \ .$$

3 Facteurs spéciaux et bispéciaux

Dans toute cette section, L est un langage factoriel prolongeable sur l'alphabet Σ. On suppose pour le moment que Σ n'a que deux éléments : $\Sigma = \{a, b\}$.

3.1 Arbres et graphes de facteurs

3.1.1 L'arbre des facteurs à droite

Il est commode de représenter le langage L sous forme d'un arbre, appelé *arbre des facteurs à droite* de L (on définit également l'arbre des facteurs à gauche). C'est un arbre infini dont les sommets sont étiquetés par les éléments de L et les arêtes par les lettres, de sorte qu'il y a une arête d'étiquette x entre les sommets u et v si et seulement si $v = ux$. On le représente conventionnellement en plaçant la racine, étiquetée par le mot vide, à gauche, si bien que quand on lit de la gauche vers la droite les étiquettes du chemin entre la racine et le sommet étiqueté u on retrouve le mot u ; on peut alors se contenter d'indiquer les étiquettes des arêtes, ou la dernière lettre des étiquettes de sommet. Par exemple, l'arbre des facteurs à droite de la suite de Fibonacci

$$\mathbf{f} = abaababaabaababaababaabaababaabaababaaba\ldots$$

(point fixe de la substitution qui transforme a en ab et b en a) est représenté sur la figure 1.

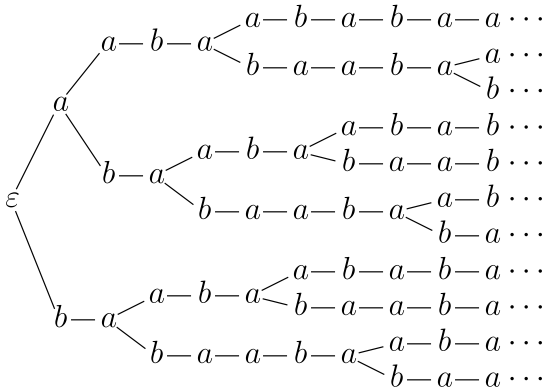

FIG. 1 – Arbre à droite des facteurs de la suite de Fibonacci

La complexité $p(n)$ est clairement le nombre de sommets de l'arbre situés à l'abcisse n. Comme le langage L est prolongeable, toutes les branches de l'arbre sont infinies, et la complexité ne peut donc qu'augmenter quand n croît. Cela se produit chaque fois que l'arbre comporte un embranchement, et s'il y a k embranchements à l'abcisse n on a précisément $p(n+1) = p(n) + k$. Autrement dit, le nombre d'embranchements à l'abcisse n est la différence première de la complexité. Ce sont les étiquettes de ces embranchements que nous nommerons *facteurs spéciaux à droite*. Dans l'exemple de la suite de Fibonacci (et pour les suites sturmiennes en général) $p(n) = n + 1$ donc $p(n+1) - p(n) = 1$ pour tout n : il y a exactement un embranchement à chaque abcisse.

3.1.2 Graphes de Rauzy

Une autre représentation agréable de L utilise non plus un arbre infini mais une famille de graphes finis. Pour tout entier n, le *graphe de Rauzy d'ordre* n de L est le graphe fini dont les sommets sont étiquetés par les mots de L de longueur n et les arêtes par les mots de L de longueur $n+1$, l'arête étiquetée w joignant le sommet étiqueté par le préfixe de longueur n de w à celui étiqueté par son suffixe de longueur n. Les premiers graphes pour la suite de Fibonacci sont représentés sur la figure 2.

À cause de la prolongeabilité de L, tout sommet a au moins une arête entrante et une arête sortante. Ceux qui ont deux arêtes sortantes sont précisément les mots qui

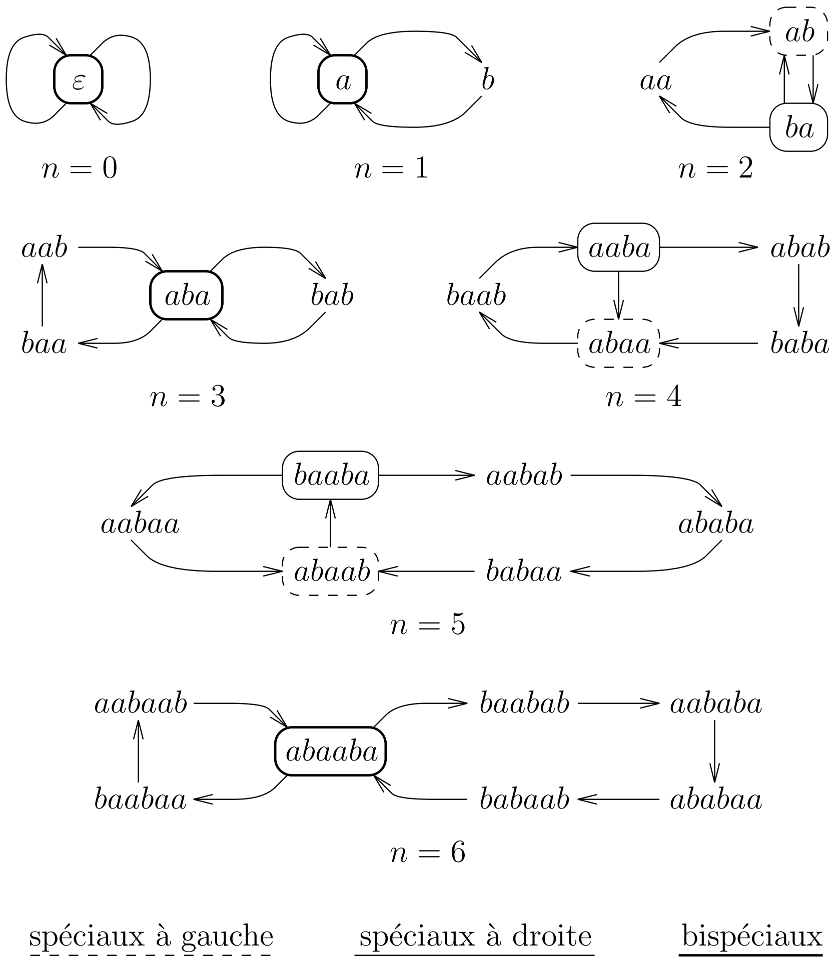

FIG. 2 – Graphes de Rauzy d'ordre 0 à 6 de la suite de Fibonacci

étiquetaient les embranchements à l'abcisse n dans l'arbre des facteurs à droite, les *facteurs spéciaux à droite*. De même, nous appellerons *facteurs spéciaux à gauche* les étiquettes des sommets où arrivent deux arêtes. Quand ces facteurs spéciaux sont peu nombreux, il est parfois possible de décrire explicitement la structure des graphes de Rauzy [3, 13], ce qui en fait un outil très efficace pour étudier les suites de faible complexité.

3.2 Facteurs spéciaux

Formellement, un mot $u \in L$ est dit *spécial à droite* pour L si les mots ua et ub sont tous deux dans L, *spécial à gauche* si au et bu sont dans L.

Proposition 3.1 *Soit L un langage factoriel prolongeable sur un alphabet binaire. Le langage L possède le même nombre de facteurs spéciaux de longueur n à gauche ou à droite, et nous noterons ce nombre $s(n)$. Il vérifie*

$$s(n) = p(n+1) - p(n) \ .$$

*Preuve***.** Soit $sd(n)$ le nombre de facteurs spéciaux à droite de longueur n. Parmi les $p(n)$ facteurs de longueur n, il y a $sd(n)$ facteurs qui se prolongent de deux manières à droite, et les autres $p(n) - sd(n)$ facteurs ne se prolongent que d'une seule manière. On obtient ainsi une fois et une seule tous les facteurs de longueur $n + 1$, donc $p(n + 1) = 2sd(n) + (p(n) - sd(n))$, d'où la relation $sd(n) = p(n + 1) - p(n)$. On démontre de même que $sg(n) = p(n + 1) - p(n)$, où $sg(n)$ est le nombre de facteurs spéciaux à gauche, et on en déduit que $sd(n) = sg(n)$. ∎

Une conséquence immédiate de cette proposition est que la complexité du langage L est une fonction croissante.

3.3 Facteurs bispéciaux

Lorsqu'un facteur u est spécial à la fois à gauche et à droite, on dit qu'il est *bispécial*. Trois cas peuvent se présenter, selon le nombre d'éléments de $L \cap \Sigma u \Sigma$: si cet ensemble a quatre éléments (le maximum), on dit que u est *bispécial strict*. S'il n'en a que trois, u est *bispécial ordinaire*. Enfin, il est possible que cet ensemble ait deux éléments ($\{aua, bub\}$ ou $\{aub, bua\}$) et on dit dans ce cas que u est *bispécial faible*.

