Université de Picardie Jules Verne
École doctorale Sciences et Santé

Contributions à la combinatoire des mots : traces et morphismes

Document de synthèse de l'activité scientifique
de

Gwénaël Richomme

présenté le 5 février 2004
en vue de l'obtention de
l'**habilitation à diriger des recherches**
en **informatique**

Devant le jury composé de :

Jean Berstel, Marne-La-Vallée (Président)
Christian Choffrut, Paris VII (rapporteur)
Volker Diekert, Stuttgart (rapporteur)
Michel Habib, Montpellier II
Yves Métivier, Bordeaux I
Patrice Séébold, Montpellier III (directeur)
Vincent Villain, Amiens (rapporteur)

Remerciements

Ce mémoire présente des travaux réalisés après ma thèse, depuis mon arrivée en 1994 à Amiens jusqu'en juillet 2003. Ces travaux n'auraient pu être réalisés sans la présence, le travail et le soutien de nombreux collègues, entre autres du LaRIA et de la Faculté de Mathématiques et d'Informatique. Je tiens à remercier plus particulièrement quelques personnes.

Ayant déjà eu l'occasion de le faire, je remercie de nouveau Yves Métivier pour avoir guidé mes premiers pas en recherche et je le remercie vivement de participer au jury.

Depuis mon arrivée à Amiens, je dois énormément à Patrice Séébold. Malgré de nombreuses autres charges, il m'a pris sous son aile, m'a communiqué son enthousiasme pour un domaine passionnant, m'a prodigué nombre de conseils, m'a consacré beaucoup de temps, ... Pour tout : Merci !

Je suis extrêmement reconnaissant à Jean Berstel de présider le jury mais surtout pour l'intérêt qu'il porte, depuis déjà quelques temps, à mes travaux.

Je tiens à exprimer toute ma gratitude à Christian Choffrut pour avoir accepté d'être rapporteur, ainsi que pour ses conseils.

Vielen Dank an Volker Diekert, der mir die Ehre erwiesen hat, mein Referent zu sein, trotz eines vollen Terminkalenders. [1]

Je remercie vivement Michel Habib de me faire l'honneur de participer au jury.

Depuis mon arrivée à Amiens, j'ai beaucoup travaillé avec Vincent Villain sur des aspects pédagogiques et administratifs et j'ai eu régulièrement avec lui des discussions informelles sur des problématiques de recherche. C'est donc un grand plaisir qu'il me fait de participer à ce jury, d'autant plus en tant que rapporteur.

Je tiens également à remercier Daniel Kirsten pour son invitation à Dresde et pour le travail réalisé ensemble.

J'ai également une pensée pour tous les étudiants qui ont travaillé avec moi, en particulier Francis et Florence.

Pour finir, j'aimerais remercier mon épouse et ma famille pour leur soutien et leur compréhension.

1. Merci à Dominique Séébold pour sa traduction en allemand.

Table des matières

1 Introduction 1

2 Monoïdes et morphismes 4

3 Traces 7
- 3.1 Monoïdes de traces . . . 7
- 3.2 Le problème de l'étoile . . . 9
- 3.3 Le problème de la puissance finie . . . 11
- 3.4 Équivalence de décidabilité . . . 14
- 3.5 Autres approches . . . 15
- 3.6 Perspectives . . . 16

4 Morphismes sans répétition 19
- 4.1 Monoïdes d'endomorphismes . . . 21
- 4.2 Ensembles de test pour morphismes sans chevauchement . . . 23
- 4.3 Le cas des morphismes uniformes . . . 28
- 4.4 Morphismes sans puissance k . . . 30
- 4.5 Perspectives . . . 33

5 Morphismes de Lyndon 36
- 5.1 Morphismes préservant l'ordre sur les mots finis . . . 36
- 5.2 Ensembles de test . . . 37
- 5.3 Monoïdes . . . 39
- 5.4 Perspectives . . . 40

6 Autour des mots et morphismes sturmiens 42
- 6.1 Mots sturmiens . . . 42
- 6.2 Morphismes sturmiens et équilibrés . . . 43
- 6.3 Morphismes épisturmiens . . . 44
- 6.4 Conjugaison de morphismes . . . 47
- 6.5 Perpectives . . . 50

7 Points fixes finis de morphismes 51
- 7.1 Ensemble des points fixes d'un morphisme . . . 51
- 7.2 Une conjecture . . . 52
- 7.3 Perspectives . . . 54

8 Conclusion 55

Références 56

Publications 62

Annexes. Encadrements et responsabilités administratives 65

1 Introduction

Le but de ce mémoire est de présenter mes travaux de recherche effectués lors des douze dernières années. Le cadre général en est l'étude des mots que ce soit du point de vue des langages formels ou d'un point de vue combinatoire. Deux thèmes particuliers ont été abordés. Le premier est l'étude d'un problème de décidabilité concernant les parties reconnaissables des monoïdes de traces. Le deuxième thème, commencé à mon arrivée à Amiens en 1994, concerne l'étude de propriétés de certains mots, et surtout de certains morphismes.

Les mots sont omniprésents en informatique (comme dans tout système de communication). Par exemple, un mot peut représenter l'exécution d'un programme. Dans ce cas, une action du programme est représentée par une lettre et l'opération de concaténation des mots permet de rendre compte de la séquentialité des opérations exécutées. L'ensemble des mots (sur un alphabet donné) possède ainsi une structure de monoïde libre. Une généralisation de cette structure, les monoïdes libres partiellement commutatifs (voir, e.g., [DR95]), aussi appelés monoïdes de traces, permet de modéliser des ensembles d'exécutions équivalentes sur un système réparti.

Pour le développement de certaines applications (par exemple des compilateurs), les langages formels (voir, e.g., [RS97]) et en particulier les langages reconnaissables jouent un rôle important. Certains problèmes simples dans le cadre d'un monoïde libre s'avèrent beaucoup plus épineux dans les monoïdes de traces. C'est le cas de la question : étant donné un ensemble reconnaissable, son itération finie est-elle aussi reconnaissable ? Dans le monoïde libre, la réponse est toujours oui. Dans le cas d'un monoïde de traces quelconque, la réponse peut être négative et la décidabilité de la question, appelée problème de l'étoile, est un problème ouvert. L'étude de ce problème était l'objet de ma thèse de doctorat soutenue en 1994. L'objectif de la partie 3 est surtout de montrer les développements ultérieurs menés en particulier avec D. Kirsten, mais aussi avec Y. Métivier et P.-A. Wacrenier.

Outre l'aspect modélisation, les mots ont un impact sur des domaines plus ou moins connexes de l'informatique. Par exemple, l'algorithmique du texte [CHL01, CR94] trouve des applications en bioinformatique. La connaissance combinatoire des mots aide alors à produire des algorithmes efficaces. En conséquence, le domaine de la combinatoire des mots [Lot83, CK97, Lot02, BK03] a pris un essor particulier depuis une vingtaine d'années.

Dans ce cadre, un outil important est la notion de morphismes entre monoïdes libres. Par définition, il s'agit d'applications préservant la structure de monoïde. Dans les parties 4, 5 et 6, nous nous intéressons de manière combinatoire à plusieurs familles de morphismes préservant une “bonne propriété”. Un exemple classique est le morphisme engendrant le mot infini de Thue-Morse [Thu12, Mor21]. Ce morphisme préserve l'absence de chevauchement (une forme de répétition de facteurs). En conséquence, le mot de Thue-Morse est un mot infini sans chevauchement. Notons que la plupart du temps les morphismes étudiés sont injectifs. Ainsi l'ensemble des images des lettres forme un code fini. Ces travaux peuvent donc aussi être considérés comme des travaux en théorie des codes [BP85].

La partie 4 présente des travaux réalisés en partie avec P. Séébold et F. Wlazinski relatifs à un problème ancien : comment savoir si un morphisme préserve les mots sans chevauchement ou les mots sans puissance k ? Plusieurs réponses sont connues dans le cas des endomorphismes binaires qui préservent les mots sans chevauchement. Tout d'abord l'ensemble de ces mots est un monoïde finiment engendré. Chercher à décomposer un morphisme sur la base de ce monoïde permet de tester s'il est sans chevauchement. Il est également possible de tester la préservation de la propriété *sans chevauchement* en analysant l'image d'un seul mot. Dans cette partie 4, nous étudions de manière systématique la généralisation de ces deux approches pour des alphabets non nécessairement binaires. Cela nous amène entre autres à introduire la notion d'ensemble de test pour morphismes sans chevauchement ou pour morphismes sans puissance k, et à étudier ces ensembles quand ils sont finis.

La partie 5 reproduit cette approche pour une autre famille de morphismes à savoir ceux préservant les mots de Lyndon. Un moyen de tester l'appartenance à cette famille de morphismes est fourni. L'étude menée dans cette partie s'avère indissociable de l'étude des morphismes préservant l'ordre lexicographique.

Au cœur de la partie 6 se trouvent les morphismes qui préservent les mots sturmiens. Il est déjà connu que l'ensemble de ces morphismes est un monoïde finiment engendré. Il existe même un mot de test pour cette famille de morphismes. Nous montrons l'optimalité de la longueur de ce mot. Les méthodes utilisées nous permettent de caractériser les morphismes qui préservent les mots équilibrés. L'essentiel de cette partie 6 reste toutefois l'étude des morphismes épisturmiens, une généralisation des morphismes sturmiens. Cela nous conduit à considérer la conjugaison des morphismes, un outil important dans l'étude des morphismes épisturmiens.

Alors que les morphismes présentés dans les parties précédentes sont plutôt utilisés pour obtenir des mots infinis avec de bonnes propriétés, la partie 7 s'intéresse à des travaux, débutés récemment avec F. Levé, concernant des mots finis points fixes de morphismes.

Terminons cette introduction par quelques remarques sur le mémoire lui-même. Tout au long du document, je me suis efforcé d'expliquer les problématiques, les résultats obtenus et les méthodes utilisées. En particulier, je présente nombre de morphismes qui permettent d'établir l'optimalité de certains résultats. A contrario par souci de ne pas être trop technique, je n'ai pas (ou très peu) parlé des nombreuses preuves combinatoires réalisées souvent par analyse de cas.

Les citations bibliographiques numériques font référence à ma liste de publications pages 62-64. Les autres citations renvoient aux références fournies pages 56-61.

2 Monoïdes et morphismes

Le but de cette partie préliminaire est d'effectuer quelques rappels sur les notions de monoïdes et morphismes beaucoup manipulées au cours de ce mémoire.

Un *monoïde* $(M, \cdot)$ est un ensemble M muni d'une opération interne $\cdot$ associative possédant un élément neutre. Nous omettrons la précision de l'opération interne en l'absence d'ambiguïté. Nous rencontrerons au fil de ce mémoire de nombreux monoïdes (monoïdes de traces, monoïdes d'endomorphismes particuliers de monoïdes libres). Nous noterons par juxtaposition les lois internes (sauf pour le monoïde $(\mathbb{N}, +)$ des entiers naturels).

Le monoïde le plus sous-jacent à tous nos travaux est bien entendu A^* l'ensemble des mots sur un alphabet (ici) *fini* A (muni de la concaténation comme loi interne et du mot vide ε comme élément neutre), un *mot* étant une suite de lettres (en l'absence de précision la suite et donc le mot seront supposés finis). Ce monoïde est un exemple de *monoïde finiment engendré* : tous les éléments se décomposent sur l'alphabet qui joue alors le rôle de *générateur*. Il est surtout un *monoïde libre* : si $n \geq 0$ et $p \geq 0$ sont des entiers, si $x_1, \ldots, x_n, y_1, \ldots, y_p$ sont des éléments du générateur A tels que $x_1 \ldots x_n = y_1 \ldots y_p$ alors $n = p$ et pour tout entier i, $1 \leq i \leq n$, $x_i = y_i$. Un autre exemple de monoïde est étudié dans la partie suivante.

Les *puissances* d'un élément x d'un monoïde M d'élément neutre ε sont les éléments x^k de M définis par $x^0 = \varepsilon$, et $x^k = x^{k-1}x$ pour $k \geq 1$ entier. En particulier dans le monoïde libre, un *carré* est un mot de la forme x^2, un *cube* un mot de la forme x^3 pour $x \neq \varepsilon$. Il sera beaucoup question de *mots sans puissance* k dans la partie 4, c'est-à-dire de mots n'ayant pas de facteur de la forme x^k (un mot u est respectivement *facteur*, *préfixe*, *suffixe* d'un mot v si respectivement $v = pus$, $v = us$, $v = pu$ pour des mots p et s).

Un *morphisme de monoïdes* est une application $f : M_1 \to M_2$ entre deux monoïdes M_1 et M_2 qui préserve la structure de monoïde de M_1. En d'autres termes, si e_1 et e_2 sont respectivement les éléments neutres de M_1 et M_2 : pour tous x, y dans M_1, $f(xy) = f(x)f(y)$ et $f(e_1) = e_2$. Nous dirons aussi que f_1 est un morphisme sur M_1 (ou sur l'alphabet A si $M_1 = A^*$) si la donnée de M_2 n'est pas importante. En particulier, quand $M_1 = M_2$, f est appelé *endomorphisme* sur M_1.

Étant donné un alphabet A, deux morphismes de A^* dans $(\mathbb{N}, +)$ sont couramment utilisés en combinatoire des mots. Il s'agit de l'application qui à tout mot u associe sa *longueur* $|u|$, et celle qui, pour une lettre a de A, associe à tout mot u le *nombre d'occurrences* $|u|_a$ de la lettre a dans u.

Les morphismes sur lesquels portent nos travaux sont les morphismes entre monoïdes libres. Ils sont souvent appelés morphismes de monoïdes libres bien que cela n'implique pas qu'ils soient injectifs, i.e., qu'ils préservent le caractère libre du monoïde. En particulier, nous rencontrerons à plusieurs reprises, des morphismes de monoïdes libres, appelés *périodiques*, qui sont les morphismes $f : A^* \to B^*$ pour lesquels il existe un mot z sur B tel que $f(a)$ est une puissance de z pour toute lettre a de A. Un (endo)morphisme *binaire* sera un (endo)morphisme de monoïdes libres défini sur un alphabet à deux lettres.

Un morphisme est entièrement défini par les valeurs des images des éléments d'un générateur. Cela est particulièrement agréable lorsque comme dans le cas du monoïde libre, nous travaillons avec un générateur fini. Pour un morphisme f entre deux monoïdes libres A^* et B^*, nous noterons $||f||$ la somme des longueurs des images des lettres de A. Cette valeur est utile pour mesurer la taille d'un morphisme en particulier pour analyser la complexité d'un algorithme cherchant une propriété de f.

Notons que l'ensemble des endomorphismes sur un monoïde M forme un monoïde pour la composition des applications. Les puissances d'un endomorphisme sont ainsi définissables. Elles peuvent servir dans le monoïde libre à engendrer des mots infinis. Un endomorphisme f sur un alphabet A est prolongeable en une lettre a de A si $f(a) = ax$ pour un mot non vide x. Pour un tel morphisme, $f^n(a)$ est un préfixe de $f^{n+1}(a)$. Quand pour tout entier n, la longueur de $f^n(a)$ est strictement inférieure à celle de $f^{n+1}(a)$, le mot infini, noté $f^\omega(a)$ ayant tous les $f^n(a)$ comme préfixes est appelé *mot engendré* par f (en a). Par exemple, le morphisme φ appelé *morphisme de Fibonacci* et défini par

$$\begin{cases} \varphi(a) = ab, \\ \varphi(b) = a, \end{cases}$$

engendre le mot infini $\varphi^\omega(a)$, appelé *mot de Fibonacci*, commençant par

$$\varphi^7(a) = abaababaabaababaababaabaababaabaab.$$

Des exemples d'endomorphismes qui n'engendrent pas de mot infini sont les *permutations*, i.e., les endomorphismes f sur A tels que $A = \{f(a) \mid a \in A\}$. Les permutations sur A forment un monoïde engendré par les *morphismes d'échange*, i.e., les permutations qui laissent invariantes toutes les lettres sauf deux.

Tout au long du mémoire, nous utiliserons les opérations rationnelles pour décrire des ensembles (de mots, de morphismes ...). Soit M un monoïde.

La première opération rationnelle est l'union ensembliste. La deuxième est l'extension de l'opération interne du monoïde : pour deux parties X et Y de M, XY est l'ensemble des xy pour x et y éléments respectivement de X et Y. L'ensemble des parties de M muni de cette extension est un monoïde d'élément neutre $\{\varepsilon\}$. En utilisant les puissances d'un ensemble, la troisième opération rationnelle est l'*itération finie* d'une partie X noté X^* et définie par une union infinie :

$$X^* = \bigcup_{i=0}^{\infty} X^i.$$

Nous utiliserons également l'itération restreinte

$$X^+ = \bigcup_{i=1}^{\infty} X^i.$$

Une *expression rationnelle* exprime comment un ensemble peut être construit à partir des générateurs du monoïde. Les ensembles constructibles de cette sorte sont appelés *rationnels*.

Dans le cas du monoïde libre, un ensemble de mots est appelé un *langage*. Le théorème de Kleene établit alors qu'un ensemble est rationnel si et seulement s'il est reconnaissable. Les ensembles *reconnaissables* peuvent être définis de plusieurs manières équivalentes. La première classique est celle des ensembles reconnaissables par automate (avec son alphabet, ses ensembles d'états, d'états initiaux, d'états terminaux et son ensemble de transitions) dont nous ne rappellerons pas ici la définition. Une autre définition équivalente de reconnaissance est celle par morphisme. Cela peut être vu comme une vision morphique des automates. Une partie X d'un monoïde M est reconnue par un triplet $[Q, h, F]$ si h est un morphisme de M dans le monoïde fini Q dont F est une partie, et si $X = h^{-1}(F)$ c'est-à-dire que X est l'ensemble des éléments x de M tels que $h(x) \in F$.

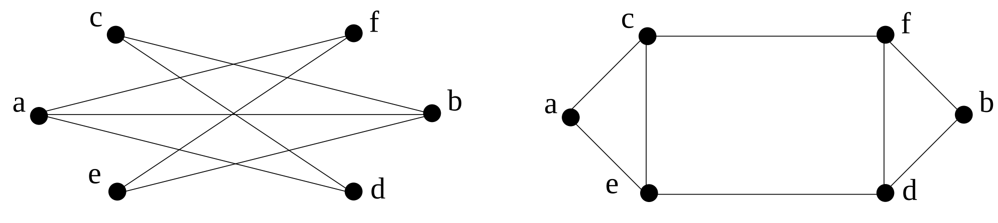
3 Traces

Le but de cette partie est essentiellement de présenter mes travaux sur les traces réalisés d'une part avec Y. Métivier et P.-A. Wacrenier (cf partie 3.5) et d'autre part avec D. Kirsten (cf partie 3.4). Ces travaux prolongeant et utilisant ceux menés pour mon doctorat, je suis amené à présenter les principaux résultats obtenus à ce moment-là.

3.1 Monoïdes de traces

Les monoïdes libres partiellement commutatifs furent introduits en 1969 à des fins combinatoires [CF69]. En 1977, A. Mazurkiewicz [Maz77] a montré la possibilité de modéliser certains aspects des systèmes distribués en utilisant de tels monoïdes. La modélisation peut s'effectuer simplement en considérant un alphabet fini A représentant des types d'actions, et une relation irréflexive et symétrique I sur cet alphabet (alors appelée *relation d'indépendance* ou de *commutation*) précisant les actions pouvant s'effectuer en parallèle. Le *monoïde libre partiellement commutatif* (aussi appelé *monoïde de traces* ou *monoïde de commutation*) sur A et I, noté $M(A, I)$, est le quotient du monoïde libre A^* par la clôture transitive, notée $\sim_I$, de la relation $\{(uabv, ubav) \mid (a, b) \in I, u, v \in A^*\}$.

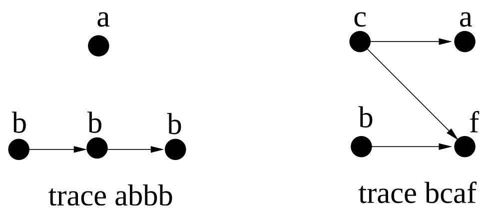
Par exemple, considérons un système constitué de deux ordinateurs X et Y pouvant communiquer par messages sur un canal bidirectionnel. Notons a (resp. b) une action interne à X (resp. à Y), c (resp. d) un envoi de message de X vers Y (resp. de Y vers X) et e (resp. f) une réception sur Y d'un message envoyé par X (resp. sur X envoyé par Y). Il est naturel de considérer que les actions indépendantes sont celles qui ne se déroulent pas sur la même machine et qui ne correspondent pas l'une à l'envoi, l'autre à la réception d'un message d'un ordinateur vers l'autre. Les actions indépendantes peuvent être représentées par le graphe (non orienté) (A, I) de la relation d'indépendance ci-après (a).



a) graphe (A,I) b) graphe (A,D)

Il est parfois plus confortable de visualiser le graphe de la relation complémentaire à I, notée D, et appelée relation de dépendance : voir figure ci-avant (b). Sur ce graphe, il est d'usage de ne pas faire apparaître la réflexivité de D.