Quand on prolonge à droite un facteur spécial à gauche u, on obtient un ou deux facteurs qui peuvent être spéciaux à gauche ou non, en fonction de la nature de u :
– aucun n'est spécial à gauche si u est bispécial faible ;
– un seul est spécial à gauche si u est non bispécial ou bispécial ordinaire ;
– les deux sont spéciaux à gauche si u est bispécial strict.

Réciproquement, un préfixe de facteur spécial à gauche est spécial à gauche, donc on peut construire l'*arbre des facteurs spéciaux à gauche*, qui est un sous-arbre de l'arbre des facteurs à droite, dans lequel on ne garde que les sommets dont les étiquettes sont spéciales à gauche. Dans cet arbre, un embranchement indique un bispécial strict et une feuille indique un bispécial faible. Cet arbre est souvent plus facile à décrire que l'arbre de tous les facteurs, le cas extrême étant celui des suites sturmiennes pour lesquelles il est filiforme. On peut bien entendu aussi construire l'arbre des facteurs spéciaux à droite, qui est orienté vers la gauche. Pour bien indiquer qu'il s'agit de facteurs spéciaux, on indique à la racine les deux lettres a et b par lesquelles les facteurs spéciaux peuvent se prolonger (figures 3 et 4).

$$\cdots\ b—a—b—a—a—b—a—a—b—a—b—a—a—b—a—\varepsilon \left| \begin{matrix} a \\ b \end{matrix} \right.$$

FIG. 3 – Arbre des facteurs spéciaux à droite pour la suite de Fibonacci

Les différents types de facteurs bispéciaux interviennent également lors de l'expansion des graphes de Rauzy : pour passer du graphe des facteurs de longueur n à celui des facteurs de longueur $n + 1$, on remplace chaque arête par un sommet et chaque sommet par une, deux, trois ou quatre arêtes selon que le mot qui étiquette ce sommet est non spécial, spécial non bispécial ou bispécial faible, bispécial ordinaire, bispécial strict (cf. figure 5). En particulier, si aucun facteur de longueur n n'est bispécial, il n'y a qu'une expansion possible.

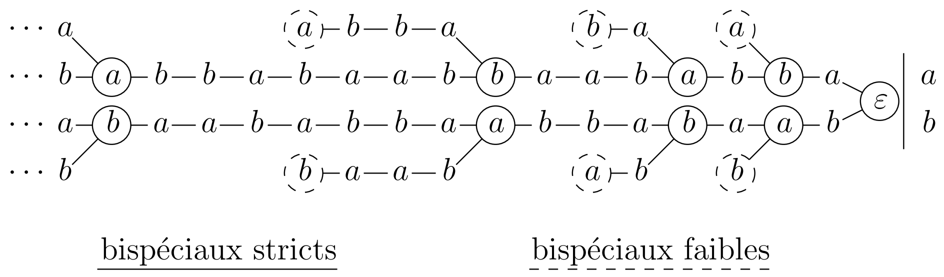

FIG. 4 – Arbre des facteurs spéciaux à droite pour la suite de Thue-Morse

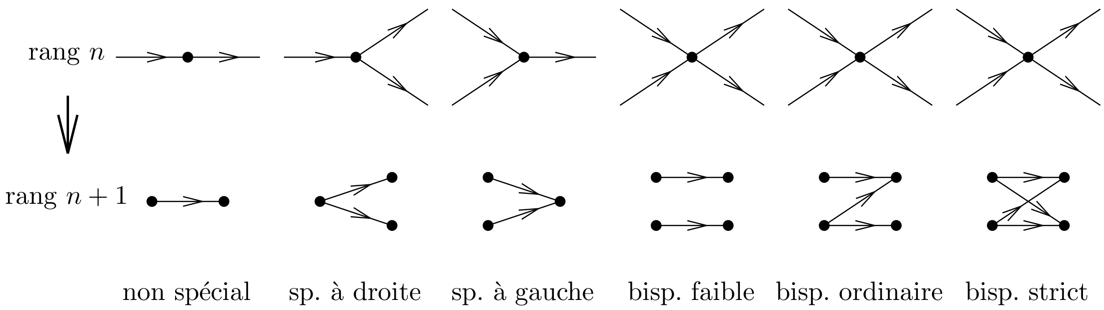

FIG. 5 – Expansion des graphes de Rauzy

Proposition 3.2 *Soit L un langage factoriel prolongeable sur un alphabet binaire, et s, bs et bf les fonctions comptant les facteurs spéciaux, bispéciaux stricts, et bispéciaux faibles de L. On a alors*

$$bs(n) - bf(n) = s(n+1) - s(n) \ .$$

*Preuve***.** Voir la construction des arbres de facteurs spéciaux ci-dessus. ∎

Pour calculer $p(n)$, il est donc suffisant de connaître $bs(n)$ et $bf(n)$ et d'appliquer les propositions 3.2 puis 3.1.

On peut également utiliser des séries génératrices : à la fonction p on associe la série $P(X) = \sum_{n=0}^{\infty} p(n)X^n$, et de même les séries $S(X)$, $BS(X)$ et $BF(X)$ sont associées à s, bs et bf. Les relations entre ces séries sont résumées dans la proposition 3.3.

Proposition 3.3 *Soit L un langage factoriel prolongeable sur un alphabet binaire, et P, S, BS et BF les séries génératrices des fonctions p, s, bs et bf associées à L. On suppose de plus que $p(1) = 2$. Alors*

$$P(X) = \frac{1}{1-X}\,(1 + XS(X))$$

et

$$S(X) = \frac{1}{1-X}\,(1 + X(BS(X) - BF(X)))$$

soit

$$P(X) = \frac{1}{(1-X)^2}\left(1 + X^2(BS(X) - BF(X))\right) \ .$$

Preuve. Il suffit de sommer les relations des propositions 3.1 et 3.2. L'hypothèse $p(1) = 2$, qui dit que les deux lettres de l'alphabet sont effectivement utilisées, assure que le mot vide est un facteur spécial et donc que $s(0) = 1$. ■

3.4 Le cas des alphabets plus grands

Quand l'alphabet Σ a plus de deux éléments, la situation se complique quelque peu, surtout en ce qui concerne les facteurs bispéciaux, car un mot peut être prolongeable par plusieurs lettres de l'alphabet sans l'être par toutes. Supposons donc que Σ a k éléments.

Soit u un élément quelconque de L. On définit l'*ordre à droite* de u, noté $m_d(u)$, par

$$m_d(u) = \#\{\ x \in \Sigma \mid ux \in L\ \} - 1\ ,$$

de sorte que $m_d(u)$ est un entier compris entre 0 et $k-1$. Un élément de L sera dit *spécial à droite* si son ordre à droite est non nul. On définit de même l'ordre à gauche $m_g(u)$ et les éléments spéciaux à gauche. Cette définition est bien compatible avec celle donnée plus haut dans le cas $k = 2$; mais alors que sur l'alphabet binaire les facteurs spéciaux sont tous d'ordre (à droite ou à gauche) 1, si l'alphabet est plus grand ils n'ont pas nécessairement tous le même ordre.

Comme avec un alphabet binaire, les facteurs spéciaux correspondent à des embranchements dans les arbres de facteurs et les graphes de Rauzy. Le nombre de branches est égal à l'ordre du facteur spécial, augmenté d'une unité.

Proposition 3.4 *Soit L un langage factoriel prolongeable. Alors*

$$\sum_{u \in L\cap\Sigma^n} m_d(u) = \sum_{u \in L\cap\Sigma^n} m_g(u) = p(n+1) - p(n)\ .$$

Cette quantité sera notée $s(n)$.

Preuve. Puisque $m_d(u) + 1$ compte les mots de longueur $|u| + 1$ de L commençant par u, la somme

$$\sum_{u \in L\cap\Sigma^n} (m_d(u) + 1)$$

compte donc tous les mots de longueur $n+1$ de L et vaut donc $p(n+1)$, et de même avec les ordres à gauche, ce qui donne les égalités annoncées. ■

Le nombre $s(n)$ peut encore être considéré comme le nombre de facteurs spéciaux à droite (ou à gauche) de L à condition de compter m fois un facteur dont l'ordre à droite est m. Aussi, nous l'appellerons parfois, par abus de langage, *nombre de facteurs spéciaux* de L, en sous-entendant que ceux-ci doivent être comptés avec multiplicité.

Un facteur u spécial à droite et à gauche est encore dit *bispécial*. Notons que les ordres à droite et à gauche d'un facteur bispécial sont tous les deux non nuls mais n'ont aucune raison d'être égaux. Nous allons encore classer les facteurs bispéciaux en fonction du nombre d'éléments de l'ensemble $L \cap \Sigma u \Sigma$. Celui-ci vaut au plus $(m_d(u)+1)(m_g(u)+1)$, et si c'est le cas on dit que u est *bispécial strict* ; il vaut au moins $\max(m_d(u), m_g(u)) + 1$, car chaque prolongement de u à droite ou à gauche donne au moins un élément de $L \cap \Sigma u \Sigma$: c'est le cas que nous avons appelé *bispécial faible* pour un alphabet binaire. Mais cette fois tous les intermédiaires sont possibles entre ces deux extrêmes : on peut quantifier cela au moyen d'un troisième entier $m(u)$, l'*ordre bilatère* du facteur bispécial, défini ainsi :

$$m(u) = \#\left(L \cap \Sigma u \Sigma\right) - m_d(u) - m_g(u) - 1 \ .$$

C'est un entier relatif, compris entre $-\min(m_d(u), m_g(u))$ et $m_d(u)m_g(u)$, donc entre $-(k-1)$ et $(k-1)^2$; si $k = 2$, on retrouve les trois catégories de facteurs bispéciaux, qui correspondent aux trois ordres (bilatères) possibles, -1, 0 et 1. Notons que si u est un élément non bispécial, on peut également définir son ordre bilatère, qui est toujours nul.

Proposition 3.5 *Soit L un langage factoriel prolongeable. Alors*

$$\sum_{u \in L \cap \Sigma^n} m(u) = s(n+1) - s(n) \ .$$

Cette quantité sera notée $b(n)$.

Preuve. Quand on prolonge par une lettre à droite un facteur spécial à gauche u, la nature de u permet de déterminer, non le nombre de facteurs spéciaux à gauche obtenus, mais la somme de leurs ordres à gauche :
- si u n'est pas bispécial, on obtient un facteur spécial à gauche de même ordre que u ;
- si u est bispécial, on obtient un certain nombre de facteurs spéciaux à gauche dont la somme des ordres à gauche est $m_g(u) + m(u)$.

Comme $s(n)$ est la somme des ordres à gauche de tous les facteurs u de longueur n, la variation de $s(n)$ est donc la somme des $m(u)$. ■

Comme dans le cas de l'alphabet binaire, les propositions 3.4 et 3.5 peuvent d'exprimer en termes de séries génératrices :

Proposition 3.6 *Soit L un langage factoriel prolongeable, et P, S et B les séries génératrices des fonctions p, s et b associées à L. Alors*

$$P(X) = \frac{1}{1-X}\left(1 + XS(X)\right)$$

et

$$S(X) = \frac{1}{1-X}\left(s(0) + XB(X)\right)$$

soit

$$P(X) = \frac{1}{(1-X)^2}\Big(1 + (p(1)-2)X + X^2 B(X)\Big) \ .$$

Preuve. La seule différence notable avec la proposition 3.3 est le terme constant de S, qui vaut $s(0) = p(1) - 1$, soit $k-1$ si toutes les lettres de l'alphabet sont utilisées. ■

Pour que la notion d'arbre de facteurs spéciaux soit utilisable, il faut un critère plus fin que l'ordre : étant donnée une partie Σ' de Σ à au moins deux éléments, on construit l'*arbre des facteurs spéciaux à gauche de type* Σ' en ne conservant dans l'arbre des facteurs à droite que ceux qui se prolongent à gauche par chacune des lettres de Σ'. L'ensemble des facteurs spéciaux à gauche est donc décrit par une famille d'arbres (un pour chaque partie Σ'), mais en pratique on peut se limiter aux arbres non vides avec Σ' maximal (i.e. aucune partie plus grande ne donne le même arbre), ce qui en général donne un petit nombre d'arbres, qu'on pourra appeler *base* de la famille des arbres des facteurs spéciaux à gauche. Dans ces arbres, tous les embranchements et toutes les extrémités de branches (feuilles) sont étiquetés par des facteurs bispéciaux, et réciproquement tous les facteurs bispéciaux d'ordre non nul donnent un embranchement ou une feuille sur au moins l'un des arbres de base : on peut donc déterminer les facteurs bispéciaux qui contribuent à la complexité par simple observation de la base, et il est alors facile de calculer leur ordre.

4 Calcul de la complexité de suites définies par morphismes

Dans cette section, nous montrons, à l'aide de quelques exemples, comment les outils que nous venons de présenter peuvent être utilisés pour calculer la complexité de certaines suites définies au moyen de morphismes de monoïdes libres, ce qui est le cas d'un grand nombre de suites classiques. Nous entendons par *suites définies par morphismes* non seulement les suites qui sont point fixe d'un morphisme, mais aussi les images de tels points fixes par un second morphisme, et plus généralement les suites définies par des *systèmes S-adiques*, c'est-à-dire par application successive d'une suite de morphismes pris dans un ensemble fini.

4.1 Suite de Thue-Morse

La suite de Thue-Morse est le point fixe

$$\mathbf{t} = abbabaabbaababbabaababbaabbabaabbaababbaabbabaababbabaabbaab\ldots$$

du morphisme

$$\begin{array}{rccc} \theta\colon & \Sigma^* & \longrightarrow & \Sigma^* \\ & a & \longmapsto & ab \\ & b & \longmapsto & ba \end{array} .$$

Comme la suite de Thue-Morse est récurrente, les constructions de la section 3 s'appliquent et il suffit de connaître ses facteurs bispéciaux pour calculer sa complexité.

Proposition 4.1 *Les facteurs bispéciaux de la suite de Thue-Morse sont*

bispéciaux stricts : ε, $\theta^m(ab)$, $\theta^m(ba)$ *pour* $m \geqslant 0$ *;*

bispéciaux ordinaires : a, b *;*

bispéciaux faibles : $\theta^m(aba)$, $\theta^m(bab)$ *pour* $m \geqslant 0$.

La figure 4, qui montre l'arbre des facteurs spéciaux à droite de la suite de Thue-Morse, illustre bien ce résultat. La preuve repose sur le lemme classique suivant :

Lemme 4.2 *Tout facteur de* $\mathbf{t}$ *est de la forme* $y = r_1.\theta(x).r_2$ *avec* $x \in F(\mathbf{t})$ *et* $r_i \in \{\varepsilon, a, b\}$. *Si* $|y| \geqslant 5$ *cette décomposition est unique.*

Preuve de la proposition. Soit y un facteur tel que $|y| \geqslant 5$. D'après le lemme, y s'écrit de manière unique $y = r_1.\theta(x).r_2$. Supposons que $r_2 = a$. Si le mot ya était facteur de $\mathbf{t}$, on pourrait aussi lui appliquer le lemme : $ya = r'_1.\theta(x').r'_2$. Comme $\theta(x')$ ne peut pas se terminer par aa, on aurait nécessairement $r'_2 = a$, donc $y = r'_1.\theta(x').\varepsilon$, ce qui contredirait l'unicité dans le lemme. Le facteur y ne peut donc se prolonger à droite que par un b. De même, si $r_2 = b$ le seul prolongement possible est ya. Si y est spécial à droite, alors nécessairement $r_2 = \varepsilon$. Le prolongement ya se décompose alors sous la forme $ya = r_1.\theta(x).a$, donc a un prolongement unique yab dont la décomposition est nécessairement $r_1.\theta(xa).\varepsilon$, donc xa est facteur de $\mathbf{t}$, et pour la même raison xb est aussi facteur de $\mathbf{t}$, donc x est spécial à droite.

On voit donc que les facteurs spéciaux à droite de longueur au moins 5 sont tous de la forme $y = r_1\theta(x)$ avec x lui-même spécial à droite. Une propriété symétrique est vérifiée à gauche, donc les facteurs bispéciaux de longueur au moins 5 sont tous des images par θ de facteurs bispéciaux, qui de plus ont même ordre. Il ne reste plus qu'à énumérer les facteurs bispéciaux courts pour pouvoir conclure : ε, ab, ba, $abba$, $baab$ sont stricts, a et b sont ordinaires, et aba et bab sont faibles. ∎

On a donc :

$$BS(X) - BF(X) = 1 + 2\sum_{n=0}^{\infty}\left(X^{2.2^n} - X^{3.2^n}\right)$$

et on en déduit immédiatement les séries $S(X)$ et $P(X)$. On retrouve ainsi les formules classiques [4] :

$$s(n) = \begin{cases} 1 & \text{si } n = 0 \\ 2 & \text{si } 0 < n \leqslant 2 \\ 4 & \text{si } 2.2^m < n \leqslant 3.2^m \\ 2 & \text{si } 3.2^m < n \leqslant 4.2^m \end{cases}$$

$$p(n) = \begin{cases} 1 & \text{si } n = 0 \\ 2 & \text{si } n = 1 \\ 4 & \text{si } n = 2 \\ 4n - 2.2^m - 4 & \text{si } 2.2^m < n \leqslant 3.2^m \\ 2n + 4.2^m - 2 & \text{si } 3.2^m < n \leqslant 4.2^m \end{cases}$$

où m est un entier positif ou nul.