Un élément d'un monoïde de traces $M(A, I)$, appelé *trace*, représente un ensemble d'exécutions équivalentes dans le système modélisé. Il y a plusieurs manières de voir une trace. Elle peut être identifiée à une classe d'équivalence de $\sim_I$. Tous les mots de cette classe d'équivalence ont même longueur et même nombre d'occurrences de chaque lettre. Cela permet de parler de longueur d'une trace ou de lettres apparaissant dans une trace. Du point de vue notation, une trace est souvent représentée par un mot quelconque de la classe (un mot partiellement commutatif). Elle peut également être représentée par son graphe de précédence qui exprime les actions qui doivent nécessairement être réalisées avant les autres. À titre d'exemple, si (A, I) est la relation d'indépendance définie précédemment, la trace $abbb$ (classe d'équivalence $\{abbb, babb, bbab, bbba\}$) et la trace $bcaf$ (classe d'équivalence $\{bcaf, bcfa, cabf, cbaf, cbfa\}$) sont représentées par les graphes de précédence suivants :

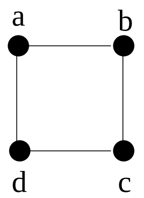


La relation $\sim_I$ étant une congruence, la loi interne du monoïde $M(A, I)$, appelée *concaténation* comme dans le monoïde libre, étend naturellement la concaténation des mots : si t_1 et t_2 sont deux traces, si w_1 et w_2 sont respectivement des mots représentant ces traces, alors le mot w_1w_2 représente la trace t_1t_2. L'élément neutre d'un tel monoïde est la trace vide correspondant à la classe d'équivalence du mot vide.

Une trace t est dite *connexe* quand son graphe de précédence est connexe, ou de manière équivalente, quand la restriction aux lettres apparaissant dans t du graphe de la relation de dépendance est un graphe connexe. Ainsi une trace n'est pas connexe quand elle peut être exprimée comme la concaténation de deux traces t_1 et t_2 (différentes de la trace vide) telles que toutes les lettres apparaissant dans t_1 sont indépendantes de toutes les lettres de t_2. Dans l'exemple précédent, la trace $bcaf$ est connexe, alors que la trace $abbb$ ne l'est pas. Pour un ensemble X, nous noterons $\mathrm{Conn}(X)$ et $\mathrm{NConn}(X)$ respectivement les ensembles de traces connexes et de traces non connexes de X.

Dans ce qui suit, nous considérons plusieurs monoïdes de traces particuliers. Quand I est la relation vide, $M(A, I)$ est le monoïde libre A^*. Le *monoïde libre totalement commutatif* correspond au cas où deux lettres quelconques sont indépendantes ($I = A \times A \setminus \{(a,a) \mid a \in A\}$).

Un *C4* dans un graphe est un sous-graphe de quatre sommets a, b, c, d dont les arêtes $(a,b), (b,c), (c,d), (d,a)$ forment un cycle sans corde.



un C4

Quand le graphe (A, I) de la relation d'indépendance d'un monoïde de traces $M(A, I)$ ne contient pas de C4, le monoïde est simplement dit *sans C4*. Un monoïde libre, un monoïde libre totalement commutatif, un monoïde dont la relation d'indépendance est transitive sont des exemples de monoïdes sans C_4. C'est également le cas des monoïdes notés $A^* \times b^*$ qui correspondent au cas où toutes les lettres de l'alphabet A sont deux à deux dépendantes et la lettre b est indépendante de toutes les lettres de A. Plus généralement étant donnés deux monoïdes de traces $M(A, I)$ et $M(B, J)$, nous noterons $M(A, I) \times M(B, J)$ le monoïde de traces $M(A \cup B, I \cup J \cup A \times B \cup B \times A)$. Un monoïde sans C4 est un monoïde dont la restriction à n'importe quel sous-alphabet $\{a, b, c, d\}$ de quatre lettres n'est pas de la forme $\{a, c\}^* \times \{b, d\}^*$. Le monoïde de traces développé en exemple page 7 n'est pas un monoïde de traces sans C_4.

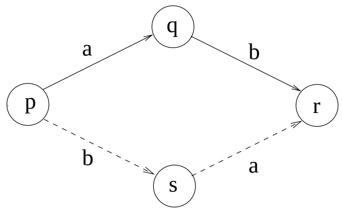
Les propriétés (combinatoires, logiques, utilisation en vérification ...) du modèle "monoïde de traces" ont été beaucoup étudiées (voir par exemple [DR95, DM97]). De nombreux travaux concernent la théorie des langages, et en particulier les langages rationnels et reconnaissables. Le problème de l'étoile s'inscrit dans ce cadre.

3.2 Le problème de l'étoile

Dans le cas d'un monoïde de traces, un ensemble X de traces est *reconnaissable* si et seulement si l'ensemble, que nous noterons[2] $[X]_I^{-1}$, des mots représentant les traces de X est un langage reconnaissable (donc reconnaissable par un automate au sens usuel du terme). L'automate minimal déterministe reconnaissant $[X]_I^{-1}$ a la *propriété* dite *de diamant* : étant données deux lettres indépendantes a et b, pour tous états p, q, r tels que (p, a, q)

2. car il s'agit tout simplement de l'image inverse de X par le morphisme canonique qui associe à un mot sa classe d'équivalent par $\sim_I$

et (q, b, r) sont des transitions de l'automate, il existe un état s tel que (p, b, s) et (s, a, r) sont des transitions de l'automate (voir figure ci-après). Réciproquement tout automate ayant la propriété de diamant peut être considéré comme un automate reconnaissant un ensemble de traces.



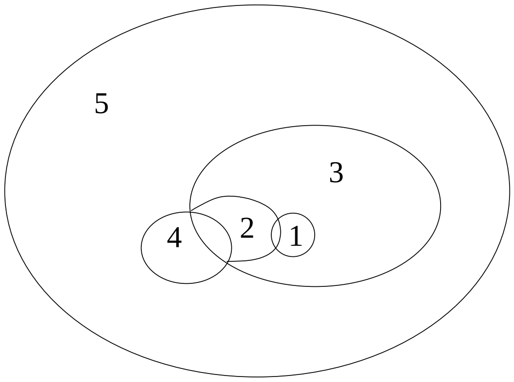
Pour un ensemble reconnaissable de traces X, toute expression rationnelle du langage $[X]_I^{-1}$ est aussi une expression rationnelle pour X. Ainsi tout ensemble reconnaissable de trace est rationnel. La réciproque n'est pas vraie. En d'autres termes, le théorème de Kleene ne s'étend pas au cas des monoïdes de traces. Sur les trois opérations rationnelles, une seule pose problème. Avec ce qui a été dit précédemment, il est vérifiable que l'union de deux ensembles reconnaissables de traces est encore un ensemble reconnaissable de traces. Même s'il existe de nombreuses preuves de natures différentes [Fli74, CP85, GRS90, GRS00, Has91], il est déjà plus difficile d'établir que la concaténation de deux ensembles reconnaissables de traces est reconnaissable. Le problème avec l'itération finie intervient sur un exemple très simple. Si a et b sont deux lettres indépendantes alors l'ensemble de traces $(ab)^*$ est rationnel mais non reconnaissable : dans ce cas, $[(ab)^*]_I^{-1} = \{w \in \{a, b\}^* \mid |w|_a = |w|_b\}$.

En 1985, E. Ochmański [Och85] a montré que les ensembles reconnaissables de traces sont exactement les ensembles qui peuvent être obtenus à partir des ensembles finis en utilisant un nombre fini de fois les opérations de concaténation, d'union finie et d'itération finie d'un ensemble de traces connexes (voir aussi [CL87, Mét86]). Notons en particulier que si X est un ensemble reconnaissable de traces connexes, alors X^* est aussi un ensemble reconnaissable.

Un ensemble rationnel pouvant être décrit par plusieurs expressions rationnelles, la caractérisation précédente d'E. Ochmański ne permet pas de décider la question suivante appelée *problème de l'étoile* : étant donné un ensemble reconnaissable X de traces, son itération finie X^* est-elle aussi reconnaissable ? Vu le théorème de Kleene dans le monoïde libre, ce problème n'apparaît que quand au moins deux lettres sont indépendantes. Depuis 1966 [GS66], sa décidabilité est connue dans les monoïdes totalement commutatifs. En 1992, J. Sakarovitch [Sak92] a montré qu'on pouvait décider la question suivante, appelée *problème de la reconnaissabilité*, si et seulement si la relation d'indépendance est transitive : étant donné un ensemble rationnel, est-il reconnaissable ? En conséquence, le problème de l'étoile est décidable dans un tel cas : si X est un ensemble reconnaissable, X, et par conséquent X^*, sont rationnels.

Quand la relation d'indépendance n'est pas transitive, il existe des cas pour lesquels le problème de l'étoile est décidable. En 1992, P. Gastin, E. Ochmański, A. Petit et B. Rozoy ont montré [GOPR92] sa décidabilité dans un monoïde de la forme $A^* \times b^*$. En généralisant leur approche (voir partie suivante), j'ai pu montrer [18] (voir aussi [5]) que *le problème de l'étoile est décidable dans tout monoïde de traces sans C4.*

Notons que la famille des monoïdes de traces sans C4 contient tous les cas de décidabilité du problème de l'étoile précédemment connus (voir schéma suivant) et qu'à ce jour aucun autre nouveau cas de décidabilité (ou d'indécidabilité) n'a été établi.



Cas connus de décidabilité du problème de l'étoile

1. monoïdes libres (y compris a^* pour a lettre)
2. monoïdes totalement commutatifs libres (y compris a^* et $a^* \times b^*$ pour a, b lettres distinctes).
3. cas de relation d'indépendance transitive
4. monoïdes de la forme $A^* \times b^*$
5. monoïdes sans C4

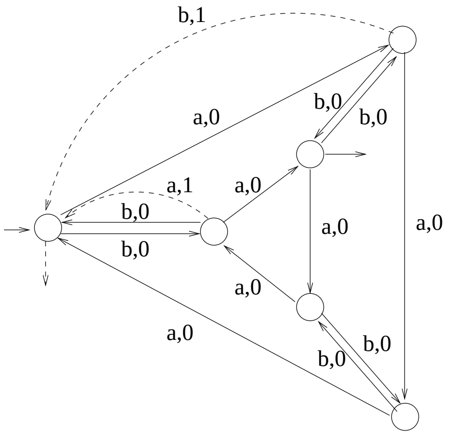
3.3 Le problème de la puissance finie

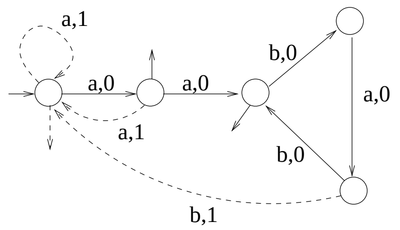
Nous présentons à présent un problème dont la prise en compte a permis des avancées significatives sur le problème de l'étoile.

Une partie d'un monoïde X possède la *propriété de puissance finie* si son itération finie est égale à l'union d'un nombre fini de ses puissances :

$X^* = \bigcup_{i=0}^{n} X^i$ pour un entier n. Le *problème de la puissance finie* est la question : étant donné un ensemble reconnaissable X, est-ce que X possède la propriété de puissance finie ? La décidabilité de ce problème a été posée dans le monoïde libre en 1967 [Brz67]. Elle a été prouvée indépendamment par I. Simon en 1978 [Sim78] et K. Hashigushi en 1979 [Has79] (voir aussi la preuve récente de D. Kirsten [Kir02a]).

K. Hashigushi a généralisé sa technique en fournissant [Has82] un moyen pour tester si un automate à coûts est limité en distance. Un *automate à coûts* est un sextuplet (A, Q, I, T, Δ, d) où (A, Q, I, T, Δ) est un automate et d est une fonction qui associe à chaque transition un coût 0 ou 1. Le coût d'un chemin dans l'automate est la somme des coûts des transitions composant le chemin. Le coût d'un mot reconnu par l'automate est le plus petit coût parmi ceux des chemins reconnaissant le mot. Un automate à coûts est *limité en distance* s'il existe un entier k tel que les coûts des mots reconnus par l'automate sont tous inférieurs à k. Le premier automate ci-dessous est 6-limité en distance, alors que le deuxième n'est pas limité en distance (voir au paragraphe suivant comment ils ont été construits).





À partir d'un automate déterministe reconnaissant un ensemble X, il est possible d'obtenir un automate reconnaissant X^* en considérant l'état initial e_0 comme terminal (si ce n'est pas déjà le cas – attention dans certains cas, il faut standardiser l'automate) et en ajoutant des transitions (p, a, e_0) à chaque fois qu'il existe une transition (p, a, q) avec q un état terminal. En associant un coût de 1 à chacune de ces transitions et un coût de 0 à chaque transition initiale, on obtient un automate à coût qui est limité en distance si et seulement si X possède la propriété de puissance finie. Les deux exemples précédents ont été obtenus respectivement à partir d'automates reconnaissant $X_1 = \{u \in \{a,b\}^* \mid |u|_a = 1 \text{ mod } 3 \text{ et } |u|_b = 1 \text{ mod } 2\}$ et $X_2 = a \cup aa(bab)^*$. Les éléments rajoutés à l'automate initial afin de reconnaître X_1^* et X_2^* sont en pointillés. En corollaire de ce qui a déjà été dit, X_1 possède la propriété de puissance finie alors que X_2 ne la possède pas.

La démarche précédente se généralise à des ensembles de traces connexes. À partir d'un automate reconnaissant un ensemble X de traces connexes, Y. Métivier a fourni [Mét86], une construction explicite d'un automate reconnaissant X^*. Il est possible d'adapter cette construction pour produire un automate à coûts qui est limité en distance si et seulement si X possède la propriété de la puissance finie. Avec Y. Métivier, cela nous a permis de montrer [8,19] que *le problème de la puissance finie est décidable pour tout ensemble de traces connexes.*

E. Ochmański [Och90] fut le premier à montrer un lien entre le problème de l'étoile et celui de la puissance finie en donnant un cas particulier d'ensembles reconnaissables de traces X et Y tels que X^* est reconnaissable si et seulement si Y possède la propriété de puissance finie. C'est en généralisant cette approche et en utilisant la décidabilité du problème de la puissance finie dans un monoïde libre, que P. Gastin, E. Ochmański, A. Petit et B. Rozoy ont pu montrer la décidabilité du problème de l'étoile dans $A^* \times b^*$. Une conséquence de leurs preuves est que pour une partie reconnaissable de $A^* \times b^*$, X^* est reconnaissable si et seulement si l'ensemble $\mathrm{Conn}(X)^* \cup \mathrm{NConn}(X)$ possède la propriété de puissance finie.

En considérant à la place du monoïde libre A^* un monoïde de traces $M(A, I)$ dans lequel le problème de l'étoile et le problème de la puissance finie sont décidables, une relecture attentive de [GOPR92] permet d'établir la décidabilité du problème de l'étoile dans le monoïde de traces $M(A, I) \times \{b\}^*$. Plus précisément, si pour E ensemble de lettres, $X_{\subseteq E}$ désigne l'ensemble des traces de X ne contenant que des lettres de E alors, pour toute partie reconnaissable X de $A^* \times b^*$, X^* est reconnaissable si et seulement si les ensembles $(X_{\subseteq A})^*$ et $(X_{\subseteq \{b\}})^*$ sont reconnaissables, et si l'ensemble $\mathrm{Conn}(X)^* \cup \mathrm{NConn}(X)$ possède la propriété de puissance finie. Ce dernier résultat technique peut alors se dériver en : une partie reconnaissable X de $A^* \times b^*$ possède la propriété de puissance finie si et seulement si X^* est reconnaissable, et $X_{\subseteq A}$, $X_{\subseteq \{b\}}$ possèdent la propriété de puissance finie. Ainsi,

Théorème 3.1 [18] (voir aussi [5]) *Étant donné un monoïde de traces $M(A, I)$ pour lequel les problèmes de l'étoile et de la puissance finie sont décidables, étant donnée une lettre b n'appartenant pas à A, les problèmes de l'étoile et de la puissance finie sont décidables dans $M(A, I) \times \{b\}^*$.*

Ce résultat est à la base d'une récurrence (sur le cardinal de l'alphabet de définition des monoïdes de traces considérés), permettant de montrer *si-*

multanément la décidabilité du problème de l'étoile et celle de la puissance finie dans tout monoïde de traces sans C4 [18] (voir aussi [5]).

La récurrence est rendue possible par le fait que, pour un monoïde $M(A, I)$ sans C4, soit le graphe de la relation de dépendance est connexe, soit $M(A, I)$ est de la forme $M(B, J) \times c^*$. Le théorème 3.1 permet de traiter le deuxième cas. Dans le premier cas, on montre que pour une partie X reconnaissable, X^* est reconnaissable si et seulement si $X^*_{\subseteq E}$ est reconnaissable pour tout sous-ensemble strict E de A : la reconnaissabilité de X^* est décidable en utilisant l'hypothèse de récurrence. Toujours dans ce cas, il est possible de construire un ensemble reconnaissable Z de traces connexes tel que X possède la propriété de puissance finie si et seulement si Z et chaque $X_{\subseteq E}$ (pour $E \subset A$, $E \neq A$) possède la propriété de puissance finie : cela est décidable en partie par hypothèse de récurrence et en partie car Z ne possède que des traces connexes.

Il est important de noter qu'à chaque étape de la récurrence, la décidabilité du problème de l'étoile puis celle de la puissance finie sont établies.

3.4 Équivalence de décidabilité

La preuve présentée dans la partie précédente fait sentir une interaction très forte entre les problèmes de l'étoile et de la puissance finie. Dès lors le résultat suivant n'apparaît pas surprenant :

Théorème 3.2 [5] *Dans un monoïde de traces, le problème de l'étoile est décidable si et seulement si le problème de la puissance finie est décidable.*

Un début de preuve de ce résultat était amorcé dans ma thèse. Utilisant une vision morphique des automates reconnaissant un ensemble de traces, après avoir établi une classification des semigroupes finis, D. Kirsten [Kir99] a montré : étant donnés deux monoïdes de traces $M(A, I)$ et $M(B, J)$ pour lesquels les problèmes de l'étoile et de la puissance finie sont décidables, le problème de l'étoile est décidable dans $M(A, I) \times M(B, J)$ si et seulement si le problème de la puissance finie est décidable dans $M(A, I) \times M(B, J)$. Le théorème 3.2 s'établit en utilisant ce résultat et la technique de récurrence décrite dans la partie précédente.

Sous-jacent à la récurrence apparaît le résultat suivant qui montre un autre lien entre les problèmes de l'étoile et de la puissance finie. Il fournit une information sur la structure des ensembles reconnaissables et généralise le résultat sous-jacent aux preuves de [GOPR92].

Proposition 3.3 [5] *L'étoile d'un ensemble reconnaissable de traces X est reconnaissable si et seulement si l'ensemble constitué des traces connexes de X^* et des traces non connexes de X possède la propriété de puissance finie.*

Notons que c'est bien $\mathrm{Conn}(X^*) \cup \mathrm{NConn}(X)$ qui doit avoir la propriété de puissance finie et non $[\mathrm{Conn}(X)]^* \cup \mathrm{NConn}(X)$. En effet, soient $A = \{a, b, c\}$, $I = \{(a, c), (c, a)\}$ et le sous-ensemble $X = \{aa, cc, b, ac\}$ de $M(A, I)$. L'ensemble X^* est reconnaissable : $[X^*]^{-1} = \{aa, ac, ca, cc, b\}^*$. On a $\mathrm{Conn}(X^*) = (aa)^* \cup (cc)^* \cup (bX^*)^*$ et $\mathrm{NConn}(X) = \{ac\}$. L'ensemble $\mathrm{Conn}(X^*) \cup \mathrm{NConn}(X)$ possède la propriété de puissance finie. En effet, $(\mathrm{Conn}(X^*) \cup \mathrm{NConn}(X))^* = (\mathrm{Conn}(X^*) \cup \mathrm{NConn}(X))^3 = \{ac, \varepsilon\}\{aa, cc\}^* \cup (bX^*)^*$. Par contre l'ensemble $[\mathrm{Conn}(X)]^* \cup \mathrm{NConn}(X) = \{aa, cc, b\}^* \cup \{ac\}$ ne possède pas la propriété de puissance finie.

3.5 Autres approches

Nous présentons à présent deux autres approches utilisées pour essayer de résoudre le problème de l'étoile.

Au début de ma thèse, je me suis intéressé au problème de l'étoile pour des ensembles finis (à noter qu'aucun ensemble fini ne possède la propriété de puissance finie). La décidabilité du problème avait déjà été obtenue pour des singletons [CM85, Has91] : pour t trace, l'ensemble $\{t\}^*$ est reconnaissable si et seulement si t est connexe. Avec Y. Métivier [8,19], nous avons montré *la décidabilité du problème de l'étoile pour des ensembles de deux, trois ou quatre traces ainsi que pour des ensembles finis de traces contenant au plus deux traces connexes.*

La démarche de preuve consistait à essayer d'établir la forme des ensembles finis qui ont une étoile reconnaissable. Pour deux et trois traces, ces formes sont assez aisées à décrire :

- pour t_1 et t_2 traces, $\{t_1, t_2\}^*$ est reconnaissable si et seulement si t_1 et t_2 sont reconnaissables.
- l'itération finie d'un ensemble de trois traces est reconnaissable si et seulement si soit les trois traces sont connexes, soit l'ensemble est de la forme $\{u^p, v^q, u^r v^s\}$ avec u et v deux traces connexes et p, q, r, s des entiers strictement positifs.

Cinq cas sont possibles pour décrire les ensembles de quatre traces ayant une itération reconnaissable. Le dernier cas fait apparaître la nécessité de tester la propriété de puissance finie.

Avec D. Kirsten [5] , nous avons continué une approche similaire pour des ensembles ayant un nombre fini de traces non connexes. Généralisant un résultat d'E. Ochmański [Och90] (obtenu dans $A^* \times B^*$), nous avons montré *la décidabilité du problème de l'étoile pour des ensembles ne contenant qu'une seule trace non connexe.*

Une autre approche a été l'étude [17] d'une procédure S introduite par Y. Métivier et B. Rozoy [MR91]. Partant d'un automate reconnaissant un ensemble de mots X, $S(X)$ calcule un automate reconnaissant X plus des mots équivalents (par $\sim_I$) à des mots de X. L'idée de base de $S(X)$ est de rajouter localement des transitions à X de manière à essayer de se rapprocher d'un automate ayant la propriété de diamant. Dans certaines conditions, en itérant un nombre fini de fois la procédure S, on obtient la clôture de X par $\sim_I$, à savoir l'ensemble $\overline{X}$ des mots équivalents à des mots de X. En d'autres termes, dans ces cas, partant d'un ensemble rationnel de trace, on prouve sa reconnaissabilité. Toutefois, même quand l'ensemble rationnel est reconnaissable, l'itération de S un nombre fini de fois ne permet pas de le prouver. Cette étude n'apporte pas d'information sur le problème de l'étoile, mais elle fournit de nouvelles preuves pour plusieurs résultats sur les ensembles reconnaissables de traces comme par exemple :

- l'étoile d'un ensemble reconnaissable de traces connexes est reconnaissable.
- la concaténation de deux ensembles reconnaissables de traces est reconnaissable.

Pour montrer que la concaténation de deux ensembles reconnaissables de traces est un ensemble reconnaissable, K. Hashigushi [Has91] utilise une notion de rang qui mesure une "distance" entre un ensemble de mots X et sa clôture par $\overline{X}$. Si un ensemble a un rang fini alors il est reconnaissable, mais la réciproque n'est pas vraie. Avec Y. Métivier et P.-A. Wacrenier [17], nous avons comparé la puissance de la procédure S à celle de la notion de rang en montrant :

- pour un ensemble X, $S(X)$ est de rang trois fois inférieur à X ;
- si un ensemble a un rang infini, alors il n'est pas possible de calculer sa clôture en utilisant S ;
- la réciproque est vraie uniquement dans les monoïdes de la forme $A^* \times B^*$.

3.6 Perspectives

Le problème de l'étoile est toujours un problème ouvert. Dès qu'on considère un monoïde qui contient un C4, on ne sait pas si le problème est décidable ou indécidable. D. Kirsten [Kir02b] a montré que le statut du problème ne dépendait que de son statut pour une famille de monoïdes de traces K_n ($n \geq 2$). Cette famille peut être définie de la façon suivante. L'alphabet est constitué de $2n$ lettres $a_1, \ldots, a_n$, $b_1, \ldots, b_n$. Pour chaque entier i, $1 \leq i \leq n$, a_i est dépendante de b_i. Toute autre paire de lettres différentes est une paire de lettres indépendantes. La résolution du problème dans K_2 (le monoïde ayant un $C4$ comme graphe de relation d'indépendance) amènera sans doute les techniques nécessaires pour conclure cette étude.

De nouvelles idées pour résoudre le problème de l'étoile pourraient peut-être être obtenues en poursuivant l'étude des ensembles finis. La proposition 3.3 donnant une expression synthétique de tous les cas pourrait peut-être permettre de repousser (au moins dans les preuves) l'explosion du nombre de cas. Notons que l'étude d'un ensemble de cinq traces se ramène (avec ce qui a été dit dans la partie précédente) à l'étude d'un ensemble de trois traces connexes et deux traces non connexes. Les techniques déjà développées pourraient permettre de décider le problème de l'étoile dans ce cas. Toutefois des difficultés nouvelles semblent apparaître lors de l'étude de la décidabilité pour un ensemble de six traces.

Dans le même ordre de démarche, se pose la question de la décidabilité du problème de l'étoile pour des ensembles finis ou infinis ayant deux traces non connexes.

Un résultat dans ma thèse [30] affirme que l'étoile d'un ensemble T de traces est reconnaissable si et seulement s'il existe un entier n tel $S^n([T]_I^*) = S^{n+1}([T]_I^*)$. Décider l'existence d'un tel n (éventuellement dans un cas plus général que $[T]_I^*$) pourrait ainsi permettre de décider le problème de l'étoile.

La collaboration avec D. Kirsten a permis d'établir des réponses négatives à une conjecture d'E. Ochmański [Och90] et une de M. Latteux (conjecture posée suite à un exposé que j'avais fait au LIFL en 1995). Cette dernière posait, pour un ensemble reconnaissable de traces X, la question de l'existence d'un entier n tel que si X possède la propriété de puissance finie alors $[(X \cup \{\varepsilon\})^n]^{-1}$ possède la propriété de puissance finie. La conjecture d'Ochmański posait la question pour un ensemble fini X de traces, de l'existence d'une trace t tel que $(X \setminus \{t\})^*$ est aussi reconnaissable. Une deuxième conjecture d'E. Ochmański [Och90] est vraie [5] pour les monoïdes de traces sans C4 mais reste ouverte dans les autres cas :

> *Conjecture.* Si X est un ensemble de traces contenant au moins deux traces et si X^* n'est pas reconnaissable alors il existe une trace t dans X telle que $(X \setminus t)^*$ n'est pas reconnaissable.

D'autres problèmes dans les monoïdes de traces existent. En particulier, malgré de récents progrès, certains problèmes concernant des morphismes entre monoïdes de traces semblent rester tout aussi mystérieux que le problème de l'étoile. Par exemple, décider l'existence d'un morphisme injectif entre deux monoïdes de traces a été montré indécidable par M. Kunc [Kun01, Kun04]. Décider si un morphisme entre deux monoïdes de traces est injectif est aussi indécidable [CR87]. Les frontières de décidabilité de ces problèmes sont encore inconnues et semblent moins régulières que celle du problème de l'étoile [Kun04, Mat97].

4 Morphismes sans répétition

Au début du vingtième siècle, A. Thue fut le premier à s'intéresser à des mots ne contenant pas certaines répétitions [Thu06, Thu12] (voir aussi [Ber92, Ber95]). Il montra en particulier l'existence d'un mot infini *sans chevauchement* sur l'alphabet binaire $\{a, b\}$, c'est-à-dire un mot infini ne contenant pas de facteur de la forme $\alpha u \alpha u \alpha$ où α est une lettre et u un mot. Ce mot, actuellement appelé *mot de Thue-Morse* (car il apparaît aussi dans les travaux de M. Morse [Mor21]), est engendré par le morphisme μ défini par

$$\begin{cases} \mu(a) = ab, \\ \mu(b) = ba. \end{cases}$$

Plus précisément, le mot de Thue-Morse est le mot :

$$\lim_{n\to\infty} \mu^n(a) = abbabaabbaababbabaab\ \ldots$$

Une manière de montrer que le mot de Thue-Morse est sans chevauchement consiste à prouver que le morphisme μ est un *morphisme sans chevauchement* c'est-à-dire un morphisme qui préserve les mots sans chevauchement (l'image de tout mot sans chevauchement est un mot sans chevauchement). Notons que le morphisme de Thue-Morse est un morphisme *uniforme*, i.e., un morphisme pour lequel les images des lettres ont toutes la même longueur, ou de manière équivalente un morphisme qui préserve l'égalité des longueurs (pour tous mots u et v de même longueur, $f(u)$ et $f(v)$ ont même longueur).

Rappelons que, pour k un entier, *un mot sans puissance* k est un mot qui ne contient pas de facteur de la forme u^k pour un mot u non vide. Par exemple, le mot de Thue-Morse est sans cube (sans puissance 3). Sur l'alphabet $\{a, b\}$ seuls les facteurs de aba et bab sont sans carré (sans puissance 2), et il n'existe donc pas de mot infini sans carré. Par contre, dans ses travaux, A. Thue [Thu12] a montré l'existence d'un mot infini sur trois lettres sans carré (et donc sans puissance k pour tout entier $k \geq 2$). Ce mot M sur l'alphabet $\{a, b, c\}$ est tel que $T = f(M)$ avec T le mot de Thue-Morse et f le morphisme défini par

$$\begin{cases} a \mapsto abb, \\ b \mapsto ab, \\ c \mapsto a. \end{cases}$$

Il est aussi engendré [Ist77] par le morphisme défini par

$$\begin{cases} a \mapsto abc, \\ b \mapsto ac, \\ c \mapsto b. \end{cases}$$

Pour tout entier $k \geq 2$, ce morphisme n'est pas sans puissance k (un *morphisme sans puissance* k est un morphisme qui préserve les mots sans puissance k). Toutefois, pour $k \geq 2$ et pour A et B deux alphabets d'au moins deux lettres (trois si $k = 2$), il existe des morphismes sans puissance k de A^* dans B^* [BEM79, Bra83, Car83].

Le problème central de cette partie est de décider les questions suivantes : étant donné un morphisme (donné par les images des lettres), est-il sans chevauchement ? Pour un entier $k \geq 2$, est-il sans puissance k ? La dernière question a déjà été résolue de manière efficace par M. Crochemore [Cro82, Cro83] pour $k = 2$. Ainsi ici nous considérerons uniquement le cas $k \geq 3$ où déjà de nombreuses conditions suffisantes ou nécessaires pour décider sont connues [BEM79, Lec85, Ker86]. La thèse de M. Leconte [Lec85] est quasi-entièrement consacrée à l'étude du problème précédent. La classe particulière des morphismes presque circulaires (dont nous ne rappellerons pas la définition ici) y est mise en avant. Pour tout morphisme f de cette classe, on peut décider s'il est sans puissance k (tout simplement en testant l'image de tous les mots sans puissance k de longueur inférieure ou égales à $2k||f|| + 1$). De ce résultat se déduisent de nombreux cas de décidabilité comme par exemple le cas des morphismes uniformes, le cas des morphismes sans cube définis sur un alphabet de deux ou trois lettres, le cas des morphismes sur A quand $k > \max\{|f(x)| \mid x \in A\}$.

Nous nous sommes intéressés à deux autres approches basées sur des techniques efficaces [Thu12, BS93] dans le cas des endomorphismes binaires sans chevauchement pour lesquelles nous précisons des possibilités de généralisation.

Dans la partie 4.1, nous répondons à la question : est-ce qu'un morphisme sans chevauchement (resp. sans puissance k) peut se décomposer sur un nombre fini de morphismes sans chevauchement (resp. sans puissance k) ?

Dans les autres parties, nous cherchons à tester (et à savoir quand c'est possible) si un morphisme est sans chevauchement ou sans puissance k en examinant l'image d'un nombre fini de mots ne dépendant pas du morphisme mais uniquement (de k et) des cardinaux des alphabets de définition du morphisme. En cas d'existence de tels ensembles de test, nous cherchons dans plusieurs cas à les déterminer avec précision afin de pouvoir en connaître les plus petits possibles.

Notons que nous nous restreignons ici à des morphismes non effaçants puisque le seul morphisme effaçant qui est sans chevauchement ou sans puissance k ($k \geq 2$ entier) est le *morphisme vide* pour lequel l'image de tout mot est le mot vide.

4.1 Monoïdes d'endomorphismes

A. Thue [Thu12] ne s'arrêta pas à montrer que le morphisme μ est sans chevauchement. Il caractérisa l'ensemble des endomorphismes non effaçants sans chevauchement sur l'alphabet $\{a, b\}$: un endomorphisme non effaçant sur $\{a, b\}$ est sans chevauchement si et seulement s'il appartient à $\{E, \mu\}^*$ où E est le morphisme d'échange défini par $E(a) = b$ et $E(b) = a$. En d'autres termes, le monoïde des endomorphismes (non effaçant) sans chevauchement sur $\{a, b\}$ est finiment engendré par $\{E, \mu\}$.

Le résultat théorique précédent permet de tester si un endomorphisme f sur $\{a, b\}$ est sans chevauchement. Puisque de plus $\mu E = E\mu$, il suffit de vérifier que f est de la forme μ^k ou $\mu^k E$. Pour cela, on commence par tester si f est l'identité ou l'échange. Sinon on regarde si $f(a)$ et $f(b)$ se décompose sur $\{ab, ba\}$. Si ce n'est pas le cas f n'est pas sans chevauchement. Si c'est le cas, on peut définir le morphisme g tel que $f = \mu g$ ($g(a)$ s'obtient à partir de a en remplaçant dans la décomposition de $f(a)$ sur $\{ab, ba\}$ une occurrence de ab par une occurrence de a et une occurrence de ba par une occurrence de b ; idem pour $g(b)$). Alors f est sans chevauchement si et seulement si g est sans chevauchement. L'algorithme termine donc en travaillant récursivement. Cet algorithme agit avec une complexité en temps en $O(||f||)$.

Étant donnée une propriété P que peuvent avoir des mots (propriété d'être sans chevauchement, sans puissance k, ...), l'ensemble des endomorphismes qui préservent cette propriété muni de la composition des morphismes a une structure de monoïde avec comme élément neutre l'identité. Le constat algorithmique du paragraphe précédent amène à se poser systématiquement la question : est-ce que le monoïde des endomorphismes qui préservent une propriété donnée est finiment engendré ?

Pour les propriétés qui nous intéressent dans cette partie, le cas des endomorphismes binaires sans chevauchement apparaît être très particulier. D'une part, tous ces morphismes sont uniformes ce qui n'est pas le cas pour des morphismes définis sur des alphabets plus grands ou pour des morphismes sans puissance k. D'autre part, le fait que le monoïde est finiment engendré est un cas isolé. En effet :

Théorème 4.1 [2] *Soit A un alphabet contenant au moins trois lettres. Le monoïde des endomorphismes sans chevauchement sur A et le monoïde des endomorphismes* uniformes *sans chevauchement sur A ne sont pas finiment engendrés.*

Théorème 4.2 [2,4] *Soient A un alphabet contenant au moins trois lettres et $k \geq 3$ un entier. Le monoïde des endomorphismes sans puissance k sur A et le monoïde des endomorphismes* uniformes *sans puissance k sur A ne sont pas finiment engendrés.*

Ce dernier théorème est aussi vrai pour $k = 2$ et un alphabet d'au moins trois lettres [Cro83].

Une idée pour démontrer des théorèmes comme les précédents consiste à déterminer un ensemble infini de morphismes irréductibles éléments du monoïde M de morphismes considéré. Un morphisme est *irréductible* si pour tous morphismes h_1 et h_2, si $f = h_1 h_2$ alors h_1 ou h_2 est une permutation. Si G est un ensemble générateur de M, alors pour tout morphisme f irréductible de M, il existe des permutations p_1 et p_2 telles que $p_1 f p_2$ appartienne à G. Comme $||f|| = ||p_1 f p_2||$, l'existence d'un ensemble infini de morphismes de M irréductibles montre que tout ensemble générateur est infini.

Pour démontrer le théorème 4.1, les morphismes uniformes f_n ($n \geq 1$ entier) présentés ci-après peuvent être utilisés : ils sont irréductibles et sans chevauchement. Soit $A_p = \{a_0, \ldots, a_{p-1}\}$ un alphabet de $p \geq 3$ lettres et $I_p = \{0, \ldots, p-1\}$. En considérant pour i entier que $a_i = a_{i \ mod \ p}$, pour $n \geq 1$, l'endomorphisme f_n sur A_p est défini par :

$$\forall i \in I_p, \quad f_n(a_i) = a_i C_i(\mu^{2n}(a))$$

où C_i est le morphisme de $\{a, b\}^*$ dans $\{a_{i+1}, a_{i+2}\}^*$ tel que $C_i(a) = a_{i+1}$ et $C_i(b) = a_{i+2}$.

Les f_n précédents étant aussi sans puissance k (quand $k \geq 3$), ils peuvent également servir pour prouver le théorème 4.2 pour des endomorphismes sur des alphabets d'au moins trois lettres. Pour les endomorphismes binaires, une autre famille de morphismes doit être utilisée. Dans [2], en notant $a = a_0$, $b = a_1$, $c = a_2$, il s'agit des endomorphismes g_n définis par $g_n = g f_n C_1$ où

$$\begin{cases} g(a) = aababb, \\ g(b) = aabbab, \\ g(c) = abbaab. \end{cases}$$

(Le morphisme g a été introduit comme exemple de morphisme uniforme sans cube par F.-J. Brandenburg [Bra83] et prouvé comme étant sans puissance k, pour $k \geq 3$ dans [4]). Une autre famille de morphismes est utilisée dans [4] (mais cette famille ne permet pas d'avoir une approche unifiée pour traiter le cas des endomorphismes sans chevauchement et celui des endomorphismes sans puissance k).

M. Leconte a montré [Lec85] qu'un endomorphisme sans puissance $k \geq 4$ est *primitif* [3] c'est-à-dire que l'image d'un mot *primitif* (mot non vide qui ne peut être exprimé comme puissance d'un autre mot) est un mot primitif. En conséquence les familles de morphismes $(f_n)_{n\geq 1}$ et $(g_n)_{n\geq 1}$ sont des ensembles de morphismes uniformes, primitifs et irréductibles. Elles permettent de répondre à une question de V. Mitrana [Mit97].

Théorème 4.3 [2] *Soit A un alphabet contenant au moins deux lettres. Le monoïde des endomorphismes primitifs sur A et le monoïde des endomorphismes* uniformes *primitifs sur A ne sont pas finiment engendrés.*

4.2 Ensembles de test pour morphismes sans chevauchement

En 1993, J. Berstel et P. Séébold [BS93] ont fourni un deuxième moyen pour tester si un endomorphisme sur $\{a, b\}$ est sans chevauchement : un endomorphisme f sur $\{a, b\}$ est sans chevauchement si et seulement si $f(abbabaab)$ est un mot sans chevauchement. L'algorithme de test consiste donc à calculer le mot $f(abbabaab)$ de longueur $4||f||$ puis à tester s'il est ou non sans chevauchement. Puisqu'il existe [Kfo88] un algorithme linéaire en temps pour tester si un mot binaire est sans chevauchement, l'algorithme sous-jacent au résultat théorique précédent agit, comme celui présenté dans la partie 4.1, avec une complexité en temps en $O(||f||)$.

Plusieurs questions se posent : pour améliorer l'algorithme obtenu dans la démarche précédente, peut-on trouver un mot plus petit que *abbabaab* jouant le même rôle ? Est-ce que cette démarche se généralise à des morphismes définis sur des alphabets quelconques ? Pour apporter des réponses à ces questions, nous avons, avec P. Séébold, introduit la notion d'ensemble de test pour morphismes sans chevauchement.

Soient A et B deux alphabets. Nous appelons *ensemble de test* [4] pour morphismes non effaçants sans chevauchement de A^* dans B^* tout ensemble T tel que pour tout morphisme non effaçant f de A^* dans B^*, f est sans chevauchement si et seulement si les images par f de tous les mots de T sont des mots sans chevauchement. Nous appelons aussi *mot de test* tout mot w tel que $\{w\}$ est un ensemble de test. Le résultat précédent de J. Berstel et

3. Cette terminologie n'est pas à confondre avec la notion de substitution primitive (voir par exemple [BFMS02])

4. La terminologie ("ensemble de test") employée ici existe pour d'autres questions. En particulier, la conjecture d'Ehrenfeucht utilise la notion d'ensemble de test pour l'égalité de deux morphismes. (voir, e.g., [CK97]).

P. Séébold peut être reformulé : *abbabaab* est un mot de test pour morphismes sans chevauchement de $\{a,b\}^*$ dans $\{a,b\}^*$.

Tout morphisme sans chevauchement est *fortement bipréfixe*, c'est-à-dire que les images de deux lettres différentes commencent et se terminent par des lettres différentes (voir [Kar81] et [11,15]). En conséquence on supposera dans la suite que B contient au moins autant de lettres que A. De plus, on observe qu'un ensemble de test ne peut contenir que des mots sans chevauchement.

Un ensemble de test peut être infini. Par exemple, l'ensemble de tous les mots sans chevauchement est un ensemble de test. Dans le but de pouvoir tester si un morphisme est sans chevauchement, nous nous sommes essentiellement intéressés aux ensembles finis de test (ensembles finis qui sont des ensembles de test), même si certaines caractérisations s'avèrent plus générales.

Avec P. Séébold, nous avons caractérisé tous les ensembles de test pour endomorphismes sans chevauchement sur $\{a,b\}$.

Théorème 4.4 [7] *Un ensemble T de mots sans chevauchement sur $\{a,b\}$ est un ensemble de test pour endomorphismes non effaçants sans chevauchement sur $\{a,b\}$ si et seulement si T possède comme facteurs*[5]

- *les mots aa et bb,*
- *au moins un des deux mots aba et bab,*
- *au moins un des quatre mots aab, bba, ababb, et babaa,*
- *au moins un des quatre mots baa, abb, bbaba, et aabab,*
- *au moins un des quatre mots aba, abba, bbab, et babb,*
- *au moins un des quatre mots bab, baab, aaba, et abaa.*

Nous rencontrerons par la suite de nombreux résultats de caractérisation d'ensembles de test. À travers leurs énoncés, le lecteur pourra comparer les effets des hypothèses, les ensembles de test (quand ils existent) étant plus ou moins contraints selon qu'on se restreigne ou non à des morphismes uniformes ou selon le type de répétitions considérées (chevauchement, puissance k).

Les techniques de preuves sont à la base les mêmes pour tous ces résultats. Tout d'abord il convient de montrer que la condition caractéristique est suffisante. Étant donné un ensemble T la vérifiant, il faut établir pour tout morphisme f (de la famille considérée) que f est sans chevauchement (ou sans puissance k) si et seulement si $f(T)$ est sans chevauchement (ou sans puissance k). Seul le cas où f n'est pas sans chevauchement (ou sans puissance k) pose problème et mène à des preuves (plus ou moins longues

5. un mot u est dit *facteur* d'un ensemble E si u est facteur d'un mot de E.

et techniques) dont nous ne parlerons pas ou très peu. Dans le cas du théorème 4.4, cette partie de la preuve est une adaptation de celle donnée par J. Berstel et P. Séébold pour établir que *abbabaab* est un ensemble de test.

Pour établir une caractérisation d'ensembles de test, il faut également montrer que la condition caractéristique est nécessaire. Cela se fait à l'aide de morphismes adéquats. La condition étant composée de plusieurs sous-conditions, un morphisme est nécessaire pour chacune de ces sous-conditions. Par exemple, si on considère le morphisme f défini par

$$\begin{cases} a \mapsto bab, \\ b \mapsto a, \end{cases}$$

on peut voir que, pour un mot w, $f(w)$ contient un chevauchement si et seulement si w contient aba ou bab comme facteur. Cela permet de montrer la deuxième partie de la condition caractéristique dans le théorème 4.4.