La suite de Thue-Morse est un cas particulièrement simple, mais de nombreuses suites définies par morphisme peuvent être traitées sur le même modèle, à condition d'avoir un *lemme de synchronisation* similaire au lemme 4.2, qui assure un découpage unique des facteurs suffisamment longs. Cela se produit notamment quand les morphismes sont *circulaires*.

4.2 Action d'un morphisme circulaire

4.2.1 Morphismes circulaires

Nous appelons *morphisme circulaire* un morphisme injectif $f\colon \Sigma^* \to \Sigma'^*$ tel que $f(\Sigma)$ est un *code circulaire*, c'est-à-dire

$$\begin{gathered}\forall u \in \Sigma^*, \forall v \in \Sigma^*, \forall x \in \Sigma, \forall t \in \Sigma'^*, \forall s \in \Sigma'^+, \\ [(f(x) = ts \text{ et } f(u) = sf(v)t) \Longrightarrow (t = \varepsilon \text{ et } u = xv)]\end{gathered}$$

(voir [6]).

Remarque. Il existe une autre définition de la circularité, plus faible, adoptée par exemple par Mignosi et Séébold [9]. Cette notion n'est utile que dans le cas de morphismes destinés à être itérés (i.e. dont on considère un point fixe), car le lemme de synchronisation correspondant ne vaut que pour les facteurs de ce point fixe. Pour des morphismes qui ne sont pas itérés, par exemple un morphisme de codage appliqué une seule fois, comme le morphisme g du théorème 5.1, elle n'est plus utilisable : on peut alors recourir à une notion plus générale de *circularité sur un langage* [6] ; nous conservons ici la définition la plus restrictive afin de simplifier l'exposé. Notons que le morphisme θ définissant la suite de Thue-Morse n'est pas circulaire selon cette définition.

Pour les morphismes circulaires, on a le lemme de synchronisation suivant [6] :

Lemme 4.3 *Soit f un morphisme circulaire. Il existe trois ensembles finis N, R_1 et R_2 de mots de Σ'^* tels que que tout mot $u \in F(f(\Sigma^*)) \setminus N$ s'écrit de manière unique $u = r_1.f(v).r_2$ avec $v \in \Sigma^*$ et $r_i \in R_i$. Par ailleurs, si $f(v') = u_1uu_2$ avec $v' \in \Sigma^*$, alors on reconnaît v dans v' : $v' = v_1vv_2$ avec $f(v_1) = u_1r_1$ et $f(v_2) = r_2u_2$.*

Il peut y avoir plusieurs choix possibles pour les ensembles R_i. Quand $f(\Sigma)$ est un code bipréfixe (i.e. si $f(u)$ est préfixe de $f(v)$, alors u est préfixe de v, et de même avec les suffixes), on peut prendre pour R_1 l'ensemble des suffixes stricts d'images de lettres, et pour R_2 l'ensemble des préfixes stricts. Cela n'est pas possible en général. Par exemple, pour le morphisme sur $\{a, b\}^*$ défini par $f(a) = ab$ et $f(b) = aba$, on peut prendre $N = \{\varepsilon, a, b, aa, ba, baa\}$, $R_1 = \{\varepsilon, a, b, ba\}$ et $R_2 = \{ab, aba, abaa\}$: R_1 ne pose pas de problème car le code est suffixe, mais par contre R_2 ne peut contenir ε et a.

Pour simplifier, nous supposerons donc que $f(\Sigma)$ est un code bipréfixe. Cette hypothèse n'est en fait pas indispensable, ce qui compte est que $f(\Sigma)$ est un *code à délai de déchiffrage borné*, et c'est le cas pour tout morphisme circulaire. Il pourra toutefois être nécessaire, quand le code n'est pas bipréfixe, de distinguer les facteurs spéciaux en fonction des mots d'une certaine longueur par lesquels ils se prolongent, et non plus seulement les lettres.

4.2.2 Action sur les facteurs bispéciaux

Soit $L \subset \Sigma^*$ un langage factoriel prolongeable, et $L' = F(f(L))$ la clôture factorielle de son image par le morphisme circulaire f (L' est également un langage factoriel prolongeable). Nous allons montrer, dans le cas où $f(\Sigma)$ est bipréfixe, qu'il est possible de décrire les facteurs bispéciaux de L' connaissant ceux de L et les couples de lettres qui peuvent les prolonger. De même, on peut construire les arbres de facteurs spéciaux de L' en fonction de ceux de L. En pratique, il est souvent plus simple de représenter d'abord ces arbres et d'y lire les facteurs bispéciaux d'ordre non nul.

Soit u un facteur bispécial de L' n'appartenant pas à N. Grâce au lemme 4.3, le mot u s'écrit de manière unique $u = r_1.f(v).r_2$ avec $r_1 \in R_1$, $r_2 \in R_2$ et $v \in \Sigma^*$. De plus, la deuxième partie du lemme implique que $v \in L$. Puisqu'on a supposé $f(\Sigma)$ bipréfixe, on peut choisir R_1 et R_2 de sorte que r_1 soit un suffixe strict et r_2 un préfixe strict d'image de lettre.

Comme u est spécial à droite, il se prolonge au moins de deux manières ua et ub. Les mots r_2a et r_2b ne peuvent être préfixes de la même image de lettre, donc v se prolonge par au moins deux lettres différentes x et y, avec r_2a préfixe de $f(x)$ et r_2b préfixe de $f(y)$, c'est-à-dire que v est spécial à droite dans L ; son ordre à droite est supérieur ou égal à celui de u. De même, v est spécial gauche dans L, donc est bispécial.

Réciproquement, si v est un facteur bispécial, l'ensemble $R_1f(v)R_2$ contient un certain nombre de facteurs bispéciaux, dont la somme des ordres est exactement l'ordre de v à condition qu'aucun d'entre eux ne soit dans N. Les couples (r_1, r_2) tels que $r_1f(v)r_2$ est bispécial, et les ordres correspondants, ne dépendent que de l'ensemble des couples de lettres (x_1, x_2) tels que $x_1vx_2 \in L$.

On voit donc que, connaissant les facteurs bispéciaux de L, on peut calculer des facteurs bispéciaux de L' en appliquant f et en ajoutant un ou plusieurs préfixes et suffixes r_1 et r_2, qui ne dépendent que des lettres par lesquelles le facteur considéré se prolonge. On obtient ainsi tous les facteurs bispéciaux de L', sauf peut-être ceux qui sont éléments de N, qu'il faut déterminer à la main, et que nous appellerons *facteurs bispéciaux exceptionnels*.

4.3 Complexité du point fixe d'un morphisme

À tout morphisme $f\colon \Sigma^* \to \Sigma^*$, on associe le langage

$$L(f) = F\left(\{\ f^n(x)\ \mid n \in \mathbb{N} \text{ et } x \in \Sigma\ \}\right)$$

(c'est la clôture factorielle d'un *DOL-langage* où toutes les lettres de l'alphabet servent d'axiomes), qui est prolongeable pourvu que f soit non-effaçant (l'image d'une lettre n'est jamais le mot vide) et que chaque lettre soit prolongeable des deux côtés, c'est-à-dire

$$\forall x \in \Sigma, \exists a, b \in \Sigma, axb \in L(f)$$

ce qui exclut les morphismes du type $(a \mapsto ab, b \mapsto bb)$. Si de plus il existe une lettre $a \in \Sigma$ telle que $f(a) \in a\Sigma^*$ et que $\bigcup_{n\in\mathbb{N}} \mathrm{Alph}\,(f^n(a)) = \Sigma$, la suite de mots $(f^n(a))$ a pour limite un mot infini $\mathbf{u} = f^\omega(a)$ dont l'ensemble des facteurs est exactement $L(f)$.