Dans ce qui suit, nous préciserons les morphismes utilisés dans l'établissement des caractérisations. Ainsi les endomorphismes sur $\{a, b\}$ définis par

$$\begin{cases} a \mapsto aa, \\ b \mapsto b, \end{cases} \quad \begin{cases} a \mapsto ab, \\ b \mapsto a, \end{cases} \quad \begin{cases} a \mapsto baab, \\ b \mapsto a, \end{cases}$$

permettent respectivement de voir que tout ensemble T de test doit avoir comme facteurs à la fois aa, l'un des mots aab, bba, $ababb$, et $babaa$, et l'un des mots aba, $abba$, $bbab$, et $babb$. Échanger le rôle de a et b dans ces morphismes permet d'obtenir des morphismes montrant la nécessité des autres facteurs dans la condition du théorème 4.4.

Un intérêt du théorème 4.4 est d'apporter une réponse positive à la première question posée au début de cette partie (existence d'un plus petit mot de test que *abbabaab*). Le mot *aababb* est un (des six) mot(s) de test sans chevauchement sur $\{a, b\}$. Par ailleurs il n'existe aucun mot de test de longueur inférieure ou égale à 5.

La réponse à la deuxième question posée en début de partie 4.2, à savoir l'existence d'un mot de test pour des morphismes sans chevauchement de A^* dans B^* pour A et B quelconques, a été étudiée en collaboration avec F. Wlazinski. Elle se décompose en deux parties selon le cardinal de A. Tout d'abord nous nous sommes intéressés à des morphismes de $\{a, b\}^*$ dans B^* quand B possède au moins trois lettres (le cas où B a deux lettres se déduit immédiatement du cas des endomorphismes). Dans ce cas, il existe des ensembles de test. Nous avons obtenu la caractérisation suivante (signalons qu'ici la première partie de la preuve n'est pas, comme pour le théorème 4.4, similaire à celle de J. Berstel et P. Séébold pour le mot de test *abbabaab*).

Théorème 4.5 [11,15] *Étant donné un alphabet B contenant au moins trois lettres, un ensemble T de mots sans chevauchement sur l'alphabet $\{a, b\}$ est un ensemble de test pour morphismes non effaçants sans chevauchement de $\{a, b\}^*$ dans B^* si et seulement si T contient comme facteurs les mots aba, bab, abba, baab.*

Bien évidemment, l'ensemble $\{aba, bab, abba, baab\}$ est un exemple d'ensemble de test dans ce cas.

Notons la plus grande simplicité de cette caractérisation par rapport à celle obtenue au théorème 4.4 dans le cas où B a deux lettres. Elle est due au fait qu'une lettre dans l'alphabet d'arrivée peut jouer un rôle de marquage permettant d'obtenir des morphismes qui préservent les mots sans chevauchement sauf pour un mot (au lieu d'un ensemble de mots comme dans la plupart des sous-conditions du théorème 4.4). Illustrons nos propos par un exemple.

Soit A un alphabet et x une lettre n'appartenant pas à A (x est la lettre servant de marqueur). Soit a une lettre particulière de A et u un mot sans chevauchement sur A. Considérons le morphisme f_u défini par

$$\begin{cases} f_u(a) = auaxa, \\ f_u(b) = b, \quad \text{pour toute lettre } b \text{ différente de } a. \end{cases}$$

Pour un mot w sans chevauchement, $f_u(w)$ contient un chevauchement si et seulement si aua est facteur de w. En conséquence, le mot aua doit être facteur de tout ensemble de test pour morphismes sans chevauchement de A^* dans B^* quand B contient strictement plus de lettres que A. Quand $A = \{a, b\}$, u ne peut prendre que trois valeurs : ε, b et bb. Cela montre, en particulier, que les mots aba et $abba$ doivent être facteurs de tout ensemble de test. En inversant le rôle de a et b, les mots bab et $baab$ doivent aussi être facteur de tout ensemble de test. On obtient ainsi la nécessité de la condition dans le théorème 4.5. La valeur ε pour u implique que aa et bb doivent aussi être facteur de tout ensemble de test. Ces mots étant facteurs respectivement de $baab$ et $abba$, ils sont automatiquement pris en compte.

La simplicité apparente de la condition dans le théorème 4.5 cache une plus grande contrainte sur les ensembles de test. En effet, il se déduit des théorèmes 4.4 et 4.5 que tout ensemble de test pour morphismes non effaçants sans chevauchement de $\{a, b\}^*$ dans B^* (avec B un alphabet contenant au moins trois lettres) est aussi un ensemble de test pour endomorphismes non effaçants sans chevauchement sur $\{a, b\}$. La réciproque n'est pas vraie. En particulier, une différence apparaît sur les mots de test. Soit B un alphabet contenant au moins trois lettres. Pour les morphismes sans chevauchement de $\{a,b\}^*$ dans B^*, il n'existe pas de mot de test de longueur 7 ou moins. Le mot $abbabaab$ est un (des huit) mot(s) de test.

En cette fin de partie, terminons de répondre à la question de l'existence d'ensembles de test en fonction des alphabets de définition des morphismes. Reconsidérons les morphismes f_u définis ci-avant en supposant cette fois-ci que l'alphabet A contient au moins trois lettres. Dans ce cas, $A \setminus \{a\}$ contient au moins deux lettres. Le mot de Thue-Morse écrit sur deux de ces lettres fournit alors une infinité de mots u sans chevauchement. Cela montre le résultat suivant quand B a strictement plus de lettres que A.

Théorème 4.6 [11,15] *Étant donnés un alphabet A contenant au moins trois lettres, et B un alphabet de cardinal supérieur ou égal à celui de A, il n'existe pas d'ensemble* fini *de test pour morphismes sans chevauchement de A^* dans B^*.*

Pour prouver le théorème quand A et B ont exactement le même nombre de lettres, il est nécessaire de considérer une autre famille de morphismes avec des propriétés similaires à celle des f_u. Nous décrivons à présent celle utilisée dans [11,15]. Soit A un alphabet contenant au moins trois lettres a, b et c. L'endomorphisme π_3 sur $\{a,b,c\}$ défini par $\pi_3(a) = abc$, $\pi_3(b) = bca$, $\pi_3(c) = cab$ généralise le morphisme μ de Thue-Morse. Ce morphisme est sans chevauchement (voir par exemple [Séé02]). Soit u un mot sur $\{a,b\}$ tel que aua est sans chevauchement (le mot de Thue-Morse contient une infinité de tels facteurs), et soit g_u l'endomorphisme sur A défini par

$$\begin{cases} g_u(a) = \pi_3(a) = abc, \\ g_u(b) = \pi_3(b) = bca, \\ g_u(c) = \pi_3(caua)caab = cababcf(u)abccaab, \\ g_u(d) = d \quad \text{pour toute lettre } d \notin \{a,b,c\}. \end{cases}$$

Pour tout mot sans chevauchement w sur A, $g_u(w)$ contient un chevauchement si et seulement si $cauac$ est un facteur de w. Le “marquage” effectué par la lettre x dans les morphismes f_u est ici plus ou moins assuré par le suffixe $ccaab$ de $g_u(c)$.

Terminons cette partie sur un tableau récapitulant les principaux résultats obtenus pour les ensembles finis de test pour morphismes sans chevauchement.

Card(A)	Card(B) ($\geq$ Card(A))	
	2	≥ 3
2	existence [BS93] caractérisation [7]	caractérisation [11,15]
≥ 3	inexistence [11,15]	

4.3 Le cas des morphismes uniformes

Le théorème 4.6 montre que les ensembles finis de test n'apportent pas de réponse dans le cas général pour tester si un morphisme est sans chevauchement. Il en va d'une toute autre manière si nous nous restreignons à des morphismes uniformes. En effet, il existe des ensembles finis de test pour morphismes uniformes sans chevauchement de A^* dans B^* quel que soit le cardinal des alphabets A et B servant à définir les morphismes (B ayant au moins autant de lettres que A). Les résultats suivants obtenus avec F. Wlazinski fournissent des caractérisations de ces ensembles de test en fonction des cardinaux de A et B. Notons la grande différence avec le cas des morphismes uniformes sans carré où dans tous les cas l'ensemble des mots sans carré de longueur 3 est un ensemble de test [Bra83, Cro82, Cro83]. En particulier, dans le cas "sans chevauchement" contrairement au cas "sans carré", la taille des ensembles de test dépend du cardinal de l'alphabet.

Théorème 4.7 [11,15] *Un ensemble T de mots sans chevauchement sur $\{a, b\}$ est un ensemble de test pour endomorphismes uniformes sans chevauchement sur $\{a, b\}$ si et seulement si T possède comme facteurs*

- *les mots ab et ba,*
- *au moins un des deux mots aa et bb,*
- *au moins un des quatre mots aab, bba, ababb, et babaa,*
- *au moins un des quatre mots baa, abb, bbaba, et aabab.*

Théorème 4.8 [11,15] *Étant donné un alphabet B contenant au moins trois lettres, un ensemble T de mots sans chevauchement sur $\{a, b\}$ est un ensemble de test pour morphismes uniformes sans chevauchement de $\{a, b\}^*$ dans B^* si et seulement si T possède comme facteurs*

- *les mots aa, bb, aba et bab,*
- *au moins un des quatre mots aab, bba, ababb, et babaa,*
- *au moins un des quatre mots baa, abb, bbaba, et aabab.*

Théorème 4.9 [11,15] *Étant donnés un alphabet A contenant au moins trois lettres et un alphabet B contenant au moins autant de lettres que A, un ensemble T de mots sans chevauchement sur A est un ensemble de test pour morphismes uniformes de A^* dans B^* si et seulement si T contient comme facteurs tous les mots des ensembles T_{U1} et T_{U2} suivants :*

$$T_{U1} = \{xw_0x \mid x \in A, w_0 \in A^*, \forall a \in A, |xw_0|_a \leq 1\},$$

$$T_{U2} = \left\{ xw_1\beta w_2y \;\middle|\; \begin{array}{l} x, y, \beta \in A, w_1, w_2 \in A^*, \\ \forall a \in A, |w_1\beta w_2|_a \leq 1, \\ |w_1| = |w_2| \geq 1, |\beta w_2|_x = 0 = |w_1\beta|_y = 0 \end{array} \right\}.$$

Du point de vue preuve, celle du théorème 4.7 est très proche de celle du théorème 4.4. Les morphismes suivants (ainsi que ceux obtenus en échangeant le rôle de a et b) permettent de montrer la nécessité des conditions :

$$\begin{cases} a \mapsto baa, \\ b \mapsto aab, \end{cases} \quad \begin{cases} a \mapsto aa, \\ b \mapsto bb, \end{cases} \quad \begin{cases} a \mapsto aba, \\ b \mapsto abb, \end{cases} \quad \begin{cases} a \mapsto aba, \\ b \mapsto bba. \end{cases}$$

Les morphismes suivants (ainsi que ceux obtenus en échangeant le rôle de a et b) permettent de montrer la nécessité des conditions dans le théorème 4.8 (dans ces morphismes la lettre c joue un rôle de marquage) :

$$\begin{cases} a \mapsto aa, \\ b \mapsto bc, \end{cases} \quad \begin{cases} a \mapsto aba, \\ b \mapsto bcb, \end{cases} \quad \begin{cases} a \mapsto ab, \\ b \mapsto ac, \end{cases} \quad \begin{cases} a \mapsto ba, \\ b \mapsto ca. \end{cases}$$

Pour la preuve du théorème 4.9, on commence par montrer de manière combinatoire que $T_{U1} \cup T_{U2}$ est un ensemble de test dans ce cas. La deuxième partie consiste à montrer que tout ensemble de test contient les mots de $T_{U1} \cup T_{U2}$. Pour cela, observons que ces mots ont une des formes suivantes :

- aba avec a, b deux lettres distinctes
- $xc_1 \cdots c_p \beta d_1 \ldots d_p y$ avec $c_1, \ldots, c_p, d_1, \ldots, d_p, \beta, \gamma$ des lettres deux à deux distinctes et avec $x \in \{c_1, \ldots, c_p, \gamma\}$, $y \in \{d_1, \ldots, d_p, \gamma\}$
- $\gamma c_1 \cdots c_p d_1 \cdots d_p \gamma$ avec $c_1, \ldots, c_p, d_1, \ldots, d_p, \gamma$ des lettres deux à deux distinctes.

Pour chacune de ces formes u, un morphisme f est donné tel que, pour tout mot w sans chevauchement, $f(w)$ contient un chevauchement si et seulement si w contient u. Afin de satisfaire le lecteur curieux, bien que d'apparence un peu techniques, nous présentons ces trois familles de morphismes.

Pour la première forme, c désignant une lettre distincte de a et b :

$$\begin{cases} f(a) = acbabca, \\ f(b) = bacbcab, \\ f(d) = dbbabbd \qquad \forall d \in A \backslash \{a, b\}. \end{cases}$$

Pour la deuxième forme, en notant $c_{p+1} = d_0 = \beta$:

$$\begin{cases} f(c_i) & = c_i\ d_{i-1}\ d_i\ \beta\ c_i\ \beta\ \beta\ c_i\ (\forall\ 1 \leq i \leq p), \\ f(d_i) & = d_i\ \beta\ c_i\ \beta\ \beta\ c_i\ c_{i+1}\ d_i\ (\forall\ 1 \leq i \leq p), \\ f(\gamma) & = \gamma\ \beta\ d_1\ \beta\ c_1\ \beta\ \beta\ \gamma, \\ f(\beta) & = \beta\ d_p\ y\ \beta\ \beta\ x\ c_1\ \beta, \\ f(a) & = a\ d_1\ d_1\ c_1\ d_1\ c_1\ c_1\ a\ (\forall a \in A \backslash \{c_1, \ldots, c_p, \beta, d_1, \ldots, d_p, \gamma\}). \end{cases}$$

Pour la troisième forme, en notant $c_0 = d_{p+1} = \gamma$:

$$\begin{cases} f(c_i) & = c_i\ d_i\ d_{i+1}\ d_{i+1}\ d_i\ d_{i+1}\ c_i\ (\forall\ 1 \leq i \leq p), \\ f(d_i) & = d_i\ c_{i-1}\ c_i\ d_i\ d_{i+1}\ d_{i+1}\ d_i\ (\forall\ 1 \leq i \leq p), \\ f(\gamma) & = \gamma\ c_p\ d_1\ c_p\ d_1\ d_1\ \gamma, \\ f(a) & = a\ c_1\ c_1\ d_1\ c_1\ c_1\ a\ (\forall a \in A \backslash \{c_1, \ldots, c_p, d_1, \ldots, d_p, \gamma\}). \end{cases}$$

Revenons à des considérations moins techniques. Après ce qui a été présenté dans la partie précédente, on peut se demander quelle est la longueur des plus petits mots de test pour morphismes uniformes sans chevauchement de A^* dans B^*. Quand $A = B = \{a, b\}$, ces plus petits mots de test sont *abba* et *baab*. Quand $A = \{a, b\}$ et B est de cardinal supérieur ou égal à 3, ce sont *aababb* et *bbabaa*. Quand A contient au moins trois lettres, la caractérisation des ensembles de test ne dépend pas du cardinal de B. La longueur d'un mot de test (s'il en existe) dans ce cadre est toujours supérieure à la factorielle du cardinal de A.

La forme compliquée de l'ensemble de test minimal $T_{U1} \cup T_{U2}$ amène à remarquer que l'ensemble des mots sans chevauchement de longueur $\mathrm{Card}(A)+2$ (resp. $\mathrm{Card}(A)+1 = 3$) est un ensemble de test pour morphismes uniformes sans chevauchement de A^* dans B^* quand $3 \leq \mathrm{Card}(A) \leq \mathrm{Card}(B)$ (resp. quand $2 = \mathrm{Card}(A) \leq \mathrm{Card}(B)$).

4.4 Morphismes sans puissance k

Les ensembles et mots de test pour morphismes sans puissance k peuvent être définis comme leurs analogues pour morphismes sans chevauchement (voir partie 4.1).

Soient A et B deux alphabets (contenant au moins deux lettres). Nous avons déjà signalé que l'ensemble des mots sans carré de longueur 3 était un ensemble de test pour morphismes uniformes sans carré. M. Crochemore [Cro82, Cro83] a aussi montré :

- quand $\mathrm{Card}(A) = 3$, un morphisme de A^* dans B^* est sans carré si et seulement si les images de tous les mots sans carré de longueur 5 sont sans carré.
- quand $\mathrm{Card}(A) \geq 4$, il n'existe pas d'ensemble de test pour morphismes sans carré de A^* dans B^*.

Avec F. Wlazinski, nous nous sommes intéressés aux ensembles de test pour morphismes sans puissance k quand $k \geq 3$. Nous avons tout d'abord obtenu un résultat similaire au théorème 4.6.

Théorème 4.10 [4,16] *Soient $k \geq 3$ un entier, A un alphabet d'au moins trois lettres et B un alphabet d'au moins deux lettres. Il n'existe pas d'ensemble* fini *de test pour morphismes sans puissance k de A^* dans B^*.*

La preuve de ce théorème utilise la même technique que celle de la preuve du théorème 4.6 : utilisation d'une infinité de morphismes ne faisant apparaître chacun une puissance k que dans l'image d'un seul mot sans puissance

k. Nous avons besoin d'une nouvelle famille de morphismes. Soient $k \geq 3$ un entier, A un alphabet, a une lettre de A et x, y, z trois lettres n'appartenant pas à $A \setminus \{a\}$ (ces trois lettres vont servir à "marquer" les images), u et v deux mots sans puissance k (avec u et v non simultanément vides si $k = 3$). Posons $Z = xzy$ et $W = uxyv$. Définissons le morphisme $f_{u,v,k}$ de A^* dans $(A \setminus \{a\} \cup \{x, y, z\})^*$ par

$$\begin{cases} f_{u,v,k}(a) = (ZW)^{k-1}Z = x(zyuxyvx)^{k-1}zy, \\ f_{u,v,k}(b) = b \quad \text{pour toute lettre } b \in A \setminus \{a\}. \end{cases}$$

Pour tout mot w sans puissance k sur A, $f_{u,v,k}(w)$ contient une puissance k si et seulement si w contient le facteur $avaua$.

Dans le théorème 4.10, A contient au moins trois lettres. Le mot de Thue-Morse nous fournit alors une infinité de mots sans chevauchement, donc sans puissance k sur $A \setminus \{a\}$. L'utilisation de la famille précédente nous permet de voir que tout ensemble de test pour morphismes sans puissance k de A^* dans B^* est infini $(\dots)$ quand $\mathrm{Card}(B) \geq \mathrm{Card}(A) + 2$.

Le lecteur aura noté que la restriction précédente n'apparaît pas dans l'hypothèse du théorème 4.10 et qu'il est même possible que A ait strictement plus de lettres que B (contrairement à la restriction générale faite lors de l'étude des morphismes sans chevauchement).

La preuve du théorème 4.10 se termine en utilisant un morphisme sans puissance k dont les images sont codées sur uniquement deux lettres. Nous avons déjà rencontré un tel morphisme. Le morphisme g page 22 défini par $g(a) = aababb$, $g(b) = aabbab$, $g(c) = abbaab$. Ce morphisme reste sans puissance k quand on le prolonge par $g(x) = x$ pour des lettres x n'appartenant pas à $\{a, b, c\}$. Le morphisme g réduit l'alphabet utilisé d'une lettre. Par composition (de copies de g obtenues en changeant éventuellement l'alphabet), on peut obtenir un morphisme h sans puissance k de A^* (A de cardinal supérieur ou égal à 2) dans $\{a, b\}^*$ (et donc dans B^* pour tout alphabet B de cardinal supérieur ou égal à 2). La famille des morphismes $hf_{u,v,k}$ permet de montrer l'inexistence d'ensemble fini de test pour morphismes sans puissance k de A^* dans B^*.

Quand A contient deux lettres, il existe des ensembles de test finis. F. Wlazinski a fourni des ensembles de test pour endomorphismes binaires sans puissance k quand $k \geq 4$ [Wla01, Wla02]. M. Leconte [Lec85] a montré que l'ensemble de tous les mots sans cube de longueur 7 en était un quand $k = 3$. Le résultat suivant obtenu avec F. Wlazinski caractérise les ensembles de test pour morphismes binaires sans cube. Remarquons que, contrairement au cas des morphismes sans chevauchement, le résultat ne dépend pas du cardinal de l'alphabet d'arrivée. Cela est dû à l'existence du morphisme h précédent.

Théorème 4.11 [4,16] *Soient a et b deux lettres et B un alphabet d'au moins deux lettres. Un ensemble T de mots sans cube est un ensemble de test pour morphismes sans cube de $\{a,b\}^*$ dans B^* si et seulement si les douze mots abbabba, baabaab, ababba, babaab, abbaba, baabab, aabba, bbaab, abbaa, baabb, ababa, babab sont facteurs de T.*

L'ensemble {*abbabba*, *baabaab*, *ababba*, *babaab*, *abbaba*, *baabab*, *aabba*, *bbaab*, *abbaa*, *baabb*, *ababa*, *babab*} est un exemple d'ensemble de test plus petit que celui constitué des 36 mots sans cube de longueur 7.

Prouver que tous les mots de cet ensemble sont facteurs de tout ensemble de test peut se faire en utilisant la famille des $f_{u,v,k}$. Ici il n'y a que 8 possibilités pour le couple (u,v) permettant de définir les morphismes : (ε, b), (ε, bb), (b, ε), (b, b), (b, bb), (bb, ε), (bb, b), (bb, bb). Cela montre que les mots *aaba*, *aabba*, *abaa*, *ababa*, *ababba*, *abbaa*, *abbaba*, *abbabba* doivent être facteurs d'un ensemble de test. Huit autres mots sont obtenus en inversant le rôle de a et de b. Les mots *aaba* et *abaa* étant facteurs du mot *baabaab* (et les mots *bbab* et *babb* étant facteur de *abbabba*), il est inutile de préciser leur présence dans un ensemble de test.