Quand f est circulaire il est possible, en itérant le procédé décrit ci-dessus, de décrire l'ensemble des facteurs bispéciaux de $L(f)$ à partir des seuls bispéciaux exceptionnels, ce qui permet de construire les arbres de facteurs spéciaux. Si f est bipréfixe, les facteurs bispéciaux de $L(f)$ sont tous de la forme

$$u = s_0f(s_1)f^2(s_2)\dots f^{k-1}(s_{k-1})f^k(v)f^{k-1}(p_{k-1})f^{k-2}(p_{k-2})\dots f^2(p_2)f(p_1)p_0$$

avec v bispécial exceptionnel, p_i plus grand préfixe commun de deux images de lettres, et s_i plus grand suffixe commun (pas nécessairement des mêmes lettres).

4.4 Un exemple non primitif

Considérons le morphisme suivant :

$$\begin{array}{rrcl} f\colon & \Sigma^* & \longrightarrow & \Sigma^* \\ & a & \longmapsto & aba \\ & b & \longmapsto & bb \end{array} .$$

Ce morphisme n'est pas primitif, car les images successives de b ne contiennent aucun a (un morphisme f est dit primitif s'il existe un entier n tel que l'image de chaque lettre par f^n contient toutes les lettres). La suite qu'il engendre,

$$f^\omega(a) = ababbababbbbababbababbbbbbbbababbababbbbababbababbbbbbbbbbbbbbbbab\ldots$$

est récurrente (tout facteur apparaît une infinité de fois), mais pas uniformément récurrente puisqu'il existe des facteurs arbitrairement longs qui ne contiennent pas de a.

Le morphisme f est bipréfixe, mais n'est pas circulaire. Toutefois, tout facteur de longueur supérieure ou égale à 4 qui n'est pas une puissance de b se synchronise. On va donc appliquer la méthode décrite en 4.3, en traitant à part les puissances de b.

Parmi les petits facteurs (longueur inférieure à 4) non puissances de b, on trouve que seul bab est bispécial : c'est donc le seul bispécial exceptionnel. Les autres facteurs bispéciaux non puissances de b seront les images itérées de ce mot par f, car $p_i = s_i = \varepsilon$ est le seul préfixe ou suffixe commun à $f(a)$ et $f(b)$. Les puissances de b sont toutes des facteurs bispéciaux (c'est d'ailleurs une propriété générale des suites qui contiennent des puissances arbitrairement grandes de b mais aussi une infinité de a), stricts quand l'exposant est une puissance de 2, ordinaires sinon.

Les facteurs bispéciaux sont donc :

<u>bispéciaux stricts :</u> $f^m(b)$, de longueur 2^m, pour $m \geqslant 0$;

<u>bispéciaux ordinaires :</u> les autres puissances de b : ε, b^3, b^5, b^6, b^7, b^9, etc. ;

<u>bispéciaux faibles :</u> $f^m(bab)$, de longueur $2^m(3+m/2)$, pour $m \geqslant 0$.

Ceci fournit une expression des séries BS et BF sous forme de sommes infinies, ainsi que les équations fonctionnelles suivantes :

$$BS(X) = BS(X^2) + X \ ;$$

$$BF(X) = \varphi(X,X) \text{ avec } \varphi(X,Y) = \varphi(XYX, YY) + YXY \ .$$

On peut enfin écrire que $s(n) = 1 + q - r$, où $q = \lceil \log_2 n \rceil$ est le nombre de facteurs bispéciaux stricts de longueur inférieure à n, et $r = \left\lceil \frac{W(128 n \log 2)}{\log 2} - 6 \right\rceil$ est le nombre de facteurs bispéciaux faibles de longueur inférieure à n ($W(x)$ est l'unique fonction analytique réelle sur $]-1/e, +\infty[$ telle que $W(x)e^{W(x)} = x$). On calcule ensuite :

$$\begin{aligned} p(n) &= 1 + \sum_{m=0}^{n-1} s(m) \\ &= 1 + n + \sum_{m=0}^{q-1} (n - 1 - 2^m) - \sum_{m=0}^{r-1} (n - 1 - 2^m(3 + m/2)) \\ &= 1 + n + (n-1)q - (2^q - 1) - (n-1)r + (2^{r+1} + 2^{r-1}r - 2) \ . \end{aligned}$$

Dans cette expression, les termes dominants sont nq et nr, soit

$$p(n) = n\frac{\log n - W(128n\log 2)}{\log 2} + O\left(n\right) \quad .$$

On peut alors, en utilisant le fait que $W(x) = \log x - \log\log x + O\left(\frac{\log\log x}{\log x}\right)$ obtenir un équivalent de la complexité de la suite étudiée : $p(n) \sim n\log_2\log_2 n$.

Notons que le théorème de Pansiot [12] permet d'établir que $p(n)$ croît comme $n\log\log n$, car f est *polynomialement divergent*, mais ne donne pas d'équivalent exact.

5 Suites de complexité affine

5.1 Suites de complexité ultimement affine

J.-P. Allouche et J. Berstel ont posé le problème suivant : pour quels entiers α et β existe-t-il des suites dont la complexité est la fonction affine $p(n) = \alpha n + \beta$? Si on veut que ce soit vrai pour tout $n \geqslant 0$, la seule possibilité est $\beta = 1$ et $\alpha = \#\,\mathrm{Alph}\,(\mathbf{u}) - 1$ (sur un alphabet binaire, cela correspond aux suites sturmiennes, de complexité $p(n) = n + 1$; sur un alphabet plus grand, de telles suites existent également, voir par exemple la suite $f^\omega(c_1)$ du théorème 5.1). On recherche donc des fonctions de complexité ultimement affines, c'est-à-dire telles qu'il existe un rang n_0 tel que pour tout $n \geqslant n_0$ on a $p(n) = \alpha n + \beta$. Pour que ce soit possible, il faut que $(\alpha, \beta) \in \{0,1\} \times (\mathbb{N} \setminus \{0\}) \cup (\mathbb{N} \setminus \{0,1\}) \times \mathbb{Z}$. Nous verrons que cette condition est suffisante, et ce sur tout alphabet ayant au moins deux éléments.

Allouche [1] a développé plusieurs techniques pour construire, à partir d'une suite de complexité ultimement $\alpha n + \beta$, des suites de complexités différentes, par exemple $\alpha n + \beta + 1$. Cependant il reste des couples (α, β) qui ne peuvent être obtenus par ces techniques en partant des quelques suites de complexité ultimement affine déjà connues. Il est donc nécessaire de construire de nouveaux points de départ. C'est dans cet objectif que cette étude a été faite, mais celui-ci a été dépassé puisque nous obtenons directement toutes les complexités ultimement affines possibles.

Pour cela, nous considérons une famille à trois paramètres de suites, définies par un morphisme itéré sur un alphabet plus ou moins grand, suivi d'un codage vers l'alphabet à deux lettres $\Sigma = \{a, b\}$.

Théorème 5.1 *Soient $j \geqslant 0$, $k \geqslant 1$ et $\ell \geqslant 1$ trois entiers. Soit $\Sigma_k = \{c_1, \ldots, c_k\}$ un alphabet à k lettres. On considère les deux morphismes f et g définis ainsi :*

$$\begin{array}{rrcll} f\colon & \Sigma_k^* & \longrightarrow & \Sigma_k^* & \\ & c_i & \longmapsto & c_1 c_{i+1} & \textit{si } i \neq k \\ & c_k & \longmapsto & c_1 & \end{array} \qquad \textit{et} \qquad \begin{array}{rrcl} g\colon & \Sigma_k^* & \longrightarrow & \Sigma^* \\ & c_i & \longmapsto & a^\ell b^{i+j} \end{array}$$

pour tout i compris entre 1 et k.

La suite $\mathbf{u} = g(f^\omega(c_1))$ a pour complexité

$$p(n) = (k-1)n - \frac{k(k-1)}{2} - j(k-2) + \ell + 1$$

pour $n \geqslant \max(\ell, j+k-1, j+1)$.

Preuve. Calculons tout d'abord les facteurs bispéciaux de $L(f)$. Le morphisme f est circulaire et suffixe (mais non bipréfixe). Dans le lemme 4.3, on peut prendre $N = R_1 = \{\varepsilon, c_2, \ldots, c_k\}$ et $R_2 = \{c_1, c_1c_2, \ldots, c_1c_k\}$. Le seul bispécial exceptionnel est ε, d'ordre 0 car il se prolonge par (c_1, c_i) et (c_i, c_1) pour $1 \leqslant i \leqslant k$. Par ailleurs, un mot de la forme $u = r_1 f(v) r_2$ ne peut être bispécial que si $r_1 = \varepsilon$ et $r_2 = c_1$. En partant de ε, on obtient c_1, puis $c_1c_2c_1$, etc., qui se prolongent par (c_m, c_i) et (c_i, c_m) pour tout i et pour un indice m fixé, et sont donc tous d'ordre 0.