Le reste de la preuve du théorème 4.11 est long et technique. Nous notons ici uniquement les deux grandes étapes de cette preuve.

1. Soit f un morphisme défini sur $\{a,b\}$ tel que les images par f des mots *abba*, *baab*, *aba* ou *aba* sont sans cube. Soit w un mot sans cube de longueur supérieure ou égale à 8 tel que $f(w)$ contient un cube. Alors w contient un facteur u de longueur au plus 7 tel que $f(u)$ contient un cube.

   Un corollaire de cette première étape est qu'un morphisme f est sans cube si et seulement si les images des mots sans cube de longueur au plus 7 sont sans cube. Comme tout mot sans cube de longueur inférieure ou égale à 7 est facteur d'un mot sans cube de longueur 7, on retrouve le résultat de M. Leconte cité précédemment.
2. La deuxième étape consiste à montrer que les images des 12 mots du théorème 4.11 sont sans cube si et seulement si les images des mots sans cube de longueur au plus 7 sont sans cube.

Remarquons à présent deux conséquences de la première étape mentionnées dans [4].

– Un moyen de tester si un endomorphisme f sur $\{a,b\}$ engendre un mot sans cube consiste à regarder si les images de *abba*, *baab*, *aba* et *bab* sont sans cube, puis à regarder si les images des facteurs de longueur 7 du mot engendré sont sans cube. Cette technique est beaucoup plus rapide qu'une technique précédemment proposée par J. Karhumäki [Kar83] consistant à regarder si le mot $f^{10}(a)$ est sans cube.

– Soient des entiers i, j, k, et les mots $X = abba$, $T = baab$. Considérons l'endomorphisme défini sur $\{a, b\}$ par

$$\begin{cases} a \mapsto X, \\ b \mapsto (TaTX^iTTX^j)^kTaT. \end{cases}$$

La technique de test précédente permet de montrer que, quand $i = j = k = 2$, le morphisme précédent engendre un mot sans cube. Ce morphisme n'est pourtant pas sans cube puisque l'image de *baabaab* contient un cube (*baabaab* est le seul mot de longueur inférieure ou égale à 7 pour lequel cela arrive, et il n'est pas facteur du mot infini engendré par le morphisme).

Par des considérations ad hoc, avec P. Séébold [9], nous avons généralisé ce résultat. Pour des entiers i, j, k tels que $\{i, j\} \neq \{1, 2\}$ et $k \geq 2$, le morphisme précédent n'est pas sans puissance $(k + 1)$ mais engendre un mot sans puissance $k + 1$. De plus, le morphisme est sans puissance $k + p$ pour tout entier $p \geq 2$. La preuve de cette généralisation utilise des considérations spécifiques aux morphismes concernés.

Terminons cette partie sur une réponse à une question maintenant naturelle quand on s'intéresse aux ensembles de test. En conséquence du théorème 4.11, tout plus petit mot de test pour morphismes sans puissance 3 de $\{a, b\}^*$ dans B^* est de longueur 24, comme par exemple :

$$aabbababbabbaabaababaabb$$

4.5 Perspectives

Décider si un morphisme est sans chevauchement ou sans puissance k (quand $k \geq 3$) reste un problème ouvert pour des alphabets d'au moins trois lettres. Nous avons montré des limites des techniques (monoïdes d'endomorphismes, ensembles de test) abordées. Le tableau suivant résume les résultats obtenus pour les ensembles de test pour morphismes sans puissance

k et morphismes sans chevauchement ($k = 2^+$).

<table>
<tr><td></td><td colspan="3">Card(A)</td></tr>
<tr><td></td><td>2</td><td>3</td><td>≥ 4</td></tr>
<tr><td>$k = 2$</td><td></td><td>existence [Cro82]</td><td>inexistence [Cro82]</td></tr>
<tr><td>$k = 2^+$</td><td>existence [BS94]
caractérisation [7,11,15]</td><td colspan="2">inexistence [11,15]</td></tr>
<tr><td>$k = 3$</td><td>existence [Lec85]
caractérisation [4]</td><td colspan="2" rowspan="2">inexistence [4]</td></tr>
<tr><td>$k \geq 4$</td><td>existence [Wla02]</td></tr>
</table>

La recherche d'ensembles de test dépendant du morphisme à tester (comme les mots de longueur inférieure ou égale à $2k||f|| + 1$ pour les morphismes presque circulaires étudiés par M. Leconte) est une suite logique aux travaux présentés. Vu que le nombre de mots sans puissance k ($k \geq 3$) sur un alphabet à deux lettres ou sans chevauchement sur trois lettres croît exponentiellement avec la longueur des mots, il y a intérêt à chercher des ensembles de test avec le moins d'éléments possible, par exemple comme cela a été obtenu pour les morphismes binaires primitifs par A. Lentin et M. P. Schützenberger [LS69] : un morphisme f sur l'alphabet $\{a, b\}$ est primitif si et seulement si les images par f des mots de $a^*b \cup ab^*$ sont des mots primitifs (il suffit de tester les mots dont l'image est de longueur inférieure ou égale à trois fois la longueur de l'image de ab).

V. Mitrana a étudié les morphismes primitifs. Il a en particulier fourni un algorithme pour tester si un morphisme est primitif. Néanmoins on peut se poser la question d'une éventuelle généralisation du résultat de A. Lentin et M. P. Schützenberger à des alphabets non binaires. V. Mitrana [Mit97] pose aussi la question : existe-t'il un ensemble fini de test pour tester si un morphisme uniforme est primitif ?

Une meilleure connaissance des morphismes sans puissance k ou sans chevauchement est sans doute nécessaire pour pouvoir les tester. Actuellement je m'intéresse à montrer qu'un morphisme sans puissance k est un morphisme primitif (après les travaux de M. Leconte [Lec85], seul le cas $k = 3$ reste à traiter).

Avec P. Séébold [9], nous avons récemment émis trois conjectures concernant les morphismes sans puissance k. Les deux premières sont :

- un morphisme binaire qui engendre un mot infini sans puissance k ($k \geq 3$) est sans puissance $k + 1$, ou n'est pas sans puissance n pour tout entier n.

- un morphisme sans puissance k ($k \geq 3$) est sans puissance $k+1$. (Il existe des morphismes sans carré qui ne sont pas sans cube, voire sans chevauchement. La question suivante est, elle aussi, ouverte : un morphisme sans chevauchement est-il nécessairement sans cube ?)

La troisième conjecture développée avec P. Séébold [9] concerne les ensembles de test pour morphismes sans puissance k quand $k \geq 4$. Nous avons déjà signalé que F. Wlazinski [Wla01, Wla02] a montré l'existence de tels ensembles pour des morphismes définis sur l'alphabet $\{a, b\}$. Aucune caractérisation de ces ensembles de test n'est à ce jour connue. Nous avons conjecturé qu'ils étaient les ensembles de mots sans puissance k contenant $S = \{ab^iab^ja, ba^iba^jb \mid 1 \leq i, j \leq k\}$.

Parmi les questions concernant les ensembles de test notons qu'il n'est pas connu s'il existe ou non (et si oui quelle peut en être une caractérisation) des ensembles de test pour morphismes uniformes sans puissance k (quand $k \geq 3$) sur des alphabets d'au moins trois lettres.

Le fait qu'un endomorphisme préserve les mots sans puissance k n'est pas une condition nécessaire pour qu'il engendre un mot infini sans puissance k. Dans [Cro83, MS93], des moyens sont fournis pour tester si un morphisme engendre un mot sans puissance k pour $k \geq 2$ entier. Qu'en est-il pour les morphismes sans chevauchement sur un alphabet d'au moins trois lettres ? (Sur deux lettres, les travaux de J.-J. Pansiot [Pan81] et P. Séébold [Séé85] montrent que si un morphisme engendre un mot sans chevauchement alors il est sans chevauchement.)

Pour finir avec les morphismes sans puissance k et les morphismes sans chevauchement, signalons que peu de choses sont connues sur les morphismes préservant des puissances non entières ou des morphismes sans puissance $k+$ (généralisation de la notion de chevauchement quand $k \geq 3$).

5 Morphismes de Lyndon

Dans cette partie, nous considérons des alphabets ordonnés. Nous notons $A_n = \{a_1 < a_2 < \cdots < a_n\}$ et $\preceq$ l'ordre lexicographique.

Les *mots de Lyndon* ont de nombreuses applications (voir par exemple [Lot83, Lot02]). Ce sont les mots non vides strictement plus petits dans l'ordre lexicographique que tous leurs suffixes propres. Un résultat fondamental (voir par exemple [Lot83]) les concernant est : tout mot peut se décomposer en un produit de mots de Lyndon décroissants. Ce résultat a été étendu aux mots infinis [SMDS94]. La décomposition d'un mot infini peut être constituée d'une infinité de mots de Lyndon finis ou se terminer par un *mot de Lyndon infini* c'est-à-dire un mot ayant un nombre infini de préfixes qui sont des mots de Lyndon. Pour obtenir la décomposition du mot de Thue-Morse, A. Ido et G. Melançon [IM97] ont utilisé des morphismes qui préservent les mots de Lyndon. Dans [Lot02, partie 4.4], un autre exemple est utilisé. Dans le but d'obtenir un outil utile pour de futures décompositions, nous avons étudié ces morphismes, que nous appellerons désormais *morphismes de Lyndon*. Notons qu'un mot infini engendré par un morphisme f de Lyndon en une lettre a est un mot de Lyndon infini puisque pour tout entier n, $f^n(a)$ est un mot de Lyndon préfixe de ce mot infini.

5.1 Morphismes préservant l'ordre sur les mots finis

Un des principaux objectifs de l'étude menée sur les morphismes de Lyndon est de proposer un moyen pour tester si un morphisme préserve ou non les mots de Lyndon. Nous verrons (Proposition 5.2) qu'un morphisme de Lyndon est nécessairement un *morphisme qui préserve l'ordre* lexicographique sur les mots finis, i.e., un morphisme f tel que, pour tous mots u et v, $u \preceq v$ implique $f(u) \preceq f(v)$.

Remarquons que les morphismes non vides qui préservent l'ordre sur les mots finis le préservent strictement, c'est-à-dire que pour tous mots u et v, $u \prec v$ implique $f(u) \prec f(v)$. La réciproque est évidemment vraie.

Muni de cette remarque, il est possible d'obtenir la caractérisation suivante :

Proposition 5.1 [10] *Soit $n \geq 2$. Un morphisme non vide sur A_n préserve l'ordre sur les mots finis si et seulement si pour tout entier i, $1 \leq i < n$, $f(a_i a_n^{k_i}) \prec f(a_{i+1})$ où k_i est le plus petit entier tel que $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$.*

La condition est immédiatement nécessaire. Pour montrer qu'elle est suffisante, la preuve dans [10] agit en quelques étapes courtes établissant successivement quelques propriétés des morphismes préservant l'ordre :

- $\forall i, j,\ 1 \leq i < j \leq n,\ f(a_i) \prec f(a_j)$,
- le morphisme f n'est pas effaçant,
- $\forall i, k,\ 1 \leq i < n,\ k \geq 0,\ f(a_i a_n^k) \prec f(a_{i+1})$,
- pour tout mot u, il existe un entier $k \geq 1$ tel que $f(u) \prec f(a_n^k)$,
- et finalement (en utilisant la définition de l'ordre lexicographique) pour tous mots u, v, si $u \prec v$ alors $f(u) \prec f(v)$.

La proposition 5.1 permet de tester si un morphisme préserve l'ordre sur les mots finis. Pour chaque entier i, il faut d'abord déterminer l'entier k_i. Il y a ensuite à vérifier la préservation de l'ordre sur $n-1$ couples de mots (dont la longueur dépend du morphisme).

Considérons à présent un morphisme f de Lyndon sur A_n. Les mots $a_i a_n^{k_i} a_{i+1}$ (k_i comme dans la proposition précédente) étant des mots de Lyndon, leurs images doivent aussi être des mots de Lyndon. En particulier, on doit avoir $f(a_i a_n^{k_i}) \prec f(a_{i+1})$. Ainsi f est un morphisme de Lyndon. Avec l'hypothèse que les images des lettres sont des mots de Lyndon, la réciproque peut être établie. En conséquence, on a :

Proposition 5.2 [10] *Soit $n \geq 2$. Un morphisme f sur A_n est un morphisme de Lyndon si et seulement si*

1. *$f(a)$ est un mot de Lyndon pour chacune des lettres a de A, et*
2. *f préserve l'ordre sur les mots finis.*

Cela est encore équivalent à :

1. *$f(a)$ est un mot de Lyndon pour chacune des lettres a de A, et*
2. *pour tout entier i, $1 \leq i < n$, $f(a_i a_n^{k_i} a_{i+1})$ est un mot de Lyndon, avec k_i le plus petit entier tel que $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$.*

Puisqu'il existe un algorithme pour tester si un mot est de Lyndon (voir par exemple [Lot]), la proposition 5.2 fournit deux moyens pour tester si un morphisme est de Lyndon (l'un d'eux utilisant la proposition 5.1).

5.2 Ensembles de test

La dernière partie de la proposition 5.2 peut être interprétée en terme d'ensemble infini de test : l'ensemble rationnel $\bigcup_{i=1}^{n-1} a_i a_n^* a_{i+1}$ est un ensemble infini de test pour morphismes de Lyndon. L'effectivité de cet ensemble de test est due au fait que seule une partie finie (déterminée en fonction du morphisme à tester) est à utiliser lors d'un test.

La question qui vient immédiatement après la remarque précédente est : existe-t'il un ensemble de test fini (ne dépendant pas du morphisme) ? Comme pour les morphismes sans chevauchement et les morphismes sans puissance k, la réponse dépend du cardinal de l'alphabet d'arrivée. Dans le cas général, elle est négative :

Proposition 5.3 [10] *Étant donnés deux alphabets A et B tels que* $\mathrm{Card}(A) \geq 3$ *et* $\mathrm{Card}(B) \geq 2$*, il n'existe pas d'ensemble fini de test pour morphismes de Lyndon de A^* dans B^*.*

La preuve de ce résultat utilise les endomorphismes f_k de A_n définis pour $k \geq 0$ par :

$$\begin{cases} f_k(a_1) = a_1, \\ f_k(a_2) = a_1 a_n^k a_2, \\ f_k(a_i) = a_i \quad \text{pour } 3 \leq i \leq n. \end{cases}$$

Quand $n \geq 3$, un tel morphisme n'est pas de Lyndon. Le plus petit mot u tel que $f_k(u)$ n'est pas un mot de Lyndon est de longueur supérieure ou égale à $|u| \geq k + 3$. En conséquence, un ensemble de test pour morphismes de Lyndon est nécessairement infini.

Remarquons que les morphismes f_k précédents permettent de montrer l'optimalité des k_i dans les propositions 5.1 et 5.2.

Quand on s'intéresse à des morphismes définis sur un alphabet à deux lettres $\{a < b\}$, il existe par contre un exemple de test fini, l'ensemble $\{a, b, ab\}$. Ceci est une conséquence de :

Proposition 5.4 *Un morphisme f sur $\{a < b\}$ est de Lyndon si et seulement si*

- *$f(a)$ et $f(b)$ sont des mots de Lyndon,*
- *$f(a) \prec f(b)$.*

Ce dernier résultat est une amélioration de la proposition 5.1 quand $n = 1$ dans la mesure où k_1 peut dans certains cas prendre la valeur 1 et pas toujours la valeur 0.

Pour finir sur les ensembles de test pour morphismes de Lyndon, notons que $A_n \cup \{a_i a_{i+1} \mid 1 \leq i < n\}$ est un ensemble de test fini pour morphismes uniformes de Lyndon définis sur A_n.

La notion d'ensemble de test peut aussi être déclinée pour les morphismes préservant l'ordre sur les mots finis. Dans ce cas, les ensembles de test sont constitués de couples de mots ordonnés. Les techniques utilisées pour les morphismes de Lyndon permettent de montrer [10] :

- Il n'existe pas d'ensemble de test fini pour morphismes préservant l'ordre définis sur un alphabet ordonné de plus de trois lettres.

– Le singleton $\{(ab, b)\}$ est un ensemble de test pour morphismes préservant l'ordre sur les mots finis sur $\{a < b\}$.
– Étant donné un alphabet A ordonné, l'ensemble $\{(a, b) \mid a < b, a, b \in A\}$ est un ensemble de test pour morphismes uniformes préservant l'ordre sur A.

5.3 Monoïdes

Nous nous intéressons à présent aux monoïdes des endomorphismes de Lyndon ou des endomorphismes préservant l'ordre sur les mots finis. Comme pour les morphismes sans chevauchement (à l'exception du cas binaire) et sans puissance k, nous avons :

Proposition 5.5 [10] *Soit A un alphabet ordonné* $\mathrm{Card}(A) \geq 2$. *Le monoïde des endomorphismes de Lyndon sur A, le monoïde des endomorphismes uniformes sur A, le monoïde des endomorphismes sur A préservant l'ordre lexicographique et le monoïde des endomorphismes uniformes sur A préservant l'ordre lexicographique sont non finiment engendrés.*

La technique utilisée est la même que dans la partie 4.1. Les endomorphismes suivant (définis pour $p \geq 2$ sur A_n) sont des endomorphismes irréductibles à la fois de Lyndon et préservant l'ordre sur les mots finis. Plus précisément, l'identité étant la seule permutation préservant les mots de Lyndon (resp. l'ordre sur les mots finis), ces morphismes ne peuvent pas se décomposer en fg avec f et g deux endomorphismes de Lyndon (resp. préservant l'ordre sur les mots finis).

$$\begin{cases} a_1 \mapsto a_1^p a_2, & \\ a_i \mapsto a_{i-1} a_i^{p-1} a_{i+1} & \forall 2 \leq i \leq n-1, \\ a_n \mapsto a_{n-1} a_n^p. & \end{cases}$$

Je me suis également intéressé à la simplifiabilité des monoïdes précédents. Les morphismes non vides préservant l'ordre sur les mots finis ou préservant les mots de Lyndon sont des morphismes injectifs. Pour de tels morphismes, f, g, h, l'égalité $hf = hg$ implique $f = g$. En d'autre termes le monoïde des endomorphismes de Lyndon et le monoïde des endomorphismes non vides préservant l'ordre sur les mots finis sont simplifiables à gauche.

Le monoïde des endomorphismes préservant l'ordre n'est pas simplifiable à droite. Si on considère les endomorphismes f, g, h définis sur A_n par

$$\begin{cases} f(a_1) = a_1 a_2 a_1, \\ f(a_2) = a_2, \end{cases} \qquad \begin{cases} g(a_1) = a_1, \\ g(a_2) = a_2 a_1 a_2, \end{cases} \qquad \begin{cases} h(a_1) = a_1 a_2, \\ h(a_2) = a_2 a_1, \end{cases}$$

et pour $i \notin \{1, 2\}$, $f(a_i) = g(a_i) = h(a_i)$, alors $fh = gh$ mais $f \neq g$.

Quand $n \geq 3$, l'exemple suivant montre que le monoïde des endomorphismes de Lyndon sur A_n n'est pas simplifiable à droite. Pour f, g, h définis sur A_n par

$$\begin{cases} f(a_1) = a_1a_1a_1a_2, \\ f(a_i) = a_1a_ia_ia_i \text{ pour } i \geq 2, \end{cases} \qquad \begin{cases} g(a_1) = a_1a_1a_2a_2, \\ g(a_i) = a_1a_ia_ia_i \text{ pour } i \geq 2, \end{cases}$$

$$\begin{cases} h(a_1) = a_2a_2a_2a_3, \\ h(a_2) = a_2a_2a_2a_3, \\ h(a_3) = a_2a_3a_3a_3, \\ h(a_i) = a_2a_ia_ia_i \text{ pour } i \geq 4, \end{cases}$$

$fh = gh$ mais $f \neq g$.

Par contre, le monoïde des endomorphismes de Lyndon sur A_2 est simplifiable à droite. La preuve donnée dans [10] utilise la caractérisation suivante de l'ensemble de test d'égalité (au sens de la conjecture d'Ehrenfeucht, voir par exemple le chapitre 7 de C. Choffrut et J. Karhumäki dans [RS97]) entre deux morphismes binaires de Lyndon.

Proposition 5.6 [10] *Étant donnés deux morphismes de Lyndon f et g sur $\{a, b\}$, l'ensemble $E(f, g) = \{w \in \{a, b\}^+ \mid f(w) = g(w)\}$ est soit l'ensemble vide, soit l'ensemble b^+, soit l'ensemble $(ab^n)^+$ pour un entier $n \geq 0$.*

Ce résultat est plus précis que le cas général (proposition 5.1 dans [RS97, chapitre 7] et [Hol2003]) : étant donnés deux morphismes non périodiques f et g sur un alphabet binaire, $E(f, g) = \emptyset$ ou $E(f, g) = \{u, v\}^+$ pour des mots u et v. À noter qu'un morphisme de Lyndon n'est pas périodique.

5.4 Perspectives

En préambule de cette partie sur les morphismes de Lyndon, nous motivions notre étude en espérant avoir un outil pour aider à décomposer des mots infinis en mots de Lyndon. Dans [10], nous illustrons cette possibilité en fournissant de nouvelles preuves du fait que $a\varphi^\omega(a)$ est un mot de Lyndon infini et de la décomposition du mot de Fibonacci $\varphi^\omega(a)$ en mots de Lyndon. Il reste à utiliser cet outil pour de nouvelles décompositions.

En prolongement de mes travaux sur les morphismes de Lyndon, une collaboration est envisagée avec P. Séébold et G. Melançon (Montpellier) pour étudier les morphismes qui préservent les mots de Lyndon infinis, les morphismes qui préservent l'ordre lexicographique sur les mots infinis, les morphismes qui préservent les décompositions en mots de Lyndon. J'ai personnellement déjà obtenu des conditions nécessaires pour l'appartenance d'un morphisme à l'une de ces familles. L'étude paraît plus délicate que pour la préservation des propriétés sur les mots finis. Par exemple un morphisme qui préserve les mots de Lyndon infinis préserve aussi l'ordre lexicographique sur les mots finis mais pas nécessairement sur les mots infinis. De plus, les images de lettres ne sont pas nécessairement des mots de Lyndon (c'est le cas pour les morphismes de Lyndon). Sur un alphabet ordonné de n lettres, l'image d'une des $n-1$ premières lettres est une puissance d'un mot de Lyndon.

L'étude des morphismes engendrant des mots infinis de Lyndon (un tel morphisme n'est pas nécessairement un morphisme de Lyndon) est aussi envisagée.

6 Autour des mots et morphismes sturmiens

6.1 Mots sturmiens

Depuis les travaux de M. Morse et G. A. Hedlund [MH40], les mots sturmiens ont été beaucoup étudiés (voir par exemple [Lot02, chapitre 2]). Ils sont en relation avec de multiples domaines comme par exemple les systèmes dynamiques ou l'arithmétique. En informatique, ils apparaissent dans des cas de complexité au pire d'algorithmes (voir, e.g., [KMP77]) ou comme solution optimale pour certains problèmes (voir, e.g., [AGH00])

Les *mots sturmiens* sont des mots binaires infinis qui peuvent être définis de multiples façons équivalentes. Deux définitions équivalentes possibles sont [CH74] :

- mots dont, pour tout entier $n \geq 0$, le nombre de facteurs de longueur n est $n+1$,
- mots infinis équilibrés et non ultimement périodiques.

Rappelons qu'un mot infini est non *ultimement périodique* s'il ne peut pas être écrit uv^ω pour des mots finis u et v. Un mot *équilibré* fini ou infini sur l'alphabet $\{a, b\}$ est un mot tel que pour tous facteurs u et v de même longueur, $-1 \leq |u|_a - |v|_a \leq 1$ (ou de manière équivalente $-1 \leq |u|_b - |v|_b \leq 1$). Le mot $abaababaabaab$ est équilibré alors que le mot $aabb$ ne l'est pas.

M. Morse et G. A. Hedlund [MH40] appellent *n-chaîne* un mot commençant et terminant par une lettre a et contenant exactement $n+1$ occurrences de la lettre a, i.e., un mot dans $a\{a, b\}^* \cap \{a, b\}^*a$ tel que $|x|_a = n+1$. Dans [13] j'ai prouvé une propriété caractéristique des mots sturmiens consistant à compter le nombre de n-chaînes pour chaque entier n. Notons $F(w)$ l'ensemble des facteurs de w.

Proposition 6.1 [13] *Un mot infini w est sturmien sur $\{a, b\}$ si et seulement si pour tout entier $n \geq 0$ les deux ensembles suivants ont chacun exactement n éléments :*

$$X_n(w) = \{x \in F(w) \cap a\{a, b\}^* \cap \{a, b\}^*a \mid |x|_a = n\},$$

$$Y_n(w) = \{x \in F(w) \cap b\{a, b\}^* \cap \{a, b\}^*b \mid |x|_b = n\}.$$

À partir des deux définitions d'un mot sturmien rappelées précédemment, la nécessité de la condition dans la précédente proposition est vérifiable. La preuve de la réciproque utilise les *morphismes sturmiens*, i.e., les morphismes qui préservent les mots infinis sturmiens. P. Séébold [Séé91] (voir aussi [Lot02, chapitre 2]) a montré que le monoïde des endomorphismes sturmiens sur $\{a, b\}$ est finiment engendré par les trois morphismes E, φ et $\tilde{\varphi}$ définis par :

$$\begin{cases} E(a) = b, \\ E(b) = a, \end{cases} \quad \begin{cases} \varphi(a) = ab, \\ \varphi(b) = a, \end{cases} \quad \begin{cases} \tilde{\varphi}(a) = ba, \\ \tilde{\varphi}(b) = a. \end{cases}$$

Si pour un mot w, les deux ensembles $X_n(w)$ et $Y_n(w)$ sont de cardinal n alors on peut voir que le mot w peut s'écrire $f(W)$ où W est un mot sturmien et f un morphisme sturmien : ceci permet de conclure que w est sturmien.

Revenons à présent sur les mots finis équilibrés sur $\{a, b\}$. Un tel mot w ne peut pas contenir simultanément aa et bb. Il se décompose (en supprimant éventuellement la première et la dernière lettre) alors sur un des ensembles $\{ab, a\}$ ou $\{ba, b\}$. Le mot w peut alors s'écrire $f(w')$ pour f un morphisme sturmien et w' un mot tel que w est équilibré si et seulement si w' est équilibré. Cette idée a conduit S. Dulucq et D. Gouyou-Beauchamps [DGB90] à donner un algorithme pour tester si un mot fini est équilibré.

La remarque précédente peut être prolongée en notant que pour un mot équilibré contenant au moins deux a ou deux b, le mot (à un préfixe et suffixe près) se décompose sur $\{ba^n, ba^{n+1}\}$ ou $\{ab^n, ab^{n+1}\}$ pour un entier $n \geq 1$. Cela conduit [6] à tester si un mot est équilibré en un temps linéaire par rapport à la longueur du mot étudié.

Les algorithmes précédents cherchent en fait à déterminer si w est facteur d'un mot $f(a)$ pour un morphisme f sturmien. En cherchant à ne pas supprimer de préfixe de w, on est amené à décomposer sur des ensembles $\{a^p ba^{n-p}, a^{p+1} ba^{n-p}\}$ ou $\{b^p ab^{n-p}, b^{p+1} ab^{n-p}\}$ pour un entier $n \geq 1$ et un entier $0 \leq p \leq n$. Ceci revient à chercher si w est préfixe d'un mot $f(a)$ pour un morphisme f sturmien. En conséquence, on retrouve des faits bien connus : tout mot équilibré est préfixe d'un mot sturmien ; tout mot équilibré peut se prolonger à droite et à gauche en des mots équilibrés plus longs.

6.2 Morphismes sturmiens et équilibrés

La partie précédente présente plusieurs utilisations des morphismes sturmiens. Le monoïde de ces morphismes étant finiment engendré, il est possible de tester si un morphisme est sturmien (voir [Lot02, chapitre 2] par exemple). Un autre moyen pour vérifier qu'un morphisme est sturmien a été donné par J. Berstel et P. Séébold [BS94]. Il s'exprime en terme de mot de test : un morphisme non périodique f sur $\{a, b\}$ est sturmien si et seulement si le mot $f(baabaababaabab)$ est équilibré. J. Berstel et P. Séébold [BS94] ont plus généralement montré que, pour $m, r \geq 1$, les mots $w_{m,r}$ et $w'_{m,r}$ suivants sont des mots de test pour morphismes sturmiens :

$$w_{m,r} = (ba^{m+1})^{r+1}ba^m(ba^{m+1})^r ba^m b,$$
$$w'_{m,r} = a(ba^m)^{r+1}ba^{m+1}(ba^m)^r ba^{m+1}b.$$

Le résultat principal de [13] établit qu'il n'existe aucun mot de test pour morphisme sturmien de longueur 13 ou moins. La preuve vient compléter l'approche de [BS94] en montrant que tout mot de test pour morphisme sturmien contient un des mots $w_{m,r}$, $w'_{m,r}$, $E(w_{m,r})$, $E(w'_{m,r})$ ou l'image miroir d'un de ces quatre mots.

Le fait que les mots de test pour morphismes sturmiens soient des mots finis équilibrés et non des mots infinis amène à se demander quels sont les morphismes qui préservent les mots finis équilibrés (appelons-les *morphismes équilibrés*). Ces morphismes sont évidemment très liés aux morphismes sturmiens. Notons l'analogie entre le résultat suivant et l'une des définitions de mots sturmiens rappelée en début de partie 6.1.

Proposition 6.2 [13] *Un morphisme est sturmien si et seulement s'il est équilibré et non périodique.*

Il existe des morphismes équilibrés qui ne sont pas sturmiens. Il s'agit [13] des morphismes périodiques dont le carré de la période est un mot équilibré.

6.3 Morphismes épisturmiens

Les morphismes sturmiens sont définis sur des alphabets à deux lettres. Nous nous intéressons dans cette partie à une généralisation introduite[6] par X. Droubay, J. Justin et G. Pirillo [DJP01] de la manière suivante.

Soit A un alphabet. Les (endo)morphismes épisturmiens de base sont les morphismes Ψ_α et $\overline{\Psi}_\alpha$ définis pour $\alpha \in A$ par

$$\Psi_\alpha : \begin{cases} \alpha \mapsto \alpha, \\ x \mapsto \alpha x \quad \forall x \in A\setminus\{\alpha\}, \end{cases} \qquad \overline{\Psi}_\alpha : \begin{cases} \alpha \mapsto \alpha, \\ x \mapsto x\alpha \quad \forall x \in A\setminus\{\alpha\}, \end{cases}$$

ainsi que les *morphismes d'échange* c'est-à-dire les morphismes E_{xy} définis pour $x, y \in A$ par

$$E_{xy} : \begin{cases} x \mapsto y, \\ y \mapsto x, \\ z \mapsto z \quad \forall z \in A\setminus\{x, y\}. \end{cases}$$

6. Initialement les morphismes épisturmiens ont été introduit comme outil fondamental pour l'étude des mots épisturmiens, une généralisation des mots sturmiens proposée dans [DJP01]. Ultérieurement, J. Justin et G. Pirillo ont montré que les morphismes épisturmiens étaient exactement les morphismes qui préservaient les mots épisturmiens [JP02].

Notons $\mathrm{Exch}(A)$ l'ensemble des morphismes d'échange sur A. Un endomorphisme sur un alphabet A est un *morphisme épisturmien* s'il appartient au monoïde

$$\mathrm{Episturm}(A) = \left(\mathrm{Exch}(A) \cup \{\Psi_\alpha, \overline{\Psi}_\alpha \mid \alpha \in A\}\right)^*.$$

Quand $A = \{a, b\}$, $E_{aa} = E_{bb}$ est l'identité et $E_{ab} = E_{ba}$ est le morphisme noté E dans la partie 6.1. De plus $\Psi_a = \varphi E$, $\Psi_b = E\varphi$, $\overline{\Psi}_a = \tilde{\varphi}E$ et $\overline{\Psi}_b = E\tilde{\varphi}$. La relation $EE = Id$ permet de vérifier que l'ensemble des morphismes épisturmiens sur $\{a, b\}$ est exactement l'ensemble des morphismes sturmiens sur $\{a, b\}$.

Notons que $\Psi_a E = E\Psi_b$ et $\overline{\Psi}_a E = E\overline{\Psi}_b$. Plus généralement, pour toute lettre z et pour tout morphisme d'échange E, les relations

$$\begin{aligned} E\Psi_z &= \Psi_{E(z)}E, \\ E\overline{\Psi}_z &= \overline{\Psi}_{E(z)}E, \end{aligned}$$

permettent de définir les morphismes épisturmiens en faisant jouer un rôle particulier à une lettre a. L'ensemble des morphismes épisturmiens est alors :

$$\mathrm{Episturm}(A) = \left(\mathrm{Exch}(A) \cup \{\Psi_a, \overline{\Psi}_a\}\right)^*.$$

Le monoïde des endomorphismes épisturmiens étant finiment engendré, il est intéressant de s'intéresser aux relations entre générateurs et donc de chercher une présentation de ce monoïde (ensemble de relations entre générateurs dont se déduisent toutes les autres). Un tel résultat avait été obtenu pour le monoïde des morphismes sturmiens par P. Séébold [Séé98]. Dans [3], on trouve la généralisation suivante.

Théorème 6.3 [3] *Soit A un alphabet contenant une lettre a. Le monoïde des morphismes épisturmiens de base* $\mathrm{Exch}(A) \cup \{\Psi_a, \overline{\Psi}_a\}$ *admet la présentation donnée par les relations (x, y, z, t sont des lettres deux à deux distinctes)*

$$\begin{aligned} E_{xy}E_{xy} &= Id, && (1) \\ E_{xy}E_{yz} &= E_{yz}E_{zx}, && (2) \\ E_{xy}E_{zt} &= E_{zt}E_{xy}, && (3) \\ E_{xy}\Psi_a &= \Psi_a E_{xy} \quad \text{quand } a \notin \{x, y\}, && (4) \\ E_{xy}\overline{\Psi}_a &= \overline{\Psi}_a E_{xy} \quad \text{quand } a \notin \{x, y\}, && (5) \\ \Psi_a E_1 \Psi_a E_2 \ldots \Psi_a E_k \overline{\Psi}_a &= \overline{\Psi}_a E_1 \overline{\Psi}_a E_2 \ldots \overline{\Psi}_a E_k \Psi_a && (6) \end{aligned}$$

où $k \geq 1$, $E_1, \ldots, E_k$ sont des morphismes d'échange tels que $E_1 \ \ldots E_k(a) = a$, et pour tout entier i, $2 \leq i \leq k$, $E_i \ \ldots E_k(a) \neq a$.

Une présentation du monoïde, en prenant comme base $\mathrm{Exch}(A) \cup \{\Psi_\alpha, \overline{\Psi}_\alpha \mid \alpha \in A\}$, peut être facilement déterminée à partir de ce théorème [10].

La connaissance de ces présentations permet de caractériser [3,10] plusieurs sous-familles de morphismes épisturmiens. La première de ces caractérisations explique (partiellement) la relation (6).

- Un morphisme épisturmien préserve les palindromes si et seulement s'il est la composition de morphismes d'échange et de morphismes $\Psi_a E_1 \Psi_a E_2 \ldots \Psi_a E_k \overline{\Psi}_a$ où $k \geq 1$, et $E_1, \ldots, E_k$ sont des morphismes d'échange tels que $E_1 \ \ldots E_k(a) = a$, et pour tout entier i, $2 \leq i \leq k$, $E_i \ \ldots E_k(a) \neq a$.
- Un morphisme épisturmien préserve l'ordre lexicographique sur les mots finis si et seulement s'il appartient à $\{\Psi_\alpha, \overline{\Psi}_\alpha / \alpha \in A\}^*$ *et* pour toutes lettres $x, y \in A$, $x \prec y$ implique que $f(y)$ n'est pas un préfixe de $f(xy)$.
- Soit A un alphabet ordonné ayant a comme plus petite lettre et b comme plus grande lettre. Un morphisme épisturmien est un morphisme de Lyndon si et seulement s'il appartient à l'ensemble
$$(\{\overline{\Psi}_\alpha / \alpha \in A \setminus \{a\}\}^* \Psi_a)^* \overline{\Psi}_b^*.$$

En corollaire de ces caractérisations, il est possible de remarquer qu'étant donné un alphabet ordonné A (d'au moins deux lettres), le monoïde des endomorphismes épisturmiens sur A qui préservent l'ordre lexicographique n'est pas finiment engendré, et que le monoïde des endomorphismes épisturmiens de Lyndon sur A est finiment engendré si et seulement si A a deux lettres (dans ce cas il s'agit de l'ensemble $\{\overline{\Psi}_b, \Psi_a\}^*$).

Revenons au théorème 6.3. La preuve que toute relation entre les éléments $\mathrm{Exch}(A) \cup \{\Psi_a, \overline{\Psi}_a\}$ se déduit des relations énoncées dans ce théorème est relativement proche de celle développée par P. Séébold dans le cas sturmien. Une des étapes de cette preuve consiste à montrer que le monoïde des morphismes sturmiens est simplifiable : pour trois morphismes sturmiens f, g, h si $hf = hg$ ou si $fh = fg$ alors $f = g$.

Réciproquement, les cinq premières relations du théorème 6.3 se vérifient simplement. Nous montrerons dans la partie suivante la méthode utilisée dans [3] pour vérifier la sixième relation.

Terminons cette partie sur une remarque. Dans le cas binaire, les morphismes sturmiens sont exactement les *morphismes inversibles* [WW94] (un tel morphisme de monoïdes libres est un morphisme dont l'extension comme morphisme de groupes libres a un inverse pour la composition des morphismes). Ce résultat n'est plus vrai quand on considère des alphabets plus grands. En effet, il a été montré indépendamment dans [WY99] et [3] que, sur un alphabet d'au moins trois lettres, le monoïde des morphismes inversibles n'est pas finiment engendré.

6.4 Conjugaison de morphismes

L'outil utilisé pour établir la sixième relation du théorème 6.3 est la conjugaison de morphismes. Un morphisme g est *conjugué* à droite d'un morphisme f s'il existe un mot w tel que pour tout mot x, $f(x)w = wg(x)$. On dit aussi que w conjugue f et que f est un conjugué à gauche de g. Un morphisme f est conjugué d'un morphisme g si c'est un conjugué à gauche ou à droite. Notons également que tout morphisme est son propre conjugué (à droite et à gauche). Considérons un exemple. Soient α et β deux lettres. Quand $\alpha \neq \beta$, le morphisme $\Psi_\alpha\overline{\Psi}_\beta$ possède quatre conjugués. Outre lui-même, $\Psi_\alpha\Psi_\beta$ et $\overline{\Psi}_\alpha\Psi_\beta$ sont ses conjugués à gauche, et $\overline{\Psi}_\alpha\overline{\Psi}_\beta$ est son conjugué à droite. Quand $\alpha = \beta$, $\Psi_\alpha\overline{\Psi}_\alpha = \overline{\Psi}_\alpha\Psi_\alpha$ a trois conjugués : le conjugué à gauche $\Psi_\alpha\Psi_\alpha$, le conjugué à droite $\overline{\Psi}_\alpha\overline{\Psi}_\alpha$ et lui-même.

Nous nous intéressons dans cette partie à de nombreux liens entre les morphismes épisturmiens et la conjugaison des morphismes. Commençons par des généralités sur la conjugaison [3].

Les morphismes périodiques jouent un rôle spécial pour la conjugaison dans la mesure où ce sont les seuls pour lesquels il existe un nombre infini de mots qui les conjuguent à droite (idem à gauche). Par contre tout morphisme f non périodique est le conjugué à droite d'un morphisme g tel qu'au moins deux images de lettres terminent par des lettres différentes (g n'a pas de conjugué à gauche). Dans ce cas, le nombre total de conjugués de f est égal au nombre de conjugués à droite de g. À noter que quand le morphisme f est épisturmien, g est aussi épisturmien (voir théorème 6.5) et pour toutes lettres distinctes a et b, $g(a)$ et $g(b)$ terminent par des lettres différentes. En d'autres termes, si $\mathrm{last}(u)$ désigne la dernière lettre d'un mot u non vide, $A = \{\mathrm{last}(g(a)) \mid a \in A\}$. Un tel morphisme épisturmien est appelé *épistandard*. L'ensemble des morphismes épistandard est l'ensemble

$$(\mathrm{Exch}(A) \cup \{\Psi_\alpha \mid \alpha \in A\})^*.$$

Dans tous les cas, les conjugués à droite (idem à gauche) d'un morphisme peuvent être totalement ordonnés selon la longueur du (plus petit) mot de conjugaison.

Un endomorphisme f a toujours un nombre fini de conjugués. Dans [3], la majoration de ce nombre par $||f||$ est établie. Cette borne est atteinte pour des morphismes effaçant toutes les lettres sauf une. Pour les endomorphismes non périodiques, le nombre de conjugués devient inférieur ou égal à $\max\{|f(xy)| \mid x, y \in A, x \neq y\} - 1$. Cette valeur est atteinte pour un morphisme f défini, pour un entier $n \geq 0$ et a, b deux lettres, par $f(a) = a$, $f(b) = a^n b$, $f(x) = \varepsilon$ pour $x \in A \setminus \{a, b\}$.

Pour certaines classes de morphismes, la borne peut encore être améliorée comme le montre le résultat suivant qui fournit par ailleurs une caractérisation des morphismes épistandard à partir de leur nombre de conjugués.

Proposition 6.4 [3] *Soit f un endomorphisme sur un alphabet A tel que $A = \{\mathrm{last}(f(a)) \mid a \in A\}$.*

- *Le morphisme f a au plus $\frac{||f||-1}{\mathrm{Card}(A)-1}$ conjugués.*
- *Il a $\frac{||f||-1}{\mathrm{Card}(A)-1}$ conjugués si et seulement s'il est épistandard.*

Le résultat précédent est la généralisation d'un résultat de P. Séébold [Séé98] dans le cas d'un alphabet binaire. Dans ce cas, l'hypothèse "$A = \{\mathrm{last}(f(a)) \mid a \in A\}$" peut être remplacée par "f non périodique". Quand A possède au moins trois lettres, il existe des morphismes f non périodiques ayant strictement plus de $\frac{||f||-1}{\mathrm{Card}(A)-1}$ conjugués. C'est le cas par exemple quand f est de la forme $f(a) = b^n c$, $f(x) = b$ pour toute lettre x différente de a.

La preuve de la proposition 6.4 peut se faire par récurrence. Les prémisses de la récurrence correspondent au cas où il n'existe pas une lettre a telle que $f(a)$ est préfixe de toutes les autres images de lettres. Dans les autres cas, le morphisme f peut être écrit $f\Psi_a$ pour une lettre a. Les relations $||f\Psi_a|| = ||f|| + (\mathrm{Card}(A) - 1)|f(a)|$, et (pour un morphisme non périodique) $\mathrm{NbR}(f\Psi_a) = \mathrm{NbR}(f) + |f(a)|$ permettent l'induction ($\mathrm{NbR}(f)$ désigne le nombre de conjugués à droite de f).