On en déduit que la base des arbres des facteurs spéciaux à gauche de $L(f)$ ne contient qu'un seul arbre, de type Σ_k, puisque chaque facteur bispécial peut être prolongé à gauche par chacune des k lettres de l'alphabet, et donc il en est de même pour chaque facteur spécial à gauche. Comme tous les facteurs bispéciaux sont d'ordre nul, cet arbre ne peut être que filiforme, et est d'ailleurs étiqueté par la suite elle-même. En quelque sorte, $f^\omega(c_1)$ est une suite de Fibonacci généralisée, et sa complexité est $(k-1)n + 1$.

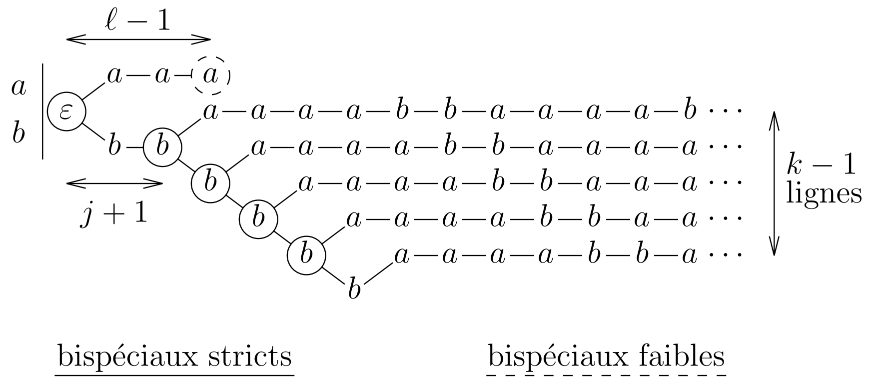

FIG. 6 – Arbre des facteurs spéciaux à gauche de $g(f^\omega(c_1))$ pour $j = 1$, $k = 6$, $\ell = 4$

On applique ensuite le morphisme g, qui est également circulaire et suffixe, pour obtenir les facteurs bispéciaux de $F(g(L(f)))$. On peut prendre

$$N = \{\ a^i b^{i'}\ \mid i < \ell, i' \leqslant j + k\ \}\ ,$$

$$R_1 = \{\ b^{i'}\ \mid i' \leqslant j + k\ \} \cup \{\ a^i b^{i'}\ \mid 1 \leqslant i < \ell, j < i' \leqslant j + k\ \}$$

et

$$R_2 = \{\ a^i\ \mid 1 \leqslant i < \ell\ \} \cup \{\ a^\ell b^{i'}\ \mid i' \leqslant j + k\ \}\ .$$

Les facteurs bispéciaux exceptionnels sont ε, d'ordre 1, $b^{i'}$ pour $1 \leqslant i' \leqslant j$, d'ordre 0, $b^{i'}$ pour $j + 1 \leqslant i' \leqslant j + k - 2$, d'ordre 1, b^{j+k-1}, d'ordre 0, a^i pour $1 \leqslant i \leqslant \ell - 2$, d'ordre 0, et $a^{\ell-1}$, d'ordre -1, sauf dans le cas $\ell = 1$ où $\varepsilon = a^{\ell-1}$ est d'ordre 0. Les autres bispéciaux sont de la forme $b^{j+i} f(v) a^\ell b^{j+i'}$ où v est un bispécial de $L(f)$ se prolongeant par (c_m, c_i) et (c_i, c_m), et $i \leqslant m$ et $i' \leqslant m$. Il sont tous d'ordre 0.

Il ne reste plus qu'à compter les facteurs bispéciaux d'ordre non nul pour obtenir la complexité cherchée : si $\ell > 1$, il y a un bispécial faible de longueur $\ell - 1$, et $k - 1$ bispéciaux stricts de longueurs 0 et $j + 1$ à $j + k - 2$, et si $\ell = 1$ il y a seulement $k - 2$ bispéciaux stricts de longueur $j + 1$ à $j + k - 2$.

On en déduit également la forme de l'arbre des facteurs spéciaux à gauche de $F(g(L(f)))$ (figure 6). Cet arbre comporte $k - 1$ branches infinies (chaque facteur spécial à gauche pour $L(f)$, d'ordre $k-1$, donne un élément dans chacune des branches) et une branche finie de longueur $\ell - 1$, terminée par le facteur bispécial exceptionnel $a^{\ell-1}$. ■

Corollaire 5.2 *Pour tout couple d'entiers* $(\alpha, \beta) \in \{0,1\} \times (\mathbb{N}\setminus\{0\}) \cup (\mathbb{N}\setminus\{0,1\}) \times \mathbb{Z}$, *il existe une suite binaire de complexité ultimement* $\alpha n + \beta$.

Preuve. La suite $\mathbf{u}$ définie dans le théorème 5.1 fournit une solution dans tous les cas, avec $k = \alpha + 1 \geqslant 1$ et j et ℓ choisis en fonction de α et β.

Si $k = 1$, la suite $\mathbf{u}$ a pour complexité ultimement $j + \ell + 1$. Si on prend $j = \left\lfloor \frac{\beta-1}{2} \right\rfloor$ et $\ell = \left\lfloor \frac{\beta}{2} \right\rfloor$, on a $p(n) = \beta$ pour $n \geqslant \left\lfloor \frac{\beta+1}{2} \right\rfloor$.

Si $k = 2$, la suite $\mathbf{u}$ a pour complexité ultimement $n + \ell$. On prend donc $\ell = \beta$ et (par exemple) $j = 0$, et on a $p(n) = n + \beta$ pour $n \geqslant \beta$.

Si $k > 2$, soit par exemple

$$j = \max\left(0, \left\lfloor \frac{-k^2 + 3k - 2 - 2\beta}{2k-4} \right\rfloor\right)$$

et

$$\ell = \beta + \frac{k(k-1)}{2} + j(k-2) - 1 \geqslant 1 \ .$$

La suite $\mathbf{u}$ a alors pour complexité $p(n) = \alpha n + \beta$ pour

$$n \geqslant \max\left(\alpha, \frac{\alpha}{2} - \frac{\beta}{\alpha - 1}, \beta + \frac{\alpha(\alpha+1)}{2} - 1\right) \ .$$

■

5.2 Complexité $\alpha n + \beta$ pour $n \geqslant 1$

La question suivante, variante du problème ci-dessus, a été posée par Pascal Alessandri : pour quelles valeurs de (α, β) existe-t-il une suite de complexité $p(n) = \alpha n + \beta$ pour tout $n \geqslant 1$? Elle peut être résolue par une construction similaire.

Théorème 5.3 *Soit* $(\alpha, \beta) \in \mathbb{N} \times \mathbb{Z}$. *Il existe une suite de complexité* $p(n) = \alpha n + \beta$ *pour tout* $n \geqslant 1$ *(et* $p(0) = 1$*) si et seulement si* $\alpha + \beta \geqslant 1$ *et* $2\alpha + \beta \leqslant (\alpha + \beta)^2$.

Preuve. Ces conditions sont évidemment nécessaires : $p(1) = \alpha + \beta$ est le nombre de lettres distinctes apparaissant dans la suite, donc est un nombre strictement positif ; et si la suite contient $\alpha + \beta$ lettres distinctes, elle ne peut contenir plus de $(\alpha + \beta)^2$ facteurs de longueur 2, donc $2\alpha + \beta = p(2) \leqslant p(1)^2 = (\alpha + \beta)^2$.

Supposons maintenant que (α, β) satisfait les deux conditions, et que $\alpha > 0$, le cas $\alpha = 0$ étant facile à résoudre avec des suites périodiques. Soit $k = \alpha + 1$ et $\ell = \alpha + \beta$: on va construire un HDOL-système $G = (\Sigma_k, f, c_1, \Sigma_l, h)$ où $\Sigma_k = \{c_1, \dots, c_k\}$ et $\Sigma_l = \{d_1, \dots, d_l\}$, de telle sorte que la suite associée $\mathbf{u} = h(f^\omega(c_1))$ aura la complexité voulue.

Le DOL-système sous-jacent (voir [6]) $G^0 = (\Sigma_k, f, c_1)$ est le même que pour le théorème 5.1. La suite $f^\omega(c_1)$ est donc une suite de Fibonacci généralisée, de complexité $\alpha n + 1$ pour tout n.

La construction du morphisme h dépend du signe de β. Supposons tout d'abord que $\ell \geqslant k$, c'est-à-dire que $\beta > 0$. Si $\ell = k$, la suite $f^\omega(c_1)$ a déjà la complexité cherchée et il suffit de prendre $h = Id$. Sinon, on définit h par $h(c_1) = d_1 d_{k+1} d_{k+2} \dots d_l$ et $h(c_i) = d_i$ pour $2 \leqslant i \leqslant k$: le morphisme h est circulaire et bipréfixe. On voit facilement que la base des arbres de facteurs spéciaux à gauche de $F(\mathbf{u})$ a deux éléments, un arbre filiforme de type $\{d_1, \dots, d_k\}$ (image par h de l'arbre de $L(f)$) et un arbre de type Σ_l contenant le seul nœud ε, seul facteur bispécial exceptionnel, et seul facteur bispécial d'ordre non nul, $k - \ell$. On en déduit immédiatement que la complexité est bien $p(n) = \alpha n + \beta$ pour $n \geqslant 1$.

Supposons maintenant que $\ell < k$, c'est-à-dire $\beta \leqslant 0$. On définit h de la manière suivante : $h(c_1) = d_1$, $h(c_i) = d_1 d_i$ pour $2 \leqslant i \leqslant \ell$, et $h(c_{\ell+1})$, ..., $h(c_k)$ sont $k - \ell$ éléments pris parmi l'ensemble $E = d_1\{d_2, \dots, d_l\}^2$, qui en compte $(\ell - 1)^2$; comme $k - \ell = 1 - \beta \leqslant (\alpha + \beta)^2 - 2\alpha - 2\beta + 1 = (\ell - 1)^2$, c'est toujours possible. L'ordre des images n'a pas d'importance.

Le morphisme h est suffixe et circulaire, de délai de synchronisation 2, grâce au rôle de marqueur joué par d_1. Nous pouvons donc déterminer les facteurs spéciaux à gauche de $F(\mathbf{u})$ en fonction de ceux de $L(f)$.

Soit u un facteur spécial à gauche pour $L(f)$: il se prolonge par chacune des lettres de Σ_k. Les suffixes communs des $f(c_i)$ sont ε et toutes les lettres d_i telles que $E \cap d_1 \Sigma_k d_i \neq \emptyset$; les facteurs spéciaux à gauche pour $F(\mathbf{u})$ obtenus sont donc $f(u)$, d'ordre $\ell - 1$ (il se prolonge par chacune des lettres de $H_1 = \Sigma_l$), et $d_i f(u)$, d'ordre $\#(E \cap d_1 \Sigma_k d_i)$ quand cet ensemble est non vide (il se prolonge par $H_i = \{d_1\} \cup \{\ d_j \mid d_1 d_j d_i \in E\ \}$). Les seuls mots non synchronisés sont ε et certains d_i, mais ils ne donnent pas de nouveaux facteurs spéciaux. Finalement, la base contient un arbre pour chaque ensemble H_i, avec une ou plusieurs branches filiformes partant de ε (une branche pour chaque j tel que $H_i \subset H_j$). Comme précédemment, ε est le seul facteur bispécial d'ordre non nul, et son ordre vaut $k - \ell$ (l'ordre de ε est toujours égal à $p(2) - 2p(1) + 1$). ■

Une extension intéressante de ces deux problèmes serait de chercher à construire des suites de complexité $\lfloor \alpha n + \beta \rfloor$, avec α (et éventuellement β) non entier. Il semble que cela ne soit pas possible.

6 Une famille continue de fonctions de complexité

Pour le moment, nous savons construire une famille dénombrable de fonctions de complexité. Nous allons montrer dans cette section que l'ensemble des fonctions de complexité a la puissance du continu.

Théorème 6.1 *Soit $\varphi \colon \mathbb{R} \longrightarrow \mathbb{R}$ une fonction deux fois dérivable et un réel strictement positif x_0 tels que*

(i) $0 \leqslant \varphi''(x) \leqslant 1$ *pour* $x \geqslant x_0$ *;*

(ii) $x \log x = o(\varphi(x))$.

Alors il existe une suite binaire $\mathbf{u}$ *dont la complexité vérifie*

$$p_{\mathbf{u}}(n) \sim \varphi(n) \ .$$

La seconde condition peut être remplacée, au prix d'une complication de la preuve, par l'hypothèse plus faible et plus naturelle que $\varphi'(x)$ tend vers $+\infty$ quand x tend vers $+\infty$.

Preuve. Pour $i \in \mathbb{N}$, soit q_i le plus petit entier supérieur ou égal à x_0 tel que $\lfloor \varphi'(q_i) - \varphi'(x_0) \rfloor = i$. D'après les propriétés de φ, ces entiers existent, sont positifs, et la suite (q_i) est strictement croissante. On définit une suite binaire $\mathbf{u}$ de la manière suivante : on part de la suite $\mathbf{v}$ sur l'alphabet dénombrable $\Sigma_\infty = \{c_0, c_1, \ldots, c_i, \ldots\}$, point fixe du morphisme f qui à c_i associe $c_0 c_{i+1}$:

$$\mathbf{v} = c_0c_1c_0c_2c_0c_1c_0c_3c_0c_1c_0c_2c_0c_1c_0c_4c_0c_1c_0c_2c_0c_1c_0c_3c_0c_1c_0c_2c_0c_1c_0c_5c_0c_1c_0c_2\ldots$$

et on lui applique le morphisme $g\colon \begin{array}{ccc} \Sigma_\infty^* & \longrightarrow & \{a,b\}^* \\ c_i & \longmapsto & ab^{q_i} \end{array}$ pour obtenir $\mathbf{u} = g(\mathbf{v})$.

Les facteurs spéciaux de $\mathbf{v}$ sont faciles à déterminer : à gauche ce sont les préfixes de $\mathbf{v}$, à droite leurs images miroir (on pourrait les appeler facteurs spéciaux d'ordre infini, car ils se prolongent par toutes les lettres de Σ_∞, sauf celles qu'ils contiennent déjà). Comme dans la preuve du théorème 5.1, on utilise le fait que g est un morphisme circulaire (et suffixe) pour en déduire la forme de l'arbre des facteurs spéciaux à gauche de $\mathbf{u}$. Il est constitué par la réunion d'une infinité de branches finies, étiquetées par $b^{q_0}ab^{q_0}$, $b^{q_1}ab^{q_0}ab^{q_1}$, $b^{q_2}ab^{q_0}ab^{q_1}ab^{q_0}ab^{q_2}$, et en général $b^{q_i}g(f^i(c_0))$ pour tout $i \in \mathbb{N}$.

Le mot infini a donc une infinité de facteurs bispéciaux stricts et faibles. Les bispéciaux stricts étiquettent les endroits où les branches se séparent, c'est-à-dire les b^{q_i} pour $i \in \mathbb{N}$, et les bispéciaux faibles sont les étiquettes des branches entières, $b^{q_i}g(f^i(c_0))$ pour $i \in \mathbb{N}$. La longueur des premiers est simplement q_i. Celle des seconds est $r_i = 2q_i + 2^i + \sum_{j=0}^{i-1} 2^{i-j-1} q_j$, que l'on peut minorer par 2^i.

Le nombre de facteurs spéciaux $s(n)$ est alors

$$s(n) = 1 + \min\{\, i \mid n \leqslant q_i \,\} - \min\{\, i \mid n \leqslant r_i \,\} \quad .$$

Le dernier terme est en $O(\log n)$. Sachant que $\lfloor \varphi'(q_i) - \varphi'(x_0) \rfloor = i$, le second terme peut être encadré par

$$\varphi'(n) - \varphi'(x_0) - 1 < \min\{\, i \mid n \leqslant q_i \,\} < \varphi'(n) - \varphi'(x_0) + 1$$

dès que $n \geqslant x_0$. En sommant une nouvelle fois, on trouve que

$$p(n) = 1 + \sum_{m=0}^{n-1} s(m) = \varphi(n) + O(n \log n)$$

et l'hypothèse (ii) nous assure que le second terme est négligeable. ∎

En quelque sorte, ce théorème montre que toutes les fonctions à croissance suffisamment régulière comprises entre $n \log n$ et n^2 sont asymptotiquement des fonctions de complexité. C'est le cas notamment des n^α pour $1 < \alpha < 2$. Par conséquent, le cardinal de l'ensemble des fonctions de complexité est égal à celui de $\mathbb{R}$ (il ne peut bien sûr pas être plus grand).

Remarque. Grillenberger [8] a montré comment construire des suites d'*entropie topologique* h pour tout réel h compris entre 0 et $\log \#\Sigma$. Cela fournit donc une autre famille continue de fonctions de complexité, qui ont une croissance exponentielle (si $h > 0$), contrairement aux nôtres qui sont à croissance polynomiale (et donc toutes associées à des suites d'entropie nulle).