P. Séébold a montré que l'opération de conjugaison de morphismes préserve le caractère sturmien des morphismes [Séé98] (voir aussi [Lot02, chapitre 2]). Ce résultat se généralise aux morphismes épisturmiens (dans ce qui précède, nous avons déjà exploité partiellement ce lien) :

Théorème 6.5 [3] *Tout morphisme conjugué d'un morphisme épisturmien est un morphisme épisturmien.*

La preuve proposée du théorème 6.5 est sensiblement différente de celle de P. Séébold dans le cas binaire. À la base, se trouve un algorithme permettant de donner une décomposition (sur la base $\mathrm{Exch}(A) \cup \{\Psi_a, \overline{\Psi}_a\}$) du premier conjugué à droite (dans l'ordre sur les conjugués) d'un morphisme épisturmien lui-même connu par sa décomposition. La proposition suivante résume cet algorithme.

Proposition 6.6 [3] *Soit A un alphabet contenant une lettre a. Soient $f_1, \dots, f_n$ des éléments de* $\mathrm{Exch}(A) \cup \{\Psi_a, \overline{\Psi}_a\}$.

Le morphisme $f = f_1 \dots f_n$ possède un conjugué à droite différent de lui-même si et seulement s'il existe un entier k tel que $f_k = \Psi_a$.

Quand c'est le cas, soit k le plus petit *entier entre 1 et n tel que $f_k = \Psi_a$. Pour chaque entier i entre 1 et $k-1$, soit g_i le morphisme défini par :*

- $g_i = \Psi_a$ si $f_i = \overline{\Psi}_a$ et la première lettre de $f_{i+1} \dots f_k(a)$ est différente de a.

- $g_i = f_i$ sinon.

Alors le premier conjugué de f est le morphisme $g_1 g_2 \dots g_{k-1} \overline{\Psi}_a f_{k+1} \dots f_n$ ($\overline{\Psi}_a f_2 \dots f_n$ quand $k = 1$).

Donnons un exemple d'utilisation de cet algorithme sur l'alphabet $\{a, b, c\}$. Soit

$$f = E_{bc} \overline{\Psi}_a E_{bc} E_{ac} \overline{\Psi}_a E_{ac} \overline{\Psi}_a E_{bc} \overline{\Psi}_a E_{ab} \Psi_a E_{ab} \Psi_a$$

c'est-à-dire $n = 13$ et $f_1 = f_3 = f_8 = E_{bc}$, $f_2 = f_5 = f_7 = f_9 = \overline{\Psi}_a$, $f_4 = f_6 = E_{ac}$, $f_{10} = f_{12} = E_{ab}$, $f_{11} = f_{13} = \Psi_a$.

Ici $k = 11$. Définissons les morphismes g_i ($1 \leq i \leq 10$). Pour $i \in \{1, 3, 4, 6, 8, 10\}$, f_i est un morphisme d'échange : $g_i = f_i$. Maintenant considérons le cas où $f_i = \overline{\Psi}_a$. Soit x_i la première lettre de $f_{i+1} \dots f_{11}(a)$ ($1 \leq i \leq 10$). Remarquons que $x_{10} = a$ et x_i est la première lettre de $f_{i+1}(x_{i+1})$ (pour $1 \leq i \leq 9$). On a $x_9 = b$, $x_8 = b$, $x_7 = c$, $x_6 = c$, $x_5 = a$, $x_4 = a$, $x_3 = c$, $x_2 = b$ et $x_1 = b$. Ainsi il vient $g_9 = g_7 = g_2 = \Psi_a$ et $g_5 = \overline{\Psi}_a$. D'après la proposition précédente, le premier conjugué de f est :

$$f = E_{bc} \Psi_a E_{bc} E_{ac} \overline{\Psi}_a E_{ac} \Psi_a E_{bc} \Psi_a E_{ab} \overline{\Psi}_a E_{ab} \Psi_a.$$

Pour le vérifier, il suffit de vérifier que le premier conjugué de $f_1 \dots f_{10} \Psi_a$ est $g_1 \dots g_{10} \overline{\Psi}_a$. Cela est vrai puisque

$$\begin{array}{ll} f_1 \dots f_{10} \Psi_a(a) = c(aac)^2 a, & g_1 \dots g_{10} \overline{\Psi}_a(a) = (aac)^2 ac, \\ f_1 \dots f_{10} \Psi_a(b) = c(aac)^3 a, & g_1 \dots g_{10} \overline{\Psi}_a(b) = (aac)^3 ac, \\ f_1 \dots f_{10} \Psi_a(c) = c(aac)^2 abac(aac)^2 a, & g_1 \dots g_{10} \overline{\Psi}_a(c) = (aac)^2 abac(aac)^2 ac. \end{array}$$

En échangeant les rôles de Ψ_a et $\overline{\Psi}_a$ dans la proposition 6.6, on obtient un algorithme pour calculer le premier conjugué à gauche d'un morphisme épisturmien.

Une application de ces deux algorithmes est la vérification de la sixième relation dans le théorème 6.3. En effet le premier conjugué à droite du premier conjugué à gauche d'un morphisme f est f lui-même. L'utilisation des algorithmes précédents montre que le premier conjugué à gauche du morphisme $\Psi_a E_1 \Psi_a E_2 \dots \Psi_a E_k \overline{\Psi}_a$ (avec les hypothèses du théorème 6.3) est

$\Psi_a E_1 \overline{\Psi}_a E_2 \dots \overline{\Psi}_a E_k \Psi_a$ et que le premier conjugué à droite de ce dernier est le morphisme $\overline{\Psi}_a E_1 \overline{\Psi}_a E_2 \dots \overline{\Psi}_a E_k \Psi_a$. D'où la relation.

Dans [1], des algorithmes pour des problèmes similaires sont étudiés : (étant donnée la décomposition d'un morphisme épisturmien f) calcul d'une décomposition d'un conjugué quelconque de f donné par son numéro d'ordre ; calcul d'une liste contenant une décomposition de chaque conjugué de f quand f est épistandard. Pour chaque problème, [1] présente plusieurs algorithmes basés sur des techniques issues de résultats différents de [3]. Une de ces techniques est l'utilisation de l'algorithme décrit dans la proposition 6.6 ainsi que de son équivalent à gauche. Une autre technique consiste à utiliser le fait que le nombre de conjugués à droite d'un morphisme épisturmien $f = f_1 \dots f_n$ avec $f_i \in \mathrm{Exch}(A) \cup \{\Psi_a, \overline{\Psi}_a\}$ (a étant une lettre donnée dans l'alphabet) vérifie :

$$\mathrm{NbR}(f) = 1 + \sum_{1 \le i \le n | f_i = \Psi_a} |f_1 \dots f_{i-1}(a)|.$$

Le but de [1] est de montrer que ces deux principes mènent à écrire des algorithmes différents mais donnant le même résultat.

6.5 Perpectives

On peut observer que les morphismes Ψ_α et $\overline{\Psi}_\alpha$ sont réductibles. Par exemple sur l'alphabet $\{a, b, c\}$, $\Psi_a = fg = gf$ où f et g sont définis par

$$f : \begin{cases} a \mapsto a, \\ b \mapsto ab, \\ c \mapsto c, \end{cases} \qquad g : \begin{cases} a \mapsto a, \\ b \mapsto b, \\ c \mapsto ac. \end{cases}$$

Un sujet intéressant est l'étude de la plus petite famille de morphismes irréductibles contenant les morphismes épisturmiens et son interaction avec les généralisations des mots sturmiens.

Les morphismes épisturmiens ne sont pas (à l'exception des permutations) des morphismes sans puissance k. La conjugaison des morphismes ne préserve pas les morphismes de Lyndon. Par contre, la question suivante se pose : le conjugué d'un morphisme sans puissance k est-il encore un morphisme sans puissance k ? Lors d'une discussion, J. Cassaigne m'a montré qu'il était aisé de prouver que la conjugaison préserve les morphismes qui préservent les mots biinfinis sans puissance k (la conjugaison préserve les facteurs apparaissant dans de tels mots biinfinis). Pour des mots infinis à droite, la question paraît plus difficile.

7 Points fixes finis de morphismes

Dans cette partie, nous présentons des travaux récemment réalisés avec F. Levé [14].

7.1 Ensemble des points fixes d'un morphisme

Les travaux sur les mots utilisant des endomorphismes considèrent ces derniers souvent comme outils pour pouvoir engendrer des mots infinis. En particulier, nous avons ainsi introduit les morphismes sans chevauchement, sans puissance k, de Lyndon. Dans ce cas, les mots engendrés peuvent aussi être considérés comme points fixes des morphismes qui les engendrent. Dans cette partie, nous nous intéressons à une conjecture concernant les mots finis points fixes de morphismes, c'est-à-dire à des mots finis w pour lesquels il existe un morphisme f différent de l'identité tel que $f(w) = w$.

Soit f un endomorphisme défini sur un alphabet A. Une première question est : quels sont les mots finis qui sont points fixes de f ? En particulier f possède-t'il un mot fini comme point fixe en plus du mot vide ? Notons que si cela est le cas, alors nécessairement f est l'identité ou un morphisme effaçant. La réponse aux questions précédentes a été apportée par T. Head [Hea81]. Cette réponse nécessite de considérer plusieurs types de lettres par rapport au morphisme f. Une lettre a est *mortelle* pour f si $f^n(a) = \varepsilon$ pour un entier n. Elle est *constante* si $f(a) = a$. Elle est *expansive* si $f(a) = uav$ pour deux mots u et v constitués uniquement de lettres mortelles et tels qu'au moins un des deux mots u et v n'est pas le mot vide. Notons M_f, C_f, E_f les ensembles de lettres respectivement mortelles, constantes et expansives.

Par exemple, pour le morphisme g défini sur l'alphabet $\{1, 2, \ldots, 8\}$ par

$$\begin{cases} g(1) = \varepsilon, \\ g(2) = g(3) = 1, \\ g(4) = 2, \\ g(5) = g(6) = 5, \\ g(7) = 3742, \\ g(8) = 28, \end{cases}$$

On a $M_g = \{1, 2, 3, 4\}$, $C_g = \{5\}$, $E_g = \{7, 8\}$. Par contre, la lettre 6 n'est ni mortelle, ni constante, ni expansive.

Appelons *exposant de mortalité*, noté $\exp(f)$, le plus petit entier tel que pour toute lettre mortelle a, $f^{\exp(f)}(a) = \varepsilon$. Dans l'exemple précédent, $\exp(g) = 3$. Nous pouvons à présent rappeler la caractérisation donnée par T. Head [Hea81] des points fixes finis de f : ce sont les éléments de l'ensemble

$$(f^{\exp(f)}(C_f \cup E_f))^*.$$

Ainsi l'ensemble des points fixes de notre exemple g est $\{5, 13742211, 128\}^*$.

7.2 Une conjecture

Dans ce qui suit, nous nous intéressons à une conjecture concernant non pas les points fixes d'un morphisme donné, mais les mots qui sont points fixes de morphismes. Plus précisément nous considérons l'ensemble FixedWords des mots w pour lesquels il existe un endomorphisme f, différent de l'identité, défini sur l'alphabet des lettres apparaissant exactement dans w et tel que $f(w) = w$. Un tel morphisme sera appelé *témoin* de w. Par exemple, le mot $abcbbcbabcbbcb$ a deux témoins t_1 et t_2 : $t_1(a) = abcbbcb$, $t_1(b) = t_1(c) = \varepsilon$ et $t_2(a) = a$, $t_2(b) = \varepsilon$ et $t_2(c) = bcb$. Notons que tout mot dans FixedWords doit contenir au moins deux lettres différentes. Notons également que tout mot dans FixedWords a un témoin f idempotent ($f^2 = f$). Ceci est très utile pour résoudre partiellement la conjecture suivante.

Si x est une lettre et w un mot, notons $\delta_x(w)$ le mot obtenu en supprimant dans w toutes les occurrences de x. Par exemple, $\delta_b(abcbbcbabcbbcb) = accacc$. En 1993 [Bil93], M. Billaud a conjecturé :

> *Conjecture.* Si, pour un mot w écrit sur au moins trois lettres différentes, tous les mots $\delta_x(w)$ (pour x lettre dans w) appartiennent à FixedWords, alors w appartient aussi à FixedWords.

Notons que la réciproque de cette conjecture n'est pas vraie. Sur n'importe quel alphabet, il existe des mots w dans FixedWords tels que pour toute lettre x apparaissant dans w, $\delta_x(w)$ n'appartient pas à FixedWords. Cela est le cas pour tous les mots écrits sur deux lettres. Sur $A_n = \{a_1, \ldots, a_n\}$ ($n \geq 3$), le mot suivant vérifie aussi cette propriété :

$$\prod_{i=1}^{n-1} (a_i a_n) \prod_{i=1}^{n-1} (a_{n-i} a_n).$$

La conjecture précédente a été résolue sur des alphabets de trois lettres par P. Zimmermann [Zim93]. Le principal résultat obtenu avec F. Levé est :

Théorème 7.1 [14] *La conjecture précédente est vraie dans le cas où pour toute lettre x, $\delta_x(w)$ a un témoin qui n'a qu'une seule lettre expansive.*

L'idée de base de la preuve de ce résultat est une récurrence menée sur le cardinal de l'alphabet du mot w considéré.

Les prémisses de la récurrence apparaissent être le cas où il existe trois lettres a, b, c telles que les mots $\delta_a(w)$, $\delta_b(w)$ et $\delta_c(w)$ ont chacun un témoin qui n'a qu'une seule lettre expansive et aucune lettre constante. Ce cas généralise le résultat obtenu par P. Zimmermann sur trois lettres. Il est relativement facile dans la mesure où l'absence de lettre constante fait qu'il existe des mots u_a, u_b, u_c tels que $\delta_a(w) \in u_a^*$, $\delta_b(w) \in u_b^*$ et $\delta_c(w) \in u_c^*$. On montre alors qu'il existe un mot u tel que $w \in u^*$. La preuve menée utilise un résultat combinatoire [CP85] emprunté à la théorie des traces mais utilisé uniquement sur des mots : deux mots u et v sur un alphabet A sont égaux si et seulement si $\pi_{ab}(u) = \pi_{ab}(v)$ pour toutes lettres a et b (π_{ab} désigne la projection sur $\{a, b\}$).

Pour montrer la partie inductive de la preuve, nous établissons le résultat technique suivant (la preuve de cette partie est très longue et très technique – 24 étapes, plus de 10 pages) :
Si un mot w, écrit sur au moins 4 lettres différentes, vérifie les hypothèses du théorème 7.1 alors l'une (au moins) des deux assertions suivantes est vérifiée :

- *w appartient à FixedWords,*
- *il existe deux ensembles non vides B et C tels que l'alphabet de w soit $B \cup C$, $B \cap C = \emptyset$, et $\forall a \in B$, $\forall b \in C$, $b \in C_{f_a}$ (les lettres constantes de f_a).*

Dans le deuxième cas, w se décompose en $w = x_0 \prod_{i=1}^{n} y_i x_i$ pour un entier $n \geq 1$, et des mots $x_0, \ldots, x_n$ sur C, $y_1, \ldots, y_n$ sur B. En conséquence, pour toute lettre a dans B et pour tout mot y dans $S = \{y_1, \ldots, y_n\}$, $f_a(\delta_a(y)) = \delta_a(y)$. Il faut alors montrer qu'il existe un morphisme f tel que pour tout mot y dans S, $f(y) = y$. En des termes moins techniques, nous nous retrouvons sur un problème similaire à celui de départ sur w mais à partir non pas d'un mot mais d'un ensemble de mots. Le gros avantage gagné au passage est que les mots $y_1, \ldots, y_n$ sont écrits sur B qui contient strictement moins de lettres que l'alphabet de w. La récurrence peut être mise en place après avoir montré que S peut être remplacé dans un premier temps par $\{y_1y_2y_2y_1, y_3, \ldots, y_n\}$ puis par induction par un seul mot w' dont l'alphabet est B.

Outre les problèmes combinatoires rencontrés dans la preuve du théorème précédent, une grande raison de la difficulté de la conjecture de M. Billaud est peut-être le fait qu'elle cache un problème sur des ensembles de mots. Cela apparaît dans la preuve précédente. Cela est encore plus clair quand on sait que cette conjecture est équivalente à la suivante [14] :

> *Conjecture.* Si, pour un ensemble de mots S écrits sur au moins trois lettres différentes, tous les ensembles $\delta_x(S)$ (pour x lettre dans S) appartiennent à FixedSets, alors S appartient aussi à FixedSets

où FixedSets est l'ensemble des ensembles S de mots pour lesquels il existe un morphisme f, différent de l'identité, défini sur les lettres de S, tel que $f(w) = w$ pour tout mot w dans S.

7.3 Perspectives

La conjecture étudiée avec F. Levé est loin d'être résolue. La résoudre sur un alphabet à 4 lettres est la prochaine étape envisagée (dans ce cas, il se peut que pour une lettre x le mot $\delta_x(w)$ ait un témoin ayant deux lettres expansives). Il faudra peut-être avant d'atteindre cette nouvelle avancée simplifier la preuve décrite dans la partie précédente.

8 Conclusion

Dans ce mémoire, j'ai décrit mes activités de recherche passées et actuelles. Les perspectives présentées à la fin de chaque partie ont précisé de nombreuses idées de recherche pour les années futures.

En particulier, même si de nombreux cas ont été résolus, plusieurs questions restent ouvertes : le problème de l'étoile (et celui de la propriété de puissance finie) est-il décidable ou non ? Existe-t'il des ensembles de test pour morphismes uniformes sans puissance k ($k \geq 3$) ? Peut-on décider si un morphisme (sur un alphabet non binaire) est sans puissance k ($k \geq 3$) ? Peut-on décider si un morphisme (sur un alphabet non binaire) est primitif ? Comment prouver la conjecture sur les points fixes finis de morphismes ? Est-ce que la conjugaison des morphismes préserve les morphismes sans puissance k ?

Par ailleurs de nouvelles études combinatoires sont proposées : sur-famille des morphismes épisturmiens, morphismes qui préservent les (ou engendrent des) mots de Lyndon infinis ?

À beaucoup des questions précédentes, se rajoute celle de la construction ou de l'utilisation d'outils permettant de simplifier des preuves existantes parfois extrêmement techniques.

Mon intérêt ne s'arrête pas aux problèmes précédents. Il existe de nombreuses autres questions intéressantes en combinatoire des mots et théorie des traces (voir par exemple [AS2003, HK03, Lot02, RS97]).

Pour finir, les projets de recherche des années futures pourraient être influencés par des domaines connexes à ceux déjà abordés comme par exemple la théorie des langages formels et la théorie des codes, mais aussi la sémantique des systèmes répartis, l'algorithmique distribuée, l'algorithmique du texte (y compris la recherche de solutions en algorithmique parallèle).

Références

[AGH00] E. Altman, B. Gaujal, and A. Hordjik. Balanced sequences and optimal routing. *Journal of the ACM*, 47:752–775, 2000.

[AS2003] J.-P. Allouche, J. Shallit. *Automatic Sequences (Theory, Applications, Generalizations)*, Cambridge University Press, 2003.

[BEM79] D. R. Bean, A. Ehrenfeucht, and G. F. McNulty. Avoidable patterns in strings of symbols. *Pacific J. of Math*, 85(2):261–294, 1979.

[Ber92] J. Berstel. Axel Thue's work on repetition in words. Technical Report 92.70, LITP, septembre 1992. 4th Conf. on Formal Power Series and Algebraic Combinatorics, Montréal 1992.

[Ber95] J. Berstel. Axel Thue's papers on repetition in words: a translation. Technical Report 20, Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec, Montréal, 1995.

[BFMS02] V. Berthé, S. Ferenczi, C. Maudit, and A. Siegel, editors. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics.* Springer-Verlag, Heidelberg, 2002.

[Bil93] M. Billaud. A problem with words. Letter in newsgroup comp.theory, 1993.

[BK03] J. Berstel and J. Karhumäki. Combinatorics on Words – A Tutorial. *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, 78:178–228, 2003.

[BP85] J. Berstel and D. Perrin. *Theory of codes.* Academic Press, 1985.

[Bra83] F.-J. Brandenburg. Uniformly growing k-th power-free homomorphisms. *Theoretical Computer Science*, 23(1):69–82, 1983.

[Brz67] J.A. Brzozowski. Roots of star events. *J. Assoc. Comput. Mach.*, 14:466–477, 1967.

[BS93] J. Berstel and P. Séébold. A characterization of overlap-free morphisms. *Discrete Applied Mathematics*, 46:275–281, 1993.

[BS94] J. Berstel and P. Séébold. A remark on morphic Sturmian words. *RAIRO Theoretical Informatics and Applications*, 28(3-4):255–263, 1994.

[Car83] A. Carpi. On the size of a square-free morphism on a three letter alphabet. *Information Processing Letters*, 16:231–235, 1983.

[CF69] P. Cartier and D. Foata. *Problèmes combinatoires de commutation et réarrangements*, volume 85 of *Lecture Notes in Mathematics.* Springer, 1969.

[CH74] E. M. Coven and G. Hedlund. Sequences with minimal block growth. *Mathematical Systems Theory*, 7:376–382, 1974.

[CHL01] M. Crochemore, C. Hancart, and T. Lecroq. *Algorithmique du texte*. Vuibert, 2001.

[CK97] C. Choffrut and J. Karhumäki. *Handbook of Formal Languages*, volume 1, chapter Combinatorics of Words. Springer, 1997.

[CL87] M. Clerbout and M. Latteux. Semi-commutations. *Information and Computation*, 73:59–74, 1987.

[CM85] R. Cori and Y. Métivier. Recognizable subsets of some partially abelian monoids. *Theoretical Computer Science*, 35:179–189, 1985.

[CP85] R. Cori and D. Perrin. Automates et commutations partielles. *RAIRO Theoretical Informatics and Applications*, 19(1):21–32, 1985.

[CR87] M. Chrobak and W. Rytter. Unique decipherability for partially commutative monoids. *Fundamenta Informaticae*, X:323–336, 1987.

[CR94] M. Crochemore and W. Rytter. *Text Algorithms*. Oxford University Press, 1994.