7 Complexité sous-affine

Dans les deux sections précédentes nous avons construit des familles de fonctions de complexité. Nous allons maintenant voir une condition nécessaire que doivent vérifier ces fonctions quand leur croissance n'est pas très rapide, et qui les empêche de croître de façon trop oscillante.

Cette section résume les résultats qui sont présentés de manière détaillée dans [7].

Notre théorème principal, conjecturé par Ferenczi, est le suivant :

Théorème 7.1 *Soit* $\mathbf{u} \in \Sigma^\omega$ *une suite sur un alphabet fini dont la fonction de complexité* $p(n)$ *est sous-affine, c'est-à-dire que* $p(n) = O(n)$*. Alors la fonction* $s(n) = p(n+1) - p(n)$ *est bornée.*

Cela signifie par exemple que la situation suivante ne peut pas se produire : $p(n+1) - p(n)$ prend des valeurs faibles sauf en des "pics" arbitrairement hauts, mais suffisamment étroits et espacés pour ne pas trop faire augmenter $p(n)$.

Une telle propriété était déjà connue pour certains types de suites, points fixes de substitutions primitives [11]. En particulier, le théorème 7.1 permet de redémontrer que si $\mathbf{u}$ est une suite automatique, alors $(p(n+1) - p(n))_{n\in\mathbb{N}}$ est aussi une suite automatique [14].

Notre preuve donne en fait un résultat plus général et plus précis :

Théorème 7.2 *Soit* $\mathbf{u} \in \Sigma^\omega$ *une suite sur un alphabet fini et* $p(n)$ *sa fonction de complexité. On suppose qu'il existe deux réels* $a > 0$ *et* $1 \leqslant \alpha \leqslant \frac{3}{2}$*, et un entier* n_0*, tels que* $p(n) \leqslant an^\alpha$ *pour tout* $n \geqslant n_0$*. Alors*

$$s(n) = p(n+1) - p(n) \leqslant Kp(1)a^3 n^{3(\alpha-1)}$$

pour tout $n \geqslant n_0$*, où* K *est une constante qui ne dépend pas de* $\mathbf{u}$*.*

La constante K de ce théorème peut être explicitée, mais cela présente peu d'intérêt tant qu'il n'est pas établi que l'exposant 3 dans $a^3 n^{3(\alpha-1)}$ est optimal. C'est à cause de cet exposant que α est limité à $\frac{3}{2}$, car au-delà de cette valeur le théorème n'apporte rien de plus que la relation triviale $s(n) \leqslant p(n+1) \leqslant a(n+1)^\alpha$. Mais nous ne connaissons aucune suite dont la complexité vérifie $p(n) = O(n^\alpha)$ mais non $s(n) = O(n^{\alpha-1})$. Il serait intéressant de construire de telles suites, ou bien sûr de montrer qu'il n'en n'existe pas, ce qui renforcerait considérablement le théorème 7.2.

Pour démontrer les théorèmes 7.1 et 7.2, on utilise le lemme suivant, dont la preuve, très technique, consiste à compter des chemins dans le graphe de Rauzy d'ordre n_1 associés à des facteurs spéciaux de longueur comprise entre n_1 et n_2 :

Lemme 7.3 *Soit* L *un langage factoriel prolongeable sur un alphabet binaire,* $p(n)$ *sa fonction de complexité, et* $n_1 < n_2$ *deux entiers. On suppose que* $s(n_1) \geqslant 1$*. Soit* $m = \max_{n_1 \leqslant n < n_2} s(n)$*. Alors la variation de complexité entre* n_1 *et* n_2 *est au moins*

$$p(n_2) - p(n_1) \geqslant \frac{n_1}{2s(n_1)^2} \Big(m - s(n_1)\big(1 + s(n_1) + s(n_2 - 1)\big)\Big) \quad .$$

De façon imagée, il signifie que si $s(n)$ présente un pic entre n_1 et n_2, alors $p(n)$ croît d'une valeur d'autant plus grande que le pic est haut et que $s(n_1)$ et $s(n_2)$ sont petits, mais qui ne dépend pas de la largeur du pic. Par un calcul relativement simple, on établit alors un résultat similaire au théorème 7.2 pour les langages factoriels prolongeables sur un alphabet binaire. Au moyen d'un codage, il peut être étendu aux langages sur un alphabet fini quelconque. Pour obtenir le théorème 7.2, il ne reste plus qu'à traiter le cas des suites non récurrentes comme indiqué en 2.2.

8 Conclusion

Nous avons vu que l'étude des facteurs spéciaux permettait de calculer plus facilement certaines complexités, et d'obtenir d'autres résultats intéressants sur les fonctions de complexité. Cela est surtout vrai pour les suites dont la complexité croît linéairement, ou à la rigueur de façon quadratique, car le nombre de facteurs bispéciaux est la différence seconde de la complexité qui est alors faible. Il serait certainement utile d'avoir une interprétation combinatoire de différences d'ordre plus élevé, mais cela ne semble pas possible avec nos techniques, car les facteurs bispéciaux n'ont plus aucune propriété de prolongeabilité, ce qui interdit de les représenter par une structure comme l'arbre des facteurs spéciaux.

Nous sommes encore très loin de pouvoir caractériser les fonctions qui sont des fonctions de complexité, même si les théorèmes 6.1 et 7.1 permettent de se faire une meilleure idée de l'endroit où se situe la limite. Il doit être possible de resserrer encore cet encadrement, en cherchant la borne exacte sur $s(n)$ dans le théorème 7.1, en étendant ce théorème de façon plus précise aux degrés plus grands, et le théorème 6.1 à d'autres classes de fonctions.

Une autre direction de recherche possible est de voir comment des restrictions sur les suites considérées influent sur les fonctions réalisables : si $\mathcal{L}$ est une famille de langages ou de suites, notons $\mathbb{P}(\mathcal{L})$ l'ensemble des fonctions de complexité des éléments de $\mathcal{L}$. Obtient-on des ensembles différents quand $\mathcal{L} = \Sigma^\omega$, $\mathcal{L} = {}^\omega\Sigma^\omega$ (suites biinfinies), $\mathcal{L}$ est l'ensembles des suites récurrentes ou uniformément récurrentes, des langages factoriels prolongeables, etc. ?

Références

[1] J.-P. Allouche, Exposé aux *Journées Montoises à Bordeaux* (1993).

[2] J.-P. Allouche, Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* **1** (1994) 133–143.

[3] P. Arnoux et G. Rauzy, Représentation géométrique de suites de complexité $2n+1$, *Bull. Soc. Math. France* **199** (1991) 199–215.

[4] S. Brlek, Enumeration of factors in the Thue-Morse word, *Discr. Appl. Math.* **24** (1989) 83–96.

[5] J. Cassaigne, Counting overlap-free binary words, in *STACS '93*, P. Enjalbert, A. Finkel et K. W. Wagner (éd.), Würzburg, *Lect. Notes Comput. Sci.* **665**, Springer-Verlag (1993) 216–225.

[6] J. Cassaigne, Motifs évitables et régularités dans les mots, *Thèse de Doctorat*, Université Paris 6 (1994). Rapport LITP TH 94-04.

[7] J. Cassaigne, Special factors of sequences with linear subword complexity, in *Developments in Language Theory*, World Scientific (1996).

[8] C. Grillenberger, Constructions of strictly ergodic systems — I. Given entropy, *Z. Wahr. verw. Geb.* **25** (1973) 323–334.

[9] F. Mignosi et P. Séébold, If a DOL language is k-power-free then it is circular, in *ICALP '93*, Lund, *Lect. Notes Comput. Sci.* **700**, Springer-Verlag (1993).

[10] M. Morse et G. A. Hedlund, Sturmian sequences, *Amer. J. Math.* **61** (1940) 1–42.

[11] B. Mossé, Notions de reconnaissabilité pour les substitutions et complexité des suites automatiques. *Preprint* (1993).

[12] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in *ICALP '84*, *Lect. Notes Comput. Sci.* **172**, Springer-Verlag (1984) 380–389.

[13] G. Rote, Sequences with subword complexity $2n$, *J. Number Th.* **46** (1994) 196–213.

[14] T. Tapsoba, Complexité de suites automatiques, *Thèse de 3e cycle*, Université d'Aix-Marseille II (1987).

Julien Cassaigne
Institut de Mathématiques de Luminy
163 avenue de Luminy, case 930
F-13288 Marseille Cedex 9 (France)
Julien.Cassaigne@iml.univ-mrs.fr