[Cro82] M. Crochemore. Sharp characterizations of squarefree morphisms. *Theoretical Computer Science*, 18:221–226, 1982.

[Cro83] M. Crochemore. Régularités évitables (thèse d'état). Technical Report 83-43, LITP, November 1983.

[DGB90] S. Dulucq and D. Gouyou-Beauchamps. Sur les facteurs des suites de Sturm. *Theoretical Computer Science*, 71:381–400, 1990.

[DJP01] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoretical Computer Science*, 255:539–553, 2001.

[DM97] V. Diekert and Y. Métivier. *Handbook of formal languages*, volume 3, chapter 8. Partial commutation and traces, pages 457–533. Springer, 1997.

[DR95] V. Diekert and G Rozenberg, editors. *The book of traces*. World Scientific, 1995.

[Fli74] M. Fliess. Matrices de hankel. *J. Math. Pures et Appl.*, 53:197–224, 1974.

[GOPR92] P. Gastin, E. Ochmanski, A. Petit, and B. Rozoy. Decidability of the star problem in $a^* \times \{b\}^*$. *Information Processing Letters*, 44:65–71, 1992.

[GRS90] G. Guaiana, A. Restivo, and S. Salemi. Complete subgraphs of bipartite graphs and applications to trace languages. *R.A.I.R.O.-Informatique Théorique et Applications*, 24(4):409–417, 1990.

[GRS00] G. Guaiana, A. Restivo, and S. Salemi. On the product of trace languages. In M. Bertoni, A.and Goldwurm and S Crespi Reghizzi, editors, *Workshop on trace theory and code parallelization*, pages 54–67, Décembre 2000. Rapport interne 263-00, Dipartimento di Scienze dell'Informazione, Universita degli Studi di Milano.

[GS66] S. Ginsburg and E. H. Spanier. Bounded regular sets. *Proceedings of the AMS*, 17(5):1043–1049, 1966.

[Has79] K. Hashigushi. A decision procedure for the order of regular events. *Theoretical Computer Science*, 8:69–72, 1979.

[Has82] K. Hashigushi. Limitedness theorem on finite automata with distance functions. *Journal of Computer and System Sciences*, 24:233–244, 1982.

[Has91] K. Hashigushi. Recognizable closures and submonoids of free partially commutative monoids. *Theoretical Computer Science*, 86:233–241, 1991.

[Hea81] T. Head. Fixed Languages and the Adult Languages of 0L Schemes. *Intern. J. Computer Math.*, 10:103–107, 1981.

[HK03] T. Harju and J. Karhumäki, editors. *Proceedings of WORDS'03,* 4^{th} *International Conference on Combinatorics on Words*, volume 27 of *Turku Centre for Computer Science, General Publications.* Turku, Finland, 2003.

[Hol2003] Š. Holub. Binary equality sets are generated by two words. *Journal of Algebra*, 259:1–42 (2003).

[IM97] A. Ido and G. Melançon. Lyndon factorization of the Thue-Morse word and its relatives. *Discret Math. and Theoret. Comput. Sci.*, 1:43–52, 1997.

[Ist77] S. Istrail. On irreducible languages and nonrationnal numbers. *Bull. Soc. Math. Roumanie*, 21:301–308, 1977.

[JP02] J. Justin and G. Pirillo. Episturmian words and episturmian morphisms. *Theoretical Computer Science*, 276(1-2):281–313, 2002.

[Kar81] J. Karhumäki. On strongly cube-free ω-words generated by binary morphisms. In *FCT'81*, volume 117 of *Lect. Notes in Comp. Sci.*, pages 182–191, 1981.

[Kar83] J. Karhumäki. On cube-free ω-words generated by binary morphisms. *Discrete Applied Mathematics*, 5:279-297, 1983.

[Ker86] V. Keränen. On the k-freeness of morphisms on free monoids. Annales Academiae Scientarium Fennicae 61, Series A, 1986.

[Kfo88] A. J. Kfoury. A linear-time algorithm to decide whether a binary word contains an overlap. *Informatique Théorique et Applications*, 22(2):135–145, 1988.

[Kir99] D. Kirsten. A connection between the star problem and the finite power property in trace monoids. In *ICALP'99*, volume 1644 of *Lect. Notes in Comp. Sci.*, pages 473–482. Springer Verlag, 1999.

[Kir02a] D. Kirsten. The finite power problem revisited. *Information Processing Letters*, 84:291–294, 2002.

[Kir02b] D. Kirsten. The star problem in trace monoids: Reductions beyond C4. *Information and Computation*, 176:22–36, 2002.

[KMP77] D. E. Knuth, J. H. Morris, and V. R. Pratt. Fast pattern matching in strings. *SIAM Journal of Computing*, 6(2):323–350, 1977.

[Kun01] M. Kunc. The trace coding problem is undecidable. In *ICALP'2001*, volume 2076 of *Lect. Notes in Comp. Sci.*, pages 603–614, Berlin, 2001. Springer.

[Kun04] M. Kunc. Undecidability of the trace coding problem and some decidability cases. *Theoretical Computer Science*, 310:393-456, 2004.

[Lec85] M. Leconte. *Codes sans répétition.* PhD thesis, LITP Université Paris 6, october 1985.

[Lot83] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics.* Addison-Wesley, 1983. Reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library, Cambridge, UK, 1997.

[Lot02] M. Lothaire. *Algebraic Combinatorics on words*, volume 90 of *Encyclopedia of Mathematics.* Cambridge University Press, Cambridge, UK, 2002.

[Lot] M. Lothaire. *Applied Combinatorics on words.* En préparation, voir : http://www-igm.univ-mlv.fr/~berstel/Lothaire.

[LS69] A. Lentin and M. P. Schutzenberger. A combinatorial problem in the theory of free monoids. In R.C. Bose and T.W. Dowling, editors, *Combinatorial Mathematics and its Applications*, pages 128–144. Univ. North Carolina Press, 1969.

[Mat97] I. Matiiassevitch. Mots et codes: cas décidables et indécidables du problème du codage pour les monoïdes partiellement commutatifs. *Quadrature*, 27:23–33, 1997.

[Maz77] A. Mazurkiewicz. Concurrent program schemes and their interpretations. Technical Report PB 78, Aarhus University, DAIMI, 1977.

[Mét86] Y. Métivier. Une condition suffisante de reconnaissabilité dans un monoïde partiellement commutatif. *RAIRO Theoretical Informatics and Applications*, 20(2):121–127, 1986.

[MH40] M. Morse and G.A. Hedlund. Symbolic Dynamics II: Sturmian trajectories. *Amer. J. Math.*, 61:1–42, 1940.

[Mit97] V. Mitrana. Primitive morphisms. *Information Processing Letters*, 64:277–281, 1997.

[Mor21] M. Morse. Recurrent geodesics on a surface of negative curvature. *Transactions Amer. Math. Soc.*, 22:84–100, 1921.

[MR91] Y. Métivier and B. Rozoy. On the star operation in free partially commutative monoids. *Int. J. of Foundation of Computer Science*, 2(3):257–265, 1991.

[MS93] F. Mignosi and Séébold. If a D0L language is k-power free then it is circular. In *ICALP'93*, volume 700 of *Lect. Notes in Comp. Sci.*, pages 507-518, Springer, Berlin, 1993.

[Och85] E Ochmanski. Regular behaviour of concurrent systems. *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, 27:56–67, 1985.

[Och90] E. Ochmanski. Notes on a star mystery. *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, 27:55–67, 1990.

[Pan81] J.-J. Pansiot. The Morse sequence and iterated morphisms. *Information Processing Letters*, 12(2):68–70, 1981.

[RS97] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages*. Springer, 1997.

[Sak92] J. Sakarovitch. The "last" decision problem for rational trace languages. In *LATIN'92*, volume 583 of *Lect. Notes in Comp. Sci.*, pages 460–473, 1992.

[Séé85] P. Séébold. Sequences generated by infinitely iterated morphisms. *Discrete Applied Mathematics*, 11:255–264, 1985.

[Séé91] P. Séébold. Fibonacci morphisms and Sturmian words. *Theoretical Computer Science*, 88:365–384, 1991.

[Séé98] P. Séébold. On the conjugation of standard morphisms. *Theoretical Computer Science*, 195:91–109, 1998.

[Séé02] P. Séébold. About some overlap-free morphisms on a n-letter alphabet. *Journal of Automata, Languages and Combinatorics* 7, p. 579-597, 2002.

[Sim78] I. Simon. Limited subsets of a free monoid. In *FOCS'19*, pages 143–150, 1978.

[SMDS94] R. Siromoney, L. Mathew, V. R. Dare, and K. G. Subramanian. Infinite Lyndon words. *Information Processing Letters*, 50:101–104, 1994.

[Thu06] A. Thue. Uber unendliche zeichenreihen. *Kristiania Videnskapsselskapets Skrifter Klasse I. Mat.-naturv*, 7:1–22, 1906.

[Thu12] A. Thue. Uber die gegenseitige Lage gleigher Teile gewisser Zeichenreihen. *Kristiania Videnskapsselskapets Skrifter Klasse I. Mat.-naturv*, 1:1–67, 1912.

[Wla01] F. Wlazinski. A test-set for k-power-free binary morphisms. *Theoretical Informatics and Applications*, 35:437–452, 2001.

[Wla02] F. Wlazinski. *Ensembles de test et morphismes sans répétition.* PhD thesis, Université de Picardie Jules Verne, LaRIA, septembre 2002.

[WW94] Z.-X. Wen and Z.Y. Wen. Local isomorphisms of invertible substitutions. *C. R. Acad. Sci. Paris*, 318, série I:299–304, 1994.

[WY99] Z.X. Wen and Z. Yiping. Some remarks on invertible substitutions on three letter alphabet. *Chin. Sci. Bulletin*, 44(19):1755–1760, 1999.

[Zim93] P. Zimmermann. Re: A problem with words. Private communications to M. Billaud, 1993.

Publications

Revue d'audience internationale avec comité de rédaction

1. G. Richomme, *Some algorithms to compute the conjugates of episturmian morphisms*, RAIRO Theoretical Informatics and Applications, vol. 37, p85-104 2003.
2. G. Richomme, *Some non finitely generated monoids of repetition-free endomorphisms*, Information Processing letters, vol. 85(2), p61-66, 2003.
3. G. Richomme, *Conjugacy and episturmian morphisms*, Theoretical Computer Science, vol. 302, p1-34, 2003.
4. G. Richomme, F. Wlazinski, *Some results on k-power-free morphisms*, Theoretical Computer Science, vol. 273, p119-142, 2002. Version complétée et largement remaniée de [16].
5. D. Kirsten, G. Richomme, *Decidability Equivalence between the Star Problem and the Finite Power Problem in Trace Monoids*, Theory of Computing Systems, vol. 34, p193-227, 2001.
6. G. Richomme, *Test-words for Sturmian morphisms*, Bulletin of the Belgian Mathematical Society, vol. 6, p481-489, 1999.
7. G. Richomme, P. Séébold, *Characterization of test-sets for overlap-free morphisms*, Discrete Applied Mathematics, vol. 98, p151-157, 1999.
8. Y. Métivier, G. Richomme, *New results on the star problem in trace monoids*, Information and Computation, vol. 119, n° 2, p240-251, 1995.

A paraître

9. G. Richomme, P. Séébold, *Conjectures and results on morphisms generating k-power-free words*, à paraître dans International Journal of Foundations of Computer Science, special issue on Combinatorics on Words (9 pages).
10. G. Richomme, *Lyndon morphisms*, à paraître dans Bulletin of the Belgian Mathematical Society, numéro spécial "Journées Montoises d'Informatique Théorique" (rapport interne 2002-15 du LaRIA, 22 pages).
11. G. Richomme, F. Wlazinski, *Overlap-free morphisms and finite test-sets*, à paraître dans Discrete Applied Mathematics (rapport interne 2002-05 du LaRIA). Version complète de [15].

En soumission

12. S. Fossé, G. Richomme, *Some characterizations of Parikh matrix equivalent binary words*, soumis à Information Processing Letters (9 pages).

Revue d'audience internationale

13. G. Richomme, *Another Characterization of Sturmian Words*, Bulletin of the European Association of Theoretical Computer Science, vol. 67, p173-175, février 1999.

Manifestations d'audience internationale avec comité de sélection et actes

14. F. Levé, G. Richomme, *On a conjecture about finite fixed points of morphisms (extended abstract)*, actes de WORDS'03 (4^{th} International Conference on Combinatorics on Words), Turku (Finland), TUCS (Turku Centre for Computer Science) General Publication 27, p198-206, 2003.
15. G. Richomme, F. Wlazinski, *Finite test-sets for overlap-free morphisms (Extended abstract)*, actes de la conférence MFCS'2002 (Mathematical Foundations of Computer Science), Varsovie (Pologne), Lecture Notes in Computer Science, vol. 2420, p605-614, 2002.
16. G. Richomme, F. Wlazinski, *About Cube-Free Morphisms (Extended abstract)*, actes de la conférence STACS'2000 (Symposium on Theoretical Aspects of Computer Science), Lille, Lecture Notes in Computer Science, vol. 1770, p99-109, 2000.
17. Y. Métivier, G. Richomme, P.-A. Wacrenier, *Computing the closure of sets of words under partial commutations*, actes de la conférences ICALP'95 (International Colloquium on Automata, Languages and Programming), Szeged, Hongrie, Lecture Notes in Computer Science, vol. 944, p75-86, 1995 (voir aussi rapport interne 836-94 du LaBRI, 1994).
18. G. Richomme, *Some trace monoids where both the Star Problem and the Finite Power Property Problem are decidable*, (extended abstract) actes de la conférences MFCS'1994 (Mathematical Foundations of Computer Science), Varsovie (Pologne), Lecture Notes in Computer Science 841, p577-586, 1994.
19. Y. Métivier, G. Richomme, *On the star operation and the finite power property in free partially commutative monoid*, (extended abstract), actes de la conférence STACS'94, (Symposium on Theoretical Aspects of Computer Science), Caen, Lecture Notes in Computer Science, vol. 775, p341-352, 1994.

Manifestations d'audience internationale avec comité de sélection

20. G. Richomme, *Morphismes de Lyndon*, actes de la 9-ème Conférence Internationale "Journées Montoises d'Informatique Théorique" (du 9 au 11 septembre 2002, à Montpellier). Un résumé (4 pages) en français des résultats est paru dans les actes. Voir aussi [10].
21. G. Richomme, *Conjugacy and episturmian morphisms*, WORDS'2001, Palerme, Sicile, septembre 2001. Voir aussi [3].
22. G. Richomme, F. Wlazinski, *Finite test-sets for overlap-free morphisms*, WORDS'2001, Palerme, Sicile, septembre 2001. Voir aussi [11,15].

Manifestations d'audience internationale

23. G. Richomme, F. Wlazinski, *Test-Sets for Cube-Free Morphisms*, Words 1999, Rouen, septembre 1999. Voir aussi [16].
24. 3-5 avril 1995. *Closure of sets of words under partial commutations* (travail avec Y. Métivier et P.-A. Wacrenier), WASAC (Workshop on Algebraic and Syntactic Aspects of Concurrency), groupe de travail ASMICS, rapport LITP 95/48, p205-212. Voir aussi [17].
25. *Équivalence de décidabilité des problèmes de l'étoile et de la propriété de puissance finie*, Journées Montoises d'Informatique Théorique, Mons (Belgique), 21-23 novembre 1994. Voir aussi [5].

Manifestation d'audience nationale sur invitation

26. Cours à l'école Jeunes Chercheurs en Algorithmique et Calcul Formel, Marne La Vallée, 31 mars-4 avril 2003 (actes contenant une copie des transparents).

Manifestations d'audience nationale et autres

27. *On the star problem and the finite power property problem*, 2-5 novembre 1998, séminaire dans le cadre d'une invitation à Dresde (Allemagne) par D. Kirsten.
28. *Le problème de l'étoile dans les monoïdes de traces*, journées Mathématiques et Informatique d'Amiens, vol. 6 (1994-1995), p12-23, 7 décembre 1994.
29. *Résultats autour de la reconnaissabilité de l'étoile d'un ensemble reconnaissable de traces*, journées LanFor (groupes Langages Formels de l'AFCET), Tignes, 5-8 avril 1993.

Thèse

30. *Contribution à l'étude de l'étoile dans les monoïdes de commutation*, thèse de doctorat de l'Université de Bordeaux I, juillet 1994.

Annexes

Encadrements de jeunes chercheurs

Thèses

- S. Fossé. *Combinatoire des mots sur la fonction matricielle de Parikh.* Thèse débutée en octobre 2002. Co-encadrement avec P. Séébold.
  Autorisation de co-encadrement par le conseil scientifique de l'UPJV (séance du 31 janvier 2003)
  Un article (*Some characterizations of Parikh matrix equivalent binary words*) concernant des problèmes de caractérisation d'un mot par ses sous-mots vient d'être soumis à la revue Information Processing Letters [12].

- F. Wlazinski. *Ensembles de test et morphismes sans répétition.* Thèse soutenue le 23 septembre 2002 à l'UPJV (au LaRIA). Co-encadrement avec P. Séébold.
  Autorisation de co-encadrement par le conseil scientifique de l'UPJV (séance du 26 novembre 1999)
  F. Wlazinski est PRAG en Mathématiques à l'antenne universitaire de Beauvais. Sa thèse a donné lieu à deux publications en revue (Theoretical Computer Science vol. 273 [4], RAIRO-Informatique Théorique et Applications vol. 35 (sans co-auteur)) et à 4 publications en conférence (MFCS'2002 [15], STACS'2000 [16], WORDS'2001 [22], WORDS'1999 [23]). Une soumission à la revue Discrete Applied Mathematics vient d'être acceptée pour publication [11].

- Membre du jury de thèse de F. Levé, novembre 2003, LaRIA, Amiens.

Mémoires de **DEA** d'Informatique (Amiens)

- Mémoire de DEA de L. Legry, *Morphismes injectifs entre deux monoïdes de traces*, 1999-2000.

- Mémoire de DEA de F. Wlazinski, *Caractérisation des morphismes sans cube*, 1997-1998. Co-encadrement avec P. Séébold.

- Mémoire de DEA de R. Lapole, *Codes abéliens (Multiset decipherable sets)*, 1994-1995. Co-encadrement avec P. Séébold.

Autres encadrements

Étude et recherche, Maîtrise d'Informatique : 45 encadrements d'étudiants en 8 ans sur des sujets essentiellement en combinatoire des mots et langages formels.

Projets, Maîtrise d'Informatique : 21 binômes d'étudiants encadrés en 6 ans.

Suivi de stagiaires en entreprise en 2002-2003 : un étudiant de DESS et un de licence professionnelle.

Responsabilités collectives

Vice-doyen (directeur adjoint) de la Faculté de Mathématiques et d'Informatique de juin 1999 à juin 2000.

Responsable de la licence d'Informatique de septembre 1995 à septembre 1999.

Responsable des rapports internes du LaRIA de juin 1999 à mars 2000.

Responsable de la RCEE, réunion consultative entre enseignants et étudiants de licence et maîtrise d'informatique, en 2001-2002 et 2002-2003.

Membre élu *de la commission de spécialistes 27ème section* de l'Université de Picardie Jules Verne depuis février 1998 (titulaire de 1998 à 2001, suppléant depuis 2002).

Membre élu *du conseil d'administration de la Faculté* de Mathématiques et d'Informatique depuis mars 1997.

Membre de la commission évaluation de l'UPJV mise en œuvre par B. Mercuzot (alors chargé de mission Évaluation) au premier semestre 2000 (membre choisi par le doyen de la Faculté de Mathématiques et Informatique).

Rédacteur en chef du site web du LaRIA depuis décembre 2001 (http://www.laria.u-picardie.fr).

Responsable du site web de la Faculté de Mathématiques et Informatique de 1998 à janvier 2002 (http://www.u-picardie.fr/FacMathInfo).

Rédacteur de la plaquette de présentation de la Faculté de Mathématiques et d'Informatique pour les éditions 1998, 1999 et 2000.

Responsable de l'organisation du stand de la Faculté de Mathématiques et Informatique au forum "du Lycéen à l'étudiant" (Mégacité, Amiens) pour les éditions 2001 et 2002.

Tuteur pédagogique de O. Cozette moniteur CIES, de septembre 1999 à juin 2002.

Index

μ, 19
$||f||$, 5
$|u|$, 4
$|u|_a$, 4

carré, 4
concaténation, 8
conjugaison de morphismes, 47
cube, 4

endomorphisme, 4
  binaire, 5
ensemble
  de test, 23
  rationnel, 6
  reconnaissable, 6, 9

facteur, 4, 24

générateur, 4

itération finie, 6

monoïde, 4
  de commutation, 7
  de traces, 7
  finiment engendré, 4
  libre, 4
  libre partiellement commutatif, 7
  libre totalement commutatif, 9
  sans C4, 9
morphisme
  binaire, 5
  conjugué, 47
  d'échange, 5, 44
  de Lyndon, 36
  de Fibonacci, 5
  de monoïdes, 4
  épistandard, 47
  épisturmien, 45
  équilibré, 44
  fortement bipréfixe, 24
  inversible, 46
  irréductible, 22
  périodique, 5
  primitif, 22
  sans chevauchement, 19
  sans puissance k, 20
  sturmien, 42
  uniforme, 19
  vide, 20
mot, 4
  de Fibonacci, 5
  de Lyndon, 36
  de test, 23
  de Thue-Morse, 19
  engendré par morphisme, 5
  équilibré, 42
  primitif, 22
  sans chevauchement, 19
  sans puissance k, 4, 19
  sturmien, 42

permutation, 5
préfixe, 4
problème
  de l'étoile, 10
  de la puissance finie, 12
propriété de puissance finie, 11
puissance, 4

relation
  d'indépendance, 7
  de commutation, 7

suffixe, 4

trace, 8
trace connexe, 8